

Bayesian LASSO Quantile Regression: An Application to the Modeling of Low Birth Weight

Ferra Yanuar^{1*}, Aidinil Zetra², Arrival Rince Putri¹, Yudiantri Asdi¹

Corresponding author: ferrayanuar@sci.unand.ac.id

¹*Department of Mathematics, Faculty of Mathematics and Natural Science, Andalas University, Kampus Limau Manis, 25163, Padang – Indonesia*

²*Department of Political Science, Faculty of Social and Political Science, Andalas University, Kampus Limau Manis, 25163, Padang – Indonesia*

Abstract. The modeling of low birth weight using ordinary least square is not appropriate and inefficient. The low birth weight data violates the normality assumption since the data is right skewed. The data usually contains outliers as well. Many researchers used quantile regression approach to model this case but this method has limitation. The limitation of this approach is need moderate to big sample size. This study aims to combine the quantile regression with Bayesian LASSO approach to model the low birth weight. Bayesian method has ability to model small sample size since it involves the information related to data (known as likelihood function) and prior information about the parameter to be estimated (prior distribution). This study demonstrated that Bayesian quantile regression and Bayesian LASSO (Least Absolute Shrinkage Selection Operator) quantile regression could yield the acceptable model of low birth weight case based on indicators of goodness of fit model. Bayesian LASSO quantile regression produced better estimated parameter values since it yielded shorter 95% Bayesian credible interval than Bayesian quantile regression.

Keywords: Bayesian LASSO quantile regression, low birth weight, normality assumption, outliers, quantile regression.

1. Introduction

Low birth weight is defined by the World Health Organization as a birth weight < 2500 grams. Low birth weight can cause the delaying of child development and carries a greater risk of early childhood mortality. Moreover, infants with low birth weight also have a significantly greater risk of infection, decreased chances of survival, higher susceptibility to childhood illnesses, and difficulties associated with psychosocial development, behavior, and learning during childhood [1], [2].

Many studies have been developing in investigating the cause of low birth weight. In modeling the birth weight and its determinants usually use linear regression approach [3].

However, this method has some limitations such as requiring the normality assumption or homogeneity of variance of data and inability to manage properly the outlier values. In this context, the quantile regression is one of the methods that can overcome these limitations. This statistical modeling technique can be implemented to evaluate the association between the dependent and independent variables in each selected conditional quantile. This method is also applicable for data with contain outlier values [4]–[9]. The classical frequentists approach to quantile regression has been developed around efficient algorithms and asymptotic theories. But this method also has limitation which is need moderate to big sample size.

Bayesian method then is combined to the quantile regression method to handle the need of big size data. Bayesian method has ability to model small size data since it involves the information related to data (likelihood function) and prior information about the parameter to be estimated in the estimation process. This study also construct the low birth model using Bayesian Lasso quantile regression. This method apply the l_1 penalty in the mixed quantile regression check function [10] which is known as LASSO.

There are not much work has been developed under the Bayesian framework especially for modeling the determinants of low birth weight. Therefore, this study will combine the quantile regression with Bayesian LASSO approach to model the low birth weight [11]. Secondary data of birth weight in Sumatera Barat at 2016 from Yanuar et al. [12] will be used. Thus, this study is also important to be done since there are no previously published manuscripts related to model the birthweight in West Sumatera using Bayesian quantile regression and Bayesian LASSO quantile method. Several low quantiles, such as quantile 0.05, 0.5 and 0.50, will be selected to identify the significant effect of low birth weight at corresponding quantiles. The aim of this study is to determine the indicators of low birthweight in province of West Sumatera, Indonesia using the Bayesian quantile regression and Bayesian LASSO quantile regression method then determine the best estimator.

2. Data and Methods

The secondary data from study of Yanuar et al. is used in this research [12]. The size of sample data is 150 respondents of mothers who just delivered a singleton live birth and were living in West Sumatera, Indonesia. Yanuar et al. collected the data from April to August 2016 in any selected Kabupaten/Kota in West Sumatera. The proposed of that study was to identify the indicators of low birth weight model in West Sumatera using quantile regression approach.

