p-ISSN 2338 – 3216
e-ISSN 2528 – 1070

JURNAL STATISTIKA
UNIVERSITAS MUHAMMADIYAH SEMARANG
10(2) 2022: 25-33

BAYESIAN ANALYSIS OF TOBIT QUANTILE REGRESSION WITH ADAPTIVE LASSO PENALTY IN HOUSEHOLD EXPENDITURE FOR CIGARETTE CONSUMPTION

Fitri Rahmawati*[1], Subanar[2]
[1] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Indonesia
[2] Department of Mathematics, Faculty of Mathematics and Natural Sciences, Gadjah Mada University, Indonesia

***e-mail**: *fitrirahma66.fr@gmail.com*

Article Info:
Received: 16 October 2022
Accepted: 11 February 2023
Available Online: 11 February 2023

Keywords:
Tobit Regression; Quantile; Adaptive Lasso; Bayesian.

Abstract: Tobit Quantile Regression with Adaptive Lasso Penalty is a quantile regression model on censored data that adds Lasso's adaptive penalty to its parameter estimation. The estimation of the regression parameters is solved by Bayesian analysis. Parameters are assumed to follow a certain distribution called the prior distribution. Using the sample information along with the prior distribution, the conditional posterior distribution is searched using the Box-Tiao rule. Computational solutions are solved by the MCMC Gibbs Sampling algorithm. Gibbs Sampling can generate samples based on the conditional posterior distribution of each parameter in order to obtain a posterior joint distribution. Tobit Quantile Regression with Adaptive Lasso Penalty was applied to data on Household Expenditure for Cigarette Consumption in 2011. As a comparison for data analysis, Tobit Quantile Regression was used. The results of data analysis show that the Tobit Quantile Regression model with Adaptive Lasso Penalty is better than the Tobit Quantile Regression.

1. INTRODUCTION

Research, especially surveys in various fields, produces outputs with various data characteristics. One of them is data that is censored at a certain value. This is what inspired Tobin [1] introducing the Tobit method in the regression model. Tobit regression is able to illustrate the relationship between a censored respon variable and its predictor variables. Unfortunately, Tobit regression is not helpful enough for data analysis that does not meet the normality assumption. Furthermore, Powell [2] introduces Tobit Quantile Regression, where the conditional mean function on parameter estimation is replaced by a function of quantiles. Tobit Quantile Regression model is able to provide a more complete explanation of the relationship between response variables and predictor variables than simple Tobit Regression. In Tobit Quantile Regression, the estimation of the regression coefficients for each quantile can be obtained..

The selection of variables in the regression model is an important thing. The selection of the right variables causes the accuracy of the prediction to increase. Tibshirani [3] introduced the LASSO (Least Absolute Shrinkage and Selection Operator) method, which is an estimation

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

method to minimize errors that depend on the sum of the absolute values of the coefficients. Lasso regression is able to eliminate the variables in the regression model by reducing the coefficient to zero. However, Zou [4] states that Lasso Regression for optimal λ results in variable selection that is not always consistent.. To solve this problem, Zou [4] introduces the Adaptive Lasso method in which an adaptive weight is applied on the l_1 penalty.

This paper will discuss about Tobit Quantile Regression with Adaptive Lasso Penalty. Adaptive weights are added to the estimation of the Tobit Quantile Regression parameter, so that the parameter estimate is the solution of minimizing the Loss Function plus the absolute sum of the adaptive weights. The form of such an optimization function cannot be solved explicitly, so the parameter estimation of this study is determined using Bayesian approach.

Bayesian analysis uses knowledge of prior distributions to find a complete conditional posterior distribution. Furthermore the joint posterior distribution can be found. This characteristic relates the Bayesian approach to the Markov Chain Monte Carlo (MCMC) method using the Gibbs Sampling algorithm. MCMC comes from a combination of two things, namely: Markov Chain and Monte-Carlo. Sampling of this method uses the Markov Chain principle, which generates a new sample based on only one previous sample. The algorithm used is the Gibbs Sampling algorithm.

