Periodicals of Engineering and Natural Sciences ISSN 2303-4521
Vol. 9, No. 1, January 2021, pp.71-79

Inference with gamma and inverse gamma prior densities in left-censored regression

Esmaeel Ali Alselmawi, Rahim Alhamzawi
Statistics Department, College of Administration and Economics, University of Al-Qadisiyah

ABSTRACT

Left-censored linear regression models are quite popular models and have been deeply considered in the last three decades. In this paper, we consider a completely Bayesian approach for making a new Markov chain Monte Carlo (MCMC) algorithm with tractable full posteriors. Simulated consequences and real data analyses depict that the new Markov chain Monte Carlo algorithm has excellent mixing property and carry out very well than the present methods based on prediction accuracy.

Keywords: Bayesian, Inference, Regularization, Variable selection, Gibbs sampler, Tobit regression

Corresponding Author:

Esmaeel Ali Alselmawi,
Statistics Department,
College of Administration and Economics,
University of Al-Qadisiyah, Al-Diwaniyah, Iraq.
Email: Esmaeel.Alselmawi@yahoo.com

1. Introduction

Left-censored linear regression models are quite popular models and have been deeply considered in the last three decades, for example see, [1], [2], [3], [4], [5], [6], [7], [8]. Suppose that the response y_i and the latent response $y_i{}^*$ are random variables connected by the following relationship

$$y_i = \max\{c, y_i{}^*\},\ \ i = 1, \dots, n,$$

$$y_i{}^* = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i \tag{1}$$

where c is a left-censored point, $\boldsymbol{x_i}$ is a vector of predictors, $\boldsymbol{\beta}$ is a vector of the regression coefficients and ε_i is an error term, The zero censored model (tobit model) is a special case from (1) and is defined as:

$$y_i = \max\{0, y_i{}^*\}, \quad i = 1, \dots, n,$$

$$y_i{}^* = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i \tag{2}$$

In high-dimensional data, we assume that one unidentified subset of predictors is active in a regression model (1). Consequently, the subset selection problem is to find these predictors. In linear regression models, regularization methods are attractive methods that has received considerable attention over the last two decades for dealing with high dimensional data, see for example, lasso and it's Bayesian version [9; 10; 11], elastic net and it's Bayesian version [12; 13; 14; 15], adaptive lasso and it's Bayesian version [13; 16; 17; 18], and so on.

In Bayesian lasso, [10] considered a scale-mixture of normal which is typically takes the form,

$$\beta_j|\sigma^2, s_j^2 \sim N(0, \sigma^2 s_j^2), \qquad j = 1,2,\dots,k,$$
$$s_j^{-1} \sim \text{inverse} - \text{Gaussian}\left(\frac{1}{2}, \sqrt{\frac{\lambda^2\sigma^2}{\beta_j^2}}, \lambda^2\right), \qquad j = 1,2,\dots,k,$$
$$\sigma^2 \sim 1/_{\sigma^2}$$

Since [10], different Bayesian regularization approaches have been proposed over the years, see for example, [19], [20], [21], [22], [23], [24], [25], [26]. Very recently, in the standard linear regression model, [27] considered normal scale mixture priors with beta prime densities for the regression coefficients. The authors noted that this prior distribution can serve as both sparse and non-sparse priors. They showed that a beta prime density is feasibly revised as a product of an independent gamma and inverse gamma densities. Specifically, [27] suggested the following prior distribution for the regression coefficients in standard linear regression model

$$\begin{aligned} &\beta_j|\sigma^2, \lambda^2 s_j^2 \sim N(0, \sigma^2\lambda^2 s_j^2), && j = 1,2,\dots,k, \\ &\lambda_j^2 \sim Gamma(a, 1) && j = 1,2,\dots,k, \\ &s_j^2 \sim inverse - Gamma(b, 1), && j = 1,2,\dots,k, \\ &\sigma^2 \sim 1/_{\sigma^2} \end{aligned} \tag{3}$$

In this paper, we use this class of priors in left-censored regression. Under the above prior distribution, we develop a new Gibbs sampler for Bayesian left censored regression.