In this data, the response variable was the baby's birth weight recorded in kilograms and eleven independent variables. Eleven indicator variables were used in this study, including continuous and categorical types. Continuous types are age, number of pregnancy problems, number of parities, number of prenatal care visits, hemoglobin (Hb) level, weight gain during pregnancy, last birth interval. Meanwhile categorical types are residence, education level, and sex of the baby [2]. Descriptives of the data can be seen in Yanuar et al. [12] and not presented here because of limited space.

In this present study, we utilize the method of Bayesian quantile and Bayesian LASSO quantile regression to model the case of low birth weight and then compare the result from both methods to find the best estimator method. Following is brief explanation about Bayesian quantile regression method.

Let y_i is the response variable, \mathbf{x}_i is a $p \times 1$ vector of p indicator variables for the i th observation. Let $q_\tau(\mathbf{x}_i)$ denote the τ th ($0 < \tau < 1$) quantile regression function of y_i given \mathbf{x}_i . It can be modeled as $q_\tau(\mathbf{x}_i) = \mathbf{x}_i^T \beta(\tau)$, where $\beta(\tau)$ is a $p \times 1$ vector of coefficients for indicator variables at specified τ . Therefore, the quantile regression model is considered as following:

$$y_i = \mathbf{x}_i^T \beta(\tau) + u_i, \quad i = 1, 2, \dots, n \quad (1)$$

where u_i is the error term with mean zero and constant variance. Then, LASSO quantile regression estimation for $\beta(\tau)$ is obtained by minimizing [13] :

$$\min \sum_i \rho_\tau(y_i - \mathbf{x}_i^T \beta(\tau)) + \lambda \|\beta(\tau)\|_1, \quad (2)$$

where $\rho_\tau(u)$ is the check function defined by $\rho_\tau(u) = u(\tau - I(u < 0))$, where $I(\cdot)$ denotes the indicator function, $\lambda \geq 0$ is a Lagrange multiplier. The second term in (2) is the so called l_1 penalty quantile regression that is crucial for the success of the LASSO, $\|\beta(\tau)\|_1 = \sum_{j=1}^k |\beta(\tau)_j|$. As pointed out already, l_1 penalty term in (2) could be interpreted as a Bayesian posterior mode estimated under independent Laplace priors for the regression coefficient

Bayesian approach to quantile regression method is in combination the minimizing problem in (2). Asymmetric Laplace error distribution is used to maximize likelihood distribution as equivalent way in minimizing the equation [10], [11], [14], [15],[16]:

$$f_\tau(u) = \tau(1 - \tau)e^{-\rho_\tau(u)}, \quad u \in R. \quad (3)$$

The posterior density for parameter estimated $\beta(\tau|y)$ is estimated by multiplying likelihood distribution and prior distribution. The likelihood distribution is normal and prior distribution used here for $\beta(\tau)$ is normal as well, denoted by :

$$\beta(\tau) \sim N(\beta(\tau_0), \beta(\tau_0)). \quad (4)$$

Since the analytical value for $\beta(\tau|y)$ is difficult to obtained, iteration process using Markov Chain Monte Carlo (MCMC) method used here [17].

3. Results and Discussions

A total sample of 150 data then analyzed and modeled using Bayesian quantile and Bayesian LASSO quantile regression method. Normality test is implemented to the data using Kolmogorov Smirnov test. Based on this test it was obtained p value is 8.771e15, this value is less than significant level 0.05. Its mean the data has non normal distribution. The Bayesian

quantile method is then used to model the data. The result of the modeling is presented in Table 1.