Cigarette consumers in Indonesia are spread across all economic levels. Based on data from the 2011 Indonesian Household Socio-Economic Survey (SUSETI) by the World Bank in Cintiani [5], household spending on cigarette consumption is quite varied. For households with no active smokers, the cigarette consumption expenditure is 0. Thus this data has censored data criteria with a lower limit of 0. Furthermore, in this paper, Tobit Quantile Regression with Adaptive Lasso Penalty will be applied in the case of Household Expenditure for Cigarette Consumption. Quantile regression was applied to determine the regression analysis for each selected quantile.

2. LITERATURE REVIEW

2.1. Loss Function

Parameter estimation in quantile regression was obtained by developing the LAD (Least Absolute Deviation) method, which is given different weights for different quantiles. In the τ -th quantile, for the underprediction case where the error is greater than or equal to zero, the weight used is τ. As for the overprediction case where the error is less than zero, the weight used is 1- τ.

According to Koenker [6], the loss function in Quantile Regression is defined:

$$\rho_\tau(\varepsilon)\begin{cases}\tau\varepsilon, \varepsilon \geq 0\\ -(1-\tau)\varepsilon, \varepsilon < 0\end{cases} \tag{1}$$

The form of the Loss Function in Alhamzawi and Yu [7] can also be expressed:

$$\rho_\tau(\varepsilon) = \frac{|\varepsilon| + (2\tau - 1)\varepsilon}{2} \tag{2}$$

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

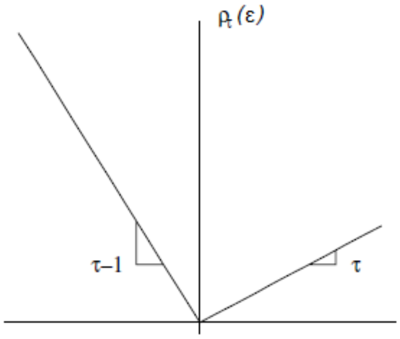

Fig 1. *Loss Function*

2.2. Tobit Regression

The censored data in Tobit regression according to Greene [8] is when the response variable values in a certain range are all changed to one value. The structure of the censored data is a discrete-scale response variable for the censored data, and a continuous scale for those that are not.

The Tobit Regression Equation is:

$$y_i^* = x_i'\beta + \varepsilon_i \tag{3}$$

$$y_i = C(y_i^*), i = 1, \dots, n \tag{4}$$

with:

$$C(y_i^*) = max\ \{y^0, x_i'\beta + \varepsilon_i\} \tag{5}$$

where:

y_i = i-th sample response variable,
y_i^* = unobserved latent variable of i-th sample,
x_i = vector k x 1 of the predictor variables for the i-th sample,
β = regression coefficient vector,
ε_i = error on i-th observation,
C(.) = link function,
y^0 = censored point.

2.3. Adaptive Lasso

LASSO stands for Least Absolute Shrinkage and Selection Operator, which is an estimation method that minimizes the sum of the squares of errors that depend on the sum of the absolute values of the coefficients. In other words, the lasso regression coefficient minimizes:

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{k} |\beta_j|. \tag{6}$$

James, et al [9] stated that Lasso Regression has the same formula as Ridge Regression, the only difference is that Ridge Regression uses the $\sum_j {\beta_j}^2$ constraint whereas Lasso Regression uses the $\sum_j |\beta_j| \le t$ constraint. Ridge Regression is able to reduce the coefficient to near zero, while Lasso Regression is able to reduce the regression coefficient to zero. As a

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

result, Lasso Regression can be used for variable selection, so that the model obtained using Lasso Regression will be easier to interpret than using Ridge Regression. To produce consistent variable selection, Zou [4] introduced the Adaptive Lasso. The adaptive weight is added to the l_1 penalty so that the equation becomes:

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2 + \lambda \sum_{j=1}^{k} \widehat{w_j}|\beta_j| \quad (7)$$

by defining $\lambda_j = \lambda\widehat{w_j}$, we get:

$$\sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{ij}\beta_j\right)^2 + \sum_{j=1}^{k} \lambda_j|\beta_j|. \quad (8)$$

