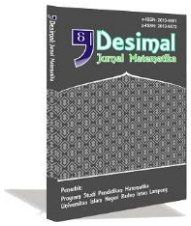
Desimal: Jurnal Matematika Vol 6 No 2 (2023) 245-250

Contents lists available at DJM

DESIMAL: JURNAL MATEMATIKA

p-ISSN: 2613-9073 (print), e-ISSN: 2613-9081 (online), DOI 10.24042/djm

http://ejournal.radenintan.ac.id/index.php/desimal/index



Comparison of quantile regression and censored quantile regression methods in the case of chicken consumption

Sarmada[1,*], Ferra Yanuar[2], Dodi Devianto[2]

[1] Jambi University, Indonesia
[2] Andalas University, Indonesia

ARTICLE INFO

Article History
Received : 28-06-2023
Revised : 29-07-2023
Accepted : 14-08-2023
Published : 30-08-2023

Keywords:
Censored Data; Quantile Regression; Censored Quantile Regression; RMSE.

*Correspondence: E-mail: sarmada@unja.ac.id

Doi: 10.24042/djm.v6i2.18949

ABSTRACT

The censored quantile regression method is a parameter estimation method that can be used to overcome censored data and BLUE (Best Linear Unbiased Estimator) assumptions that are not met. This research aims to compare the quantile regression method and the censored quantile regression method on data on chicken consumption cases in West Sumatra. The smallest RMSE (Root Mean Square Error) is an indicator of the goodness of the model. This research proves that the censored quantile regression method tends to produce smaller RMSE values than the quantile regression method. So it is concluded that the censored quantile regression method is the appropriate method for estimating parameters with censored data.

http://ejournal.radenintan.ac.id/index.php/desimal/index

INTRODUCTION

The BLUE estimator (Best Linear Unbiased Estimator) is a good parameter. A BLUE estimator will be obtained if it fulfills all the assumptions of the linear model. The Least Squares Method (MKT), as a parameter estimation method that is often used, is not always a BLUE estimator. In real life, there are cases where the assumptions of normality and homoscedasticity are not met when there is data that has a value of zero or outlier data. In such cases, the Least Squares Method cannot be used (Saputri, Yanuar, & Devianto, 2018).

The quantile method emerged as a parameter estimation method that can overcome unfulfilled assumptions such as normality and homoscedasticity. This method uses a parameter estimation approach by separating or dividing data into quantiles, namely by using a conditional quantile function on a data distribution and minimizing the weighted absolute residual, which is not symmetrical (Sarmada & Yanuar, 2020).

The next problem is that data is often found with the independent variable present, but the response variable is y, with the y value being below c (left sensor), and the y value being above c (right sensor), with c being the threshold (sensor point). In most cases, the sensor threshold is zero. Multiple censoring is also possible if the value of the response variable is only available for observations where $c < y < d$, c and d are the two thresholds (Davino, Furno, & Vistocco, 2014). The censored quantile regression method is a parameter estimation method that can overcome problems with censored data and unfulfilled BLUE assumptions.

Research related to censored quantile regression by Gustavsen & Rickertsen (2013) carried out a censored quantile regression approach in research on adjusting VAT rates to promote healthy diets in Norway. Censored regression models were also carried out by Leiker (2012), who studied the comparison of censored regression method estimators with MLE. Chernozhukov, Fernandez-Val, & Kowalski (2011) also studied regression quantiles with censorship and endogeneity. The topic of study in this research is to carry out a comparison between the quantile regression method and censored quantile regression using the smallest RMSE value as the best estimator indicator. This research uses secondary data, namely household expenditure data for chicken meat consumption in West Sumatra..

Chicken meat is one of the most important sources of protein that needs to be consumed for physical and mental development, as well as for a healthy and balanced diet (Saçlı, 2018). Adequate nutritional needs, especially animal protein, in society, will be more efficient if increased food consumption is sourced from livestock commodities, especially broiler chicken meat. According to the Central Statistics Agency (BPS), people's consumption patterns for purebred chicken meat tend to increase every year from 2014 to 2018 (Badan Pusat Statistik, 2019). For this reason, it is necessary to research what factors influence chicken consumption in rural and urban communities in West Sumatra.

Household expenditure for chicken meat consumption (in rupiah) in several cities and districts in West Sumatra in 2018, with the independent variables in this study consisting of the total number of individuals in the household (X_1), total per capita household expenditure (X_2), and age of the head of household (X_3).

