Lattice Journal : *Journal of Mathematics Education and Applied*
Vol. 6 No. 1, Januari-Juni 2026 pp. 16-29
e-ISSN: 2798-4621 p-ISSN: 2798-6470
DOI:http://dx.doi.org/10.30983/lattice.v6i1.11113

A Comparative Study of Quantile Regression and Bayesian Quantile Regression for Modelling Tuberculosis Cases in West Sumatra

Lilis Harianti Hasibuan[1]*
[1]Universitas Islam Negeri Imam Bonjol, Padang, Indonesia

Informasi Artikel

Diterima Redaksi: 12 April 2026
Revisi Akhir: 18 Juni 2026
Diterbitkan *Online*: 30 Juni 2026

Kata Kunci

Bayesian
Quantile Regression
Tuberculosis (TB)
Markoc Chain Monte Carlo (MCMC)
Mean Square Error (MSE)

Korespondensi

E-mail: lilisharianti@uinib.ac.id

A B S T R A C T

This study examines modeling the number of TB in West Sumatra with Quantile regression and Bayesian quantile regression. Quantile regression are statistical methods that look at the pattern of data distribution in each quantile. The Bayesian method is combined with quantile regression to produce more robust parameters. Both methods are used to model data on the number of Tuberculosis (TB) sufferer in West Sumatra. The data used amounted to 76 data obtained from the Central Bureau of Statistics (BPS) of West Sumatra. The dependent variable, the number of TB sufferer, is assumed to be influenced by five independent variables, namely Percentage of Poor Population, Number of People with HIV, Percentage of Population Age 35-44 who Smoke, Number of Nurses, and Population Density. Error data is assumed to follow an asymmetric Laplace distribution. Parameter estimation in quantile regression is performed on each quantile while the Bayesian method estimates parameters by finding the mean of the posterior distribution. This study found the best modeling for Number TB at quantile 0.50 with an MSE value of 11,942.

1. Introduction

Tuberculosis (TB) An infectious disease that continues to be a global health issue, including in Indonesia [1]. According to data from the Indonesian Ministry of Health, Indonesia is the country with the second largest number of TB cases in the world after India [2]. West Sumatra as one of the provinces in Indonesia also faces great challenges in controlling this disease. According to 2019 data from the Central Bureau of Statistics (BPS) of West Sumatra, tuberculosis (TB) ranks as the second most prevalent disease in Indonesia, with a total of 11,790 reported cases. The case detection rate (CDR) in West Sumatra remains relatively low at 48.1% [3]. The prevalence of TB in West Sumatra is quite high, and many factors influence its spread, ranging from people's lifestyles, sanitation conditions, to the availability of health facilities [4]. The increasing number of TB cases also influenced by the percentage of poor people [5]. The Central Bureau of Statistics reports that the poverty rate in West Sumatra in March 2023 was around 6.69%, which means that it still needs attention from the government. In addition to the percentage of poor people [5], the number of TB sufferer is also influenced by the number of HIV sufferer [6]. Not only that, the number of TB sufferer is also influenced by factors such as smoking [7], the number of health workers and population density [8].

Tuberculosis is caused by infection with Mycobacterium tuberculosis, a bacterium that primarily affects the lungs and is transmitted through airborne droplets when an infected individual coughs, sneezes, or speaks. At the global level, the persistence of tuberculosis is strongly associated with socioeconomic determinants such as poverty, malnutrition, overcrowded living conditions, and limited access to quality healthcare services. These factors weaken the immune system and facilitate transmission, particularly in densely populated and low-income regions. In addition, co-infection with HIV significantly increases the risk of developing active tuberculosis, as it compromises the body's immune defenses. The emergence of drug-resistant tuberculosis due to incomplete or improper treatment further exacerbates the global burden of the disease.

In Indonesia, tuberculosis remains a major public health challenge driven by a combination of structural, environmental, and behavioral factors. High population density, especially in urban areas, accelerates the spread of the disease, while limited awareness and delayed diagnosis contribute to continued transmission within communities. Risk factors such as smoking, poor ventilation in housing, and inadequate nutrition further increase susceptibility to infection. Moreover, disparities in healthcare access and adherence to treatment regimens play a critical role in sustaining the incidence of tuberculosis. As a result, comprehensive strategies that address both medical and social determinants are essential to effectively control and reduce the burden of tuberculosis in Indonesia.

Classical linear regression is a fundamental statistical method used to model the relationship between a dependent variable and one or more independent variables by estimating the conditional mean of the response[9]. This approach assumes a linear relationship between variables and relies on several key assumptions, including normality, homoskedasticity, and independence of the error terms[10], [11], [12]. By minimizing the sum of squared residuals, ordinary least squares (OLS) provide parameter estimates that best fit the observed data. Despite its simplicity and wide applicability, classical regression may be limited in capturing heterogeneous effects across different parts of the distribution, particularly in the presence of outliers or non-constant variance [13].

Regression analysis is a widely used statistical method for assessing the relationship between one or more explanatory (independent) variables and a response (dependent) variable [14],[15],[10],[16]. In regression analysis, there are four assumptions that must be met by the data that we will model, namely the assumptions of normality, homoscedasticity, no multicollinearity and no autocorrelation [17]. Sometimes, however, field data may violate these assumptions, which can be overcome with quantile regression [18].

Quantile regression is an extension of classical regression that models the relationship between independent variables and a dependent variable at different points (quantiles) of the conditional distribution, rather than focusing solely on the mean[19], [20]. This approach allows researchers to examine how the effects of explanatory variables vary across the lower, median, and upper parts of the distribution. As a result, quantile regression provides a more comprehensive understanding of the data structure, particularly when the distribution is skewed or when extreme values play an important role. Unlike ordinary least squares (OLS), which minimizes the sum of squared residuals, quantile regression estimates parameters by minimizing an asymmetric loss function that assigns different weights to positive and negative deviations. This makes the method more robust to outliers and suitable for handling heteroskedastic data without requiring strict distributional assumptions[21], [22]. Consequently, quantile regression is widely applied in various fields where capturing distributional heterogeneity is essential for accurate modeling and interpretation [23].

