

PAPER • OPEN ACCESS

## Bayesian Quantile Regression Method to Construct the Low Birth Weight Model

To cite this article: Ferra Yanuar *et al* 2019 *J. Phys.: Conf. Ser.* **1245** 012044

View the [article online](#) for updates and enhancements.

You may also like

- [Interval Parameter Estimation of Quantile Regression Using Bca-Bootstrap Approach with Application to Open Unemployment Rate Study](#)  
Solehatul Ummah, Vita Ratnasari and Dedy Dwi Prastyo
- [Analysis for instability of parameter in quantile regression with Lagrange multiplier: Is the dependent and independent variable relationships have changed?](#)  
TJ Parmaningsih, S Haryatmi and Danardono
- [Association between airborne PM<sub>2.5</sub> chemical constituents and birth weight—implication of buffer exposure assignment](#)  
Keita Ebisu, Kathleen Belanger and Michelle L Bell



**ECS** The Electrochemical Society  
Advancing solid state & electrochemical science & technology

**241st ECS Meeting**

Vancouver, BC, Canada. May 29 – June 2, 2022

ECS Plenary Lecture featuring  
**Prof. Jeff Dahn,**  
**Dalhousie University**

**Register now!**

# Bayesian Quantile Regression Method to Construct the Low Birth Weight Model

Ferra Yanuar<sup>1</sup>, Aidinil Zetra<sup>2</sup>, Catrin Muharisa<sup>3</sup>, Dodi Devianto<sup>4</sup>, Arrival Rince Putri<sup>5</sup>, Yudiantri Asdi<sup>6</sup>

<sup>1,3,4,5,6</sup>Mathematics Department, Andalas University, Indonesia

<sup>2</sup>Political Science Department, Andalas University  
Kampus Limau Manis Padang, 25163, Indonesia

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

**Abstract.** This study aims to implement Bayesian quantile regression method in constructing the model of Low Birth Weight. The data of Low Birth Weight is violated of nonnormal assumption for error terms. This study considers quantile regression approach and use Gibbs sampling algorithm from Bayesian method for fitting the quantile regression model. This study explores the performance of the asymmetric Laplace distribution for working likelihood in posterior estimation process. This study also compare the result of variable selection in quantile regression and Bayesian quantile regression for Low Birth Weight model. This study proved that Bayesian quantile method produced better model than just quantile approach. Bayesian quantile method proved that it can handle the nonnormal problem although using moderate size of data.

## 1. Introduction

In the classical linier models interests on the conditional mean function, that is the function that estimates how the mean of response variable changes with the vector of covariates. Sometimes, some researchers do not only need mean values but also need additional information requires about the whole conditional distribution of the response variable.

Quantile regression extends the classical mean regression to conditional quantiles of the response variable. It is more robust to nonnormal error distributions and outliers, and provides more complete information on the relationship between the response variable and the covariates than classical theory of linear models [1].

The main problem in any regression is the selection of appropriate covariates, as in quantile regression problem. Excluding important covariates could result biased estimator meanwhile including spurious covariates may lead to loss in estimation efficiency. Quantile regression combined with Bayesian approaches have more powerfull to handle the problem and often more competitive for small or moderate data sets with a low signal-to-noise ratio [2]. This approaches has received considerable attention in recent literatures. [3], [4] explored stochastic search variable selection (SSVS) to estimate the parameter based on quantile regression using latent variable. Meanwhile Oh et al. [5] proposed the



method in Bayesian variable selection using the Savage-Dickey density ratio. Oh et al. [2] described the Bayesian variable selection in binary quantile regression.

In this paper we implement the Bayesian variable selection in quantile regression in the construction the Low birth weight model. The United Nations Children's Fund (UNICEF) defined Low birth weight babies as newborns weighing less than 2,500 grams with the measurement taken within the first hour of life. The data of Low Birth Weight is violated of nonnormal assumption for error terms which is introduced in Section 2 together with the method. This study considers quantile regression model and use Gibbs sampling algorithm from Bayesian approach for fitting the quantile regression model. This study explores the performance of the asymmetric Laplace distribution for working likelihood in posterior estimation process. This study also compare the result of quantile regression and Bayesian quantile regression at selected quantile in the Low Birth Weight case in Section 3. We end with concluding remarks in Section 4.

