African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 1, 2023 (pp. 1-11)
www.abjournals.org

QUANTILE REGRESSION FOR COUNT DATA AS A ROBUST ALTERNATIVE TO NEGATIVE BINOMIAL REGRESSION

Nwakuya Maureen T. and Nkwocha Chibueze C.

[1]Department of Mathematics and Statistics, University of Port Harcourt

E-mail: maureen.nwakuya@uniport.edu.ng

[2]Department of Mathematics and Statistics, University of Port

Cite this article:

Nwakuya M.T., Nkwocha C.C. (2023), Quantile Regression for Count Data as a Robust Alternative to Negative Binomial Regression. African Journal of Mathematics and Statistics Studies 6(1), 1-11. DOI: 10.52589/AJMSS-CLQ73EUZ

Manuscript History

Received: 21 Dec 2022

Accepted: 18 Jan 2023

Published: 2 Feb 2023

ABSTRACT: *The study investigated the robustness of Quantile regression of count data over negative binomial regression, when there is overdispersion and presence of outlier. The study made use of a complete data and the data with 30% missing data which was imputed using Multiple Imputation by Chain Equation (MICE) in R and also an outlier was injected into the data during imputation of missing values. The Quantile Regression and Negative Binomial Regression estimates were compared and their model fits were also compared. Results showed that the quantile regression for count data provided a better model estimate with both complete data and data with multiple imputed value with comparison to the negative binomial regression in terms of AIC, BIC RMSE and MSE. Hence, Quantile Regression is better than the negative binomial regression when the researcher is interested in the effect of the independent variable on different points of the distribution of the response variable and when there is overdispersion and presence of an outlier.*

KEYWORDS: Count Data, Quantile Regression, Negative Binomial Regression, Missing Data, Overdispersion and Outlier.

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 1, 2023 (pp. 1-11)
www.abjournals.org

INTRODUCTION

On a daily basis, count data is used in transactions. Some statistical analysis or modelling is essential to gain a deeper understanding of such data and to extract the data's key information. Different count data may have unique properties that prevent their use with specific count data models. A foundation for the study of count data is provided by the Poisson regression model. Many practitioners use the Poisson model when analyzing count data, frequently without verifying that all of its underlying assumptions are correct. In place of fitting a model naively, one should examine whether the model's assumptions are met before choosing a model to match a particular set of data. There are instances where these presumptions are broken, necessitating the use of an alternative model. Negative binomial and hurdle models are a couple of the different models that can be taken into account when modelling count data. It is challenging to decide which of the statistical models to use for processing count data based solely on intuition.

During the data collection procedure, incomplete data commonly occurs. They may result from incomplete records, such as when people refuse to participate in a census or survey and withhold information. The use of partial data has negative effects on the reliability of the findings and the validity of the conclusions (Graham, 2009). Additionally, many complete-data analysis techniques are not relevant when there are missing data, and current software may not function properly. Imputing the missing values to complete the datasets can be advantageous. Preferably data collectors use this imputation technique rather than data analysts since they may have additional information that increases the precision of imputation models (Allison, 2000). When the data collector only fills in the missing values once, data analysts use the imputations as though they were actual values, which underestimate the level of uncertainty in the inferences (Allison, 2000). Rubin suggests multiple imputation, which preserves the benefits of imputation techniques while also enabling analysts to take uncertainty into account throughout the imputation process (Allison, 2003). In multiple imputation, the data collector takes a selection of various values from a predictive distribution for the missing elements and makes repeated releases of the finished datasets. Data collectors can provide secondary data analyzers with principled interval estimates in this way (Allison, 2000). Multiple imputation (MI) addresses the drawbacks of single imputation by adding a second type of error known as "between imputation error" based on variations in parameter estimations across the imputation. Each missing item is replaced with two or more acceptable values to illustrate a range of possibilities. MI is a simulation-based approach. Instead of trying to replicate each missing value as closely as feasible to the actual ones, its aim is to handle missing data so that reliable statistical inference can be performed (Schafer, 1997).

In removing some of its restrictions, multiple imputation has the same ideal qualities as maximum likelihood (ML). Any type of data and model can be used with it while using conventional software. Every time you use it, it generates a different estimate (ideally just slightly different), which can result in scenarios where various researchers obtain various results from the same data while using the same methodology.