Quantile regression provides a framework for assessing the influence of independent variables on a dependent variable across different segments (quantiles) of the distribution [21],[24].
In the context of modeling the number of TB cases, quantile regression can provide more comprehensive information about the relationship between predictor factors (such as population density, access to health services, nutritional status, and environmental factors) and the number of TB cases at various quantiles. The advantage of quantile regression over ordinary regression is its ability to handle non-normal distributions of data or data with many outliers (e.g., areas with a very high number of TB cases). Quantile regression allows researchers to model a more complete distribution of TB cases, focusing not only on the mean value, but also on the distribution at the bottom or top of the data. It turns out that regression has its drawbacks, as it requires a fairly large amount of data. Bayesian quantile regression extends quantile regression by using a Bayesian approach, which allows the determination of probability distributions for model parameters with less data [25]. Bayesian studied by [26],[27],[28],[29],[30] widely studied by researchers.
This approach is particularly useful in situations where the data is not large enough or the parameter distributions cannot be estimated directly. This method provides advantages in terms

http://dx.doi.org/10.30983/11113

of modeling uncertainty and accounting for the probabilistic distribution of the parameters used in quantile regression. In the context of modeling the number of TB sufferer in West Sumatra, Bayesian quantile regression allows for accounting for uncertainty in the model estimates as well as incorporating a prior information prior knowledge about the distribution of risk factors or the distribution of TB in a particular region. This provides more robust results that can be interpreted in terms of probability distributions, which is very useful in health policy decision-making. With proper modeling, authorities can plan and assess more effective TB control policies. For example, modeling that identifies high-risk areas can help local governments allocate resources more appropriately, or provide a clearer picture of the effectiveness of implemented health interventions. Quantile regression and Bayesian quantile regression can help provide a better understanding of the distribution of TB cases and the factors that influence them, both under average and extreme conditions.
Despite the extensive use of regression-based approaches in modeling tuberculosis (TB) incidence, most existing studies primarily rely on classical methods such as ordinary least squares (OLS), which focus on the conditional mean and often fail to capture heterogeneity across different levels of the distribution. While quantile regression (QR) has been introduced as a more flexible alternative capable of analyzing the effects of predictors at various quantiles, its application in TB modeling remains relatively limited. In particular, few studies have explored how socioeconomic, behavioral, and healthcare-related factors influence TB incidence differently across lower, middle, and upper quantiles, leaving a gap in understanding the distributional dynamics of the disease.

Furthermore, although Bayesian Quantile Regression (BQR) offers additional advantages such as incorporating prior information and providing more robust estimation under uncertainty comparative studies between QR and BQR in the context of TB modeling are still scarce. Existing research rarely evaluates the relative performance of these two approaches using consistent datasets and evaluation metrics, such as mean squared error (MSE), across multiple quantiles. This lack of comprehensive comparison creates a methodological gap, particularly in identifying which approach yields more accurate and stable estimates for TB data. Therefore, this study aims to address this gap by systematically comparing QR and BQR to provide deeper insights into their effectiveness in modeling tuberculosis incidence across the entire distribution.

2. Methodology

This study proceeds to the methodology section, which outlines the analytical framework used to compare Quantile Regression (QR) and Bayesian Quantile Regression (BQR) in modeling tuberculosis cases. This section describes the data sources, variable specifications, and the modeling procedures applied under both approaches, including the estimation techniques and underlying assumptions. By systematically presenting these components, the methodology aims to ensure clarity, reproducibility, and a rigorous basis for evaluating the comparative performance of QR and BQR in capturing the distributional characteristics of tuberculosis incidence.

2.1. Data Set

This dataset section describes the source, characteristics, and structure of the data used in the study, including information on the observed variables, the number of observations, and the data collection period. It also outlines the initial data preprocessing steps, such as handling missing values, detecting outliers, and applying necessary transformations to ensure data quality prior to further analysis. This description aims to provide a clear understanding of the empirical foundation of the study and to support transparency and reproducibility.

The Data used in this study are Number of People with TB from BPS Sumatra Barat. The data used was 76 data from 2019 until 2023. Variable X_1(Percentage of Poor Population), X_2(Number of People with HIV) ,X_3 (Percentage of Population Age 35-44 who Smoke), X_4 (Number of Nurses), X_5(Population Density). Dependent Variable Y is Number of People with TB. variable information can be seen in the following table 1:

http://dx.doi.org/10.30983/11113

Table1. Variable research, description and Unit

Variable	Description	Unit	Type Data
Y	Number of People with TB	Person	Numeric
X_1	Percentage of Poor Population	Percent (%)	Numeric
X_2	Number of People with HIV	Person	Numeric
X_3	Percentage of Population Age 35-44 who Smoke	Percent (%)	Numeric
X_4	Number of Nurses	Percent (%)	Numeric
X_5	Population Density	Per (km²	Numeric

2.2. Quantile Regression (QR)

Quantile regression is a statistical method used to model the relationship between independent variables and a dependent variable at various points (quantiles) of the dependent variable's distribution [10], [31]. Unlike ordinary linear regression, which estimates only the mean of the dependent variable, quantile regression enables researchers to examine how independent variables influence different parts of the distribution, such as the median (0.5 quantile), lower quantiles (e.g., 0.25), or upper quantiles (e.g., 0.75). This approach provides a more comprehensive understanding of the relationships between variables, particularly when the data are not symmetrically distributed or contain outliers[32].

Moreover, quantile regression is robust to outliers and heteroskedasticity, as it does not rely on the assumption of normally distributed errors required in classical linear regression. It employs an asymmetric loss function to estimate parameters at each specified quantile. Therefore, quantile regression is highly valuable across various research fields, including economics, health sciences, and social sciences, especially when the objective is to understand how the effects of independent variables vary across the entire distribution of the dependent variable, rather than focusing solely on its mean.

Suppose $\boldsymbol{y} = (y_1, y_2, \dots, y_n)'$is a vector the dependent variable and $\boldsymbol{x} = (x_1, x_2, \dots, x_k)'$ is the independent variable, then the τ^{th} is quantile, for $0 < \tau < 1$, if n samples and k predictors for $i = 1,2, \dots, n$ is formula below [33]:

$$y_i = \beta_{0\tau} + \beta_{1\tau}x_{i1} + \beta_{2\tau}x_{i2} + \dots + \beta_{k\tau}x_{ik} + \varepsilon_i \quad (2.\ 1)$$

With $\boldsymbol{\beta}(\tau)$ as the parameter vector and $\boldsymbol{\varepsilon}$ as the residual vector. The conditional quantile function τ^{th} is $Q_\tau(\boldsymbol{y}|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}_\tau$, to estimate $\widehat{\boldsymbol{\beta}_\tau}$ ekuivalent with minimize the equation below[19]:

$$\sum_{i=1}^{n} \rho_\tau(y_i - \boldsymbol{x}_i'\boldsymbol{\beta}_\tau). \quad (2.2)$$

for $\rho_\tau(u) = u(\tau - I(u < 0))$is the *loss function* with the equation [31]:

$$\rho_\tau(\epsilon) = \varepsilon\big(\tau I(\varepsilon \geq 0) - (1-\tau)I(\varepsilon < 0)\big). \quad (2.\ 3)$$

$I\ (\ .\ )$ is an indicator function, which has a value of 1 when $I\ (\ .\ )$ is true dan 0 otherwise.