Table 1. Estimated Parameter of Birth Weight Model Using Bayesian Quantile Regression

Quantile	Variables	Estimate
0.05	β_1 (Middle)	2.124
	β_2 (High)	-0.996
	β_3 (Parity)	0.196
	β_4 (One problem)	5.365
	β_5 (No problem)	7.540*
0.25	β_1 (Middle)	2.493*
	β_2 (High)	2.027*
	β_3 (Parity)	0.616*
	β_4 (One problem)	1.125
	β_5 (No problem)	2.073*
0.50	β_1 (Middle)	0.845
	β_2 (High)	0.781
	β_3 (Parity)	0.250
	β_4 (One problem)	0.897
	β_5 (No problem)	1.167

*Significant at 5% level

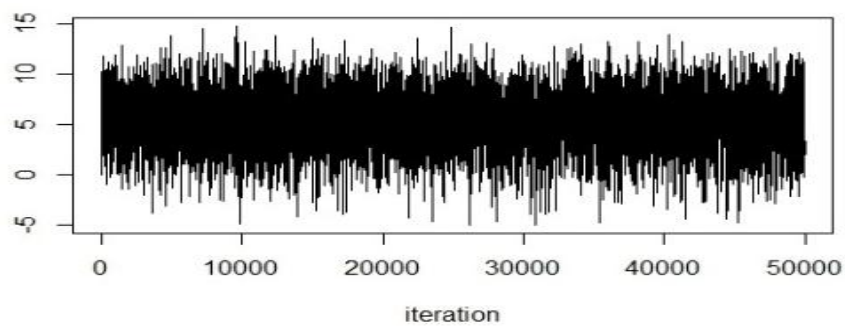
Then, we implement the Bayesian LASSO quantile regression to the same data set. Following Table 2 presents the results.

Table 2. Estimated Parameter of Birth Weight Model Using Bayesian LASSO Quantile Regression

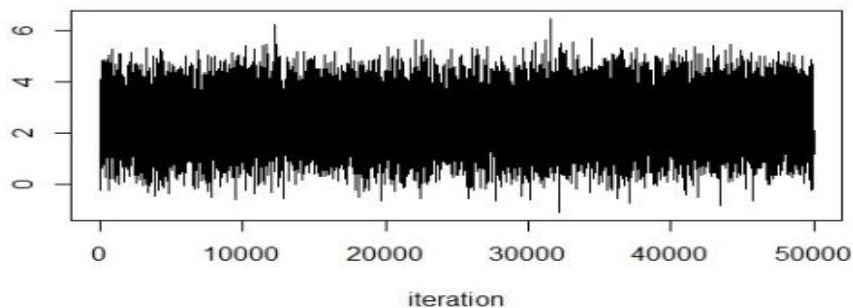
Quantile	Variables	Estimate
0.05	β_1 (Middle)	2.033
	β_2 (High)	-0.456
	β_3 (Parity)	0.202
	β_4 (One problem)	5.209
	β_5 (No problem)	6.638*
0.25	β_1 (Middle)	2.315*
	β_2 (High)	2.112*
	β_3 (Parity)	0.723*
	β_4 (One problem)	0.912
	β_5 (No problem)	2.123*
0.50	β_1 (Middle)	0.835
	β_2 (High)	0.725
	β_3 (Parity)	0.277
	β_4 (One problem)	0.844
	β_5 (No problem)	1.134

*Significant at 5% level

The next process in Bayesian approach is convergence test of the parameter estimated. The convergence is assessed using a variety of diagnostics as detailed in the CODA package, plotting the time series to assess the quality of the individual parameters with different starting values graphically, and provide a diagnosis based on the trace plots[17]. The proposed parameter estimated is converge if the plots lie within two parallel horizontal lines. Following figure 1 shows two selected trace plot of “one problem” at quantile 0.05 and “middle” at quantile 0.25.



(a)



(b)

Figure 1. Trace plot of (a) One problem at quantile 0.05 and (b) Middle at quantile 0.25

Based on those trace plots, it could be concluded that parameter estimated have converge and proposed model could be accepted. Thus, proposed model for Low Birth Weight case in West Sumatera could be modeled as following :

$$\hat{y} = 2.493 \text{ Middle} + 2.027 \text{ High} + 0.616 \text{ Parity} + 1.125 \text{ One problem} + 2.073 \text{ No problem.}$$

The proposed model informs us that to minimize the case of low birth weight, mother has to improve their educational level, minimize the sum of parities and have no problem while pregnancy.