2. Methods

2.1. Model hierarchy and prior distributions

Based on (1) and (4), the hierarchical representation is originated as follows:

$$\begin{aligned} &y_i = max\{c, y_i^*\}, && i = 1,2,\dots,n, \\ &y_i^*|\beta, \sigma^2 \sim N(\boldsymbol{x}_i'\beta, \sigma^2), && i = 1,2,\dots,n, \\ &\beta_j|\sigma^2, \lambda_j^2, s_j^2, \sim (0, \sigma^2\lambda_j^2 s_j^2), && j = 1,2,\dots,k, \\ &\lambda_j^2 \sim Gamma(a, 1), && j = 1,2,\dots,k, \\ &s_j^2 \sim inverse - Gamma(b, 1), && j = 1,2,\dots,k, \\ &\sigma^2 \sim 1/_{\sigma^2} \end{aligned} \tag{4}$$

2.2. Full conditional posterior distributions

Following [27], the full conditional posterior distributions are given as

- Updating y_i^*, $\quad i = 1,2,\dots n$.
  Let $\delta(y_i)$ denotes to a degenerate distribution, then y_i^* has a conditional distribution given by:

$$y_i^*|y_i, \boldsymbol{x}_i, \beta, \boldsymbol{\lambda}^2, \boldsymbol{s}^2, \sigma^2 \sim \begin{cases} \delta(y_i), & if\ y_i > c, \\ N(\boldsymbol{x}_i'\beta, \sigma^2) I(y_i^* \le c), & otherwise, \end{cases}$$

where, $\boldsymbol{\lambda}^2 = (\lambda_1^2, \lambda_2^2, \dots, \lambda_k^2$and $s^2 = (s_1^2, s_2^2, \dots, s_k^2)$.

- Updating β,
  The full conditional distribution of β is $N_k(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where

$$\boldsymbol{\mu} = (X'X + \boldsymbol{\Omega}^{-1})^{-1} X'\boldsymbol{y}^*$$

and

$$\boldsymbol{\Sigma} = \sigma^2 (X'X + \boldsymbol{\Omega}^{-1})^{-1},$$

where

$$\boldsymbol{\Omega} = (\lambda_1^2 s_1^2, \dots, \lambda_k^2 s_k^2)'.$$

- Updating ${\lambda_j}^2$ from Generalized Inverse Gaussian (GIG) distribution as follows:
  ${\lambda_j}^2 | {s_j}^2, \beta_j, \sigma^2 \sim GIG(\frac{{\beta_j}^2}{\sigma^2 {s_j}^2}, 2, a - \frac{1}{2}).$

- Updating ${s_j}^2$ from Inverse Gamma (IG) distribution as follows:
  ${s_j}^2 | \beta_j, \sigma^2 {\lambda_j}^2 \sim IG(b + \frac{1}{2}, \frac{{\beta_j}^2}{2\sigma^2 {\lambda_j}^2} + 1).$

- Updating σ^2
  $\sigma^2 | \boldsymbol{y}, \boldsymbol{y}^*, , \beta \sim \text{Inverse−Gamma} \left(\frac{n-1}{2}, \frac{1}{2}(\boldsymbol{y}^* - X\beta)'(\boldsymbol{y}^* - X\beta)\right)$

The above hierarchical model can be adopted for an exact Gibbs sampler that begins at primary guesses for β, σ^2, λ and $\boldsymbol{s}$ iterates the above steps.

3. Simulation Studies

Based on the simulated studies, the median of mean squared errors (MMSE) is considered where the median is taken over the 150 replications. In each replication, a training set with 100 observations is generated and a testing set with 200 observations. Models have been fitted based on the training set, and MSE's have been computed based on the test set. Methods in the comparison include: Tobit regression (TR), Lasso tobit regression (LTR), Bayesian Lasso tobit regression (BLTR) and our proposed method (referred to as 'NewBTR').

3.1. Simulation 1

In this example, the data are simulated from the following model

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

$$ {y_i}^* = \boldsymbol{x_i}'\boldsymbol{\beta} + \varepsilon_i \qquad (5)$$

A very sparse model is considered with a strong level of correlation (ρ = 0.95). The true regression coefficients have been set, involving the intercept term, $\boldsymbol{\beta}$ = (1,2,0,0,0,0,0,0,0), $\sigma^2 = \{1,2,3,4,5\}$ and $e_i \sim N(0, \sigma^2)$. The predictors' matrix X is simulated from a multivariate Gaussian distribution with mean 0, variance 1 and pairwise correlations among x_i and x_j equal to ρ.

