



Simulation Study of the Implementation of Quantile Bootstrap Method on Autocorrelated Error

Ovi Delviyanti Saputri¹, 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: ovidelviyanti@gmail.com, ferrayanuar@sci.unand.ac.id, ddevianto@sci.unand.ac.id

*Corresponding Author E-mail: ferrayanuar@sci.unand.ac.id

ABSTRACT

Quantile regression is a regression method with the approach of separating or dividing data into certain quantiles by minimizing the number of absolute values from asymmetrical errors to overcome unfulfilled assumptions, including the presence of autocorrelation. The resulting model parameters are tested for accuracy using the bootstrap method. The bootstrap method is a parameter estimation method by re-sampling from the original sample as much as R replication. The bootstrap trust interval was then used as a algorithm consistency test constructed on the estimator by the quantile regression method. And test the uncommon quantile regression method with bootstrap method. The data obtained in this test were 10 times replication data. The biasness was calculated from the difference between the quantile estimate and bootstrap estimation. Quantile estimation methods were said to be unbiased if the standard deviation bias was less than the standard bootstrap deviation. This study proved that the estimated value with quantile regression was within the bootstrap percentile confidence interval and proved that 10 times replication produces a better estimation value compared to other replication measures. Quantile regression method in this study was also able to produce unbiased parameter estimation values.

Keywords: Quantile Regression; Bootstrap Method; Autocorrelation Error

INTRODUCTION

Regression analysis is a statistical analysis that is often used in all fields of science. This analysis aims to model the relationship between two types of variables, the dependent variable (Y) with one or more independent variables (X) in a system. The relationship between these variables is expressed in a regression model which is generally stated as $Y = f(x) + e$, by e stating the error component. The model connects independent and dependent variables through a parameter named as a regression parameter, denoted by β .

Regression model can be obtained by estimating the model parameters. To estimate the value of this regression parameter is usually used the least squares method (OLS). This OLS method is applied if some assumptions are met, such as there is no autocorrelation error, normal distribution error, there is no multicollinearity between the independent variables and homogeneous error [1]. All assumptions must be fulfilled so that the BLUE parameter estimators can be obtained (Best Linear Unbiased Estimator). But this method becomes no longer BLUE and is inefficient if there are unfulfilled assumptions.

In many cases, not all assumptions are often met. The problem then is how to overcome this violated assumption. In this study the focus was on how to overcome the problem of data whose model errors were autocorrelated. Autocorrelation is if the

remainder ε_i , $i = 1, 2, \dots, n$, correlated with each other. Time-dependent data such as inflation, Rupiah exchange rate, monthly exports, daily rainfall habits often have autocorrelation [2].

Quantile regression method then appears to overcome the weaknesses in MKT. The quantile regression method was first introduced by Koenker and Bassett [3]. This method uses the parameter estimation approach by separating or dividing the data into quantiles by assuming the conditional quantile function on a distribution of the data and minimizes the absolute error of an asymmetrical weight. Quantile regression analysis is used to overcome unfulfilled assumptions, including autocorrelation, normal assumptions, no multicollinearity and homogeneity of variance [4]. Furthermore, the parameter values of the models produced using the quantile regression method must be tested for accuracy, which is to ensure that the values obtained have produced true values.

As for previous researchers such as Feng [5] studied the wild bootstrap in the quantile regression method on errors containing heteroskedacity. Oktafia [6] examined the quantile regression parameter estimator simulation with the bootstrap method on errors containing heteroscedacity. This bootstrap method was a method of parameter estimation by re-sampling from the original sample [7]. The bootstrap method could generate statistical values to create a bootstrap trust interval [8]. This bootstrap confidence interval was then used as a consistency test statistic algorithm constructed on the estimator using the quantile regression method. Thus, in this study bootstrap method was used in quantile regression to overcome errors that contained autocorrelation problems and applied on simulation data.

METHODS

1. Data

Simulation study was done by generating a set of data using R software. The data used in this study consisted of two independent variables (x_1 and x_2) and one dependent variable (y). The data for the independent variables (x_1 dan x_2) were each distributed according to the normal distribution ($x_1 \sim N(0,1)$ and $x_2 \sim N(0,1)$), whereas the dependent variable (y) was set to the value $y_t = 0.5x_{1t} + 2x_{2t} + \varepsilon$ where in $\varepsilon_t = \sin(\text{seq}(0.1\pi, 18.3\pi, 0.1\pi)) + Z_t$ with $Z_t \sim N(0,0.1)$, $t = 1, 2, 3, \dots, 183$. Each variable was 183 pieces of data.