## 2. Data and Methods

This study used primary data collected by distributing the questionnaires from March to July 2017. We limited the sample to mothers who just had singleton live birth and living in West Sumatera, Indonesia. There were 92 respondents with complete information included in the analysis. The response variable are birth weight, recorded in kilograms. There were 11 indicator variables used in this study, which consist of continuous and categorical types, i.e., mother's education, mother's job, residence, the number of pregnancy problems, mother's age, the number of maternal parity, the number of prenatal care, mother's weight gain during pregnancy, mother's hemoglobin (Hb) level, last birth interval and sex of the baby [6]. Mother's education was divided into three levels; low level, middle level and high level, where low level was as reference category so coefficients were interpreted relative to this category. Mother's job was classified into three categories, i.e., government employee, housewife and others. Residence is categorized into urban and rural. The number of pregnancy problems was categorized into three types, i.e., more than one problems, one problem and no problem. The category more than one problem is noted as reference category. Meanwhile mother's age, the number of maternal parity, the number of prenatal care, mother's weight gain during pregnancy, mother's hemoglobin level and last birth interval are represented as continuous variables.

Figure 1 (a) presents the histogram for the dependent variable of 92 birth weight. Based on the figure, it could be seen that distribution of the data is skewed to the right. It informs us that more data at the lower values, thus the distribution of the data is not normal. Figure 1 (b) shows normal Q-Q plot for the data. This figure also proves that normality assumption is violated in this birth weight data and any outliers are in the data.

To model the low birth weight, quantile regression approach then implemented in this present study. Quantile regression is based on an idea as following.

Consider a linear model [6] [7]:

$$y_i = x_i^T \beta + e_i, \quad i = 1, \dots, n. \quad (1)$$

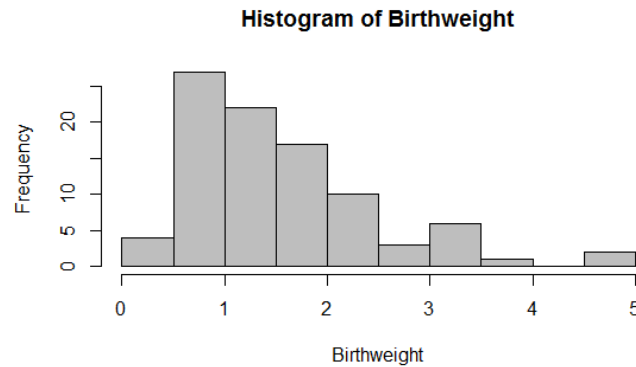
where  $y_i$  is the  $i$ th observation,  $x_i$  is the  $i$ th independent and identically distributed random variables in  $\mathcal{R}^n$  and  $e_i$  is an independent error variable with probability density  $f_i$ . For identifiability, we assume that, for a quantile level  $\tau \in (0,1)$  of interest, the conditional  $\tau$ th quantile of  $e_i$  given  $x_i$  is zero. The conditional quantile regression as follows :

$$Q_Y(\tau|\mathbf{x}) = \mathbf{x}^T \beta(\tau), \quad (2)$$

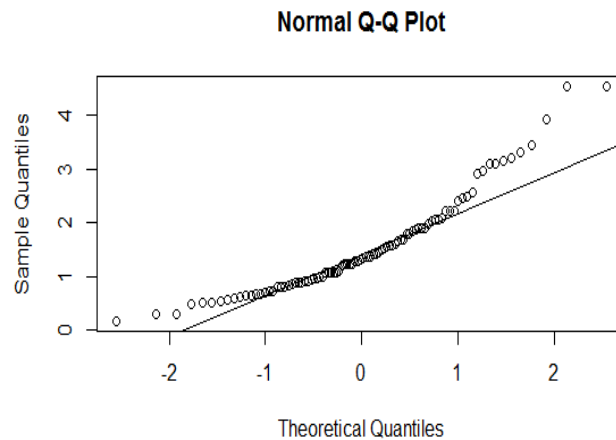
where  $Q_Y(\tau|\mathbf{x})$  represent the  $\tau$ th conditional quantile of the response  $Y$  given  $\mathbf{x}$  and parameter  $\beta(\tau)$  is an unknown functional vector. A point estimate  $\hat{\beta}(\tau)$ , of the parameter  $\beta(\tau)$  is obtained by minimizes the objective function.

