



Simulation Study of the Using of Bayesian Quantile Regression in Non-normal Error

Catrin Muharisa¹, Ferra Yanuar^{2*}, Dodi Devianto³

^{1,2,3} Department of Mathematics, Faculty of Mathematics and Natural Science, Andalas University, Kampus Limau Manis, 25163, Padang - Indonesia

Email: catrinmuharisa@gmail.com, ferrayanuar@sci.unand.ac.id,
ddevianto@sci.unand.ac.id

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

ABSTRACT

The purposes of this paper are to introduce the ability of the Bayesian quantile regression method in overcoming the problem of the non-normal errors. In this research we do simulation study to apply the proposed method. We generate data and assume error by asymmetric Laplace distribution. In this research, we solve the non-normal problem using quantile regression method and Bayesian quantile regression method and then we compare. The quantile regression approach we divide the data into any quantiles, then estimated the conditional quantile function and minimize absolute error that is asymmetrical. Bayesian regression method used the asymmetric Laplace distribution in likelihood function. Markov Chain Monte Carlo method using Gibbs sampling algorithm is applied then to estimate the parameter in Bayesian regression method. Convergence and confidence interval of parameter estimated are also checked. Bayesian quantile regression method results has more significance parameter and smaller confidence interval than quantile regression method. The best regression equation model is Bayesian quantile method used quantile value 0.50. This study proves that Bayesian quantile regression method can produce acceptable parameter estimate for non-normal error.

Keywords: Quantile regression; Asymmetric Laplace distribution; Gibbs sampling; Markov Chain Monte Carlo; Bayesian quantile regression

INTRODUCTION

Regression analysis is a tool in experienced statistics development and used in many areas of life. The analysis has a purpose to estimate the relationship between dependent variable with independent variable [1]. There are several methods used to estimate parameters in the regression equation, one of the most commonly used is the Ordinary Least Square (OLS) method. The used of OLS method is based on several assumptions, one of which is the assumption of normality. Furthermore, developed Quantile regression method which is generally used in case of econometrics. In this quantile regression approach, we divide the data into any quantiles, then estimated the conditional quantile function and minimize absolute error that is asymmetrical. In the method of quantile regression usually requires large data size.

Bayes introduced a method to estimate parameters by utilizing initial information called prior distribution. This method is known as the Bayesian method. Prior distribution can be derived from the prior research data or based on the researcher's intuition[2]. Prior information from the distribution of parameters is then combined with information from data obtained from sampling or so-called likelihood function so that posterior

distribution is obtained. The mean and variance of this posterior distribution or the posterior mean and posterior variance then can be estimated.

Bayesian Method uses MCMC algorithm (Markov Chain Monte Carlo). MCMC can be easily used to obtain posterior distribution even in that situation complex [3]. In this research will be estimated model parameters with combining the quantile regression method and the Bayesian regression method called the Bayesian quantile regression method.

METHODS

In this section, we will introduce two models, quantile regression and Bayesian quantile regression. Here, we denote Y as the dependent variable, X is the independent variable.

1. Quantile Regression

Quantile function denoted by Q_θ where $\theta, 0 \leq \theta \leq 1$. Given Y be a random variable with a cumulative distribution function $F_Y = P(Y \leq y)$. The Quantile regression θ^{th} of Y can be simply written as follows [4]:

$$Q_\theta(Y) := F_Y^{-1}(\theta) = \inf\{y: F_Y(y) \geq \theta\} \quad (1)$$

Given Y is dependent variable, and X is independent variable have p dimension. Let $F_Y(y|X = x) = P(Y \leq y|X = x)$ conditional cumulative function notation of Y given $X = x$. The conditional quantile θ^{th} of Y defined as bellows [4]:

$$Q_\theta(Y|X = x) = \inf\{y: F_Y(y|x) \geq \theta\}. \quad (2)$$

Based on the median concept of estimate for β from the quantile regression θ^{th} obtained by minimizing the absolute number of errors by weighting θ for positive error and weighting $1 - \theta$ for negative error [5] :

$$\operatorname{argmin}_{x_i^T \beta_0 \in \mathbb{R}^\theta} \sum_{i=1}^n \rho_\theta (y_i - Q_\theta(Y|X)) \quad (3)$$

where θ is quantile indeks $\in (0, 1)$ and ρ_θ is asymmetric loss function for $Q_\theta(Y|X) = X^T \beta$.

