Available online at http://scik.org
Commun. Math. Biol. Neurosci. 2023, 2023:108
https://doi.org/10.28919/cmbn/7976
ISSN: 2052-2541

BAYESIAN REGULARIZED TOBIT QUANTILE TO CONSTRUCT STUNTING RATE MODEL

FERRA YANUAR[1,*], ATHIFA SALSABILA DEVA[1], RUDIYANTO[2], AIDINIL ZETRA[3], CHYNTIA DWI YAN[1], ARFARANI ROSALINDARI[1], HAZMIRA YOZZA[1]

[1]Department of Mathematics and Data Science, Andalas University, Padang, 25163, Indonesia

[2]Department of Crop Science, Faculty of Fisheries and Food Science, Universiti Malaysia Terengganu, 21030 Kuala Nerus, Terengganu, Malaysia

[3]Department of Political Science, Andalas University, Padang, 25163, Indonesia

Abstract: This study aims to identify the best model for the stunting rate by applying and comparing several methods based on the Tobit quantile regression method's modification. The stunting rate dataset is left censored and violated with linear model assumptions; thus, Tobit quantile approaches are used. The Tobit quantile regression is adjusted by combining it with the Bayesian approach since the Bayesian method can produce the best model in small-size samples. Three kinds of modified Tobit quantile regression methods considered here are the Bayesian Tobit quantile regression, the Bayesian Adaptive Lasso Tobit quantile regression, and the Bayesian Lasso Tobit quantile regression. This article implements the skewed Laplace distribution as the likelihood function in Bayesian analysis. This study used the data of 3534 stunting children obtained from the Health Departments of several districts and municipals in West Sumatra, Indonesia. The result of this study indicated that Bayesian Lasso quantile regression performed well compared to the

[*]Corresponding author
E-mail address: ferrayanuar@sci.unand.ac.id
Received March 30, 2023

other two methods. Criteria of better method are based on a smaller absolute bias and a shorter Bayesian credible interval which are obtained from the simulation study and empirical study. This study also found that exclusive breastfeeding give impact to stunting rate only at middle quantiles, while comorbidity tend to affect all distribution of stunting rate.
Keywords: stunting rate; bayesian tobit quantile regression; bayesian adaptive lasso tobit quantile regression; bayesian lasso tobit quantile regression.

2020 AMS Subject Classification: 62F15.

1. INTRODUCTION

One-third of all deaths in children under five are caused by malnutrition [1]. Malnutrition has serious health, social, and economic consequences throughout one's life and across generations, making it as the leading risk factor among children under the age of five worldwide [2]. Low height-for-age, also known as stunting, is a key indicator of chronic malnutrition because it reflects a failure to reach linear growth potential. Globally, depending on the precise definition and estimate, between 171 million and 314 million children under five are currently classified as stunted, with 36 African and Asian countries bearing 90% of this burden. West Sumatra, a province in Indonesia, has a higher stunting rate than the WHO's tolerance, which is above 20%. Therefore, the stunting problem has become a priority issue by the government of West Sumatra that has to be handled and solved soon.

The stunting rate variable is a so-called limited dependent variable whose distribution is mostly continuous but has a point mass at one or more specific values, such as zero. The Tobit model is one of statistical approach to models limited dependent variables[3].Tobit regression has become one of the most commonly used statistical tools utilized by researchers to describe the relationship between a non-negative response variable and a set of covariates [4–6]. The Tobit regression has been routinely applied in medicine, biology, ecology, economics, and social sciences [7–11]. This model can be viewed as a linear regression model with a latent continuous response $y*$. Consider the standard Tobit regression model

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

$$y_i^* = x_i^{'}\beta + \varepsilon_{i,} \tag{1}$$

where ε_i 's are residuals with $\varepsilon_i \sim N(0,\sigma^2), \boldsymbol{x_i} = (x_{i1},\dots,x_{ik})'$, and $\beta = (\beta_1,\dots,\beta_k)'$. The observed stunting rate is assumed to be related to the latent value by the following:

$$y_i = \begin{cases} y_i^*, if\ y_i^* > 0, \\ 0, otherwise. \end{cases}$$

The response variable can be written as $y_i = max\{0, y_i^*\}$, unknown parameter β is estimated using the maximum likelihood method. Although the asymptotic theory for maximum likelihood has been well studied in Tobit models [12], a Bayesian method produces exact inference even if *n* is small [13–15]. Alhamzawi and Yu [6] applied a Bayesian approach to the Tobit regression model using the normal density for the residuals and generating β from its full conditional posterior distribution using a Gibbs sampler. Alhamzawi and Ali [16] proposed adaptive lasso in Tobit quantile regression using the Bayesian technique. Alhamzawi and Yu [6] suggested a Bayesian technique for coefficient estimation in (TobitQReg) model utilizing g-prior distribution with ridge parameter. Alhamzawi [4,5] proposed a Bayesian elastic net penalty in (TobitQReg). Mallick and Yi [11] provided a new technique for achieving Bayesian Lasso in a traditional regression model by scale mixture of uniform formulation of the Laplace density.

The objective of this paper was to find the association between demographic, socioeconomic, and health factors of the stunting rate of children under 3 years in West Sumatra, Indonesia by applying Bayesian Tobit quantile regression and its modified techniques. The current analysis expects to improve the structure of successful intervention measures designed to tackle the stunting rate or reduce the prevalence of stunting and improve child health.