2.3. Bayesian Quantile Regression (BQR)

Furthermore, quantile regression does not rely on the assumptions of homoskedasticity or normally distributed errors, making it more flexible in handling data with non-constant variance. Parameter estimation in quantile regression is carried out by minimizing an asymmetric loss function, which assigns different weights to positive and negative deviations relative to the specified quantile. This approach enables a more comprehensive analysis of the data distribution[34].

http://dx.doi.org/10.30983/11113

The Bayesian approach is a statistical inference framework that combines prior information with evidence from observed data (likelihood) to produce a posterior distribution of the parameters [22]. In this framework, parameters are treated as random variables with probability distributions rather than fixed but unknown quantities, as in classical methods. This perspective allows for a more natural incorporation of uncertainty and prior knowledge into the estimation process [35].

One of the key advantages of the Bayesian approach is its ability to accommodate complex models and quantify parameter uncertainty through posterior distributions. When closed-form posterior distributions are unavailable, computational methods such as Markov Chain Monte Carlo (MCMC) can be used to generate samples from the posterior distribution [36]. Bayesian inference summarizes parameter uncertainty using credible intervals, which should be distinguished from confidence intervals in the frequentist framework. A 95% confidence interval is constructed using a procedure that, under repeated sampling, would contain the true fixed parameter in 95% of repeated samples. In contrast, a 95% Bayesian credible interval represents an interval containing 95% of the posterior probability of the parameter, conditional on the observed data and the specified prior distribution. Accordingly, this study reports 95% confidence intervals (CI) for Quantile Regression (QR) and 95% credible intervals (CrI) for Bayesian Quantile Regression (BQR).

Bayesian Quantile Regression is an integration of quantile regression and the Bayesian framework to estimate model parameters at different quantiles of the distribution. In this approach, the likelihood function is typically constructed using the Asymmetric Laplace Distribution (ALD), which is closely related to the loss function used in quantile regression. By combining the prior and likelihood, the posterior distribution of the parameters at each quantile can be obtained.

Thus, Bayesian Quantile Regression offers several advantages, including flexibility, robustness to outliers, and the ability to incorporate prior information into the estimation process. This method is particularly valuable in various research fields when the objective is to gain deeper insights into how the effects of independent variables vary across the entire distribution of the dependent variable, supported by a comprehensive inferential framework. Bayesian quantile regression (BQR) is a combination of quantile regression method with Bayesian approach The Bayesian concept is to minimize the loss function, which is the same as maximizing the likelihood function of the Asymmetric Laplace Distribution (ALD) [37],[38],[39]. Function of error (ε) with ALD distribution with likelihood density function $f(\epsilon)$, namely:

$$f_\tau(\epsilon) = \tau(1-\tau)\exp\left(-\rho_\tau(\varepsilon)\right) \qquad (2.4)$$

With $0 < \tau < 1$ and ρ_τ denotes the loss function, in which the input corresponds to the estimation errors and *I(.)* is an indicator function. Suppose Z is exponentially distributed $(Z \sim exp\ (1))$ and U follows a standard normal distribution $U \sim N\ (0,1)$. If ε is an ALD distributed random variable then ε can be written as:

$$\varepsilon = \theta z + pu\sqrt{z}. \qquad (2.5)$$

With $\theta = \frac{1-2\tau}{(1-\tau)\tau}$ and $p^2 = \frac{2}{(1-\tau)\tau}$ [38]. From the equation (2.5) , the likelihood function used in parameter $\boldsymbol{\beta}$ estimation for the τ^{th} quantile in the Bayesian quantile regression analysis is formulated in equation (2.6) as follows [40],[13]:

$$L(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{\sigma},\boldsymbol{v}) = (\prod_{i=1}^{n}(\sigma v_i)^{-\frac{1}{2}})\left(\exp\left(-\frac{y-(\boldsymbol{x}_i'\boldsymbol{\beta}_\tau+\theta v_i))^2}{2p^2\sigma v_i}\right)\right) \qquad (2.6)$$

Where $\sigma > 0$ and $v_i = \sigma z_i$ spreading exp *(σ)* distribution. Where $\beta_\tau \sim N(b_0, B_0)$, $v_i \sim \exp(\sigma)$, and $\sigma \sim IG(a, b)$is the prior. Estimation for mean of distribution posterior is obtained as follows:

$$(\boldsymbol{\beta}|, \sigma, v, \boldsymbol{y}) \sim N[(\boldsymbol{B_0^{-1}} + \boldsymbol{x_i}(p^2\sigma v)^{-1}\boldsymbol{x_i'})^{-1}(\boldsymbol{B_0^{-1}b_0} + \boldsymbol{x_i}(p^2\sigma v)^{-1}x_i')^{-1}y_i - x_i(p^2\sigma v)^{-1}\theta v_i), (\boldsymbol{B_0^{-1}} + \boldsymbol{x_i}(p^2\sigma v)^{-1}\boldsymbol{x_i'})^{-1}] \quad (2.7)$$

,

$$(v_i|\boldsymbol{\beta}, \sigma, \boldsymbol{y}) \sim GIG(\tfrac{1}{2}, \left(\frac{(y_i - x_i'\beta_\tau)^2}{p^2\sigma}\right), (\tfrac{2}{\sigma} + \tfrac{\theta^2}{p^2\sigma})),$$

$$(\sigma|\boldsymbol{\beta}, \boldsymbol{v}, \boldsymbol{y}) \sim IG(a + \tfrac{3n}{2}, (b + \textstyle\sum_{i=1}^n v_i + \sum_{i=1}^n + \left(\frac{y_i - (x_i'\beta_\tau + \theta v_i))^2}{2p^2\sigma}\right)).$$

Bayesian Quantile Regression provides a powerful and flexible analytical framework for understanding the relationship between independent and dependent variables across different points of the distribution [22],[36]. By combining the robustness of quantile regression to outliers with the Bayesian approach's ability to incorporate uncertainty and prior information, this method yields more comprehensive and informative estimates than conventional approaches [41].

