Al-Qadisiyah Journal for Administrative and Economic Sciences ISSN$_{Online}$: 2312-9883
QJAE, Volume 25, Issue 4 (2023) ISSN$_{Print}$: 1816-9171

Safe Bayesian Inference of Lasso Censored Regression Based on Multinomial Distribution

Asaad Naser Hussein Mzedawee
asaad.nasir@qu.edu.iq
University of Al-Qadisiyah – College of Administration & Economics, Iraq
Corresponding Author : Asaad Naser Hussein Mzedawee

Abstract **:** In this work the author investigate a Bayesian inference in lasso censored regression with more flexible hierarchical prior model. Learning rate parameter proposed to have multinomial distribution which is a prior distribution of the target parameter, so updated hierarchical prior expression developed and Gibbs sample algorithm have independent based on the proposed hierarchical prior model to generate samples from full conditional posterior distribution. Simulation example have conducted to test the performance of the suggested model and real data analysis presented. The result summary of simulation studies and real raw data show that the proposed model perform better than some existing methods **.**

Keywords: safe Bayesian, Lasso, censored, multinomial distribution, Gibbs sample algorithm

1- Introduction

The Censored model is also known as the Tobit structure pioneered by Tobin in 1958 to model the association between the censored response data and some economic predictor variables. The observed response variable in tobit (censored) mode is defined by

$$y = \max\{ y^0, y^* \}$$

Here y^* is the latent response variable, and y^0 which represents a censored point which is usually set to zero in the tobit model . There are a large literature on tobit regression model. See Amemiya T. (1984) for more details about the tobit model .

Recently, many authors employed the tobit model in variable selection methods, like, lasso, adaptive lasso, elastic net (combined),…etc. Yu and Stander (2007) developed a Bayesian estimators for quantile regression with Tobit by proposing that the error term follows the asymmetric Laplace distribution. Flaih and Al-Saadony (2020) proposed lasso , on the tobit regression based on new scale mixture that represent Laplace distribution as prior distribution from Bayesian perspective Odah et al. (2017) suggested the Tobit regression to analyzed the Iraqi bank loans. Alhusseini et al.(2020) introduced the variable selection procedure in Bayesian adaptive lasso , tobit regression based on a new scale mixture of Rayleigh distribution. Abbas and Alhamzawi (2019) investigate the lasso Tobit regression model from Bayesian point of view, new scale mixture of uniform distribution. Alhusseini and Georgescu (2018) composite model employed with Tobit quantile regression based on (probabilistic) Bayesian estimation with scale mixture of uniform distribution .

However, all these works mentioned above are focused on Bayesian inference. This paper have focused on safe Bayesian inference for the lasso tobit regression. We assume that the learning rate parameter of likelihood function have multinomial prior distribution. Following Heide (2016) we suggest a safe Bayesian (probabilistic) hierarchical model by employing the multinomial distribution as prior density for the learning rate parameter that equipped the likelihood function of response variable. New generalized posterior have developed to generate the sample of interested parameter. In the second section, we identified our safe Bayesian (probabilistic) Tobit model and selects of the prior hierarchical model. Full conditional generalized posterior distribution and Gibbs sampling algorithm steps will be given. Third section demonstrate simulation results. In forth section our model is illustrated on creatinine level in the blood. We conclude our proposed model in section fifth.

2- Model Formulation

Suppose that the general structure formula for Tobit (censored) model follows,

$$y_i = \begin{cases} y_i^* & ; \quad y_i > 0 \\ \\ 0 & ; \quad y_i \le 0 \end{cases} \tag{1}$$

Where y_i is the observed response variable and y_i^* is the unobserved (latent) response variable.

Now consider we have a random sample of $y_1^*,\dots\dots\dots, y_n^*$ and the corresponding predictor variables X , then based on (1), we can define:

$$y^* = X\beta + e \tag{2}$$

Here we set β as vector of unknown interested parameters .

Based on variable section method (lasso) , the estimator of unknown interested parameter in model (2) defined by

$$\hat{\beta}_{\text{lasso}} = \arg\min_{\underline{\beta}} \left\| y^* - X\,\underline{\beta} \right\|^2 + \lambda \sum_{j=1}^{k} \left|\beta_j\right| \tag{3}$$

Where the shrinkage parameter $\lambda \ge 0$.