$$\sum_{i=1}^n \rho_{\tau}(y_i - x_i'\beta)$$

where  $\rho_{\tau}(u) = u(\tau - I(u < 0))$  is the quantile loss function and  $\psi_{\tau}(u) = \tau - I(u < 0)$  is the score function.



(a)



(b)

**Figure 1.** (a) Histogram and (b) Normal QQ Plot for Birth weight Data

The goodness of fit for these quantile regression are assessed using  $R^2$  values. The formulation of  $R^2$  index for quantile regression differs from OLS regression because it is based on the minimization of an absolute weighted sum (not an unweighted sum of squares as in OLS). The  $R^2$  formulation for quantile regression is represented in terms of the complement of 1 of the ratio between the residual sum of squares and the total sum of squares of the dependent variable [1], usually called as *pseudo* $R^2$  formulated as follows:

$$PseudoR^2 = 1 - \frac{RASW_{\tau}}{TASW_{\tau}} \tag{3}$$

where  $RASW_{\tau}$  is residual absolute sum of weighted differences between the observed dependent variable and the estimated quantile conditional distribution in the more complex model, and  $TASW_{\tau}$  is

total absolute sum of weighted differences between the observed dependent variable and the estimated quantile in the simplest model.

Yu & Moyeed [8] proposed combination of Bayesian approach to quantile regression method in the minimizing problem. They used asymmetric Laplace error distribution to maximize likelihood distribution as equivalent way in minimizing this equation [9, 10, 11, 12]. They assumed that error term follows an independent asymmetric Laplace distribution:

$$f_{\tau}(u) = \tau(1 - \tau)e^{-\rho_{\tau}(u)}, \quad u \in R, \quad (4)$$

The mode of  $f_{\tau}(u)$  is the solution to (2), thus the asymmetric Laplace distribution is closely related to quantile regression. However, the posterior density for parameter estimated  $\beta(\tau)$  is not simple to obtained due to the complexity of the likelihood function, then Markov Chain Monte Carlo (MCMC) method is applied to sample from the approximate posterior distribution. [12] used a random walk Metropolis algorithm with a Gaussian density centred at the current parameter value. Meanwhile [13] developed a Gibbs samplig algorithm based on a location-scale mixture representation of the asymmetric Laplace distribution.

Given  $\mathbf{y} = (y_1, y_2, \dots, y_n)$ , where the prior distribution of  $\beta$  is  $p(\beta)$ . The prior distribution taken in this research is prior informative those originating from previous research. Determination of prior distribution parameters are very subjective, depending on the researcher's intuition. A variable  $Y$  is said to follow Asymmetric Laplace Distribution with the density function of the probability is as follows:

$$f_p(y) = p(1 - p)\exp\{\rho_{\theta}(y_i - \mu)\} \quad (5)$$

and likelihood function as follows :

$$L(y|\beta) = p^n(1 - p)^n \exp\{-\sum_i \rho_{\theta}(y_i - x_i^T \beta)\}. \quad (6)$$

Then the posterior distribution of  $\beta$ ,  $f(\beta|y)$  is given by

$$\begin{aligned} f(\beta|y) &\propto L(y|\beta) p(\beta) \\ &\propto p^n(1 - p)^n \exp\{-\sum_i \rho_{\theta}(y_i - x_i^T \beta)\} p(\beta) \end{aligned}$$