Moreover, the ability of Bayesian Quantile Regression to model complex data that do not satisfy classical assumptions makes it highly relevant for a wide range of empirical research contexts. This approach also enables deeper interpretation of results through posterior distributions and credible intervals, allowing researchers to gain clearer insights into the variation of effects at each analyzed quantile. Therefore, the application of Bayesian Quantile Regression not only enhances statistical analysis but also opens opportunities for further methodological development. In the future, this method can be integrated with other techniques, such as regularization or more efficient computational approaches, to improve estimation accuracy and efficiency, as well as to expand its applicability across diverse research fields.

2.4. Bayesian Implementation and Model Evaluation

To estimate the Bayesian Quantile Regression (BQR) model, the likelihood function was specified using the Asymmetric Laplace Distribution (ALD), which is widely adopted in Bayesian quantile regression because its likelihood is directly linked to the quantile loss function. Let y_i denote the response variable and $\boldsymbol{x_i}$ the vector of explanatory variables. The conditional distribution of the response at quantile τ is assumed to follow an ALD with location parameter $\boldsymbol{x_i'\beta_\tau}$ scale parameter σ and quantile parameter τ.
For the regression coefficients, weakly informative normal prior distributions were assumed [42]:

$$\beta \sim N(0,1000),\ j=0,1,\ldots,p,$$

to reflect limited prior information while allowing the observed data to dominate the posterior inference. The scale parameter was assigned a proper positive prior to ensure posterior propriety.
Posterior inference was obtained using Markov Chain Monte Carlo (MCMC) simulation. The posterior mean was used as the Bayesian point estimator for each regression coefficient, while the uncertainty of the estimates was summarized using 95% Bayesian credible intervals. Unlike classical quantile regression, which reports confidence intervals, Bayesian quantile regression provides credible intervals that directly represent the posterior probability of the parameters.
Convergence of the posterior samples was assessed visually through trace plots and posterior density plots for each parameter. Stable trace plots without systematic trends and unimodal posterior density functions indicate satisfactory convergence of the MCMC algorithm [43]. An example of the convergence diagnostics for the median quantile 0.05 is presented in Figure 3.1.
To compare the predictive performance of Quantile Regression (QR) and Bayesian Quantile Regression (BQR), the Mean Squared Error (MSE) was calculated for each quantile [40]. Since the objective of quantile regression is to estimate conditional quantiles rather than conditional means, the interpretation of model performance is discussed together with parameter stability and interval estimation obtained from both approaches.

http://dx.doi.org/10.30983/11113

3. Result and Discussion

This stage, parameter estimation will be carried out on both methods. Quantile regression by minimizing the loss function, while Bayesian quantile regression by estimating the mean value of the posterior distribution. Tabel 3.1 shows the performance from two methods. Estimation result, width confidence interval width of 95% for quantile 0.05; 0.25; 0,50; 0.75; 0,90.

Table 2. Estimated Value of β for each quantile – τ using QR and BQR

Independent Variables	QR		BQR	
	Parameter $(\widehat{\beta})$	QR CI Width 95%	Parameter $(\widehat{\beta})$	BQR CrI Width 95%
$\boldsymbol{\tau = 0.05}$				
Intercept	-106	422	-92.482	**181**
X_1(Percentage of Poor Population)	-4.750*	**5.100**	-3.598	6.900
X_2(Number of People with HIV)	0.002	**0.001**	-0.039	0.140
X_3 (Percentage of Population Age 35-44 who Smoke),	5.310	19.200	4.504	**8.095**
X_4 (Number of Nurses)	0.001	0.003	0,010	**0.002**
X_5(Population Density)	-0.004	0.005	-0.003	**0.001**
$\boldsymbol{\tau = 0.25}$				
Intercept	-323*	**298**	-330*	300
X_1(Percentage of Poor Population),	-7.420*	8,670	-7.35	**17.400**
X_2(Number of People with HIV)	0.008	**0.001**	-0.046	21.800
X_3 (Percentage of Population Age 35-44 who Smoke),	15.100*	18.800	15.200	**12.300**
X_4 (Number of Nurses)	0.001	0.178	0.027	**0.106**
X_5(Population Density)	-0.004	**0.001**	-0.048	0.002
$\boldsymbol{\tau = 0.50}$				
Intercept	-481,674*	**275,811**	-484*	336
X_1(Percentage of Poor Population),	-8.528*	18.780	-9.500*	**26.100**
X_2(Number of People with HIV)	-0.049	1.418	-0.056	**0.298**
X_3 (Percentage of Population Age 35-44 who Smoke),	21.929*	8.492	22.300*	**6.200**
X_4 (Number of Nurses)	0.054*	0.478	0.057*	**0.167**
X_5(Population Density)	-0.008	0.025	-0.0083	**0.001**
$\boldsymbol{\tau = 0.75}$				
Intercept	-681*	**491**	-763*	510
X_1(Percentage of Poor Population),	-17.700	31.100	-13.400	37.300
X_2(Number of People with HIV)	-0.235	**0.001**	-0.177	0.633
X_3 (Percentage of Population Age 35-44 who Smoke),	31.500*	**15.600**	33.600*	17500
X_4 (Number of Nurses)	0.023*	1.450	0.205*	**0.441**
X_5(Population Density)	-0.012	0.003	-0.009	**0.001**
$\boldsymbol{\tau = 0.90}$				
Intercept	-191*	0.502	-177*	**0.501**
X_1(Percentage of Poor Population),	-20.700	**2.860**	-19.700	5.310
X_2(Number of People with HIV)	-1.450	0.002	-1.300	**0.001**
X_3 (Percentage of Population Age 35-44 who Smoke),	70.500*	**0.596**	65.900*	22.200
X_4 (Number of Nurses)	1,240*	4.002	1.200*	**0.332**

http://dx.doi.org/10.30983/11113

X_5(Population Density)	0.004	1.003	0.041	**0.092**
***significantcy $\alpha = 0,05$.**				

Note: *CI denotes the 95% confidence interval for QR estimates, whereas CrI denotes the 95% Bayesian credible interval for BQR posterior estimates. The asterisk (*) indicates statistical significance for QR or posterior evidence for BQR according to the inferential criterion specified in the analysis.*

In the table 2 above, at the lower quantile (τ = 0.05), the model captures the dynamics within the group with the lowest response values. The variable percentage of poor population (X_1) shows a negative and significant coefficient in QR (−4.750*), indicating that an increase in poverty is associated with a decrease in the dependent variable at the lower tail of the distribution. Although the coefficient remains negative in BQR (−3.598), the wider uncertainty (larger confidence interval) suggests a less precise estimate compared to QR. Other variables, such as the number of people with HIV (X_2) and the number of nurses (X_4), exhibit relatively small effects. However, in BQR, X_4 shows a notably narrow credible interval (0.002), indicating a more stable estimate. Overall, at this quantile, structural factors such as poverty appear to play a more dominant role in explaining variation at the lower extreme than healthcare-related variables.