2. Bootstrap Quantile Regression

Quantile regression was an approach in regression analysis introduced by Koenker and Bassett [10]. This approach assumed the various quantile functions of a Y distribution as a function of X . The use of this regression method was done by dividing or splitting data into several groups which having different estimations on quantile results.

Quantile function was denoted by Q_θ with value θ , $0 \leq \theta \leq 1$. For example Y was dependent variable, dan X was independent variable dimension p . For example $F_Y(y|X = x) = P(Y \leq y|X = x)$ was cumulative distribution function from Y given $X = x$. The quantile function was defined as:

$$Q_\theta(Y|X = x) = \inf\{y: F_Y(y|x) \geq \theta\} \quad (1)$$

The estimation by the θ -th quantile regression method was obtained by minimizing the weighted absolute sum of deviations. That was weighted sum of absolute

deviations, where a $(1 - \theta)$ weight was assigned to the negative deviations and a θ weight was used for the positive deviations. The previous minimization problem became:

$$Q_\theta = \operatorname{argmin}_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\theta(y_i - Q_\theta(Y|X)) \quad (2)$$

Can be written in the following equation:

$$Q_\theta = \operatorname{argmin}_{x'_i \beta_\theta \in \mathbb{R}} \left\{ \sum_{i \in I | y_i \geq x'_i \beta_\theta} \theta |y_i - x'_i \beta| + \sum_{i \in I | y_i < x'_i \beta_\theta} (1 - \theta) |y_i - x'_i \beta| \right\} \quad (3)$$

Where ρ_θ denoted an asymmetric loss function from quantile regression and θ was quantile indeks $\in (0,1)$.

Quantile regression estimation analysis was performed by bootstrap method. The bootstrap method was a method used to estimate an unknown population distribution where the empirical distribution was obtained from the repeating process as much as R replication [6]. Bootstrap method was a nonparametric approach to estimating various statistical quantiles such as mean, standard deviation, and estimation bias or to form confidence intervals by following certain algorithms. Based on [5] the mean values of bootstrap parameters were:

$$\bar{\hat{\beta}}(\theta_q) = \frac{1}{R} \sum_{b=1}^R \hat{\beta}_b(\theta_q) \quad (4)$$

where $\hat{\beta}_b(\theta_q)$ with $b = 1, 2, \dots, R$.

While the confidence interval used in this research was the percentile confidence interval. Percentile method was based on the lower limit $(\hat{\beta}_b(\theta_q)_{low})$ percentile α -th and based on the upper limit $(\hat{\beta}_b(\theta_q)_{up})$ percentile $(1 - \alpha)$ - the cumulative distribution function of the estimated bootstrap vector parameters were:

$$(\hat{\beta}(\theta)_{low}, \hat{\beta}(\theta)_{up}) = (\hat{F}(\alpha), \hat{F}(1 - \alpha)) \quad (5)$$

The indicators for the goodness of fit for parameter estimation were based on the smallest of confidence intervals and biasness.

RESULTS AND DISCUSSION

Simulation study was done by generating a set of data using R software. The data used in this study consisted of two independent variables (x_1 and x_2) and one dependent variable (y). The data for the independent variables (x_1 dan x_2) were each distributed according to the normal distribution ($x_1 \sim N(0,1)$ and $x_2 \sim N(0,1)$), whereas the dependent variable (y) was set to the value $y_t = 0.5x_{1t} + 2x_{2t} + \varepsilon$ for autocorrelated errors, the error setting becomes $\varepsilon_t = \sin(\operatorname{seq}(0.1\pi, 18.3\pi, 0.1\pi)) + Z_t$ with $Z_t \sim N(0,0.1)$, $t = 1, 2, 3, \dots, 183$. Each variable was 183 pieces of data. After the data is generated, the model parameters estimation with OLS will be carried out on the simulation data and test the significance of each result parameter estimate. The following are the results of estimating the model parameters with OLS. The resulting model parameters are tested for accuracy using the bootstrap method. The bootstrap method is a parameter estimation method by re-sampling from the original sample as much as R replication. The bootstrap trust interval was then used as a algorithm consistency test constructed on the estimator by the

quantile regression method. And test the uncommon quantile regression method with bootstrap method

In the initial stage, the estimation of the model parameters with OLS will be carried out on the simulation data and testing the significance of each result parameter estimate. The following are the results of estimating the model parameters with OLS.