2. PRELIMINARIES

2.1. Bayesian Tobit Quantile Regression

Given a sample of independent observations $\boldsymbol{y} = (y_1,\dots,y_n)$ and associated k covariates $= (x_1,\dots,x_k)$, the latent variable y_i^* is modeled as follows:

$$y_i^* = \eta_\tau(x_i|\theta) + u_{\tau i}, \quad u_{\tau i} \sim F_{\tau i}, \text{ subject to } F_{\tau i}(0|x_i) = \tau, \tag{2}$$

where $\eta_\tau(.|\theta)$ is the τth quantile conditional of y_i^* given x_i with the parameters $\theta \in \boldsymbol{\Theta}$, and the random error $u_{\tau i}$ follows a cumulative distribution function $F_{\tau i}$ whose τth quantile conditional on x_i equals zero. Assuming linear model $\eta_\tau(x_i|\theta) = x_i^{'}\beta_\tau(\beta_\tau \in \mathbb{R}^k)$, an intuitive estimator for the Tobit quantile is:

$$\arg min_{\beta_\tau} \sum_{i=1}^{n} \rho_\tau\left(y_i - max\left\{0, x_i^{'}\beta_\tau\right\}\right) \tag{3}$$

where $\rho_\tau(s) = \begin{cases} s\tau, & s \geq 0, \\ s(\tau - 1), & s < 0. \end{cases}$ (4)

Since Eq. (3) is not differentiable at the origin, there is no closed form solution for $\boldsymbol{\beta}$ and the minimization of (2) can be achieved by a linear programming algorithm [17]. However, in high dimensional censored data problems, the algorithm of [17] might be inefficient [18–20]. From a Bayesian perspective, Yu and Stander [19] suggested a Bayesian formulation of Tobit quantile regression employing a skewed Laplace distribution (SLD) for the errors as a "working model". The SLD connects the Bayesian analysis to standard frequentist tobit quantile regression, which proceeds semiparametrically using ρ_τ as a loss function [21,22]. Let u_i follow a $SLD\ (0, \theta, \tau)$, where the parameters are the location, precision, and skewness, respectively. The density of the SLD for the error term (u_i) is written explicitly as

$$f(u_i) = \tau(1-\tau)\theta exp\{-\theta\rho_\tau(u_i)\}. \tag{5}$$

Under the above density, the joint distribution of $\boldsymbol{y}^* = (y_1^*, \dots, y_n^*)'$ given $\boldsymbol{X} = (x_1, \dots, x_n)'$ is

$$f(\boldsymbol{y}^*|\boldsymbol{X}, \beta, \theta) = \tau^n(1-\tau)^n\theta^n exp\left\{-\theta\sum_{i=1}^{n}\rho_\tau\left(y_i^* - x_i^{'}\beta\right)\right\}. \tag{6}$$

2.2. Bayesian Tobit Quantile Regression with Adaptive Lasso and Lasso

The Bayesian approach for the Tobit regression model using the normal density for the residuals and generating β from its conditional posterior distribution using a Gibbs sampler has been proposed by Alhamzawi & Ali [16]. According to the Tobit model, only an unknown subset of predictors is important in the model, so the problem of covariate selection is to select these active covariates. Various approaches to dealing with covariate selection in quantile regression models have been proposed recently. Because of its susceptibility to overfitting issues, the least absolute shrinkage and selection operator (Lasso) method [23,24] has received much attention over the years. The Lasso is obtained for the quantile model by minimizing the following formula:

$$min_\beta \sum_{i=1}^{n} \rho_\tau\left(y_i - max\left\{0, x_i'\beta_\tau\right\}\right) + \lambda \sum_{j=1}^{k}\left|\beta_j\right|, \quad (7)$$

where $\lambda \geq 0$. Rather than minimizing the above regularization problem for the Tobit model, we solve it by constructing a Bayesian framework. Here, $\lambda \sum_{j=1}^{k}\left|\beta_j\right|$ is called the penalty for the selection and estimation of quantile coefficients. Meanwhile, we also consider Tobit quantile regression with the adaptive Lasso penalty, which solves the following [25–27]:

$$min_\beta \sum_{i=1}^{n} \rho_\tau\left(y_i - max\left\{0, x_i'\beta_\tau\right\}\right) + \sum_{j=1}^{k} \lambda_j\left|\beta_j\right|, \quad (8)$$

where λ_j are non-negative adaptive weight and $\lambda_j\left|\beta_j\right|$ is known as the adaptive penalty for selecting and estimating quantile coefficients. As the penalty parameters $(\lambda_j, j = 1, \dots, k)$ increase the Tobit quantile regression coefficients of independent variables are continuously shrunk toward 0 and due to the adaptive penalty form $\left(\sum_{j=1}^{k} \lambda_j\left|\beta_j\right|\right)$, some coefficients of independent variables can be set exactly to 0.

Now, if we assume the error ε, follow the ALD with a scale parameter $\theta(\theta > 0)$ is:

$$f(\varepsilon|\tau) = \frac{\tau(1-\tau)}{\theta} exp\{-\theta^{-1}\rho_\tau(u_i)\}. \quad (9)$$

We also assign a Laplace prior distribution for $\pi(\beta_\tau|\lambda_1, \lambda_2, \dots, \lambda_k) = \prod_{j=1}^{k} \frac{\lambda_j}{2} exp\left\{-\lambda_j|\beta_\tau|\right\}$ on the regression coefficients, then the conditional distribution of the regression coefficients is:

$$P(\beta_\tau|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{\lambda}) \propto exp\left\{-\sum_{i=1}^{n} \rho_\tau\left(y_i - max\left\{0, x_i'\beta_\tau\right\}\right) + \sum_{j=1}^{k} \lambda_j\left|\beta_j\right|\right\}, \quad (10)$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_k)'$. Under this setting, maximizing the posterior estimator of β_τ in Eq. (10) is equivalent to minimizing Eq. (8).

3. SIMULATION STUDIES

In this section, the performance of the Bayesian Tobit Quantile Regression (BTQR) and its modifications, i.e; Bayesian Adaptive Lasso Tobit Quantile Regression (as BALTQR) and

Bayesian Lasso Tobit Quantile Regression (BLTQR) are investigated and compared by simulations. The goal of this simulation study here is to reveal the performance of all three proposed methods and their associated algorithm in recovering the true parameters. The methods are evaluated based on the median of mean absolute deviations, referred to as $MMAD$, and the standard deviation of $MMAD$. $MMAD$ is estimated using this formula: $MMAD = median\left(mean\left(\left|\boldsymbol{X}\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta}^{true}\right|\right)\right)$, where $\widehat{\boldsymbol{\beta}}$ is the posterior mean of $\boldsymbol{\beta}$. $MMAD$ and its standard deviation are estimated over 200 replications. Model selection performance is evaluated based on the credible intervals for the approaches in the comparison.