3.2. Simulation 2

In this example, we consider the sparse case = $(1,0.25,0.25,0.25,0.25,0.25,0.25,0.25)'$, leaving another setup exactly same as model simulation 1.

3.3. Simulation 3

In this example, we consider the dense case = $(1,0.5,0.5,0.5,0.5,0.5,0.5,0.5)'$ and ρ = 0.5, leaving other setups precisely identical as model Simulation 1. Summary statistics of median mean squared error (MMSE) based on 100 replications for each simulation study have explained in Table 1, 2 and 3 that obviously suggest that a new Bayesian regression method for left censored data (NewBTR) outperform the other methods across all simulation studies. We can observe that the NewBTR produce the smallest MMSE. These results show that the NewBTR exhibit promising performance in terms prediction accuracy.

Table 1. Median mean squared error (MMSE) for 100 replications in Example 1. In the parentheses are standard deviations of the MSEs

n	σ^2	Lasso	aLasso	NewBTR
25	1	1.647 (0.392)	1.502 (0.453)	0.876 (0.333)
25	2	2.247 (0.610)	2.079 (0.683)	1.213 (0.459)
25	3	1.819 (0.646)	1.692 (0.523)	1.077 (0.545)
25	4	2.003 (0.679)	2.042 (0.761)	1.147 (0.729)
25	5	2.119 (1.123)	2.073 (1.236)	1.252 (0.619)
50	1	1.564 (0.278)	1.562 (0.283)	0.854 (0.271)
50	2	1.594 (0.337)	1.578 (0.367)	0.984 (0.284)
50	3	1.733 (0.505)	1.669 (0.584)	1.078 (0.378)
50	4	1.769 (0.415)	1.751 (0.412)	1.083 (0.465)
50	5	1.887 (0.341)	1.850 (0.424)	1.229 (0.422)
100	1	1.532 (0.194)	1.514 (0.187)	0.713 (0.203)
100	2	1.551 (0.211)	1.525 (0.227)	0.892 (0.226)
100	3	1.680 (0.252)	1.645 (0.263)	1.015 (0.155)
100	4	1.700 (0.288)	1.655 (0.305)	1.142 (0.326)
100	5	1.750 (0.309)	1.750 (0.394)	1.215 (0.276)
200	1	1.508 (0.146)	1.468 (0.151)	0.674 (0.164)
200	2	1.529 (0.172)	1.525 (0.171)	0.815 (0.133)
200	3	1.606 (0.291)	1.576 (0.286)	0.932 (0.237)
200	4	1.653 (0.197)	1.653 (0.202)	1.094 (0.183)
200	5	1.639 (0.258)	1.646 (0.293)	1.122 (0.256)

Table 2. Median mean squared error (MMSE) for 100 replications in Example 2. In the parentheses are standard deviations of the MSEs

n	σ^2	Lasso	aLasso	NewBTR
25	1	1.390 (0.212)	1.420 (0.321)	0.392 (0.184)
25	2	1.357 (0.350)	1.413 (0.274)	0.513 (0.306)

n	σ^2	Lasso	aLasso	NewBTR
25	3	1.364 (0.566)	1.457 (0.589)	0.567 (0.488)
25	4	1.440 (0.365)	1.336 (0.586)	0.754 (0.347)
25	5	1.670 (0.601)	1.704 (0.714)	0.820 (0.604)
50	1	1.205 (0.175)	1.236 (0.166)	0.274 (0.133)
50	2	1.369 (0.238)	1.374 (0.275)	0.422 (0.194)
50	3	1.397 (0.198)	1.403 (0.225)	0.567 (0.187)
50	4	1.496 (0.310)	1.436 (0.360)	0.792 (0.363)
50	5	1.448 (0.486)	1.510 (0.584)	0.820 (0.301)
100	1	1.239 (0.152)	1.234 (0.141)	0.302 (0.104)
100	2	1.266 (0.122)	1.305 (0.124)	0.351 (0.102)
100	3	1.323 (0.129)	1.347 (0.138)	0.536 (0.131)
100	4	1.458 (0.241)	1.429 (0.229)	0.672 (0.217)
100	5	1.374 (0.173)	1.389 (0.194)	0.694 (0.185)
200	1	1.223 (0.090)	1.207 (0.095)	0.311 (0.062)
200	2	1.238 (0.107)	1.240 (0.106)	0.443 (0.083)
200	3	1.263 (0.096)	1.283 (0.102)	0.490 (0.093)
200	4	1.310 (0.150)	1.363 (0.185)	0.582 (0.189)
200	5	1.292 (0.141)	1.284 (0.144)	0.692 (0.160)