At the lower-middle quantile (τ = 0.25), the pattern of relationships becomes more complex. The variable X_1 (poverty) remains significant and negatively associated in both QR (−7.420*) and BQR (−7.35), with a larger magnitude than at τ = 0.05, suggesting that the impact of poverty intensifies at higher points of the distribution. Additionally, X_3 (percentage of population aged 35–44 who smoke) becomes significant in QR (15.100*), indicating that behavioral factors begin to play an important role in explaining variation in the dependent variable. BQR supports this finding with similar estimates but more controlled uncertainty. Meanwhile, variables such as X_2 (HIV) and X_5 (population density) show instability (very narrow confidence intervals in QR but much wider in BQR), suggesting sensitivity to the estimation approach. Overall, it can be concluded that at higher quantiles, the determinants are no longer purely structural but shift toward a combination of social and behavioral factors, with BQR tending to provide more robust estimates in capturing parameter uncertainty.

At the median quantile (τ = 0.50), the model reflects the central tendency of the distribution, where several variables exhibit strong and consistent effects. The percentage of poor population (X_1) remains negative and significant in both QR (−8.528*) and BQR (−9.500*), indicating that poverty continues to play a substantial role in reducing the outcome variable even at the median level. In contrast, the percentage of population aged 35–44 who smoke (X_3) shows a positive and significant effect (QR: 21.929*, BQR: 22.300*), suggesting that behavioral factors become increasingly influential in shaping the response variable. Additionally, the number of nurses (X_4) is significant and positive in both approaches, implying that improved healthcare capacity contributes positively to the outcome. Notably, BQR generally produces narrower credible intervals (e.g., X_2 and X_4), indicating more stable and reliable parameter estimates compared to QR.

At the upper quantile (τ = 0.75), the influence of variables becomes more differentiated, highlighting effects in the higher end of the distribution. The percentage of population aged 35–44 who smoke (X_3) emerges as the most dominant factor, with a strong positive and significant coefficient in both QR (31.500*) and BQR (33.600*), indicating that behavioral risk factors are particularly critical in higher outcome levels. Meanwhile, the percentage of poor population (X_1), although still negative, loses its statistical significance, suggesting that poverty plays a less decisive role at the upper tail. The number of nurses (X_4) remains positively significant, especially in BQR (0.205*), reinforcing the importance of healthcare resources. Furthermore, BQR again demonstrates tighter credible intervals (e.g., X_4 and X_5), reflecting greater estimation precision. Overall, these results indicate a clear shift from structural factors (such as poverty) toward behavioral and service-related factors (such as smoking and healthcare availability) as one moves to higher quantiles, with BQR providing more robust inference in capturing parameter uncertainty.

http://dx.doi.org/10.30983/11113

At the upper quantile (τ = 0.90), the model highlights determinants associated with the highest levels of the response variable, where the effects become more selective and pronounced. The percentage of population aged 35–44 who smoke (X_3) emerges as the most dominant factor, with a large positive and statistically significant coefficient in both QR (70.500*) and BQR (65.900*). This indicates that smoking behavior plays a critical role in driving higher outcomes, particularly in regions already situated at the upper end of the distribution. The magnitude of this effect is substantially larger compared to lower quantiles, suggesting an escalating influence of behavioral risk factors as the outcome increases.

In addition, the number of nurses (X_4) remains positively significant in both models (QR: 1.240*; BQR: 1.200*), implying that healthcare capacity continues to contribute meaningfully even at the highest quantile. In contrast, the percentage of poor population (X_1) and the number of people with HIV (X_2) exhibit negative but statistically insignificant coefficients, indicating that structural and epidemiological factors are less influential at this level. Similarly, population density (X_5) shows no significant effect in either approach. Notably, BQR provides tighter credible intervals for several variables (e.g., X_2, X_4, and X_5), reflecting more stable and reliable parameter estimates. Overall, these findings suggest a clear shift in dominance toward behavioral and healthcare-related factors at the upper tail, with Bayesian estimation offering improved robustness in capturing parameter uncertainty.

It can be informed that the variables that significantly affect the number of TB suffefer in West Sumatra for both methods are the percentage of pupulation age 35-44, the number of nurses, the percentage of poor pupulation in the quantile 0,50; 0,75; and 0,90. The variable number of people with HIV and population density is not significant in all quantiles in the two methods. Meanwhile, when viewed from the estimation of 95% confidence interval using BQR produces a smaller interval width than QR. Then for the selection of the best model of modeling the number of TB suffefer, Mean Square Error (MSE) is used. The MSE value in each quantile using both methods can be seen in Table 3.below:

Table 3. MSE Value QR and BQR

Quantile	**MSE**	
	QR	**BQR**
0,05	20,751	21,295
0,25	17,438	17,227
0,5	15,065	**11,942**
0,75	12,228	12,232
0,90	44,053	42,354

In Table 3, at the lower quantile (τ = 0.05), the MSE values show that QR (20.751) is slightly lower than BQR (21.295). This indicates that, at the extreme lower tail of the distribution, QR provides marginally more accurate predictions than BQR. From a sharper perspective, this suggests that in lower-end conditions often dominated by noise or random variation the classical approach is still able to capture the data pattern efficiently without the added complexity of Bayesian priors. In other words, the advantages of the Bayesian framework have not yet fully emerged at this quantile.

At the lower-middle quantile (τ = 0.25), a subtle but important shift occurs, where BQR (17.227) slightly outperforms QR (17.438). Although the difference is small, it signals the beginning of Bayesian advantage as we move away from the extreme lower tail. This reflects a transitional phase in which the data become sufficiently informative for the prior structure in BQR to contribute to improved estimation accuracy and stability.

At the median quantile (τ = 0.50), the difference becomes much more pronounced, with BQR (11.942) significantly lower than QR (15.065). This provides strong evidence that, at the center of

http://dx.doi.org/10.30983/11113

the distribution where most observations are concentrated the Bayesian approach delivers substantially better predictive performance. Analytically, this indicates that the combination of likelihood and prior information leads to more precise estimates, particularly when the underlying data structure is relatively stable. This also highlights the superiority of BQR in capturing the core pattern of the data.

At the upper quantile (τ = 0.75), the performance of both methods is nearly identical (QR: 12.228; BQR: 12.232), suggesting no clear dominance of one approach over the other. This implies that, in this part of the distribution, both methods are equally capable of modeling the data pattern. However, the inherent stability of BQR remains an added advantage, even if it is not strongly reflected in the MSE values.