2.4. Tobit Quantile Regression with Adaptive Lasso Penalty

Mosteller and Tukey [10] stated that in order to get a more complete explanation of the data, several different regression curves can be calculated according to various percentage points of the data distribution. Quantile regression was first introduced by Koenker and Basset [11]. In this model, various quantile functions are estimated from a distribution of Y as a function in X. One of the developments of quantile regression is tobit quantile regression. Based on Powell [2], the Tobit Quantile Regression coefficient estimator can be estimated by finding a solution that minimizes β_τ from:

$$\sum_{i=1}^{n} \rho_\tau(y_i - C(x_i\beta_\tau)) \quad (9)$$

where $\rho_\tau(.)$ is the *Loss Function.*

Alhamzawi [12] added the adaptive lasso penalty to the tobit quantile regression in equation (9) so that the τ -th quantile regression is the solution to minimize β from:

$$\sum_{i=1}^{n} \rho_\tau(y_i - C(x_i\beta_\tau)) + \sum_{j=1}^{k} \lambda_j|\beta_{j\tau}| \quad (10)$$

This equation contains a Loss Function which is not differentiable at point 0. Therefore, this minimization problem cannot be solved explicitly. The method used in this paper is the Bayesian approach.

2.5. Bayesian Analysis of Tobit Quantile Regression with Adaptive Lasso Penalty

Kozumi and Kobayashi [13] stated that minimizing the Loss Function is equivalent to finding the posterior maximum of the estimator with an error following the Asymmetric Laplace distribution. Furthermore ε_i, i=1,2,...,n are assumed to have ALD distribution where

$$f(\varepsilon_i|\tau) = \frac{\tau(1-\tau)}{\sigma} exp\left\{ -\frac{\rho_\tau(\varepsilon_i)}{\sigma}\right\}. \quad (11)$$

The form of the Laplace asymmetric distribution in (11) can be expressed as a mixture of the normal and exponential distribution families. We use the equations of Andrews and Mallows [14], that is, for a, b > 0, applies:

$$exp\{-|ab|\} = \int_0^\infty \frac{a}{\sqrt{2\pi v}}\ exp\left\{-\frac{1}{2}(a^2v + b^2v^{-1}\right\} dv. \quad (12)$$

As a result, the likelihood of the data is stated as:

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

$$L(y|\beta_\tau,\sigma,x) = \prod_{i=1}^{n} \int_0^\infty \frac{\tau(1-\tau)}{\sigma\sqrt{4\sigma\pi v_i}} exp\left\{-\frac{(y_i - C(x'_i\beta_\tau) - (1-2\tau)v_i)^2}{4\sigma v_i} - \frac{\tau(1-\tau)v_i}{\sigma}\right\} dv_i \qquad (13)$$

with $y_i^*|\beta_\tau, v_i \sim N(x'_i\beta_\tau + (1-2\tau)v_i, 2\sigma v_i)$ and $v_i \sim Eksponensial\left(\frac{\tau(1-\tau)}{\sigma}\right)$.

The prior distribution for other parameters follows Alhamzawi [12], is:

$$\beta_{j\tau}|s_j \sim N(0, s_j),$$

$$s_j \sim Eksponensial\left(\frac{\lambda_j^2}{2}\right),$$

$$\lambda_j^2 \sim Gamma(a, b^{-1}),$$

$$\sigma \sim Invers\ Gamma(c, d^{-1}),$$

where $y_i = \max(y^0, x_i\beta_\tau)$.