Tibshirani 1996 suggest the β_j have Laplace distribution from Bayesian point view. Many authors investigate the Bayesian estimation of problem (3) , Park and Casella (2008) used the scale mixture of normal to represent Laplace density. Mallick and Yi (2014) used the scale mixture of uniform to Laplace distribution , Flaih et al.( ) used the scale mixture of Rayeigh to represent Laplace distribution. In this paper we will use Mallick and Yi (2014) to redefine the Laplace distribution as follows :

$$Laplace \quad pdf = \frac{\lambda}{2} e^{-\lambda\left|\beta_j\right|} = \int_{w>\left|\beta_j\right|} \frac{1}{2w} \frac{\lambda^2}{\Gamma 2} w^{2-1} e^{-\lambda w} dw \tag{4}$$

Conditioning (4) on σ^2 and follow Mallick and Yi(2014) proposition the scale mixture in (4) can rewritten as follows,

$$\frac{\lambda}{2\sqrt{\sigma^2}} e^{\frac{-\lambda|\beta|}{\sqrt{\sigma^2}}} = \int_{u>\frac{|\beta|}{\sqrt{\sigma^2}}} Un(\beta;-u\sqrt{\sigma^2}, u\sqrt{\sigma^2}) * Gamma(u,2,\lambda) \tag{5}$$

3- Safe Bayesian Hierarchical prior model

Using the tobit formula of regression in (1) , (2) and the scale mixture in (5)the hierarchical prior model can be formulate as follows :

$$\left[y_{n\times 1}^* \mid X, \beta, \sigma^2 \sim \mathrm{N_n}(X\beta, \sigma^2 I_n\right]^{\alpha},$$

$$\beta_{\mathrm{k}\times 1} \mid u, \sigma^2 \sim \prod_{\mathrm{j=1}}^{\mathrm{k}} \mathrm{uniform}(\text{-}\mathrm{u_j}\sqrt{\sigma^2}, u_j\sqrt{\sigma^2}),$$

$$U_{\mathrm{k}\times 1} \mid \lambda \sim \prod_{\mathrm{j=1}}^{\mathrm{k}} \mathrm{Gamma}(2,\lambda) \tag{6}$$

$$p(\alpha \mid p) \propto uni[0,1]$$

$$\sigma^2/\mathrm{c,d} \sim inverse\text{-}\mathrm{Gamma(c,d)}$$

Where α is the learning rate parameter .

3-1 Conditional posterior distribution

Based on Mallick and Yi (2014), the following are the full conditional posterior distribution considering the safe Bayesian technique. The joint density with likelihood powered to the learning rate parameter (α) becomes

$$f(y^* \mid \beta, \sigma^2, \alpha) \cdot \pi(\sigma^2) \cdot \prod_{j=1}^{k} \pi(\beta_j \mid u, \sigma^2) \cdot \pi(u \mid \lambda) \cdot \pi(\lambda) \cdot \pi(\alpha)$$

$$= \left(\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y^*-X\beta)^T (y^*-X\beta)}\right)^{\alpha} \prod_{j=1}^{k} \frac{1}{2u_j\sqrt{\sigma^2}}$$

$$\cdot \prod_{j=1}^{k} I\{\beta_j < u_j\sqrt{\sigma^2}\} \cdot \prod_{j=1}^{k} e^{-\lambda u_j} I\{u_j > \frac{|\beta_j|}{\sqrt{\sigma^2}}\} \cdot \frac{b^a}{\Gamma a}\lambda^{a-1}e^{-\lambda b}$$

$$\cdot \prod_{j=1}^{k} multinomial(\alpha, n, p) \cdot \frac{d^c}{\Gamma c}(\sigma^2)^{-c-1} e^{-\frac{d}{\sigma^2}}$$

Then

- The full conditional posterior distribution of β ,

$$\beta \sim \text{multi-Normal}(\alpha A^{-1} x^T y^*, \sigma^2 A^{-1}) \prod_{j=1}^{k} I\{|\beta_j| < u_j\sqrt{\sigma^2}\}$$