METHOD

This research uses secondary data, namely data obtained from the West Sumatra Central Statistics Agency. The data used is household expenditure data for chicken meat consumption (in rupiah) in several cities and districts in West Sumatra in 2018.

Meanwhile, the independent variables in this research consist of:

1. Total number of individuals in the household (X_1)
2. Total household expenditure per capita (X_2)
3. Age of head of household (X_3).

The data analysis methods used are quantile regression analysis and censored quantile regression analysis, while the calculations are carried out with the help of R software. The flowchart for this research is shown in Figure 1.

The data analysis steps for chicken consumption are as follows:

a) Collecting data, where the data used is household expenditure data for chicken meat consumption for 36 rural areas totaling 507 data, urban areas totaling 867 data, and a combination of rural and urban areas totaling 1374 pieces of data, which will be input into R software, where the variables are as follows:
    i. The dependent variable (Y)

states the consumption of purebred chicken meat in thousands of rupiah.
$y_i = 0$ if the household does not incur costs for consuming purebred chicken meat.
$y_i = y_i^*$ if the household incurs costs for consuming purebred chicken meat.

ii. Total number of individuals in the household (X_1).

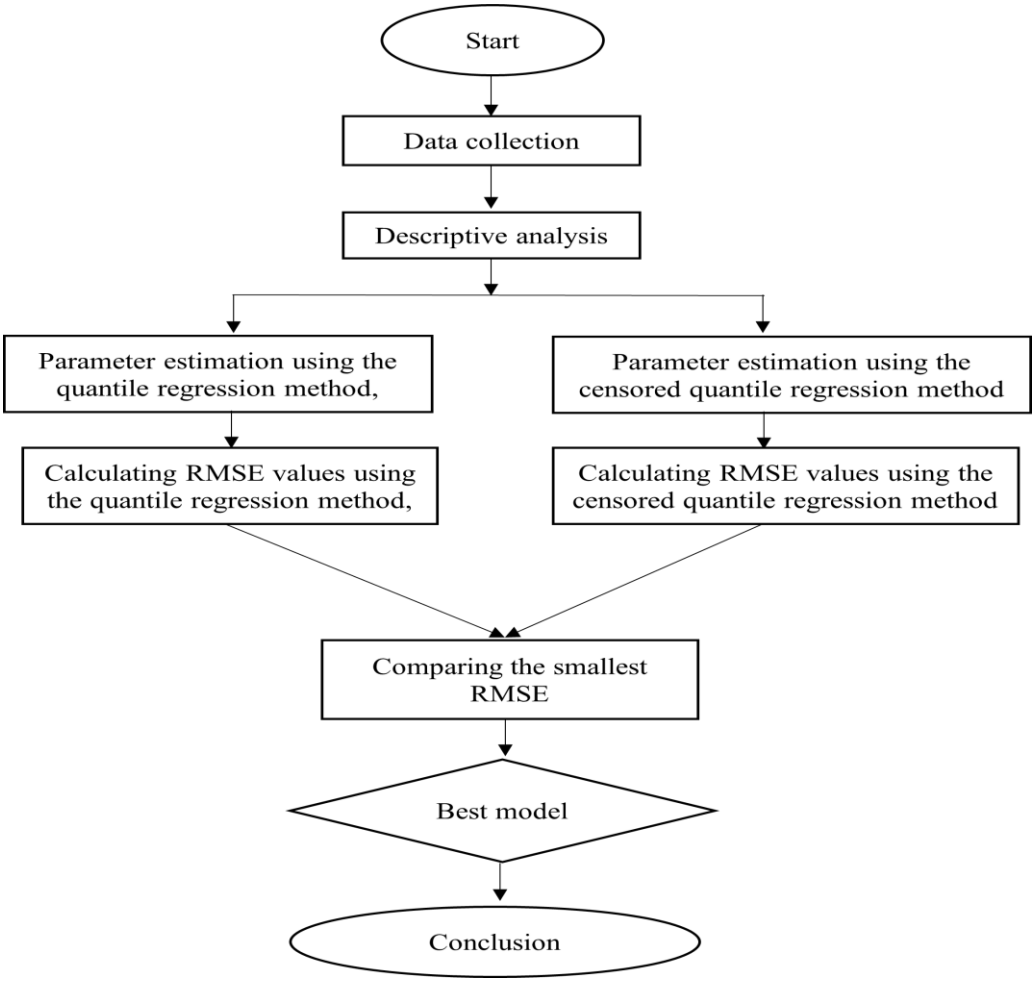


Figure 1. Research Flow

iii. Total household expenditure per capita (X_2).
iv. Age of head of household (X_3).