2. Bayesian Quantile Regression

The Bayesian method uses Markov Monte Carlo chain (MCMC) to estimate the posterior distribution. Given $\mathbf{y} = (y_1, y_2, \dots, y_n)$, where the prior distribution of β is $p(\beta)$. The prior distribution taken in this research is prior informative those originating from previous research [6]. Determination of prior distribution parameters are very subjective, depending on the researcher's intuition. A variable Y is said to follows Asymmetric Laplace Distribution with the density function of the probability as follows [2]:

$$f_p(y) = p(1 - p)\exp\{\rho_\theta(y_i - \mu)\} \quad (4)$$

and likelihood function as follows :

$$L(\mathbf{y}|\beta) = p^n(1 - p)^n \exp\{-\sum_i \rho_\theta (y_i - x_i^T \beta)\}. \quad (5)$$

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

$$f(\beta|y) \propto L(y|\beta) p(\beta) \\ \propto p^n(1-p)^n \exp\{-\sum_i \rho_\theta(y_i - x_i^T \beta)\} p(\beta) \tag{6}$$

RESULTS AND DISCUSSION

In this study data generated by software R that consists of two independent variables (X_1, X_2) and one response variable (Y), each independent variable (X_1, X_2) spreads according to the normal distribution ($X_1 \sim N(0, 1)$) and ($X_2 \sim N(0, 1)$). While the response variable (Y), is set value $y = 0,7 x_1 + x_2 + \varepsilon$, where $\varepsilon \sim ALD(0, 1, 0.75)$. Each variable measured 150 sample data.

Estimation results of each model parameter for each quantile using the quantile regression method can be seen in the following Table 1

Table 1. Estimated parameter model of quantile regression method

Quantile (θ^{th})	β_1	Se	β_2	Se
0.05	0,97237 (0,30104)	0,9369	4,43896 (0,000055**)	1,06125
0.25	0,7057 (0,11758)	0,44829	0,92898 (0,06935)	0,50779
0.50	1,00287 (0,00031**)	0,27177	1,25697 (0,00007**)	0,30784
0.75	1,20465 (0,0000**)	0,19633	1,22291 (0,0000**)	0,30784
0.95	1,12467 (0,1943)	0,86254	1,27023 (0,1956)	0,97702

**Significant on level $\alpha=0.05$ Se=Standart error

Based on Table 1, the parameter of β_1 significant in quantities to 0.50 and 0.75. While the parameter β_2 is significant in quantiles to 0.05, 0.50 and 0.75. Hence, the coefficient parameter of each quantile satisfied when used quantile 0.50.

Estimation results for each quantile using the Bayesian quantile regression method can be seen in the Table 2

Table 2. Estimated parameter model of Bayesian quantile regression method

Quantile (θ^{th})	β_1	β_2
0.05	0,527	3,499
0.25	0,749	1,037
0.50	1,04	1,22

0.75	1,077	1,195
0.95	1,08	1,48

Based on the coefficient parameter of each quantile satisfied when used quantile 0.25. Based on the point estimated value of the Bayesian quantile regression closer to the beta value than the quantile regression. Estimation of model parameters with quantile regression method and Bayesian quantile regression method as previously obtained. Next will be compared by using the confidence interval. The comparison results are shows in the following Table 3

Table 3. Confidence interval of quantile regression and Bayesian quantile regression

Quantile (θ^{th})		β_1		β_2	
		QR	BQR	QR	BQR
0.05	Lower Limit	-1,84973	-0,819	-2,8179	1,663
	Upper Limit	2,54512	1,17	4,63321	4,77
	Difference	4,39485	1,989	7,45111	3,107
0.25	Lower Limit	-0,07067	0,189	0,34168	0,462
	Upper Limit	1,51635	1,33	1,84061	1,63
	Difference	1,58702	1,141	1,49893	1,168
0.50	Lower Limit	0,7346	0,706	0,69714	0,801
	Upper Limit	1,53139	1,41	1,54094	1,61
	Difference	0,79679	0,704	0,8438	0,809
0.75	Lower Limit	0,49631	0,601	0,87919	0,795
	Upper Limit	1,46396	1,5283	1,4644	1,5898
	Difference	0,96765	0,9273	0,58521	0,7948
0.95	Lower Limit	0,6571	0,369	0,58451	0,337
	Upper Limit	1,64607	1,72	2,69848	2,7
	Difference	0,98897	1,351	2,11397	2,363