The complete conditional posterior distribution of each parameter is then searched using the Box-Tiao method. The results obtained are:

a. The conditional posterior distribution of y_i^*
$$y_i^*|y_i,\beta_\tau,v_i,\sigma \sim \begin{cases} \delta(y_i), & \text{if } y_i > y^0 \\ N(x'_i\beta_\tau + (1-2\tau)v_i, 2\sigma v_i)I(y_i^* \le y^0) & \text{other} \end{cases}.$$
b. The conditional posterior distribution of β_τ is normal multivariat distribution with $mean\ \mu_{\beta\tau} = \ \Sigma_{\beta\tau}X'V(y-(1-2\tau)v)$,
$variance\ \Sigma_{\beta\tau} = (X'VX + S)^{-1}$
where:
$V = diag((2\sigma v_1)^{-1}, (2\sigma v_2)^{-1}, \dots, (2\sigma v_n)^{-1})$,
$S = diag(s_1^{-1}, s_2^{-1}, \dots, s_k^{-1})$.
c. The conditional posterior distribution of v_i^{-1} is Wald distribution with $mean\ \mu = \ |y_i^* - x'_i\beta_\tau|^{-1}$ and shape parameter $v = (2\sigma)^{-1}$
d. The conditional posterior distribution of s_j^{-1} is invers Gaussian distribution with $mean\ \mu = \ \frac{\lambda_j^2}{\beta_j^2}$ and shape parameter $v = \lambda_j^2$.
e. The conditional posterior distribution of λ_j^2 is Gamma distribution with parameters a+1 and $(\frac{s_j}{2} + b)^{-1}$.
f. The conditional posterior distribution of σ is invers gamma distribution with parameters $\frac{3n}{2} + c$ and $\left(\sum_{i=1}^{n}\left(\frac{(y_i^* - x'_i\beta_\tau - (1-2\tau)v_i)^2}{4v_i} + \tau(1-\tau)v_i\right) + d\right)^{-1}$.

Based on the complete conditional posterior distribution obtained, then the Gibbs Sampling process is carried out to estimate the parameter.

2.6. MCMC Gibbs Sampling

MCMC comes from a combination of two things, namely: Markov Chain and Monte-Carlo. Ravenswaaij, et al. [15] stated that Monte-Carlo is a method of estimating the parameters of the distribution by taking a random sample from the distribution. Markov Chain in MCMC

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

is the idea that random samples are generated from sequential processes with certain rules. Each random sample is used as a stepping stone to generate the next random sample (forming a chain). The rule of Markov Chain is that each new sample only depends on one previous sample. Walsh [16] states that Gibbs Sampling aims to build a Markov Chain that converges to the targeted distribution. The key of Gibbs Sampling is to use a univariate complete conditional distribution, where the random variables have a fixed value except for one variable.

The steps of the Gibbs Sampling process carried out in this study are as follows:

a. Determine the τ -th quantile to be estimated.
b. Take the initial value for each parameter $y^*, \beta_\tau, v_i, {\lambda_j}^2, s_j, \sigma$ suppose $y^{*(0)}, {\beta_\tau}^{(0)}, v^{(0)}, {\lambda^2}^{(0)}, {s_j}^{(0)}, \sigma^{(0)}$.
c. Generated sample $y_i^{*(1)}$, $y_i^{*(1)} \sim \pi(y_i^* | {\beta_\tau}^{(0)}, {v_i}^{(0)}, {\lambda_j}^{2^{(0)}}, {s_j}^{(0)}, \sigma^{(0)}$,y)
d. Generated sample ${\beta_\tau}^{(1)}$, ${\beta_\tau}^{(1)} \sim \pi(\beta_\tau | y_i^{*(1)}, {v_i}^{(0)}, {\lambda_j}^{2^{(0)}}, {s_j}^{(0)}, \sigma^{(0)}$,y)
e. Generated sample ${v_i}^{(1)}$, ${v_i}^{(1)} \sim \pi(v_i | y_i^{*(1)}, {\beta_\tau}^{(1)}, {\lambda_j}^{2^{(0)}}, {s_j}^{(0)}, \sigma^{(0)}$,y)
f. Generated sample ${\lambda_j}^{2^{(1)}}$, ${\lambda_j}^{2^{(1)}} \sim \pi({\lambda_j}^2 | y_i^{*(1)}, {\beta_\tau}^{(1)}, {v_i}^{(1)}, {s_j}^{(0)}, \sigma^{(0)}$,y)
g. Generated sample ${s_j}^{(1)}$, ${s_j}^{(1)} \sim \pi(s_j | y_i^{*(1)}, {\beta_\tau}^{(1)}, {v_i}^{(1)}, {\lambda_j}^{2^{(1)}}, \sigma^{(0)}$,y)
h. Generated sample $\sigma^{(1)}$, $\sigma^{(1)} \sim \pi(\sigma | y_i^{*(1)}, {\beta_\tau}^{(1)}, {v_i}^{(1)}, {\lambda_j}^{2^{(1)}}, {s_j}^{(1)}$,y)
i. Step c until h repeated as many iterations used.
j. Obtained a sample with joint posterior distribution $\pi(y^*, \beta_\tau, v_i, {\lambda_j}^2, s_j, \sigma$|y)