At the extreme upper quantile (τ = 0.90), BQR (42.354) once again outperforms QR (44.053). This is particularly important because the upper tail of the distribution is typically characterized by high variability and potential outliers. The improved performance of BQR in this region indicates that the Bayesian approach is more robust in handling extreme uncertainty and is able to produce more accurate predictions under such conditions. Overall, this pattern clearly demonstrates that while QR performs slightly better at the lowest quantile, BQR consistently shows superior performance across most parts of the distribution especially in the middle and upper quantiles making it a more reliable choice for analyses requiring both stability and high predictive accuracy.

From the Table 3 above, BQR has the smallest MSE value compared to quantile regression (QR), namely at quantile 0,5 in the BRQ method of 11,942. So that the best model chosen is:

$$\hat{Y} = -484 - 9{,}500X_1 - 0{,}056X_2 + 22{,}300X_3 + 0{,}057X_4 - 0{,}0083X_5. \quad (3.1)$$

Based Equation (3.1) tells us that if the Percentage of Poor Population increases by 1 percent (%), the number of TB sufferer in the 0,50 quantile will decrease by about 9,5 or 10 people. If the Number of People with HIV decreases by 1 person, the number of TB sufferer decreases by about 0,056 percent, if the Percentage of Population Age 35-44 who Smoke increases by 1 percent, the number of TB sufferer increases by about 22,3 people. Then continued if the Number of Nurses increases by 1 percent below the 0,5 quantile, the number of TB sufferers decreases by about 0,057 sufferers or almost no TB sufferers. Furthermore, the number of TB sufferer will decrease below the 0,5 quantile by 0,0083 people if the population density increases by 1 person. Then the convergence of each parameter of the independent variable estimated from both methods can be seen from the trace plot and density plot results. The figure below is only shown for the 0.5 quantile. The picture is presented in **Figure 3.1** below:

Figure 3.1 shows density plots are often used to describe the posterior distribution of model parameters after a sampling process and the purpose of trace plots is to check the convergence process of a sampling algorithm to visualize the behavior of the sample set throughout the iterations. Parameter in quantile 0,5 is convergent. From figure 3.1 also give information that trace plot and density plot give model has satisfied the criterion of convergence to normal distribution. All statistical analyses were performed using R statistical software. Quantile Regression (QR) and Bayesian Quantile Regression (BQR) were implemented using the corresponding R packages. The Bayesian estimation was conducted through MCMC sampling, while posterior convergence was evaluated using trace plots and posterior density plots. Model performance measures and graphical diagnostics were also obtained using R.

http://dx.doi.org/10.30983/11113

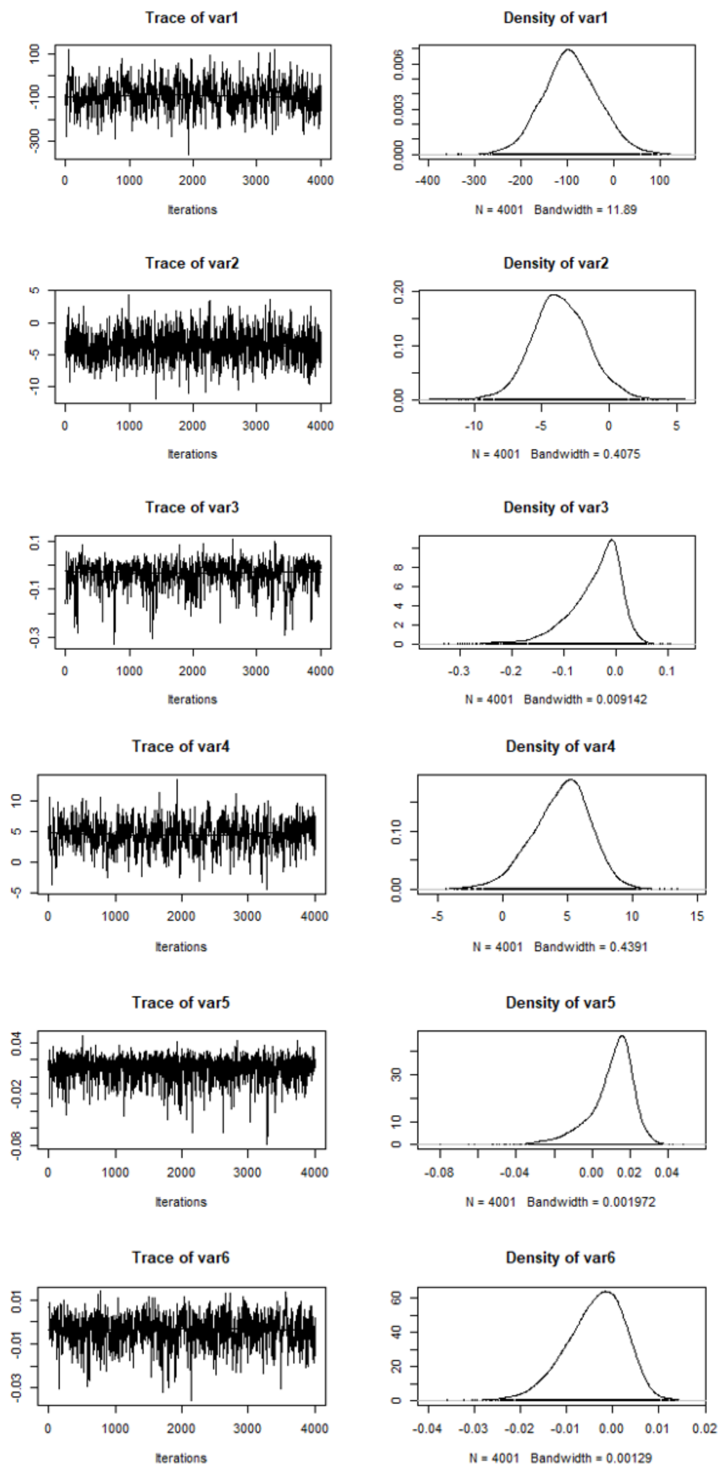

Figure 3.1. Trace Plot and density of parameter $\boldsymbol{\beta}$ for τ = 0,5

4. Conclusion

This study compared Quantile Regression (QR) and Bayesian Quantile Regression (BQR) for modeling the number of tuberculosis cases in West Sumatra across different levels of the conditional distribution. The results demonstrate that the effects of the investigated determinants are heterogeneous across quantiles, indicating that factors associated with TB burden cannot be fully characterized by focusing only on the conditional mean. The percentage of the population aged 35–44 years who smoke, the number of nurses, and the percentage of the poor population showed important associations with TB cases at several quantiles. Among the evaluated models, BQR at quantile 0.05 produced the lowest MSE (11.942), indicating better predictive performance at the median of the conditional distribution.