The analysis then continued to do comparison between results of Bayesian quantile and Bayesian LASSO quantile method by looking at length of 95% confidence interval. Following Table 3 presents the results.

Table 3. The 95% Confidence Interval for Bayesian Quantile Regression (BQR) and Bayesian LASSO Quantile Regression (BLQR).

Quantile	Variables	95% CI for BQR and BLQR	Length of 95% CI for BQR and BLQR
0.05	β_1 (Middle)	(-1.083 ; 5.520)	6.603
		(-1.090 ; 5.030)	6.120
	β_2 (High)	(-4.111 ; 2.450)	6.561
		(-4.123 ; 2.320)	6.443
	β_3 (Parity)	(-1.080 ; 1.170) (-0.998 ; 1.110)	2.251 2.108
β_4 (One problem)	(-0.223 ; 9.850)	10.073	
	(-0.198 ; 5.032)	5.230	
β_5 (No problem)	(4.204 ; 10.980)*	6.776	
	(4.098 ; 8.873)*	4.775	
0.25	β_1 (Middle)	(0.937 ; 4.250)*	3.313
		(0.936 ; 4.119)*	3.183
	β_2 (High)	(0.078 ; 3.910)*	3.832
		(0.023 ; 3.878)*	3.855
	β_3 (Parity)	(0.144 ; 1.030)*	0.885
(0.289 ; 0.990)*		0.701	
β_4 (One problem)	(-0.758 ; 3.550)	4.308	
	(-0.722 ; 3.121)	3.843	
β_5 (No problem)	(0.309 ; 4.310)*	4.001	
	(0.298 ; 4.010)*	3.712	
0.50	β_1 (Middle)	(-0.851 ; 2.365)	3.216
		(-0.298 ; 2.290)	2.588
	β_2 (High)	(-1.023 ; 2.397)	3.421
(-0.934 ; 2.333)		3.267	
β_3 (Parity)	(-0.074 ; 0.375)	0.449	

		(-0.020 ; 0.323)	0.343
	β_4 (One problem)	(-0.498 ; 2.224) (-0.155 ; 2.199)	2.722 2.354
	β_5 (No problem)	(-0.026 ; 1.557) (-0.010 ; 1.490)	1.583 1.500

*Significant at 5% level

We look at Table 3 that last column informs us that all 95% confidence interval from Bayesian LASSO quantile regression yielded smaller interval than Bayesian quantile regression for all parameter at different quantiles. This information means that Bayesian LASSO quantile regression resulted better model than Bayesian quantile regression with the model at quantile 0.25 as the best model. Therefore the significant indicators of low birth weight are education level of the mother and number of parities that mother had.

4. Conclusions

The main purpose of this article is to model the low birth weight using Bayesian quantile regression and Bayesian LASSO (Least Absolute Shrinkage Selection Operator) quantile regression. This present study then do comparison between both methods. Quantile regression is used to construct the low birth weight model since we want to model the data at the lower quantile conditional for all data. The empirical data which is used in this study is right skewed or not normal distribution. The appropriate method to model such this data is quantile regression approach [3][12]. We then combine the Bayesian approach and quantile regression since the size of data is moderate, 150 sample only. We also construct the low birth weight model use Bayesian LASSO quantile regression, a modification to the Bayesian quantile regression method. After fitting the data to the hypothesis model, this study demonstrated that Bayesian quantile regression and Bayesian LASSO quantile regression could yield the acceptable model of low birth weight case based on indicators of goodness of fit model. Bayesian LASSO quantile regression produced better estimated parameter values since it yielded shorter 95% Bayesian credible interval than Bayesian quantile regression.

Acknowledgments

This research was support by a grant from Andalas University research grant 2019. The authors wish to thank to all the examiners and anonymous reviewers whose suggestions made this manuscript much improved.