### 3. Results and Discussion

The model hypothesis in this study is presented in the Birth weight's equation as following :

$$\begin{aligned} \text{Birth weight}_i &= \beta_1 \text{Age}_i + \beta_2 \text{Education (Middle)}_i + \beta_3 \text{Education (High)}_i + \beta_4 \text{Parity}_i \\ &+ \beta_5 \text{Last birth interval}_i + \beta_6 \text{Weight gain}_i + \beta_7 \text{Problems (One problem)}_i \\ &+ \beta_8 \text{Problems (No problem)}_i + \beta_9 \text{Hb}_i + \beta_{10} \text{Rural}_i + \beta_{11} \text{Female}_i + e_i; \end{aligned}$$

**Table 1.** Coefficient Estimated for Low Birth Weight Model Using Quantile Regression (QR)

| Indicator Variable      | Estimate of QR (Standard Error) |                |                |
|-------------------------|---------------------------------|----------------|----------------|
|                         | $\tau = 0.05$                   | $\tau = 0.25$  | $\tau = 0.50$  |
| $\beta_1$ (Middle)      | 1.369 (2.370)                   | 2.338 (0.896)* | 1.255 (0.845)  |
| $\beta_2$ (High)        | -0.904 (2.455)                  | 2.065 (0.982)  | 0.981 (0.875)  |
| $\beta_3$ (Parity)      | 0.703 (0.702)                   | 0.777 (0.265)* | 0.232 (0.250)* |
| $\beta_4$ (One problem) | 4.790 (2.595)                   | -0.026 (0.981) | 1.000 (0.925)  |
| $\beta_5$ (No problem)  | 8.273 (2.271)*                  | 1.282 (0.859)* | 1.145 (0.809)* |
| Pseudo R <sup>2</sup>   | 0.26                            | 0.69           | 0.19           |

\* Significant at 10% level

**Table 2.** Coefficient Estimated for Low Birth Weight Model Using Bayesian Quantile Regression (Bayes QR)

| Indicator Variable      | Estimate of Bayes QR |               |               |
|-------------------------|----------------------|---------------|---------------|
|                         | $\tau = 0.05$        | $\tau = 0.25$ | $\tau = 0.50$ |
| $\beta_1$ (Middle)      | 2.124                | 2.493*        | 0.845         |
| $\beta_2$ (High)        | -0.996               | 2.027*        | 0.781         |
| $\beta_3$ (Parity)      | 0.196                | 0.616*        | 0.250*        |
| $\beta_4$ (One problem) | 5.365                | 1.125         | 0.897         |
| $\beta_5$ (No problem)  | 7.540*               | 2.073*        | 1.167*        |

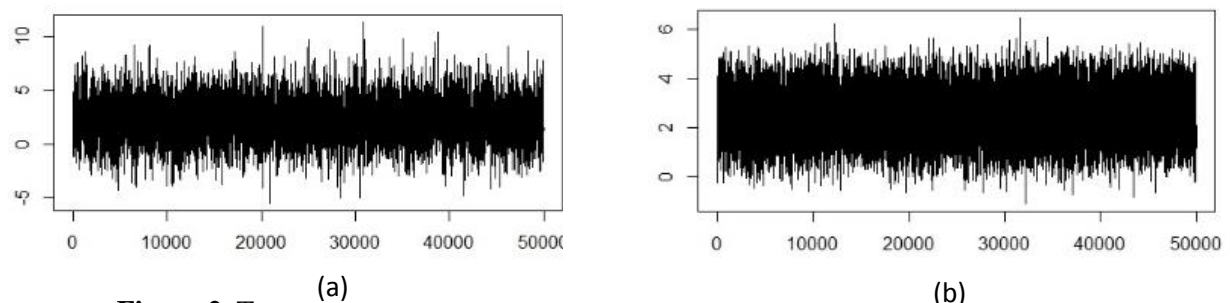
\* Significant at 10% level

The model hypothesis is then fitted to the birth weight data. After fitting, four indicator variables statistically significant to give effect to the response. The variable problems were excluded from the model since those variables are not statistically significant in any of the constructed equations. Following Table 1 presents the results of conditional quantile regression for low quantiles (at quantiles 0.05, 0.25, and 0.50) for significant variables only.

