Simulation Study The Implementation of Quantile Bootstrap Method on Autocorrelated Error

Ovi Delviyanti Saputri1, Ferra Yanuar2\*, 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 test consistency test algorithm 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 is data replication 10 times. The biasness is calculated from the difference between the quantile estimate and bootstrap estimation. Quantile estimation methods are said to be unbiased if the standard deviation bias is less than the standard bootstrap deviation. This study proves that the estimated value with quantile regression is within the bootstrap percentile confidence interval and proves that 10 times replication produces a better estimation value compared to other replication measures. Quantile regression method in this study is 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 , by  stating the error component. The model connects independent and dependent variables through a parameter named as a regression parameter, denoted by.

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. 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 be 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 is how to solve the problem of data error model berorokorelasi. Autocorrelation is if the remainder  , i = 1,2, ..., n, correlate with each other. Time-dependent data such as inflation, Rupiah exchange rate, monthly exports, daily rainfall habits often have autocorrelation [8].

Quantum regression method then appears to overcome the weaknesses in MKT. The quantil regression method was first introduced by Koenker and Bassett [10]. This method uses the parameter estimation approach by separating or dividing the data into quantiles by assuming the quantile function is conditional 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 [10]. 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 [7] studied the wild bootstrap in the quantile regression method on errors containing heteroskedacity. Octave [12] examines the quantil regression parameter estimator simulation with the bootstrap method on errors containing heteroscedacity. This bootstrap method is a method of parameter estimation by re-sampling from the original sample. The bootstrap method can generate statistical values to create a bootstrap trust interval. This bootstrap confidence interval is then used as a consistency test statistic algorithm constructed on the estimator using the quantile regression method. Thus, in this study bootstrap method will be used in quantile regression to overcome errors that contain autocorrelation problems and applied to simulation data.

## Methods

## 2.1. Data

## Simulation study is done by generating a set of data using software R. The data used in this study consists of two independent variables ( and ) and one dependent variable (). The data for the independent variables ( dan ) each distribute according to the normal distribution ($∼$ and $∼$), whereas the dependent variable () is set to the value wherein  with dengan $∼$, . Each variable is 183 pieces of data.

## 2.2. Bootstrap Quantile Regression

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

Quantile function is denoted by $Q\_{θ}$ with value $θ$, $0\leq θ\leq 1$. For example $Y$ is dependent variable, dan $X$ adalah independent variable dimension *p*. For example $F\_{Y}\left(X=x\right)=P\left(X=x\right)$ is cumulative distribution function from $Y$ given $X=x$. The quantile function is defined as:

$Q\_{θ}\left(X=x\right)=inf\left\{y:F\_{Y}\left(x\right)\geq θ\right\}$ (1)

The estimation by the $θ$-th quantile regression method is obtained by minimizing the weighted absolute sum of deviations. That is weighted sum of absolute deviations, where a $\left(1-θ\right)$ weight is assigned to the negative deviations and a $θ$ weight is used for the positive deviations. The previous minimization problem becomes:

 $Q\_{θ}=argmin\_{βϵR^{p}}\sum\_{i=1}^{n}ρ\_{θ}\left(y\_{i}-Q\_{θ}\left(X\right)\right)$ (2)

Can be written in the following equation:

$Q\_{θ}=argmin\_{x\_{i}^{'}β\_{θ}\in R}\left\{\sum\_{iϵi|y\_{i}\geq x\_{i}^{'}β\_{θ}}^{}θ\left|y\_{i}-x\_{i}^{'}β\right|+\sum\_{iϵi|y\_{i}\geq x\_{i}^{'}β\_{θ}}^{}\left(1-θ\right)\left|y\_{i}-x\_{i}^{'}β\right|\right\}$ (3)

Where $ρ\_{θ}$ denotes an asymmetric loss function from quantile regression and $θ $is quantile indeks $\in \left(0,1\right)$.

 Quantile regression estimation analysis was performed by bootstrap method. The bootstrap method is a method used to estimate an unknown population distribution where the empirical distribution is obtained from the repeating process as much as R replication [6]. Bootstrap method is 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 are:

 (4)

where  with .

While the confidence interval used in this research is the percentile confidence interval. Percentile method is based on the lower limit  percentile-th and based on the upper limit  percentile -th cumulative distribution function of the estimated bootstrap vector parameters are:

 (5)

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

# Results and Discussion

The estimation of the parameters by quantile regression based on equation (3) can be seen in Table 1. This bootstrap quantile regression analysis is aided by the R program.

**Table 1**. Quantile regression parameter estimation results



In Table 1. It is known that the estimation of the regression parameters for and  significant at the level  for all quantiles. The value of the regression coefficient for  and  all is positive on each selected quantity means that these two independent variables significantly influence .