Table 1. Absolute Bias of Posterior Mean for the Simulated Data in Simulation 1, $\varepsilon \sim N(0,1)$.

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
β^{True}	1	0	1.2	0	0.8	0	0	0
				Quantile $\tau = 0.10$				
BTQR	0.2418	0.1438	**0.0585**	0.2274	0.3349	0.3449	0.1051	0.1204
BALTQR	**0.0132**	0.0445	0.2418	0.1604	0.4176	0.2117	**0.0999**	**0.0641**
BLTQR	0.0176	**0.0432**	0.2786	**0.1579**	**0.4155**	**0.2049**	0.1265	0.0752
				Quantile $\tau = 0.25$				
BTQR	0.3293	0.0175	0.1476	0.2286	**0.1920**	0.3568	0.0357	0.1738
BALTQR	0.1344	**0.0065**	0.0948	**0.1721**	0.2746	0.2235	**0.0349**	0.1079
BLTQR	**0.0723**	0.0347	**0.0689**	0.1735	0.3033	**0.2015**	0.0418	**0.0952**
				Quantile $\tau = 0.50$				
BTQR	0.2034	0.0147	0.0525	0.1273	**0.0765**	0.2184	0.0485	0.1428
BALTQR	0.0928	0.0078	**0.0435**	0.0662	0.1465	0.1567	**0.0292**	0.1050
BLTQR	**0.0543**	**0.0042**	0.0933	**0.0603**	0.1896	**0.1470**	0.0330	**0.1032**
				Quantile $\tau = 0.75$				
BTQR	0.0908	0.0104	0.0302	0.1570	0.0314	0.0565	0.0733	0.1546
BALTQR	0.0428	0.0043	**0.0272**	0.1083	**0.0206**	0.0327	0.0510	0.1125
BLTQR	**0.0095**	**0.0009**	0.0657	**0.0973**	0.0541	**0.0285**	**0.0481**	**0.1020**
				Quantile $\tau = 0.90$				
BTQR	**0.0622**	0.0699	0.1900	0.0942	0.0724	0.0382	0.1262	0.0230
BALTQR	0.1189	0.0607	0.1944	**0.0768**	0.0693	**0.0252**	**0.1101**	0.0192
BLTQR	0.1239	**0.0552**	**0.1452**	0.0808	**0.0460**	0.0337	0.1242	**0.0169**

In this section, eight predictors $x_1, \ldots, x_8$ were simulated independently from $N_8(\mathbf{0}, \boldsymbol{\Sigma})$. We simulated 100 observations from the model $y_i = max\{0, y_i^*\}$, where the response variable y_i^* is generated from $y_i^* = x_i^{'}\boldsymbol{\beta} + e_i$. Three different distributions for e_i were simulated from the following distributions: $N(0,1), t_{(3)}$ distribution with three degrees of freedom, and $Laplace\ (0.5\ , 1)$. Three cases for $\boldsymbol{\beta}$ were considered:

a. Simulation 1 (sparse case): $\boldsymbol{\beta} = (1, 0, 1.2, 0, 0, 8, 0, 0, 0, 0)'$
b. Simulation 2 (very sparse case): $\boldsymbol{\beta} = (3, 0, 0, 0, 0, 0, 0, 0)'$
c. Simulation 3 (dense case): $\boldsymbol{\beta} = (0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80)'$

We consider four choices of θ, 0.10, 0.25, 0.50, and 0.75. Under the three error distributions $N(0,1), t_{(3)}$ and $Laplace\ (0.5\ , 1)$ the censored levels of $\boldsymbol{y}$ were 30%, 50%, and 30%, respectively.

Table 2. Absolute Bias of Posterior Mean for the Simulated Data in Simulation 2, $\varepsilon \sim N(0,1)$.

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
β^{True}	3	0	0	0	0	0	0	0
				Quantile $\tau = 0.10$				
BTQR	0.4670	**0.1066**	0.3714	0.1632	0.2714	0.3854	0.2657	0.1281
BALTQR	0.2685	0.1124	0.2741	**0.1158**	**0.1812**	0.2768	**0.1300**	0.0400
BLTQR	**0.2142**	0.1198	**0.2117**	0.1483	0.2014	**0.2691**	0.1878	**0.0115**
				Quantile $\tau = 0.25$				
BTQR	0.1927	0.0560	0.2282	**0.0221**	0.1400	0.2898	0.1147	0.0270
BALTQR	0.0545	**0.0003**	0.1704	0.0227	0.0545	0.2355	**0.0530**	**0.0227**
BLTQR	**0.0239**	0.0008	**0.1484**	0.0246	**0.0540**	**0.2135**	0.0692	0.0313
				Quantile $\tau = 0.50$				
BTQR	0.0570	0.1078	0.2156	0.0195	0.0335	0.1209	**0.0198**	0.0473
BALTQR	**0.0245**	0.0535	**0.1591**	**0.0116**	0.0239	0.1079	0.0265	0.0291
BLTQR	0.0361	**0.0466**	0.1632	0.0163	**0.0086**	**0.0866**	0.0337	**0.0161**
				Quantile $\tau = 0.75$				
BTQR	0.0789	0.1537	0.1701	0.0020	0.1626	0.0536	0.0586	0.0930
BALTQR	0.0217	0.1108	**0.1364**	**0.0006**	**0.1195**	0.0488	**0.0340**	0.0843
BLTQR	**0.0205**	**0.1038**	0.1439	0.0129	0.1375	**0.0511**	0.0384	**0.0725**
				Quantile $\tau = 0.90$				
BTQR	**0.1882**	0.1488	0.1909	0.0211	0.2972	0.0753	0.0174	0.1073
BALTQR	0.2114	0.1232	0.1820	0.0278	**0.2552**	0.0717	**0.0098**	**0.0727**
BLTQR	0.2311	**0.1206**	**0.1652**	**0.0105**	0.2786	**0.0710**	0.0171	0.0872