The estimation of the Tobit Quantile Regression parameter with the Adaptive Lasso Penalty sought is the mean of the posterior distribution in the j-th step obtained by the Gibbs Sampling process.

3. METHODOLOGY

3.1. Research Data

This study uses secondary data, namely data published by the World Bank through the results of the Indonesian Household Socio-Economic Survey (SUSETI) in 2011 which had previously been used in Cintiani (2017). The sample used in the study was 1009 households with one response variable and four predictor variables. The response variable used is Cigarette Consumption (y), while the predictor variables are: Income (x1), Number of Family Members (x2), Expenditure per Capita (x3), and Age (x4).

3.2. Data Analysis Methods

In an effort to build a sample into the targeted distribution, in this study a large iteration was used. Burn (number of Gibbs Sampling iterations before the results are used) is selected as 1000. As for run (number of Gibbs Sampling iterations performed) is 11000. As stated in Walsh [16], the first 1000 - 5000 iterations of Gibbs Sampling are usually selected as burn . In Gibbs Sampling there is no rejection criteria for samples, all samples obtained in the iteration process can be used.

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

This study analyzed the regression model on three quantiles, τ = 0.25, τ = 0.5, and τ = 0.75. Each quantile will be modeled using Tobit Quantile Regression and Tobit Quantile Regression with Adaptive Lasso Penalty. Tobit Quantile Regression and Tobit Quantile Regression with Adaptive Lasso Penalty were analyzed using the Brq package in R software.

4. RESULT AND DISCUSSION

4.1. Estimation of Regression Coefficient Parameter

Tobit Quantile Regression (BTQR) and Tobit Quantile Regression with Adaptive Lasso Penalty (BALTQR) will be analyzed using Bayesian approach. After the complete conditional prior and posterior distributions are determined, then the solution is carried out using the MCMC Gibbs Sampling method. The following is a summary of the regression coefficient parameter values in both models.

Table 1. Parameters of Regression Coefficient

Quantile	Model	β_0	β_1	β_2	β_3	β_4
0,25	BTQR	-7,2057	0,000196	5,01602	0,000675	-0,137506
	BALTQR	-2,5024	0,000452	4,5765	0,0000942	-0,183185
0,5	BTQR	10,78099	0,0051228	3,99758	0,00788	0,09874
	BALTQR	7,20972	0,0052409	4,19531	0,00807	0,1532913
0,75	BTQR	12,549104	0,0108402	6,3792894	0,0148299	0,2289465
	BALTQR	7,7889775	0,010823	6,68544	0,015354	0,2963156

4.2. Comparison of Tobit Quantile Regression and Tobit Quantile Regression with Adaptive Lasso Penalty

Comparison of the goodness of the model in this study will be carried out by finding the Mean Absolute Deviation (MAD) in each quantile, which is formulated:

$$\frac{\sum_{i=1}^{n} |y_i - \widehat{y_i}|}{n} \tag{14}$$

The smaller the MAD value, the more accurate the model will be. MAD was calculated using R software. Summary of MAD for each quantile in both models is shown in Table 2 below.

Table 2. *Mean Absolute Deviation* (MAD)

Quantile	Model	MAD
0,25	BTQR	45,23
	BALTQR	44,95
0,5	BTQR	35,59
	BALTQR	35,59
0,75	BTQR	43,85
	BALTQR	43,81

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

Obtained at the 0.25th and 0.75th quantiles, the Tobit Quantile Regression with Adaptive Lasso Penalty model has a smaller MAD than the Tobit Quantile Regression without using a penalty. As for the 0.50th quantile, both MAD values are the same.