b) Conduct a descriptive analysis of all independent variables and dependent variables used in this research.
c) Estimating quantile regression model parameters.
d) Estimating the parameters of a censored quantile regression model by minimizing the function.

$$\hat{\beta}(\theta) = \min_{\beta \in R} \sum_{i=1}^{n} \rho_\theta (y_i - \max(\tau, x_i^\tau \beta))$$

Parameter significance test in both methods.
e) Calculate the RMSE value for both methods.
f) Comparing RMSE values for both methods.

RESULTS AND DISCUSSION

This research uses case data using R Software version 1.1463. The data used in

this research consists of three independent variables (X_1, X_2, and X_3) and a dependent variable (Y). The case data used in this research is secondary data obtained from the West Sumatra Central Statistics Agency. The case data raised in this research is data on household expenditure for chicken meat consumption (in rupiah) in several cities and districts in West Sumatra. The independent variables in this study are the total number of individuals in the household (X_1), the total per capita household expenditure (X_2), and the age of the head of the household (X_3). The dependent variable (Y) in this study is the consumption of purebred chicken meat in thousands of rupiah.

The parameter estimates in the censored quantile regression of household expenditure for chicken consumption are divided into 3 cases: the first for urban areas, the second for rural areas, and the last for combined areas (urban and rural). The sample in this study amounted to 1374 samples for combined areas, 867 data for urban areas, and 507 data for rural areas. Parameter estimates were chosen for 4 quantile points, namely 0.25, 0.5, 0.75, and 0.9 quantiles. The analyses used are quantile regression analysis and censored quantile regression analysis. Below, we will present estimates using both methods for the case of chicken meat consumption (in thousands) in urban areas in Table 1.

Table 1. Parameter Estimates for Urban Areas

Quantile	Quantile Regression			Censored Quantile Regression		
	β_1	β_2	β_3	β_1	β_2	β_3
0,25	3,292*	0,252*	0,172*	3,112*	0,240*	0,163*
0,5	3,696*	0,328*	0,097*	3,573*	0,313*	0,093*
0,75	3,435*	0,529*	0,118*	3,305*	0,517*	0,106*
0,9	5.643*	0,761*	0,085	5,709*	0,763*	0,087

Based on Table 1, the parameter estimates for the β_2 parameter in quantile regression and censored quantile regression tend to increase as the quantile increases. This shows that the influence of the variable gets bigger as the quantile value increases. Meanwhile, the parameters β_1 and β_3 show varying values at each quantile. The variable that has a big influence on the high consumption of chicken meat in urban areas is the amount of household expenditure per capita. Because of its own estimation, it amounts to 5.643 in the regression quantile and 5.709 in the regression quantile censored.

Next, it will be estimated for each parameter using both methods in rural areas. The estimation results are presented in Table 2.

Table 2. Parameter Estimates for Rural Areas

Quantile	Quantile Regression			Censored Quantile Regression		
	β_1	β_2	β_3	β_1	β_2	β_3
0,25	2,616*	0,421*	-0,210*	1,984*	0,334*	-0,154*
0,5	2,133*	0,736*	-0,137*	2,133*	0,736*	-0,138*
0,75	2,574*	1,207*	-0,077	2,495*	1,243*	-0,071
0,9	2,967*	1,659*	-0,245	2,618*	1,288*	-0,211

Based on Table 2, it is explained that the effect of each parameter has varying values at each quantile. The estimation results for β_2 obtained tend to increase as the quantile increases. This shows that the influence of the variable gets bigger as the quantile value increases. Parameter estimates in β_1 and β_3 show varying values at each quantile. The variable that has a large influence on rural areas is the

amount of per capita household expenditure because the own mark estimation amounts to 2.967 in the regression quantile and 2.618 in the regression quantile censored.

Next, it will be estimated for each parameter using both methods in the combined area. The estimation results are presented in Table 3.