For each error distribution, we simulate 200 data sets assuming the sample size is $n = 100$. We fit the models at four different quantiles, $\tau = 0.10, 0.25, 0.50, 0.75$, and 0.95. The MCMC algorithms are run for 17,000 iterations, discarding the first 2000 as burn-in. Methods are evaluated based on the smallest value of absolute bias of parameter models. The results for each simulation at selected quantiles for each parameter from Normal distribution are presented in Tables 1, 2, and 3. Other results are saved by the author provided by request.

Table 3. Absolute Bias of Posterior Mean for the Simulated Data in Simulation 3, $\varepsilon \sim N(0,1)$

Method	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
β^{True}	0.80	0.80	0.80	0.80	0.80	0.80	0.80	0.80
Quantile $\tau = 0.10$								
BTQR	0.0476	0.1471	0.1022	0.4039	0.2952	**0.1167**	0.1646	0.0533
BALTQR	0.0015	0.1665	**0.0357**	0.4063	**0.2388**	0.1304	0.0981	**0.0159**
BLTQR	**0.0008**	**0.0914**	0.0882	**0.3886**	0.3051	0.1339	**0.0899**	0.0187
Quantile $\tau = 0.25$								
BTQR	0.1062	0.1474	0.2044	0.3016	0.2697	0.1998	0.1854	0.0046
BALTQR	0.0463	0.1550	0.2063	**0.2821**	0.2606	**0.1567**	**0.1040**	0.0344
BLTQR	**0.0336**	**0.0527**	**0.1015**	0.3015	**0.1770**	0.1954	0.1059	0.0347
Quantile $\tau = 0.50$								
BTQR	0.1294	0.0585	0.1660	0.0346	0.1118	0.1635	0.1920	**0.0054**
BALTQR	0.0858	0.0604	0.1692	0.0294	**0.0403**	**0.1023**	0.0905	0.0546
BLTQR	**0.0704**	**0.0089**	**0.0830**	**0.0050**	0.1221	0.1640	**0.0891**	0.0524
Quantile $\tau = 0.75$								
BTQR	0.1142	0.1240	**0.1143**	0.0511	0.0862	0.1430	0.1269	0.0542
BALTQR	0.1031	**0.0680**	0.2114	**0.0174**	0.0437	0.0872	0.1278	**0.0228**
BLTQR	**0.0932**	0.1204	0.2075	0.0181	**0.0352**	**0.0847**	**0.1256**	0.0544
Quantile $\tau = 0.90$								
BTQR	0.0303	0.1930	**0.0720**	0.0675	0.2351	0.1934	0.0768	0.1640
BALTQR	0.0047	0.1997	0.1704	0.0269	0.2025	0.1502	**0.0261**	**0.0277**
BLTQR	**0.0044**	**0.1718**	0.1353	**0.0040**	**0.1883**	**0.1354**	0.0389	0.0660

Clearly, the biases due to the three approaches are more or less the same (very similar values). However, the BLTQR generally behaves much better than the other approaches (BTQR and BALTQR) in terms of absolute bias. Across the three simulations, it can be seen that the absolute bias obtained from the BLTQR method is much smaller at selected quantiles than the competing approaches. Most noticeably, when $\tau = 0.75$ the absolute bias generated by all three methods for all parameters is much smaller than the absolute bias at a smaller quantile. But, for the most extreme quantile ($\tau = 0.90$), the values of absolute bias are generally larger than quantile $\tau = 0.75$. We then check for the results of the median of mean absolute deviations (MMAD) and the standard deviations (SD) of the MAD as presented in Table 4.

Table 4. MADs and Standard Deviations (SD) of MADs for Simulations 1, $\beta = (1, 0, 1.2, 0, 0.8, 0, 0, 0, 0)'$

Quantile τth	Methods	MMAD (s.d)		
		$\varepsilon \sim \mathbb{N}(0,1)$	ε~t (n-1)	$\varepsilon \sim Laplace(0.5\ , 1)$
$\tau = 0.10$	BTQR	1.881 (0.286)	3.886 (0.801)	2.044 (0.298)
	BALTQR	1.712 (0.244)	3.133 (0.626)	1.869 (0.253)
	BLTQR	**1.666 (0.240)**	**2.897 (0.553)**	**1.865 (0.234)**
$\tau = 0.25$	BTQR	1.423 (0.231)	3.234 (0.687)	1.504 (0.230)
	BALTQR	1.310 (0.196)	2.596 (0.532)	1.394 (0.201)
	BLTQR	**1.270 (0.192)**	**2.377 (0.471)**	**1.370 (0.184)**
$\tau = 0.50$	BTQR	0.883 (0.142)	1.974 (0.495)	0.930 (0.149)
	BALTQR	0.838 (0.114)	1.510 (0.348)	0.887 (0.122)
	BLTQR	**0.822 (0.112)**	**1.303 (0.300)**	**0.870 (0.115)**
$\tau = 0.75$	BTQR	0.730 (0.104)	1.413 (0.317)	0.805 (0.112)
	BALTQR	0.710 (0.102)	1.221 (0.246)	0.771 (0.101)
	BLTQR	**0.689 (0.099)**	**1.124 (0.224)**	**0.757 (0.096)**
$\tau = 0.90$	BTQR	0.882 (0.130)	2.561 (0.517)	1.193 (0.230)
	BALTQR	0.865 (0.124)	2.462 (0.484)	1.130 (0.206)
	BLTQR	**0.862 (0.119)**	**2.328 (0.435)**	**1.099 (0.198)**

From Tables 4, 5, and, 6 we can observe that for MMADs and SD criteria, the method Bayesian Lasso Tobit Quantile Regression (BLTQR) generally performs better than the other methods for all the distributions under consideration. In Simulation 1 and 2, the BLTQR method has the smallest MMAD in all 15 simulation setups. In Simulation 3, the BLTqr method has the smallest MMAD in 14 out of 15 simulation setups. In general, Bayesian Lasso quantile regression performs well compared to two other methods, BTQR and BALTQR.