Where $A = \alpha X^T X$

- The full conditional posterior distribution of u is defined by

$$\pi(u / y^*, X, \beta, \tau, \sigma^2) \propto \prod_{j=1}^{k} \exp onential(\lambda) \cdot I\{u_j > \frac{|\beta_j|}{\sqrt{\sigma^2}}\}$$

- The full conditional posterior distribution of σ^2 is defined by

$$\sigma^2/(y^*, X, \beta, u, \lambda, \alpha) \sim inverse\text{-}Gamma \quad (\text{shape parameter and scale parameter})$$

$$= \alpha\frac{n}{2} + c$$

$$= \frac{\alpha}{2}(y^* - X\beta)^T (y^* - X\beta) + d$$

- The full conditional posterior distribution of regularization coefficient λ is defined by

$$\lambda \mid y^*, X, \beta, \sigma^2 \propto Gamma(a + 2k, b + j = 1\sum^{k} |\beta_j|$$

5- The full conditional posterior distribution of α is defined by

$$\alpha \mid y^* \propto \prod_{j=1}^{k} \left[\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}(y^*-X\beta)^T (y^*-X\beta)}\right]^{\alpha_j} \cdot Uniform[0,1]$$

$$\propto multinomial\ distribution(1, \hat{p}_1, \ldots\ldots\ldots, \hat{p}_k)$$

Where $\hat{p}_i = (\frac{1}{(2\pi\sigma^2)^{n/2}} \exp\{\frac{1}{2\sigma^2}(y^* - X\beta)^T (y^* - X\beta)\ \}$

4- Simulation

This section describe our proposed approach to safe Bayesian censored Tobit (SBT) regression through simulation examples . We proceed by adopting the Gibbs sample algorithm of the full conditional posterior that considered in section (3). Furthermore we use the median of means absolute deviation criteria (MMAD) to assess the performance of the proposed model (SBT)with tobit regression model (TRM) , Bayesian tobit regression model (BTRM) and Bayesian tobit quantile regression process (BLTQR). R language have used to implement the above packages (see Kleiber and Zeileis (2017), Morten et.al. (2018) and Alhamzawi (2018) for more details about packages). Now let MMAD defined by

$$\text{MMAD=Me. [mean} \left| X\hat{\beta} - X\beta^{true} \right| \text{]}$$

4-1 Simulation Example first

In this example we generates 200 observations of each predictor variables of the standardized these observation , the correlation coefficient = $\rho^{|i-j|}$, $\rho = 0.5$ the vector of true values of β is defined by $\beta = (3\ ,1.5, 0, 0, 2, 0, 0, 0)^T$, The regression model is $y^* = \beta^T X + e$ we set $\sigma \in (2,5)$ see Tibshirani (1996) for more information. Also, we assumed that the learning rate parameter $\alpha \in (0.2, 0.9)$.
The model is defined as follows

$$y^* = 3x_1 + 1.5x_2 + 2x_5 + e$$

Table (1) Shows the MMAD value and its standard deviation over 200 simulations from the above model.

Table (1) : MMAD and SD for example (1)

α	method	σ	MMAD	SD
0.2	SBT		**0.21213**	**0.1712**
	TRM	2	0.37450	0.2915
	BTRM		0.38212	0.2834
	BLTQR		0.34210	0.2021
0.6	SBT		**0.30154**	**0.1954**
	TRM	6	0.38151	0.3214
	BTRM		0.39651	0.2989
	BLTQR		0.36128	0.2516

The proposed model (Safe Bayesian Tobit)regression model performs the best because it gain the least MMAD values over all other methods under $\alpha = 0.2$.

After implementing the Gibbs sample algorithm , we must check and diagnosis the convergence of the algorithm and that can be obtained by drawing the following trace plot of generating process for each interested parameter .