QR=Quantile Regression BQR=Bayesian Quantile Regression

Based on Table 3, it can be seen in the quantile 0.05 by using the quantile regression method on β_1 parameter the result is not significant, where Bayesian quantile regression method is significant. On the parameter of β_2 to the quantile 0.05 using the quantile regression method and Bayesian quantile regression method not significant. In quantiles 0.25, 0.50, 0.75, and 0.95 it can be seen that using the Bayesian quantile and quantile regression methods has been significant. From the analysis, the best regression equation model is obtained when using Bayesian quantile regression analysis with 0.50 quantile value.

The next step in Bayesian quantile regression approach is convergence test of convergence of model parameters that have been estimated parameter. Test is using history trace plot and density plot [3]. Figure 1 and Figure 2 presents a trace plot and density plot for some selected parameters.

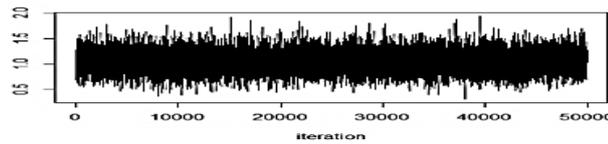


Figure 1. Trace Plot of parameter β_1 for $\theta = 0.50$

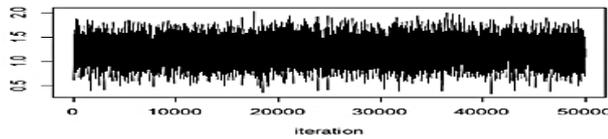


Figure 2. Trace Plot of parameter β_2 for $\theta = 0.50$

Figure 1 and Figure 2 it can be concluded that the assumption of convergence is related. Data distribution has been stable as it is between two parallel horizontal lines.

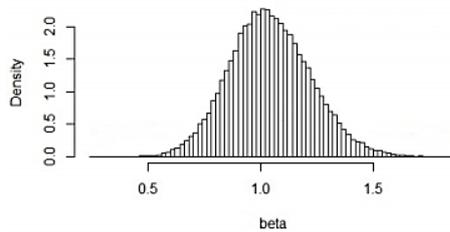


Figure 3. Density plot of parameter β_1 for $\theta = 0.50$

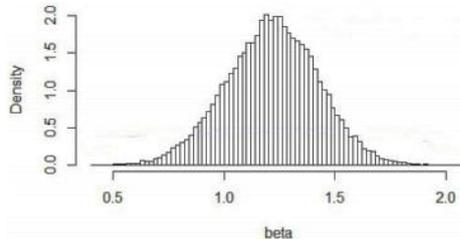


Figure 4. Density plot of parameter β_2 for $\theta = 0.50$

In Figure 3 and Figure 4, the density plot for some parameters show the normal distributed curve. This result informs us that the selected parameters model is convergent. Based on the convergence examination of the trace plot and density plot, it can be concluded that the alleged model has satisfied the criterion of convergence.

CONCLUSIONS

In this research, we use the analysis quantile and Bayesian quantile regression to analyze simulation study in non-normal error. Obtained the best regression equation model is Bayesian quantile method used quantile value 0.50. Bayesian quantile regression method using Gibbs algorithm sampling better estimator than quantile regression method. Because the Bayesian method has more parameter significance and confidence interval values smaller.

REFERENCES

- [1] Walpole, R.E and Myers, R. H. 1995. *Ilmu Peluang dan Statistika untuk Insinyur dan Ilmuwan* Edisi ke-4. ITB : Bandung.
- [2] Yu, K. and Moyeed, R. 2001. Bayesian Quantile Regression. *Statistics & Probability Letters*, 54(4), 437-447.
- [3] Ntzoufras, I. 2009. *Bayesian Modeling Using WINBugs*. John Wiley Sons, Inc: New Jersey.
- [4] Davino, C., Furno, M. and Vistocco, D. 2014. *Quantile Regression Theory and Applications*. John Wiley and Sons, Ltd.
- [5] Koenkar, R and Basset, G.Jr. 1978. Regression Quantiles. *Econometrica*, 46: 33-50
- [6] Benoit, D.F and Van den Poel, D. 2017. BayesQR: A Bayesian Approach to Quantile Regression. *Journal of Statistical Software*, 76(7), 1-32.