Table 5. MADs and Standard Deviations (SD) of MADs for Simulations 2, $\beta = (3, 0, 0, 0, 0, 0, 0, 0)'$

Quantile τth	Methods	MAD (SD)		
		$\varepsilon \sim \mathbb{N}(0,1)$	ε~t (n-1)	$\varepsilon \sim Laplace(0.5\,, 1)$
$\tau = 0.10$	BTQR	2.103 (0.296)	3.923 (0.755)	2.490 (0.410)
	BALTQR	1.984 (0.268)	3.462 (0.592)	2.322 (0.357)
	BLTQR	**1.936 (0.254)**	**3.287 (0.539)**	**2.312 (0.340)**
$\tau = 0.25$	BTQR	1.591 (0.245)	3.254 (0.665)	1.754 (0.284)
	BALTQR	1.470 (0.214)	2.814 (0.530)	1.654 (0.258)
	BLTQR	**1.446 (0.204)**	**2.649 (0.462)**	**1.643 (0.241)**
$\tau = 0.50$	BTQR	1.032 (0.159)	2.060 (0.495)	1.207 (0.232)
	BALTQR	1.009 (0.133)	1.764 (0.360)	1.146 (0.186)
	BLTQR	**1.000 (0.130)**	**1.634 (0.304)**	**1.119 (0.177)**
$\tau = 0.75$	BTQR	0.943 (0.125)	1.749 (0.374)	1.099 (0.185)
	BALTQR	0.929 (0.128)	1.605 (0.336)	1.051 (0.169)
	BLTQR	**0.913 (0.122)**	**1.496 (0.297)**	**1.029 (0.169)**
$\tau = 0.90$	BTQR	1.125 (0.156)	2.517 (0.492)	1.400 (0.249)
	BALTQR	1.095 (0.149)	2.374 (0.433)	1.335 (0.222)
	BLTQR	**1.088 (0.147)**	**2.282 (0.417)**	**1.307 (0.212)**

4. MODELING STUNTING RATE

All three methods then are applied to construct a model of the stunting rate in West Sumatra, Indonesia. The data obtained from Health Office in several districts and cities in West Sumatra is regarding the determinants of stunting in August 2021 and February 2022. The response variable represents the stunting rate of 3534 stunting children (in cm) from August 2021 to February 2022, the summary statistics for the response are provided in Figure 1. The mean stunting rate is 3.42 cm and the standard deviation is 3.758. Since the data is related to stunting children, some children have zero height gain, thus censored here is about zero.

While this study assumed ten predictor variables as factors influencing the stunting rate based on previous studies. The indicator variables consist of nine categorical variables, as presented in Table 7, and one numerical variable, i.e., birth weight (X_2). After fitting the linear regression model using the ordinary least square (OLS) method, it is necessary to check whether the normality assumption of the residuals is held or not. To do so, the Chi-square test was performed and the test shows that the normality assumption is not held with $p - value = 3.56 \times 10^{-8}$. Additionally, the histogram and the Q-Q plot show that the distribution of the residual may be poor. Similar to the simulation studies, all three methods are then compared : BTQR, BALQR, and BTQR. For each method, the MAD is recorded, where the $MAD = \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{n}$.

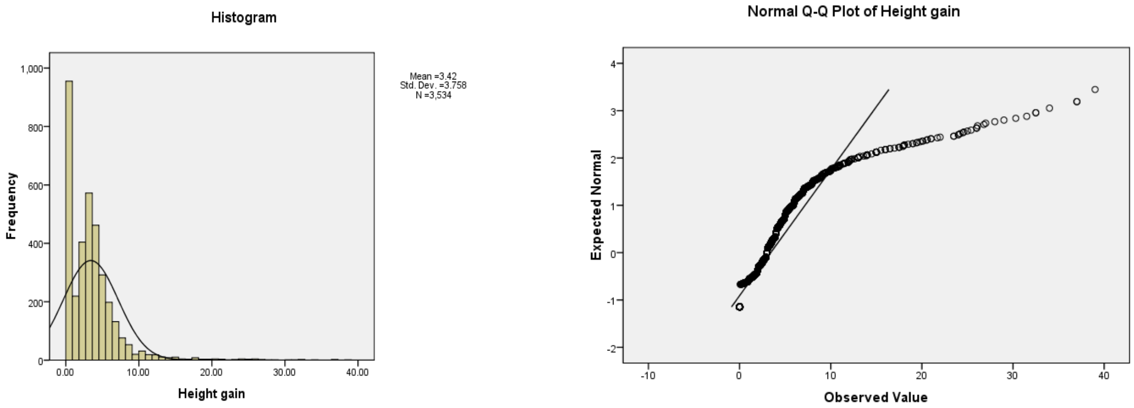
Table 6. MADs and Standard Deviations (SD) of MADs for Simulations 3