Table 3. Median mean squared error (MMSE) for 100 replications in Example 3. In the parentheses are standard deviations of the MSEs

n	σ^2	Lasso	aLasso	NewBTR
25	1	1.647 (0.392)	1.502 (0.453)	0.876 (0.333)
25	2	2.247 (0.610)	2.079 (0.683)	1.213 (0.459)
25	3	1.819 (0.646)	1.692 (0.523)	1.077 (0.545)
25	4	2.003 (0.679)	2.042 (0.761)	1.147 (0.729)

n	σ^2	Lasso	aLasso	NewBTR
25	5	2.119 (1.123)	2.073 (1.236)	1.252 (0.619)
50	1	1.564 (0.278)	1.562 (0.283)	0.854 (0.271)
50	2	1.594 (0.337)	1.578 (0.367)	0.984 (0.284)
50	3	1.733 (0.505)	1.669 (0.584)	1.078 (0.378)
50	4	1.769 (0.415)	1.751 (0.412)	1.083 (0.465)
50	5	1.887 (0.341)	1.850 (0.424)	1.229 (0.422)
100	1	1.532 (0.194)	1.514 (0.187)	0.713 (0.203)
100	2	1.551 (0.211)	1.525 (0.227)	0.892 (0.226)
100	3	1.680 (0.252)	1.645 (0.263)	1.015 (0.155)
100	4	1.700 (0.288)	1.655 (0.305)	1.142 (0.326)
100	5	1.750 (0.309)	1.750 (0.394)	1.215 (0.276)
200	1	1.508 (0.146)	1.468 (0.151)	0.674 (0.164)
200	2	1.529 (0.172)	1.525 (0.171)	0.815 (0.133)
200	3	1.606 (0.291)	1.576 (0.286)	0.932 (0.237)
200	4	1.653 (0.197)	1.653 (0.202)	1.094 (0.183)
200	5	1.639 (0.258)	1.646 (0.293)	1.122 (0.256)

4. Real data

In this section, we use a sample of 100 randomly selected patients on 9 variables from the Medical Alaietimad Laboratory in the city of Kut (Iraq) to measure Estimated Glomerular Filtration Rate (eGFR). The response variable is the change in levels of eGFR. The other eight variables are explanatory variables as follows: gender (x_1), age (x_2), percentage of urea in the blood (x_3), the percentage of creatinine in the blood (x_4), calcium level in the blood (x_5), the percentage of potassium in the blood (x_6), the percentage of sodium in the blood (x_7) and the percentage of phosphate in the blood (x_8).

In Table 4, we compare the Mean squared prediction errors by using our proposed method to those obtained using the Lasso and adaptive Lasso. It can be seen that the new method outdoes both Lasso and aLasso using Mean squared prediction errors. The trace plot presented in Figure 1 shows that the Gibbs draw jumps to the stationary distribution in relatively few steps. We also see that the histograms in Figure 2 based on 10,000 posterior samples disclose that the conditional posteriors are the required stationary distributions. Therefore, both the simulation investigations along with the real data consequences exhibit strong support for using the proposed method.

Table 4. Mean squared prediction errors for three methods: Lasso, aLasso and NewBTR