Table 1. Estimation of model parameters with OLS

Parameter	coefficient	Standar Deviation	t-Value	P-Value
β_0	0.0083	0.0528	0.1580	0.8750
β_1	0.5335	0.0505	10.5730	0.0000
β_2	1.9608	0.0527	37.2130	0.0000

The F test value on this simulation data is 732.5 with a value $p < 0.05$. Therefore, the p value is less than the real level $\alpha = 0.05$, so there is at least one significant independent variable in the model. The analysis continued with the t test where the value is presented in table 1. Based on table 1 it can be seen that the estimated regression coefficients are for β_1 and β_2 significant at the level $\alpha = 0.05$, because the p -value is smaller than $\alpha = 0.05$.

Next estimation of the parameters by quantile regression based on equation (3) can be seen in Table 2.

Table 2. Quantile regression parameter estimation results

Quantile	Quantile Regression Parameter Estimation			
	β_1	P-Value	β_2	P-Value
0.1	0.5513	0.0000	2.0149	0.0000
	0.0366*		0.0382*	
0.25	0.5714	0.0000	1.9681	0.0000
	0.0733*		0.0766*	
0.5	0.5883	0.0000	1.8656	0.0000
	0.1083*		0.1131*	
0.75	0.4868	0.0000	2.0523	0.0000
	0.0581*		0.0606*	
0.9	0.4797	0.0000	1.9829	0.0000
	0.0201*		0.0210*	

*SD (Standar Deviation Value)

In Table 2. It was found that the estimation of the regression parameters for β_1 and β_2 was significant at the level $\alpha = 0.05$ for all quantiles. The value of the regression coefficient for x_1 and x_2 all was positive on each selected quantity means that these two independent variables significantly influenced y .

The analysis continued with the Durbin Watson test, where the DW value in the 0.1th quantile in this simulation data was 2.0355 with a value $p < 0.05$. In the DW table with $N = 183$ and $K = 2$ and $D_L = 1.7249$ with $D_U = 1.7915$. Based on the Durbin Watson test if the DW value is located between D_U and $4 - D_U$ stated that $D_U < 4 - D_U$ the model error does not contain autocorrelation. Likewise, with the quantile 0.25, 0.5, 0.75 and 0.9.

Then the accuracy of the quantile regression method was tested in the quantiles 0.1, 0.25, 0.5, 0.75 and 0.9 using the bootstrap method for cases of autocorrelation errors. Estimation of the value of the quantile regression parameter for each data by using bootstrap method, where the number of replications used was 10 times, 25 times, 50 times and 100 times. Then do the bootstrap percentile confidence interval for each replication was according to equation (5).

Table 3. The result of estimation of quantile regression parameter with bootstrap method on quantile 0.1

Replication	Parameter	Quantile Regression Estimation Value	Mean of Bootstrap	Confidence Interval of Bootstrap Percentile	
				Lower Limit	Upper Limit
10	β_1	0.5513	0.4741	0.3657	0.5906
	β_2	2.0149	1.9582	1.7967	2.1501
25	β_1	0.5513	0.4653	0.1746	0.6177
	β_2	2.0149	1.9189	1.6560	2.1652
50	β_1	0.5513	0.5184	0.2128	0.7587
	β_2	2.0149	1.8993	1.6544	2.1432
100	β_1	0.5513	0.5143	0.2650	0.8283
	β_2	2.0149	1.9257	1.6997	2.2117

Based on Table 3. it was found that the value of the quantile regression estimation for all model parameters was within the confidence interval of the bootstrap percentile for each replication. Furthermore, the estimation value of quantile regression parameter β_1 and β_2 for 10 times replication in a row that was 0.5513 and 2.0149 was close enough to the initial population model value that was set to 0.5 for the parameter value β_1 and 2 for the parameter value β_2 , as well as 25 times, 50 times and 100 times . Similar results were also obtained on the 0.25, 0.5, 0.75 and 0.9 quantiles.

Then the difference between the two intervals from the bootstrap percentile confidence interval was calculated for each replication to select the best replication. According to [8] the best replication is the replication that has the smallest confidence interval difference and close to the replication value that has the smallest confidence interval gap and closer to the actual parameter estimation value.