Based on the result of quantile regression model as presented in Table 1 above, the value of Pseudo  $R^2$  for proposed model at quantiles 0.05 and 0.50 are 0.26 and 0.19 respectively. These values indicate that the goodness of fit for proposed model are not good enough, thus both proposed model could not be accepted. Meanwhile the value of Pseudo  $R^2$  for proposed model at quantile 0.50 is 0.69, therefore the proposed model at this quantile could be accepted.

For the next analysis, we consider the Bayesian quantile regression method to construct the Low Birth Weight model and fitted to the same data for  $\tau = 0.05, 0.25,$  and  $0.50$  as well. To assess the sampling efficiency of the proposed algorithm we calculated Monte Carlo standard error [14, 15].

After obtaining the values for parameter model, the analysis is then continued to do the convergence test for all selected parameters. Convergence test in Bayesian method is done by check its trace plot by running the Gibbs sampler for 50,000 iterations with an initial burnin of 5000 iterations. Following are the figure of trace plot for selected parameters ("Middle" at quantiles 0.05th and 0.25th).



**Figure 2.** Trace plot for parameter "Middle" at quantile (a)  $\tau = 0.05,$  (b)  $\tau = 0.25$

Based on the figures, it can be seen that the distribution of selected parameter at any selected quantiles lie within two parallel horizontal lines. These indicated that the parameter have converged [14, 15].

We then do the comparison of difference for 95% confidence interval based on both approaches, quantile regression and Bayesian quantile regression. The 95% confidence interval and the difference are presented in Table 3.

It is clear from Table 3 that Bayesian quantile regression models almost yield shorter 95% confidence interval than those found under quantile regression, except for any indicator variables. This result indicates that Bayesian quantile regression yield better proposed model than quantile regression. This is not surprising due to the extra information brought by the prior distribution in Bayesian quantile estimation method.

Table 3 also informs us that significant variables for each quantiles are different. At quantile 0.05th, the significant variable is  $\beta_5$  (No problem), while at quantile 0.25th are  $\beta_1$  (Middle),  $\beta_3$  (Parity) and  $\beta_5$  (No problem). Meanwhile the significant variables at quantile 0.50th are  $\beta_3$  (Parity) and  $\beta_5$  (No problem). All three models resulted from Bayesian quantile approach at selected quantiles are good enough and could be accepted.

**Table 3.** Comparison the 95% Confidence Interval for Parameter Estimated Using QR and Bayes QR.

| Quantile      | Indicator Variable      | 95% Confidence Interval / Difference |                           |
|---------------|-------------------------|--------------------------------------|---------------------------|
|               |                         | QR                                   | Bayes QR                  |
| $\tau = 0.05$ | $\beta_1$ (Middle)      | (-3.070 ; 4,797)/ 7.867              | (-1.083 ; 5.520) / 6.603  |
|               | $\beta_2$ (High)        | (-2.980 ; 4,797)/ 6.561              | (-4.111 ; 2.450) / 6.561  |
|               | $\beta_3$ (Parity)      | (-1.999 ; 2,910) / 2.250             | (-1.080 ; 1.170) / 2.250  |
|               | $\beta_4$ (One problem) | (-0.320 ;10,797) / 10.073            | (-0.223 ; 9.850) / 10.073 |
|               | $\beta_5$ (No problem)  | (4.465 ;10,797)* / 6.332             | (4.204 ; 9.980)* / 5.776  |
| $\tau = 0.25$ | $\beta_1$ (Middle)      | (1.161 ; 4,659)* / 3.498             | (0.735 ; 4,250)* / 3.512  |
|               | $\beta_2$ (High)        | (-0.706 ; 4,694) / 5.401             | (0.077 ; 3,910)* / 3.832  |
|               | $\beta_3$ (Parity)      | (0.005 ; 0,992)* / 0.987             | (0.144 ; 1,030)* / 0.885  |
|               | $\beta_4$ (One problem) | (-0.498 ;10,905) / 11.404            | (-0.758 ; 3,550) / 4.308  |
|               | $\beta_5$ (No problem)  | (0.848 ; 8,843)* / 7.994             | (0.309 ; 4,310)* / 4.000  |
| $\tau = 0.50$ | $\beta_1$ (Middle)      | (-1.123 ; 2,386) / 3.509             | (-0.851 ; 2,365) /3.216   |
|               | $\beta_2$ (High)        | (-1.180 ; 2,647) /3.828              | (-1.023 ; 2,397) /3.420   |
|               | $\beta_3$ (Parity)      | (0.022 ; 0,462)* /0.439              | (0.074 ; 0,575)* / 0.501  |
|               | $\beta_4$ (One problem) | (-0.840 ; 2,007) /2.848              | (-0.498 ; 2,224) / 2.722  |
|               | $\beta_5$ (No problem)  | (0.011 ; 1,948)* /1.936              | (0.026 ; 1,357)* / 1.331  |