From a practical perspective, the findings highlight the importance of considering socioeconomic, behavioral, and healthcare-related characteristics simultaneously in regional TB control strategies. The strong association of smoking with higher levels of TB cases emphasizes the relevance of integrating tobacco-control interventions into TB prevention programs. The variation in the effects of poverty and healthcare resources across quantiles further suggests that TB interventions should be adapted to the characteristics and burden levels of different regions rather than relying on a uniform strategy. From a statistical perspective, BQR provides a useful framework for capturing heterogeneous relationships while explicitly quantifying parameter uncertainty through posterior inference. These findings may therefore support more targeted and evidence-based TB control planning in West Sumatra. So, the model obtained from modeling the number of TB in West Sumatra is:

$$\hat{Y} = -484 - 9.500X_1 - 0.056X_2 + 22.300X_3 + 0.057X_4 - 0.0083X_5.$$

References

[1] R. O. Nyarko, A. Prakash, N. Kumar, P. Saha, and R. Kumar, “Tuberculosis a globalized disease,” *Asian J. Pharm. Res. Dev.*, vol. 9, no. 1, pp. 198–201, 2021.

[2] J. R. A. Tangkilisan, F. L. F. G. Langi, and A. F. C. Kalesaran, “Angka penemuan kasus tuberkulosis paru di indonesia tahun 2015-2018,” *KESMAS*, vol. 9, no. 5, 2020.

[3] Y. Masnarivan and A. Haq, “Pemodelan Faktor Risiko Tuberkulosis Paru Di Sumatera Barat,” *JAMBI Med. JOURNAL" J. Kedokt. dan Kesehatan"*, vol. 10, no. 1, pp. 68–80, 2022.

[4] E. D. Kartiningrum, L. T. Alberta, D. Puspitaningsih, and Y. L. H. Kusuma, “Konsep dasar keperawatan komunitas,” *E-b. Penerbit STIKes Majapahit*, pp. 1–152, 2017.

[5] N. N. Rizki Noventy, “PENGARUH INDEKS PEMBANGUNAN MANUSIA, JUMLAH PENDUDUK MISKIN, PRODUK DOMESTIK REGIONAL BRUTO DAN INDEKS KUALITAS UDARA TERHADAP ANGKA TUBERCULOSIS DI INDONESIA (Studi Kasus 6 Provinsi di Pulau Jawa),” 2023.

[6] H. Hikmawati, S. Muhsina, and S. P. Amandaty, “Knowledge and Perceptions of Coastal Adolescents on TB and HIV/AIDS,” *J. Ilm. Kesehat. Sandi Husada*, vol. 12, no. 2, pp. 488–493, 2023.

[7] V. Maurya, K. Vijayan V, and A. Shah, “Smoking and tuberculosis: an association overlooked,” *Int. J. Tuberc. Lung Dis.*, vol. 6, no. 11, pp. 942–951, 2002.

[8] S. C. Smeltzer, D. S. Suddarth, B. G. Bare, L. S. Brunner, and S. C. Smeltzer, *Brunner and Suddarth’s Textbook of Medical-surgical Nursing: Units 8-16*. Lippincott Williams & Wilkins, 2000.

[9] L. H. Hasibuan and S. Musthofa, “Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang,” *JOSTECH J. Sci. Technol.*, vol. 2, no. 1, pp. 85–95, Mar. 2022, doi: 10.15548/jostech.v2i1.3802.

[10] L. H. Hasibuan, F. Yanuar, D. Devianto, and M. Maiyastri, “Quantile Regression Analysis; Simulation Study With Violation of Normality Assumption,” *JOSTECH J. Sci. Technol.*, vol. 4, no. 2, pp. 133–142, 2024, doi: https://dx.doi.org/10.15548/jostech.v4i2.9643.

[11] A. S. Sholih, L. H. Hasibuan, and I. D. Rianjaya, “Analisis Regresi Logistik Ordinal Terhadap Faktor-Faktor Yang Mempengaruhi Predikat Kelulusan Mahasiswa Sarjana Uin Imam Bonjol Padang,” *MAp (Mathematics Appl. J.*, vol. 6, no. 2, pp. 99–110, 2024, doi: https://doi.org/10.15548/map.v6i2.10075.

[12] A. Apriona, L. H. Hasibuan, and E. Asfa’ani, “MODEL REGRESI KUANTIL SPLINE KUADRATIK DALAM MENGANALISIS PERUBAHAN LAMA RAWAT INAP PASIEN DEMAM BERDARAH DENGUE (DBD) DI RSUP M. DJAMIL PADANG,” *MAp (Mathematics Appl. J.*, vol. 7, no. 2, pp. 96–107, 2025.

[13] L. H. Hasibuan, F. Yanuar, D. Devianto, and Maiyastri, “Bayesian binary quantile