References

- [1] J. Abrevaya and C. M. Dahl, "The effects of birth inputs on birthweight : evidence from quantile estimation on panel data," *J. Bus. Econ. Stat.*, vol. 26, pp. 379–397, 2008.
- [2] A. Hulmán, D. R. Witte, and Z. Kerényi, "Heterogeneous effect of gestational weight

gain on birth weight : quantile regression analysis from a population-based screening,” *Ann. Epidemiol.*, vol. 25, no. 2, pp. 133–137, 2015.

- [3] M. Geraci, “Estimation of regression quantiles in complex surveys with data missing at random: An application to birthweight determinants,” *Stat. Methods Med. Res.*, vol. 25, no. 4, pp. 1393–1421, 2016.
- [4] X. Feng and L. Zhu, “Estimation and Testing of Varying Coefficients in Quantile Regression,” *J. Am. Stat. Assoc.*, vol. 111, no. 513, pp. 266–274, 2016.
- [5] M. Oh, J. Choi, and E. Sug, “Bayesian variable selection in quantile regression using the Savage – Dickey density ratio,” *J. Korean Stat. Soc.*, vol. 45, no. 3, pp. 466–476, 2016.
- [6] H. J. Wang, Z. Zhu, and J. Zhou, “Quantile regression in partially linear varying coefficient models,” *Ann. Stat.*, vol. 37, pp. 3841–3866, 2009.
- [7] M. Oh, E. Sug, and B. So, “Bayesian variable selection in binary quantile regression,” *Stat. Probab. Lett.*, vol. 118, pp. 177–181, 2016.
- [8] S. Tu, M. Wang, and X. Sun, “Bayesian variable selection and estimation in maximum entropy quantile regression,” *J. Appl. Stat.*, vol. 44, no. 2, pp. 253–269, 2016.
- [9] O. D. Saputri, F. Yanuar, and D. Devianto, “Simulation Study of the Implementation of Quantile Bootstrap Method on Autocorrelated Error,” *Cauchy - J. Mat. Murni dan Apl.*, vol. 5, no. November, pp. 95–101, 2018.
- [10] R. Alhamzawi and K. Yu, “Bayesian Lasso-mixed quantile regression,” *J. Stat. Comput. Simul.*, vol. 84, pp. 868–880, 2012.
- [11] F. Yanuar, H. Yozza, and A. Zetra, “Bayesian Quantile Regression Methods in Handling Non-normal and Heterogeneous Error Term,” *Asian J. Sci. Res.*, vol. 12, no. 3, pp. 346–351, 2019.
- [12] F. Yanuar, H. Yozza, I. Rahmi, and A. Zetra, “Applying bootstrap quantile regression for the construction of a low birth weight model,” *Makara J. Heal. Res.*, vol. 23, no. 2, pp. 90–95, 2019.
- [13] D. F. Benoit, R. Alhamzawi, and K. Yu, “Bayesian lasso binary quantile regression,” *Comput. Stat.*, vol. 28, no. 6, pp. 2861–2873, 2013.
- [14] H. Kozumi and G. Kobayashi, “Gibbs sampling methods for Bayesian quantile regression,” *J. Stat. Comput. Simul.*, vol. 9655, no. August, 2011.
- [15] K. Yu and R. A. Moyeed, “Bayesian quantile regression,” *Stat. Probab. Lett.*, vol. 54, no. January, pp. 437–447, 2001.
- [16] F. Yanuar, A. Zetra, C. Muharisa, and D. Devianto, “Bayesian Quantile Regression Method to Construct the Low Birth Weight Model,” *J. Phys. Confrence Ser.*, vol. 1245, pp. 1–7, 2019.
- [17] F. Yanuar, K. Ibrahim, and A. Aziz Jemain, “Bayesian structural equation modeling for the health index,” *J. Appl. Stat.*, vol. 40, no. 6, pp. 1254–1269, 2013.