$\beta = (0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80, 0.80)'$

Quantile τth	Methods	MAD (s.d)		
		$\varepsilon \sim \mathbb{N}(0,1)$	ε~t (n-1)	$\varepsilon \sim Laplace(0.5, 1)$
$\tau = 0.10$	BTQR	1.991 (0.322)	3.901 (0.749)	2.293 (0.431)
	BALTQR	1.847 (0.289)	3.179 (0.605)	2.058 (0.347)
	BLTQR	**1.842 (0.287)**	**3.102 (0.538)**	**2.032 (0.350)**
$\tau = 0.25$	BTQR	1.501 (0.252)	3.229 (0.660)	1.674 (0.307)
	BALTQR	1.375 (0.224)	2.644 (0.516)	1.507 (0.258)
	BLTQR	**1.371 (0.224)**	**2.513 (0.459)**	**1.497 (0.251)**
$\tau = 0.50$	BTQR	0.972 (0.164)	1.957 (0.472)	1.066 (0.198)
	BALTQR	0.919 **(0.135)**	1.550 (0.329)	0.984 **(0.153)**
	BLTQR	**0.917** (0.140)	**1.407 (0.294)**	**0.969** (0.158)
$\tau = 0.75$	BTQR	0.827 (0.139)	1.534 (0.340)	0.903 (0.159)
	BALTQR	0.789 (0.129)	1.312 (0.250)	0.843 (0.137)
	BLTQR	**0.776 (0.128)**	**1.234 (0.235)**	**0.836 (0.136)**
$\tau = 0.90$	BTQR	1.002 (0.153)	2.536 (0.495)	1.244 (0.238)
	BALTQR	**0.966** (0.146)	2.432 (0.467)	1.188 (0.206)
	BLTQR	0.968 (**0.145**)	**2.316 (0.420)**	**1.183 (0.206)**

In this data, we considered four quantiles, these were 0.25, 0.50, 0.75, and 0.90. We ran algorithms for 30.000 iterations, discarding the first 1000 as burn-in. The results of the parameter estimated and the width of the 95% confidence interval at each quantile for all three methods are provided in Table 8.

After the three methods, BTQR achieves the best prediction accuracy. The width of the 95% confidence interval for BLTQR is lower than that of BTQR and BALQR when $\tau = 0.10, 0.25, 0.50,$ and 0.75. Besides, BLTQR does not almost as well as BTQR and BALQR when $\tau = 0.90$. Table 9 informs us that for the stunting rate dataset, it can be seen that the MAD of the BTQR was about 2.08% and 0.86% lower than that of BTQR and BALQR when $\tau = 0.25$, respectively.



(a) (b)

Figure 1. Normality assumption checking. (a) Histogram of the OLS residuals for the stunting rate dataset, (b) Q−Q plot of the OLS residuals for the stunting rate dataset

5. CONCLUSIONS

In this article, we construct the model of stunting rate in selected cities and districts in West Sumatra, Indonesia using Bayesian Tobit quantile regression and its generalized methods. This study compares the result of BTQR, BALTQR, and BLTQR methods using a simulation study and an empirical study. The Bayesian Tobit quantile regression and its generalized methods not only accommodate the messy attributes of the stunting rate response but also provides a complete picture of the covariate effects on the stunting rate distribution. Furthermore, it successfully selects and models the important categorical predictors. Our findings are summarized below. First, exclusive breastfeeding affects the stunting rate only at the middle quantile, at $\tau = 30, 50$. Exclusive breastfeeding seems not to be an important factor claimed for the high stunting rate. Comorbidity tend to be an important factor in stunting rates not only at lower quantiles but also at higher quantiles. The analysis of simulation studies and stunting rate dataset shows strong support for the use of Bayesian Lasso Tobit quantile regression to inference for Tobit quantile regression models. The proposed method generally behaves much better than the other approaches in terms of a width of 95% Bayes confidence interval and absolute bias. The work presented in this paper opens the door to new research directions for subset selection and coefficient estimation in quantile regression models with right-censored or interval-censored responses.

Table 7. Summary Statistics of Category Variables

Variable	Frequency	Percentage (%)
Sex (X_1)		
Female	1550	43.9
Male	1984	56.1
Exclusive breastfeeding (X_3)		
Yes	2393	67.7
No	1141	32.3
Healthy toilet (X_4)		
Yes	2089	59.1
No	1445	40.9
Clean water (X_5)		
Yes	2897	82.0
No	637	18.0
Health and social security (X_6)		
Yes	1270	35.9
No	2263	64.1
Worms (X_7)		
Yes	76	2.2
No	3458	97.8
Immunization (X_8)		
Yes	2645	74.8
No	888	25.1
Smoking (X_9)		
Yes	3169	89.7
No	365	10.3
Comorbidity (X_{10})		
Yes	240	6.8
No	3294	93.2

Table 8. Estimates of Model Parameters For The Stunting Data Set

Independent Variable	Bayesian Tobit QR		Bayesian *Adaptive* LASSO Tobit QR		Bayesian LASSO Tobit QR	
	$\hat{\beta}$	Width of 95% CI	$\hat{\beta}$	Width of 95% CI	$\hat{\beta}$	Width of 95% CI
$\tau = 0.10$						
Intercept	-1.2687	3.8435	-1.1763	0.6120	-1.1045	**0.5688**
X_1	0.0361	1.4296	0.0260	0.6240	0.0182	**0.6063**
X_2	0.1826	0.9825	0.1450	0.6594	0.1243	**0.6154**