\*Significant at 10% level

#### 4. Conclusions

In this paper, we implement the quantile regression and Bayesian quantile regression for the construction of Low Birth Weight model in West Sumatera. This study found that Bayesian quantile regression yield better proposed model than quantile regression. In the quantile regression produces acceptable model only at quantile 0.25 with indicator for goodness of fit for the model, Pseudo  $R^2 = 0.69$ . This method produced that Low birth weight is affected by middle education level, parity and

having no problem. Whereas Bayesian quantile regression method could produce acceptable model at all selected quantiles with difference significant variables for each quantile.

This study also did the comparison of the length of the 95% confidence intervals associated with the parameters obtained from both methods. This present study found that the length of the 95% confidence intervals associated with the parameters obtained from Bayesian quantile method are generally shorter than with those of the quantile method. This fact also support the conclusion that Bayesian quantile method result better model than classical quantile method [14].

## References

- [1] C Davino, M Furno and D Vistocco 2014 *Quantile Regression Theory and Applications*
- [2] M. Oh, E Sug and B So 2016 Bayesian variable selection in binary quantile regression *Stat. Probab. Lett.* **118** pp 177–181
- [3] R Alhamzawi and K Yu 2012 Variable selection in quantile regression via Gibbs sampling *J. Appl. Stat.* **39**(4) pp 799–813
- [4] R Alhamzawi and K Yu 2013 Conjugate priors and variable selection for Bayesian quantile regression *Comput. Stat. Data Anal.* **64** pp 209–219
- [5] M Oh, J Choi and E Sug 2016 Bayesian variable selection in quantile regression using the Savage – Dickey density ratio *J. Korean Stat. Soc.*
- [6] H. J. Wang, Z. Zhu, and J. Zhou 2009 Quantile regression in partially linear varying coefficient models *Ann. Stat.* **37** pp 3841–3866
- [7] I Spyroglou, S Gunter, C E AC, R AG, and E Paraskakis 2018 A Bayesian Logistic Regression approach in Asthma Persistence Prediction *Epidemiol. Biostat. Public Heal.* **15**(1) pp 1–14
- [8] K Yu and R A Moyeed 2001 Bayesian quantile regression *Stat. Probab. Lett.* **54** 437–447
- [9] D F Benoit BayesQR 2017 A Bayesian Approach to Quantile Regression *J. Stat. Softw.* **76**(7)
- [10] S Chib 2001 Markov Chain Monte Carlo Methods: Computation and Inference *Handbook of Econometrics* vol 5 (North Holland, Amsterdam) pp 3569–3649.
- [11] H M Choi and J P Hobert 2013 Analysis of MCMC algorithms for Bayesian linear regression with Laplace errors *J. Multivar. Anal.* **117** pp 32–40
- [12] D B Dunson and J A Taylor 2005 Approximate Bayesian inference for quantiles *J. Nonparametr. Stat* **17**(3) pp 385–400
- [13] H Kozumi and G Kobayashi 2011 Gibbs sampling methods for Bayesian quantile regression *J. Stat. Comput. Simul.* **9655**
- [14] F Yanuar, K Ibrahim, and A A Jemain 2013 Bayesian structural equation modeling for the health index *J. Appl. Stat.* **40**(6) pp 1254–1269
- [15] A Rahmadita, F Yanuar and D Devianto 2018 The Construction of Patient Loyalty Model Using Bayesian Structural Equation Modeling Approach *Cauchy - J. Mat. Murni dan Apl.* **5**(2) pp 73–79