Table 3. Parameter Estimates for Combined Areas

Quantile	Quantile Regression			Censored Quantile Regression		
	β_1	β_2	β_3	β_1	β_2	β_3
0,25	3,161*	0,279*	0,040	2,250*	0,193*	0,001
0,5	2,958*	0,415*	0,019	2,851*	0,383*	0,034
0,75	3,298*	0,692*	0,011	3,195*	0,691*	-0,004
0,9	3,718*	0,849*	-0,039	4,015*	0,927*	-0,090

Based on Table 3, the estimated values produced by each parameter for combined areas (urban and rural) have varying values at each quantile. The estimated parameters β_1 and β_2 in quantile regression and censored quantile regression tend to increase as the quantile increases. This shows that the influence of the variable gets bigger as the quantile value increases. Parameter estimates in β_3 show varying values. Variables that have quite a large influence on chicken meat consumption in the combined area are the total number of individuals in the household and the total per capita expenditure of the head of the household.

The best model can be shown with the smallest RMSE value. Table 4 below presents the RMSE values for both methods. Numbers in bold indicate smaller RMSE values between corresponding regions.

Table 4. Quantile Regression RMSE and Censored Quantile Regression RMSE

Quantile	Quantile Regression			Censored Quantile Regression		
	Urban	Rural	Combined	Urban	Rural	Combine
0,25	23,010	8,776	16,357	**21,752**	**7,834**	**10,450**
0,5	16,639	**10,767**	12,301	**15,903**	10,790	**12,142**
0,75	17,038	**12,358**	14,805	**16,217**	12,561	**14,336**
0,9	**24,601**	15,405	**16,109**	24,905	**15,362**	16,600

Based on Table 4, it shows that at the 0.25 quantile, the RMSE value for urban, rural, and combined areas with censored quantile regression is smaller than the quantile regression, namely values of 21.752, 7.834, and 10.450, respectively. Meanwhile, at the 0.50th quantile produced for urban and combined areas, the RMSE value with censored quantile regression tends to be smaller than quantile regression. However, the results are different in rural areas, where the rural RMSE quantile regression (10.767) is smaller than the censored quantile regression (10.790). The same analysis can be carried out for the 0.75th and 0.90th quantiles, with the RMSE value in the censored quantile regression tending to be smaller. As seen in Table 4, there are 4 RMSE values from the quantile regression method that are smaller, and in the censored quantile regression method there are 8 smaller RMSE values, so it can be concluded that the censored quantile regression RMSE values at each quantile tend to produce smaller RMSE values from quantile regression. This shows that censored quantile regression is a better estimation method than quantile regression.

CONCLUSIONS AND SUGGESTIONS

This research aims to find estimated parameter values using quantile

regression and censored quantile regression with censored data. This research is applied to the case of chicken meat consumption in West Sumatra, which is divided into three, namely urban, rural, and combined areas. The results obtained show that the parameter estimation values in censored quantile regression tend to be better than in quantile regression. This is indicated by the RMSE value in the censored quantile regression, which tends to be smaller. So, it can be concluded that the censored quantile regression method is better than the quantile method.

The research only focuses on estimating model parameters using the censored quantile regression method. In future research, it is recommended to carry out hybridization using the Bayesian method or the LASSO (Least Absolute Shrinkage Selection Operator) method to produce a better method.

REFERENCES

Badan Pusat Statistik. (2019). *Pola konsumsi makanan penduduk provinsi sumatera barat 2017-2018*. Sumatera barat.

Chernozhukov, V., Fernandez-Val, I., & Kowalski, A. (2011). Quantile regression with censoring and endogeneity. *Journal of Econometrics*, *186*(1), 201–221.

Davino, C., Furno, M., & Vistocco, D. (2014). *Quantile regression: Theory and applications*. John Wiley & Sons, Ltd.

Gustavsen, G. W., & Rickertsen, K. (2013). Adjusting vat rates to promote healthier diets in norway: A censored quantile regression approach. *Food Policy*, *42*, 88–95. https://doi.org/10.1016/j.foodpol.2013.07.001

Leiker, A. (2012). *The estimation in tobit regression models* (Thesis Master of Science). Kansas State University, Kansas.

Saçlı, Y. (2018). Evaluation of socio-economic factors affecting chicken meat consumption in turkey. *Journal of Poultry Research*, *15*(2), 47–52.

Saputri, O. D., Yanuar, F., & Devianto, D. (2018). Simulation study of the implementation of quantile bootstrap method on autocorrelated error. *CAUCHY: Jurnal Matematika Murni Dan Aplikasi*, *5*(3), 95–101. https://doi.org/10.18860/ca.v5i3.5349

Sarmada, & Yanuar, F. (2020). Quantile regression approach to model censored data. *Science and Technology Indonesia*, *5*(3). https://doi.org/10.26554/sti.2020.5.3.79-84