Difference in bootstrap percentile confidence interval on autocorrelated error data for each quantile of the calculation results was presented in the following table

Table 4. Difference between bootstrap percentile confidence intervals in quantile 0.1

Replication	Parameters	Difference Between Bootstrap Percentile Confidence Intervals
10	β_1	0.2249
	β_2	0.3533
25	β_1	0.4430
	β_2	0.5092
50	β_1	0.5460
	β_2	0.4888
100	β_1	0.5633
	β_2	0.5120

Based on Table 4. It was found that the smallest bootstrap percentile confidence interval was on 10 times replication was 0.2249. Thus, the best replication of the 0.1-quantile for data with an autocorrelated error was 10 times replication. The same thing was done for each quantile, so it was obtained that for quantile 0.25, the best replication was 10 times replication, for quantile 0.5, the best replication was 10 times replication, for quantile 0.75, the best replication was 10 times replication, and for quantile 0.9, the best replication was 10 times replication. It can be concluded that the confidence interval of the smallest bootstrap percentile census tended to be 10 times replication.

Furthermore, test for the unbiased quantile regression method was done with bootstrap method. As for the data obtained in this test was the 10 times replication data, because the replication tended to produce the best estimation value. Bias was calculated from the difference between quantile estimation and bootstrap estimation. Quantile estimation method was said to be unbiased if the standard deviation bias was less than the standard bootstrap deviation.

Table 5. Estimation of quantile regression parameters by determining unbiased

Quantiles	SD* bootstrap of Parameters		Bias of Parameters		SD (Bias)
	β_1	β_2	β_1	β_2	
0.1	0.0756	0.1133	0.0772	0.0567	0.0145
0.25	0.0882	0.1236	0.0529	0.0327	0.0143
0.5	0.0538	0.0657	0.0186	0.0072	0.0057
0.75	0.0720	0.0527	0.0157	0.1727	0.1110
0.9	0.1442	0.1748	0.0822	0.0956	0.0095

Based on Table 5 It can be informed that the standard deviation bias on the 0.1 quantile was 0.0145. This value was less than the value of the bootstrap standard deviation value for β_1 (0.0756) and also β_2 (0.1133). This meant that the quantile estimation and bootstrap estimation resulted in an unbiased value in the 0.1 quantile. The

same way can be seen for the 0.25, 0.5, 0.75 and 0.9 quantiles. From the results of the analysis, it was found that the quantile estimation method produced an unbiased value. Thus, it can be said that in this case, the quantile estimation method was able to produce unbiased and accurate values.

CONCLUSIONS

Quantile regression method used to model simulation data was accurate based on the accuracy test with the bootstrap method. This study proved that the estimated value with quantile regression was within the bootstrap percentile confidence interval. Of several replication measures selected, 10, 25, 50 and 100 replications. This study proved that 10 times replication resulted in a better estimation value compared to the other replication measures. Quantile regression method in this study was also able to produce unbiased parameter estimation values. Based on the unbiased test, it was found that the standard deviation estimation result of the quantile regression method was smaller than the standard deviation of the bootstrap simulation result.

REFERENCES

- [1] Bain, L. J., and Engelhardt, M. 1992. *Introduction to Probability and Mathematical Statistics Second Edition*. Duxbury Press, California.
- [2] Gujarati, D. N. 2003. *Basic Econometrics*. 4th Edition. The McGraw-Hill Companies, New Jersey.
- [3] Koenker, R., and Basset, G. J. 1978. *Regression Quantiles*. *Econometrica*, 46, 33-50.
- [4] Koenker, R. 2005. *Quantile Regression*. Cambridge University Press, New York.
- [5] Feng, X. 2011. *Wild Bootstrap for Quantile Regression*. *Journal Biometrika* 98, 4, 995-999.
- [6] Oktafia, M., Yanuar, F., and Maiyastri. 2015. *Simulasi Penduga Parameter Regresi Kuantil dengan Metode Bootstrap*. *Jurnal Matematika UNAND*, 5(1), 125-130.
- [7] Efron, B., and Tibshirani, J. R. 1993. *An Introduction to the Bootstrap*. New York: Chapman and Hall, Inc.
- [8] Mansyur, A., Efendi, S., Firmansyah, and Togi. 2010. *Estimator Parameter Model Regresi Linear dengan Metode Bootstrap*. *Jurnal Bulletin of Mathematics*, 03, 189-202.