 Continue to test the accuracy of the quantile regression method on the 0.1, 0.25, 0.5, 0.75 and 0.9 quantile regression by using the bootstrap method for the case of an autocorrelated error. Estimation of the value of the quantile regression parameter for each data by using bootrsap method, where the number of replication used is 10 times, 25 times, 50 times and 100 times. Then do the bootstrap percentile confidence interval for each replication according to equation (5).

**Table 2**. The result of estimation of quantile regression parameter with bootstrap method on quantile 0.1



Based on Table 2. it is found that the value of the quantile regression estimation for all model parameters is within the confidence interval of the bootstrap percentile for each replication. Furthermore, the estimation value of quantile regression parameter and  for replication 10 times respectively that is 0.5513 and 2.0149 have close enough to the value of the initial population model that is defined is 0.5 for the parameter value  and 2 for the parameter value , as well as 25 times, 50 times and 100 times . The same results were also obtained on the 0.25, 0.5, 0.75 and 0.9 quantiles.

Next calculated the difference between the two hoses from the bootstrap percentile confidence interval for each replication to select the best replication. According to [11] 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.

The confidence interval of the bootstrap percentile credentials on the data that is error-berorocorrelated for each quantile result is presented in the following table

 **Table 3**. Difference between bootstrap percentile confidence intervals in quantile 0.1



Based on Table 3. It is found that the smallest bootstrap percentile confidence interval is on replication 10 times that is 0.2249. Thus, the best replication of the 0.1-quantile for data with an autocorrelated error is replication 10 times. The same is done for each quantile, so that for the best 0.25 replication quantile is replicated 10 times, for the best replication 0.5th quantile it is replicated 10 times, for the best 0.75 replication the replication is at 10 replications and the best 0.9 replication quantile is at replication 100 times. It can be concluded that the confidence interval of the smallest bootstrap percentile census tends to be at replication 10 times.

Next, test the uncommon quantile regression method with bootstrap method. The data obtained in this test is the replication data 10 times, because the replication tends to produce the best estimation value. The biasness is calculated from the difference between the quantitative estimate and bootstrap estimation. Quantile estimation method is said to be unbiased if the standard deviation bias is less than the standard bootstrap deviation.

**Table 4.** Estimation of the quantile regression parameter by determining unbiased



 Based on Table 4. It can be informed that the standard deviation bias on the 0.1 quantile is 0.0145. This value is less than the value of the standard deviation value of bootstrap, for (0.0756) and also (0.1133). This means that the bootstrap’s quantile and estimation estimates result 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 analysis result obtained that the method of quantile estimation yields an unbiased value. Thus it can be said that in this case, the quantile estimation method is capable of producing an unbiased value and yielding an accurate value.

**Conclusions**

Quantile regression method used to model the simulation data has been accurate based on the accuracy test by bootstrap method. This study proves that the estimated value with the quantile regression is within the confidence interval of the bootstrap percentile. Of the selected replication sizes are 10, 25, 50 and 100 replication. This study proves that 10 times replication produces a better estimation value compared to other replication measures. Quantile regression method in this study is also able to produce unbiased parameter estimation values. Based on the test of the biasness it is found that the standard deviation value estimation result of the quantile regression method is 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. Capinski, M., and Kopp, E. 2003. *Measure, Integral, and Probability*. Second Edition. Springer-Verlag, New York.
3. Casella, G., and Berger, R. L. 2002. *Statistical Inference*. Ed. Ke-2, Pasific Grove, California.
4. Chung, K. L. 2001. *A Course In Probability Theory*. Third Edition. Academy Press, Sandiego.
5. Davino, C., Furno, M., and Vistocco, D. 2014. *Quantile Regression Theory and Aplications*. Pondicherry, India.
6. Efron, B., and Tibshirani, J. R. 1993. *An Introduction to the Bootstrap*. New York: Champman and Hall, Inc.
7. Feng, X. 2011. *Wild Bootstrap for Quantile Regression*. Journal Biometrika 98, 4, 995-999.
8. Gujarati, D. N. 2003. *Basic Econometrics*. 4th Edition. The McGraw-Hill Companies, New Jersey.
9. Koenker, R., and Basset, G. J. 1978. *Regression Quantiles*. Econometrica, 46, 33-50.
10. Koenker, R. 2005. *Quantile Regression*. Cambridge University Press, New York.
11. Mansyur, A., Efendi, S., Firmansyah, and Togi. 2010. *Estimator Parameter Model Regresi Linear dengan Metode Bootstrap*. Jurnal Bulletin of Mathematics, 03, 189-202.
12. Oktafia, M., Yanuar, F., and Maiyastri. 2015. *Simulasi Penduga Parameter Regresi Kuantil dengan Metode Bootstrap*. Jurnal Matematika UNAND, 5(1), 125-130.