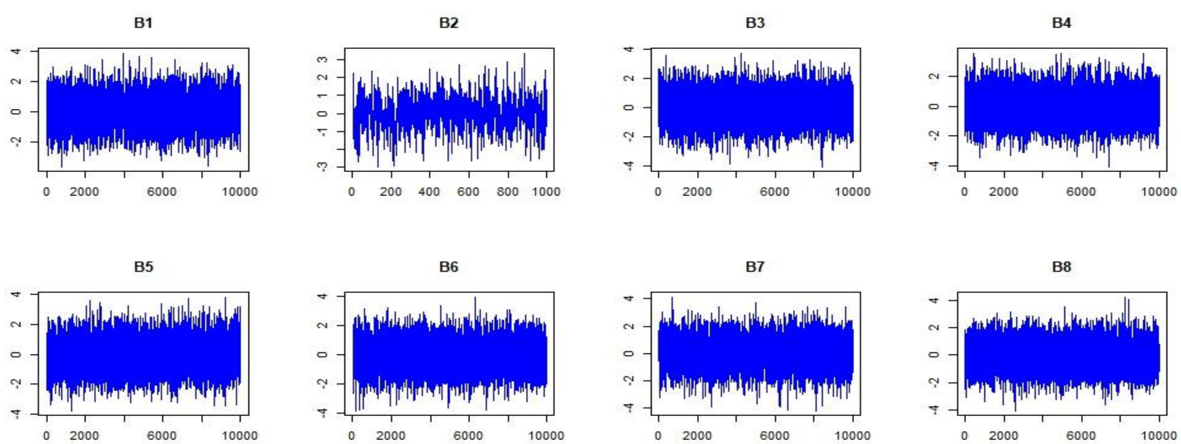

Figure (1):Trace plots of simulation first for $\beta_1 - \beta_8$

Figure (1) show that trace plot for every parameter estimator $(\beta_1 - \beta_8)$, clearly that the Gibbs sample algorithm generating sample from the proposed posterior distribution which are converge to the target distribution. Furthermore the Gibbs sample algorithm showing no slow mixing between the iterations .

4-2 Simulation Example second

This is the same as simulation example first , but the true vector of β is defined by

$$\beta = (0.85)^T_{8\times 1}$$

With $\sigma \in (2,5)$. The results are explained in Table (2)

Table (2) : MMAD and SD for example (2)

α	method	σ	MMAD	SD
0.2	SBT		**0.1925**	**0.1534**
	TRM	2	0.2954	0.2702
	BTRM		0.3034	0.2683
	BLTQR		0.2857	0.1846
0.6	SBT		**0.2287**	**0.1745**
	TRM	6	0.3059	0.3072
	BTRM		0.3281	0.2751
	BLTQR		0.2956	0.2367

Table (2) suggest that proposed model preforms better comparing with the other estimation method by capturing the least MMAD and SD values under $\alpha = 0.2$

After implementing the Gibbs sample algorithm , we must draw the following trace plot of generating process for each interested parameter two check and diagnosis the convergence of the algorithm.