Lasso	aLasso	NewBTR
1.0372	1.0388	0.9736

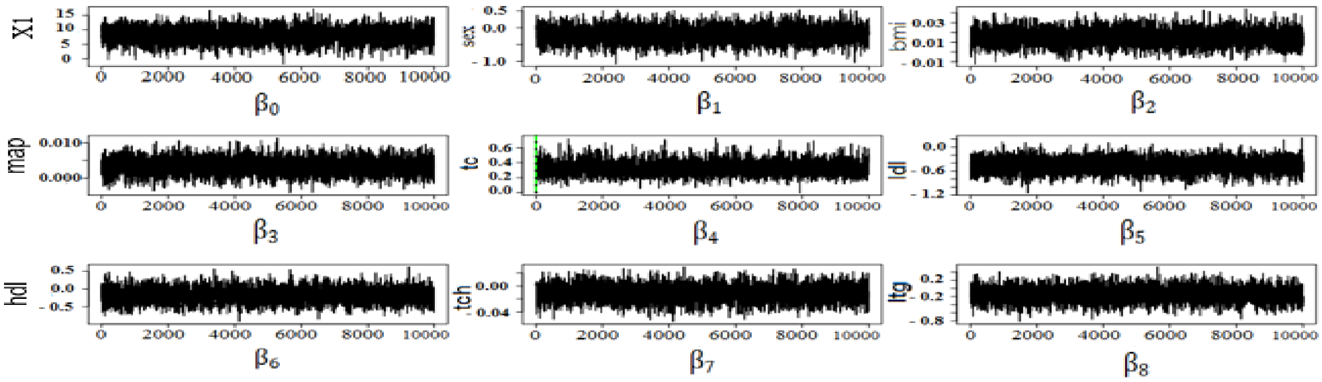

Figure 1. Trace plots of tobit regression parameters.

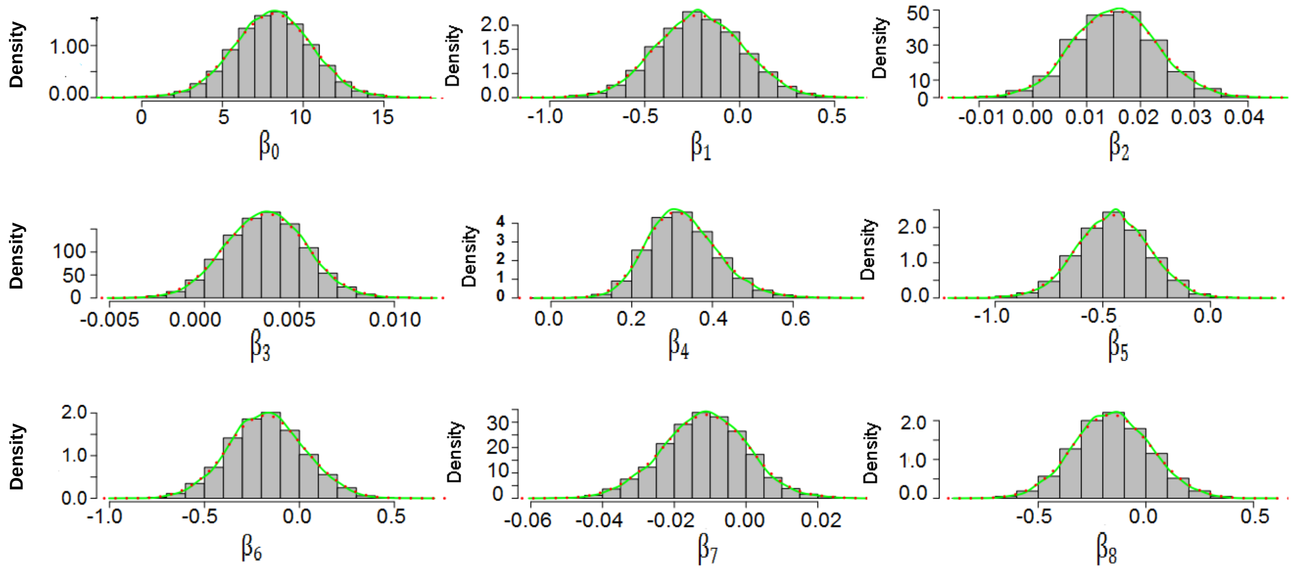

Figure 2. Histograms based on posterior samples of Medical alaietimad Laboratory

5. Conclusions

In this study, we have introduced a new hierarchical model for variable selection and estimation in Left-censored linear regression models. We have proposed a completely Bayesian approach for generating a new Markov chain Monte Carlo (MCMC) algorithm with tractable full posteriors. The proposed model is then demonstrated through simulated samples and a real data set. Consequences depict that the projected method performs very well compared to other existing methods.

References

[1] J. L. Powell, "Censored regression quantiles", *Journal of econometrics*, vol.32, no.1, pp.143-155, 1992.

[2] S. Chib, "Bayes inference in the Tobit censored regression model", *Journal of Econometrics*, vol.51, no.1-2, pp.79-99, 1992.

[3] M. Buchinsky, & J. Hahn, "An alternative estimator for the censored quantile regression model", *Econometrica*, pp.653-671, 1998.