X_3	-0.3477*	1.0713	-0.3064*	**0.5078**	-0.2817	0.6978
X_4	-0.3136	1.2417	-0.2541	0.5434	-0.2310	**0.5221**
X_5	0.4737	1.4367	0.3971	0.5654	0.3701	**0.5324**
X_6	-0.1363	1.2311	-0.1008	0.5123	-0.0944	**0.5023**
X_7	-0.2830	3.6124	-0.1464	0.5435	-0.1006	**0.5135**
X_8	-0.2244	1.3107	-0.1944	0.5100	-0.1729	**0.5020**
X_9	-0.3404	1.9674	-0.2767	**0.5311**	-0.2614	0.6123
X_{10}	-0.4988*	5.3213	-0.4212*	0.9845	-0.3541*	**0.9012**
$\tau = 0.25$						
Intercept	1.5957*	0.5641	-0.0578	0.5991	-0.0378	**0.5377**
X_1	-0.0619	0.7694	-0.0215	0.5343	-0.0283	**0.4935**
X_2	0.1433	0.6095	0.1098	0.5295	0.0899	**0.4925**
X_3	-1.1233*	1.2828	-0.5028*	**0.6006**	-0.4755*	0.6093
X_4	-0.8537*	0.6480	-0.3704*	0.6939	-0.3183	**0.6692**
X_5	1.8984*	0.4883	0.6802*	0.6495	0.6243*	**0.6450**
X_6	-0.5210	1.2545	-0.2249	0.5711	-0.2143	**0.5553**
X_7	-1.5750	3.0846	-0.1444	1.2772	-0.0960	**0.9864**
X_8	-0.8007*	1.3449	-0.3466*	0.5834	-0.3365*	**0.5695**
X_9	-1.5706*	1.5987	-0.4666*	**0.4847**	-0.4550*	0.4911
X_{10}	-2.3296*	1.1956	-0.5424*	0.7752	-0.4827*	**0.7141**
$\tau = 0.50$						
Intercept	2.9126*	3.7640	2.3559*	0.6500	2.3403*	**0.6280**
X_1	0.0697	1.1307	0.0320	0.5106	0.0267	**0.4853**
X_2	0.0253	0.7921	0.0181	0.6831	0.0131	**0.6335**
X_3	-0.6038*	1.1028	-0.2623*	0.5078	-0.2533*	**0.4977**
X_4	-0.3603	1.2037	-0.1418	0.5592	-0.1235	**0.5342**
X_5	1.1556*	1.9496	0.3928*	0.7362	0.3728*	**0.7223**
X_6	-0.2836	1.2647	-0.1110	0.5690	-0.1018	**0.5475**
X_7	-1.1941	3.3788	-0.1351	1.2954	-0.1058	**0.9136**
X_8	-0.3353	1.1343	-0.1376	0.4672	-0.1329	**0.4519**
X_9	-0.6103	1.6955	-0.1708	0.4902	-0.1607	**0.4773**
X_{10}	-1.4470	4.3602	-0.3178	1.0286	-0.2971	**0.9785**
$\tau = 0.75$						
Intercept	4.1110*	3.7435	4.3342*	0.7151	4.2692*	**0.6977**
X_1	0.1472	1.2296	0.0655	0.5599	0.0590	**0.5443**
X_2	0.1317	0.8825	0.1051	0.7680	0.0944	**0.7486**
X_3	-0.3291	1.2713	-0.1291	0.5476	-0.1300	**0.5328**
X_4	-0.6301	1.3417	-0.2613	0.6229	-0.2449	**0.6162**
X_5	0.6141	1.5367	0.1964	0.5637	0.1917	**0.5610**
X_6	-0.1377	1.2460	-0.0621	0.5566	-0.0589	**0.5465**
X_7	0.3871	4.5054	0.0357	1.1257	0.0237	**1.0560**

X_8	0.1336	1.2907	0.0325	0.5180	0.0259	**0.5119**
X_9	-0.4827	1.9484	-0.1327	0.5439	-0.1342	**0.5382**
X_{10}	3.8867*	5.4623	0.8300*	1.4124	0.7100*	**1.3604**
$\tau = 0.90$						
Intercept	5.8011*	4.3136	6.7327	0.9600	6.7022*	**0.9572**
X_1	0.1916	1.6364	0.0684	**0.7508**	0.0736	0.7618
X_2	0.4978	1.1053	0.3494	**1.0001**	0.3671	1.0077
X_3	0.0060	1.6422	-0.0093	**0.6922**	-0.0115	0.7133
X_4	-1.7255*	1.9737	-0.7378	1.0268	-0.7330*	**0.9941**
X_5	1.5331*	1.8645	0.5280	0.7699	0.5312*	**0.7639**
X_6	-0.5466	1.6593	-0.2025	**0.7603**	-0.2087	0.7684
X_7	1.1635	6.0405	0.0736	**1.1796**	0.0849	1.2411
X_8	-0.0051	1.6999	-0.0099	**0.6909**	-0.0092	0.7128
X_9	-1.2002	2.9985	-0.2826	**0.8527**	-0.3000	0.8753
X_{10}	5.9451*	3.7166	1.4563	0.9367	1.4322*	**0.9302**

Table 9. MAD, MSE, and RMSE

	MAD	MSE	RMSE
$\tau = 0.10$			
BTQR	3.8183	25.4302	5.0428
BALTQR	3.7834	24.1319	4.9124
BLTQR	**3.6516**	**20.3420**	**4.5102**
$\tau = 0.25$			
BTQR	3.1183	20.4606	4.5233
BALTQR	3.0811	20.1629	4.4903
BLTQR	**3.0546**	**19.9430**	**4.4658**
$\tau = 0.50$			
BTQR	2.1415	11.6511	3.4134
BALTQR	**2.1389**	**11.5841**	**3.4035**
BLTQR	2.1395	11.5916	3.4046
$\tau = 0.75$			
BTQR	2.6486	13.1781	3.6302
BALTQR	2.6251	12.9311	3.5960
BLTQR	**2.5989**	**12.7069**	**3.5647**
$\tau = 0.90$			
BTQR	4.2395	26.0259	5.1016
BALTQR	4.1829	25.2056	5.0205
BLTQR	**4.1584**	**24.9575**	**4.9957**

ACKNOWLEDGEMENT

This research was funded by Directorate of Resources Directorate General of Higher Education, Ministry of Education, Culture, and Research and Technology of Indonesia, in accordance with Contract Number T/23/UN.16.17/PT.01.03/PDKN-Kesehatan/2022.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] N. Fenske, J. Burns, T. Hothorn, E.A. Rehfuess, Understanding child stunting in India: a comprehensive analysis of socio-economic, nutritional and environmental determinants using additive quantile regression, PLoS ONE. 8 (2013), e78692. https://doi.org/10.1371/journal.pone.0078692.