http://dx.doi.org/10.30983/11113

regression adaptive Lasso for modelling hypertension: a bootstrap based consistency analysis," *Commun. Stat. Case Stud. Data Anal. Appl.*, pp. 1–24, 2026, doi: https://doi.org/10.1080/23737484.2026.2671866.
[14] L. Harianti Hasibuan, S. Musthofa, P. Studi Matematika, and U. Imam Bonjol Padang, "Journal of Science and Technology Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Harga Beras di Kota Padang," *J. Sci. Technol.*, vol. 2, no. 1, pp. 85–95, 2022.
[15] L. H. Hasibuan, F. Yanuar, V. P. Harahap, and L. Qalbi, "Performance Quantile Regression and Bayesian Quantile Regression in Dealing with Non-normal Errors (Case Study on Simulated Data)," *Numer. J. Mat. Pendidik. Mat.*, vol. 8, 2024, doi: https://doi.org/10.25217/numerical.v8i2.4922.
[16] L. H. Hasibuan, D. M. Putri, and M. Jannah, "Simple Linear Regression Method to Predict Cooking Oil Prices in the Time of Covid-19," *Logaritma J. Ilmu-ilmu Pendidik. dan Sains*, vol. 10, no. 01, pp. 81–94, 2022.
[17] L. H. Hasibuan and S. Musthofa, "Penerapan Metode Regresi Ridge untuk Mengatasi Masalah Multikolinearitas pada Kasus Tingkat Pengangguran Terbuka di Pulau Sumatera," *JOSTECH J. Sci. Technol.*, vol. 6, no. 1, pp. 97–110, 2026.
[18] K. Yu and J. Stander, "Bayesian analysis of a Tobit quantile regression model," *J. Econom.*, vol. 137, no. 1, pp. 260–276, 2007.
[19] C. Davino, M. Furno, and D. Vistocco, *Quantile regression: theory and applications*, vol. 988. John Wiley & Sons, 2013.
[20] F. Yanuar, L. Hasnah, and D. Devianto, "The Simulation Study to Test the Performance of Quantile Regression Method With Heteroscedastic Error Variance," vol. 5, no. May, pp. 36–41, 2017.
[21] L. H. Hasibuan, F. Yanuar, D. Devianto, M. Maiyastri, and R. Rudiyanto, "Comparison Between Bayesian Quantile Regression And Bayesian Lasso Quantile Regression For Modeling Poverty Line With Presence Of Heteroscedasticity In West Sumatra," *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 19, no. 3, pp. 1587–1596, 2025, doi: https://doi.org/10.30598/barekengvol19iss3pp1587-1596.
[22] Z. Ma, Z. Han, and M. Wang, "Bayesian variable selection and estimation in binary quantile regression using global-local shrinkage priors," *Commun. Stat. Comput.*, pp. 1–15, 2023, doi: https://doi.org/10.1080/03610918.2023.2293638.
[23] F. Yanuar, A. S. Deva, and A. Zetra, "Length of hospital stay model of COVID-19 patients with quantile Bayesian with penalty LASSO," *Commun. Math. Biol. Neurosci.*, vol. 2023, p. Article-ID, 2023.
[24] L. H. Hasibuan, F. Yanuar, and D. Devianto, "Hypertension Status Model with Bayesian Adaptive LASSO Binary Quantile Regression Using Asymmetric Laplace Distribution.," *IAENG Int. J. Comput. Sci.*, vol. 52, no. 8, 2025.
[25] G. Kobayashi and H. Kozumi, "Bayesian analysis of quantile regression for censored dynamic panel data," *Comput. Stat.*, vol. 27, pp. 359–380, 2012, doi: https://doi.org/10.1007/s00180-011-0263-3.
[26] F. Yanuar, C. Mukti, U. Andalas, and K. L. Manis, "Komparasi Model Pertambahan Tinggi Badan Balita Stunting Dengan Metode Regresi Kuantil dan Regresi Kuantil Bayesian," vol. 20, no. 2, pp. 165–177, 2023.
[27] R. Alhamzawi, "Model selection in quantile regression models," *J. Appl. Stat.*, vol. 42, no. 2, pp. 445–458, Feb. 2015, doi: 10.1080/02664763.2014.959905.
[28] A. Afnanda, M. Maiyastri, and D. Devianto, "Model Volatilitas Return Index Saham Syariah Indonesia Melalui Pendekatan Bayesian Markov Switching GARCH," *Lattice J. J. Math. Educ. Appl.*, vol. 4, no. 1, pp. 14–26, 2024, doi: 10.30983/lattice.v4i1.8381.
[29] V. Acnesya and D. Devianto, "Pemodelan Return Harga Emas Dengan Pendekatan Inferensi Bayesian ARFIMA," vol. 5, no. 1, pp. 1–13, 2025.
[30] R. Anwar, L. Rassiyanti, and K. Kunci, "Analisis Komparasi Model Peramalan Prophet Dan Arima Dalam Memprediksi Harga Saham Penutupan PT ANTM," vol. 5, no. 1, pp. 57–74, 2025.
[31] R. Koenker and K. F. Hallock, "Quantile regression," *J. Econ. Perspect.*, vol. 15, no. 4, pp. 143–156, 2001, doi: DOI: 10.1257/jep.15.4.143.

http://dx.doi.org/10.30983/11113

[32] A. Khadka *et al.*, "Quantile regressions as a tool to evaluate how an exposure shifts and reshapes the outcome distribution: a primer for epidemiologists," *Am. J. Epidemiol.*, vol. 194, no. 7, pp. 2075–2084, 2025.
[33] F. Yanuar, "The Simulation Study to Test the Performance of Quantile Regression Method With Heteroscedastic Error Variance," *Cauchy J. Mat. Murni dan Apl.*, vol. 5, no. 1, pp. 36–41, 2017.
[34] Q. Li, R. Xi, and N. Lin, "Bayesian regularized quantile regression," *Bayesian Anal.*, vol. 1, no. 1, pp. 1–26, 2004.
[35] R. J. T. Al-Hamzawi, "Prior elicitation and variable selection for bayesian quantile regression," 2013, *Brunel University, School of Information Systems, Computing and Mathematics*.
[36] R. Alhamzawi and H. T. M. Ali, "Brq: An R package for Bayesian quantile regression," *Metron*, vol. 78, no. 3, pp. 313–328, 2020.
[37] L. H. Hasibuan, F. Yanuar, D. Devianto, M. Maiyastri, and A. Apriona, "Bayesian Method for Linear Regression Modeling; Simulation Study with Normality Assumption," *JOSTECH J. Sci. Technol.*, vol. 5, no. 1, pp. 81–92, 2025.
[38] K. Yu and R. A. Moyeed, "Bayesian quantile regression," *Stat. Probab. Lett.*, vol. 54, no. 4, pp. 437–447, 2001, doi: https://doi.org/10.1016/S0167-7152(01)00124-9.
[39] L. H. Hasibuan, F. Yanuar, D. Devianto, and M. Maiyastri, "Modeling Classification Of Stunting Toddler Height Using Bayesian Binary Quantile Regression With Penalized Lasso," *Mathline J. Mat. dan Pendidik. Mat.*, vol. 10, no. 2, pp. 373–385, 2025, doi: https://doi.org/10.31943/mathline.v10i2.928.
[40] D. F. Benoit, R. Alhamzawi, and K. Yu, "Bayesian lasso binary quantile regression," *Comput. Stat.*, vol. 28, pp. 2861–2873, 2013, doi: https://doi.org/10.1007/s00180-013-0439-0.
[41] R. Alhamzawi, K. Yu, and D. F. Benoit, "Bayesian adaptive Lasso quantile regression," *Stat. Modelling*, vol. 12, no. 3, pp. 279–297, 2012, doi: https://doi.org/10.1177/1471082X1101200304.
[42] R. Alhamzawi and K. Yu, "Conjugate priors and variable selection for Bayesian quantile regression," *Comput. Stat. Data Anal.*, vol. 64, pp. 209–219, 2013.
[43] R. Alhamzawi and K. Yu, "Variable selection in quantile regression via Gibbs sampling," *J. Appl. Stat.*, vol. 39, no. 4, pp. 799–813, 2012.

http://dx.doi.org/10.30983/11113