[4] S. A. Murphy, A. W. Van der Vaart, & J. A. Wellner, "Current status regression", 1990.

[5] Y. Bilias, S. Chen, & Z. Ying, "Simple resampling methods for censored regression quantiles", *Journal of Econometrics*, vol.99, no.2, pp.373-386, 2000.

[6] K. Y. Chay, & J. L. Powell, "Semiparametric censored regression models", *Journal of Economic Perspectives*, vol.15, no.4, pp.29-42, 2001.

[7] V. Chernozhukov, & H. Hong, "An MCMC approach to classical estimation", *Journal of Econometrics*, vol.115, no.2, pp.293-346, 2003.

[8] P. Müller, & S. Van de Geer, "Censored linear model in high dimensions", *Test*, vol.25, no.1, pp.75-92, 2016.

[9] R. Alhamzawi, & H. Taha Mohammad Ali, "A new Gibbs sampler for Bayesian lasso",*Communications in Statistics-Simulation and Computation*, vol.49, no.7, pp.1855-1871, 2020.

[10] T. Park, & G. Casella, "The bayesian lasso", *Journal of the American Statistical Association*, vol.103, no.482, pp.681-686, 2008.

[11] 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.

[12] R. Alhamzawi, A. Alhamzawi, & H. T. M. Ali, "New Gibbs sampling methods for bayesian regularized quantile regression", *Computers in biology and medicine*, vol.110, pp.52-65, 2019.

[13] R. Alhamzawi, & H. T. M. Ali, "The Bayesian adaptive lasso regression", *Mathematical biosciences*, vol.303, pp.75-82, 2018.

[14] Q. Li, & N. Lin, "The Bayesian elastic net", *Bayesian analysis*, vol.5, no.1, pp.151-170, 2010.

[15] H. Zou, & T. Hastie, "Regularization and variable selection via the elastic net'‘, Journal *of the royal statistical society: series B (statistical methodology)*, vol.67, no.2, pp.301-320, 2005.

[16] R. Alhamzawi, K. Yu, & D. F. Benoit, "Bayesian adaptive Lasso quantile regression", *Statistical Modelling*, vol.12, no.3, pp.279-297, 2012.

[17] C. Leng, M. N.Tran, & D. Nott, "Bayesian adaptive lasso", *Annals of the Institute of Statistical Mathematics*, vol.66, no.2, pp.221-244, 2014.

[18] H. Zou, "The adaptive lasso and its oracle properties", *Journal of the American statistical association*, vol.101, no.476, pp.1418-1429, 2006.

[19] C. M. Carvalho, N. G. Polson, & J. G. Scott, "Handling sparsity via the horseshoe", In *Artificial Intelligence and Statistics*, pp.73-80, 2009, April.

[20] C. M. Carvalho, N. G. Polson, & J. G. Scott, "The horseshoe estimator for sparse signals", *Biometrika*, vol.97, no.2, pp.465-480, 2010.

[21] J. E. Griffin, & P. J. Brown, "Inference with normal-gamma prior distributions in regression problems'‘, Bayesian *analysis*, vol.5, no.1, pp.171-188, 2010.

[22] R. Alhamzawi, & K. Yu, "Bayesian Lasso-mixed quantile regression", *Journal of Statistical Computation and Simulation*, vol.84, no.4, pp.868-880, 2014.

[23] A. Bhattacharya, D. Pati, N. S. Pillai, & D. B. Dunson, "Dirichlet–Laplace priors for optimal shrinkage", *Journal of the American Statistical Association*, vol.110, no.512, pp.1479-1490, 2015.

[24] R. Alhamzawi, "Model selection in quantile regression models", *Journal of Applied Statistics*, vol.42, no.2, pp.445-458, 2015.

[25] A. Bhattacharya, A. Chakraborty, & B. K. Mallick, "Fast sampling with Gaussian scale mixture priors in high-dimensional regression", *Biometrika*, pp. asw042, 2016.

[26] R. Alhamzawi, "Inference with three-level prior distributions in quantile regression problems", *Journal of Applied Statistics*, vol.44, no.11, pp.1947-1959, 2017.

[27] R. Bai, & M. Ghosh, "On the Beta Prime Prior for Scale Parameters in High-Dimensional Bayesian Regression Models", *arXiv preprint arXiv:1807.06539*, 2018.