5. CONCLUSION

Based on the value of Mean Absolute Deviation, in the case of Household Expenditure for Cigarette Consumption, Tobit Quantile Regression with Adaptive Lasso Penalty produces better parameter estimates than Tobit Quantile Regression at quantiles 0.25 and 0.75. Further research can develop the model by replacing the Adaptive Lasso penalty with another penalty such as Ridge or Elastic Net and using a larger data set.

REFERENCES

[1] J. Tobin, "Estimation of Relationships for Limited Dependent Variables," *Econometrica*, vol. 26, no. 1, pp. 24–36, 1985, [Online]. Available: https://www-jstor-org.ezproxy.ugm.ac.id/stable/1907382?seq=1.

[2] J. L. Powell, "Censored regression quantiles," *J. Econom.*, vol. 32, no. 1, pp. 143–155, 1986, doi: https://doi.org/10.1016/0304-4076(86)90016-3.

[3] R. Tibshirani, "Regression Shrinkage and Selection Via the Lasso," *J. R. Stat. Soc. Ser. B*, vol. 58, no. 1, pp. 267–288, 1996, doi: 10.1111/j.2517-6161.1996.tb02080.x.

[4] H. Zou, "The Adaptive Lasso and Its Oracle Properties," *J. Am. Stat. Assoc.*, vol. 101, no. 476, pp. 1418–1429, 2006, doi: 10.1198/016214506000000735.

[5] Cintiani, "Pemodelan Regresi Kuantil (Studi Kasus Pengeluaran Rumah Tangga untuk Konsumsi Rokok)," Program Magister Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, 2017.

[6] R. Koenker, *Quantile Regression*. Cambridge: Cambridge University Press, 2005.

[7] R. Alhamzawi and K. Yu, "Conjugate Priors and Variable Selection for Bayesian Quantile Regression," *Comput. Stat. Data Anal.*, vol. 64, pp. 209–219, 2012, doi: https://doi.org/10.1016/j.csda.2012.01.014.

[8] W. H. Greene, *Econometric Analysis Internastional Edition*. UK: Pearson Education, 2012.

[9] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning - with Applications in R*. New York: Springer, 2013.

[10] F. Mosteller and J. Tukey, *Data Analysis and Regression: A Second Course in Statistics*. Boston, USA: Addison-Wesley, 1977.

[11] R. Koenker and G. Bassett, "Regression Quantiles," *Econometrica*, vol. 46, no. 1, p. 33, 1978, doi: 10.2307/1913643.

[DOI: 10.26714/jsunimus.10.2.2022.25-33]

[12] R. Alhamzawi, “Tobit Quantile Regression with the adaptive Lasso penalty,” no. 1681 6870, pp. 1–19, 2013, [Online]. Available: https://www.researchgate.net/publication/258834905_Tobit_Quantile_Regression_with_the_adaptive_Lasso_penalty.

[13] H. Kozumi and G. Kobayashi, “Sampling Methods for Bayesian Quantile Regression,” *J. Stat. Comput. Simul.*, vol. 81, no. 11, pp. 1565–1578, 2011, doi: https://doi.org/10.1080/00949655.2010.496117.

[14] D. F. Andrews and C. L. Mallows, “Scale Mixtures of Normal Distributions Author,” *Society*, vol. 36, no. 1, pp. 99–102, 2010.

[15] D. van Ravenzwaaij, P. Cassey, and S. D. Brown, “A simple introduction to Markov Chain Monte–Carlo sampling,” *Psychon. Bull. Rev.*, vol. 25, no. 1, pp. 143–154, 2018, doi: 10.3758/s13423-016-1015-8.

[16] B. Walsh, “Markov Chain Monte Carlo and Gibbs Sampling,” *Lect. Notes EEB 581*, vol. 581, no. April, p. 24, 2004, [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.131.4064.

[DOI: 10.26714/jsunimus.10.2.2022.25-33]