[2] R. Martorell, A. Zongrone, Intergenerational influences on child growth and undernutrition, Paediatric Perinatal Epid. 26 (2012), 302-314. https://doi.org/10.1111/j.1365-3016.2012.01298.x.

[3] Y.R. Yue, H.G. Hong, Bayesian Tobit quantile regression model for medical expenditure panel survey data, Stat. Model. 12 (2012), 323-346. https://doi.org/10.1177/1471082x1201200402.

[4] R. Alhamzawi, Bayesian Elastic net Tobit quantile regression, Commun. Stat. - Simul. Comput. 45 (2014), 2409-2427. https://doi.org/10.1080/03610918.2014.904341.

[5] R. Alhamzawi, H.T.M. Ali, The Bayesian elastic net regression, Commun. Stat. - Simul. Comput. 47 (2017), 1168-1178. https://doi.org/10.1080/03610918.2017.1307399.

[6] R. Alhamzawi, K. Yu, Bayesian Tobit quantile regression usingg-prior distribution with ridge parameter, J. Stat. Comput. Simul. 85 (2014), 2903-2918. https://doi.org/10.1080/00949655.2014.945449.

[7] R. Alhamzawi, Model selection in quantile regression models, J. Appl. Stat. 42 (2014), 445-458. https://doi.org/10.1080/02664763.2014.959905.

[8] R. Alhamzawi, A. Alhamzawi, H.T. Mohammad Ali, New Gibbs sampling methods for Bayesian regularized quantile regression, Computers Biol. Med. 110 (2019), 52-65. https://doi.org/10.1016/j.compbiomed.2019.05.011.

[9] M.K. Cowles, B.P. Carlin, J.E. Connett, Bayesian Tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness, J. Amer. Stat. Assoc. 91 (1996), 86-98. https://doi.org/10.1080/01621459.1996.10476666.

[10] H. Hashem, V. Vinciotti, R. Alhamzawi, K. Yu, Quantile regression with group lasso for classification, Adv. Data Anal. Classif. 10 (2015), 375-390. https://doi.org/10.1007/s11634-015-0206-x.

[11] H. Mallick, N. Yi, A new Bayesian lasso, Stat. Interface. 7 (2014), 571-582. https://doi.org/10.4310/sii.2014.v7.n4.a12.

[12] A. Alhamzawi, Tobit regression with Lasso penalty, J. Phys.: Conf. Ser. 1664 (2020), 012046. https://doi.org/10.1088/1742-6596/1664/1/012046.

[13] T.R. Mokalla, V.V. Rao Mendu, Application of quantile regression to examine changes in the distribution of Height for Age (HAZ) of Indian children aged 0–36 months using four rounds of NFHS data, PLoS ONE. 17 (2022), e0265877. https://doi.org/10.1371/journal.pone.0265877.

[14] F. Yanuar, H. Yozza, A. Zetra, Modified quantile regression for modeling the low birth weight, Front. Appl. Math. Stat. 8 (2022), 890028. https://doi.org/10.3389/fams.2022.890028.

[15] F. Yanuar, A. Zetra, The performance of bayesian analysis in structural equation modelling to construct the health behaviour during pandemic COVID-19, Pak. J. Stat. Oper. Res. (2022), 575-587. https://doi.org/10.18187/pjsor.v18i3.4096.

[16] R. Alhamzawi, H.T.M. Ali, Bayesian Tobit quantile regression with penalty, Commun. Stat. - Simul. Comput. 47 (2017), 1739-1750. https://doi.org/10.1080/03610918.2017.1323224.

[17] R. Koenker, V. d'Orey, Remark AS R92: A remark on algorithm AS 229: computing dual regression quantiles and regression rank scores, Appl. Stat. 43 (1994), 410-414. https://doi.org/10.2307/2986030.

[18] B. Santos, H. Bolfarine, Bayesian quantile regression analysis for continuous data with a discrete component at zero, Stat. Model. 18 (2017), 73-93. https://doi.org/10.1177/1471082x17719633.

[19] K. Yu, J. Stander, Bayesian analysis of a Tobit quantile regression model, J. Econometrics. 137 (2007), 260-276. https://doi.org/10.1016/j.jeconom.2005.10.002.

[20] K. Zhao, H. Lian, Bayesian Tobit quantile regression with single-index models, J. Stat. Comput. Simul. 85 (2014), 1247-1263. https://doi.org/10.1080/00949655.2013.873041.

[21] Y. Feng, Y. Chen, X. He, Bayesian quantile regression with approximate likelihood, Bernoulli. 21 (2015), 832-850. https://doi.org/10.3150/13-bej589.

[22] M.A. Taddy, A. Kottas, A Bayesian nonparametric approach to inference for quantile regression, J. Business Econ. Stat. 28 (2010), 357-369. https://doi.org/10.1198/jbes.2009.07331.

[23] R. Tibshirani, Regression Shrinkage and Selection Via the Lasso, J. R. Stat. Soc.: Ser. B (Methodol.) 58 (1996), 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[24] F. Yanuar, A.S. Deva, A. Zetra, et al. Length of hospital stay model of COVID-19 patients with quantile Bayesian with penalty LASSO, Commun. Math. Biol. Neurosci. 2023 (2023), 23. https://doi.org/10.28919/cmbn/7881.

[25] R. Alhamzawi, K. Yu, D.F. Benoit, Bayesian adaptive Lasso quantile regression, Stat. Model. 12 (2012), 279-297. https://doi.org/10.1177/1471082x1101200304.

[26] F. Yanuar, A. Zetra, C. Muharisa, et al. Bayesian quantile regression method to construct the low birth weight model, J. Phys.: Conf. Ser. 1245 (2019), 012044. https://doi.org/10.1088/1742-6596/1245/1/012044.

[27] F. Yanuar, A. Zetra, Length-of-stay of hospitalized COVID-19 patients using bootstrap quantile regression, IAENG Int. J. Appl. Math. 51 (2021), 40.