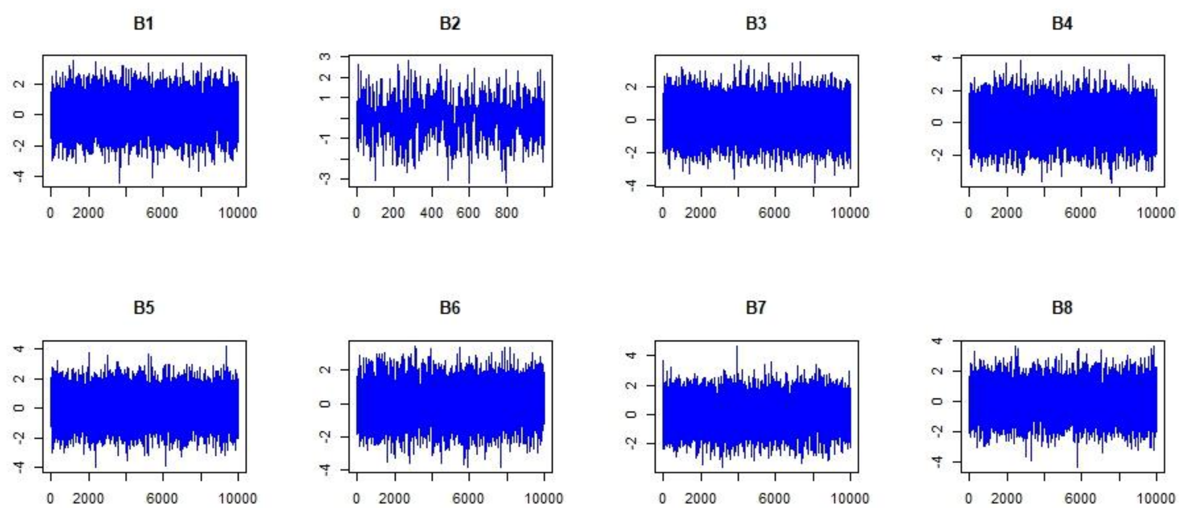

Figure (2): Trace plots of simulation second for $\beta_1 - \beta_8$

Figure (1) show that trace plot for every parameter estimator $(\beta_1 - \beta_8)$, Gibbs sample algorithm showing no slow mixing between the iterations .

• Conclusions

We developed new safe Bayesian tobit regression model hierarchical prior model and new full conditional posterior distribution. Our new proposed model out per forms overall standard Bayesian methods and non-Bayesian method. Also, the proposed Gibbs sample algorithm is very time – intensive. The optimal value of α - learning rate parameter is learnt from generated samples. Consequently, the proposed model performs comparably with other methods

6-Reference

[1] Amemiya T. Tobit models : A survey . J Econ. 1984; 24 3-61 .

[2] Amemiya T. The regression analysis when the dependent variable is truncated normal . Econometrica :. 1973; 41 : 997 -1016 .

[3] Amemiya T. The estimation of a simultaneous equation generalized probit model . Econometrica :. 1981; 49 : 505-513 .
[4] Alhusseini , F. H. H. and Georgescu , V. (2018) . Bayesian composite tobit quantile regression . journal of applied statistics , 45(4), 727-739 .
[5] Fadel Hamid Alhusseini , Ahmad Naeem Flaih and Taha Alshaybawee , (2020) . Bayesian extensions on lasso and adaptive lasso tobit regressions , Periodicals of engineering and Nutural sciences Vol8 . No2 : 1131-1140 .
[6] Flaih , A. M. Al-saadony , H . Elsalloukh . New scale mixture for Bayesian Adaptive lasso tobit regression . Journal of Al-Rafidain university college , 2020 , 46 : 494-505
[7] Greene W. On the asymptotic bias of the ordinary least squares estimator of the tobit model.
[8] Greene W. Estimation of limited dependent variable models by ordinary least squares and the method of moments . J Econ. 1983; 21: 195-212 .
[9] Heckman J. The common structure of statistical models of truncation , sample selection and limited dependent variables and a simple estimator for such models . Ann Econ Soc Meas . 1976 ; 5 : 475 -492 .
[10] Haider Kadhim Abbas . Bayesian lasso tobit regression . journal Al-Qadisiyah for computer science and mathematics 2019 ; 2 : 1-13 .
[11] Heide , R. de. (2016) . The safe – Bayesian lasso . master thesis , universiteit Leiden .
[12] Mallick , H., and Yi , N. (2014). A new Bayesian lasso . statistics and its interface , 7(4), 571 .
[13] Meshal Harbi odah , Ali Sadig mhhommed and Bahr Kadhim mohammed . Tobit Regression analysis Applied on Iraqi Bank leans . American journal of mathematics and statistics 2017 ; 7(4): 179-182 .
[14] Park , T., and Casella , G. (2008) . The Bayesian lasso . journal of the American statistical Association , 103 (482) , 681-686 .
[15] Tobin J. Estimation or relationships for limited dependent variables . Econometrica . 1958;26;24-36 .
[16] *Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society Series B: Statistical Methodology, 58(1), 267-288.*
[17] Wales T , Woodland A. Sample selectivity and the estimation of labor supply functions . Int Econ Rev. 1980 ; 21 : 437-468 .
[18] Wendelin S. Likelihood estimation for censored random vectors . Econ Rev . 2005 ; 24(2): 195-217 .
[19] Yu , K., and Stander , J. (2007) . Bayesian analysis of a tobit quantile regression model . Journal of Econometrics , 137 (1) , 260-276 .