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 1, 2023 (pp. 1-11)
www.abjournals.org

Quantile Regression

The midpoint of the population of interest is the limit of the predictor effects in linear regression analysis. This constraint may only give a partial picture, which may lead to potentially incorrect inferences when the assumptions of the conventional linear regression are broken. Koenker and Bassett (1978) proposed QR, which involves estimation of the entire distribution of the response variable dependent on selected linear predictors, as a way to improve regression analysis. This implies that the computation of a comprehensive collection of values that represent the conditional quantiles will take the place of the computation of a single value that represents the conditional mean. These conditional quantiles can provide a more complete picture of the current connection.

The conditional quantile regression approach is the foundation of the QR framework, which has been widely adopted in the literature on applied economics. The conditional regression approach is used to determine how a predictor will change at a certain location on the response distribution depending on the values of other predictors. Typically, conditional quantile regression may produce results that are difficult to understand in the context of policy or population, contrary to the unconditional quantile regression technique, which marginalizes the predictor impact over the distributions of other predictors in the model and yields results that are easier to comprehend (Borah & Basu, 2013).

The existing quantitative evidence from more typical mean-based analyses might be expanded and deepened using QR approaches (Wei et al., 2019). Chernozhukov et al. (2022) detail a variety of innovative procedures for accelerating quantile regression computations when the necessity to estimate a high number of different quantiles is of interest. Nwakuya (2020) used a Bayesian ordinal quantile regression approach to evaluate the mental health of undergraduate students based on age. Quantile regression has been used in various research fields.

Koenker and Bassett (1978) proposed a method to calculate the quantile, given as:

$$\widehat{Q_Y}(\tau) = argmin_{\xi_\tau \in R} \left\{ \sum_{i \in \{i|Y_i \geq \xi_\tau\}} \tau |Y_i - \xi_\tau| + \sum_{i \in \{i|Y_i < \xi_\tau\}} (1 - \tau)|Y_i - \xi_\tau| \right\} \qquad (1)$$

Koenker and Bassett (1978) remarked that "the case of the median (τ = 1/2) is, of course, well known, but the general result has languished in the status of curiosum." We use the indicator function (I(A) = 1 if A is true, and I(A) = 0 if otherwise) to introduce the so-called check function:

$$\rho_\tau(e) = e\big(\tau - I(e < 0)\big) \qquad 0 < \tau < 1 \qquad (2)$$

The check function allows us to reformulate the objective function of (3.1) as a single expression:

$$\widehat{Q_Y}(\tau) = argmin_{\xi_\tau \in R} \sum_i \rho_\tau(Y_i - \xi_\tau) \qquad (3)$$

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

Notice that since $\widehat{Q_Y}(\tau)$ is not a continuous function of $\hat{q}$ at $\tau = \hat{q}$, Slutsky's theorem cannot be invoked to claim the consistency for $\widehat{Q_Y}(\tau)$ because of the discrete nature of our data.

According to Machado (2005), the De Moivre-Laplace Theorem, for large n, the probability $Pr\ (\widehat{Q_Y}(\tau) = 0)$ can be approximated by 1- $\phi\left(\sqrt{n}\frac{\tau-q}{\sqrt{q(1-q)}}\right)$.

Because the estimated quantiles are random variables with the same support as Y, their asymptotic distribution cannot be normal when the data is discrete.

Machado (2005) considered a new random variable Z which is constructed by adding to Y a random variable U uniformly distributed in an interval (0,1); this is known as jittering, that is, Z = Y + U. Obviously, Z is a continuous random variable with positive density on the interval (0, 2) and $\widehat{Q_Z}(\tau)$ is defined by:

$$\widehat{Q_Z}(\tau) = \{\frac{\tau}{q} \qquad q > \tau \; 1 + \frac{\tau-q}{(1-q)} \qquad q < \tau \tag{4}$$

Interestingly the feature of the random variable Z is such that there is a one-to-one relationship between $\widehat{Q_Y}(\tau)$ and $\widehat{Q_Z}(\tau)$. $Q_Y(\tau) = \lceil Q_Z(\tau) - 1\rceil$ where $\lceil\tau\rceil$ denotes the ceiling function which returns the smallest integer greater than, or equal to τ. Therefore, data about the quantile of Y, the variate of interest, can be acquired from $\widehat{Q_Z}(\tau)$. The minimization of the absolute deviations, $\sum_{i=1}^{n} \quad |y_i - X_i\beta|$. This eases the impact of outliers in the response data providing a better fit for the majority of observations.

LITERATURE/THEORETICAL UNDERPINNING

Nwakuya and Nwabueze (2022) applied Poisson regression and negative binomial regression on count data from road accident fatality during COVID-19 hit era in Nigeria. The negative binomial regression was seen to perform better due to the presence of overdispersion, rendering Poisson regression inadequate. The comparison was based on mean–variance relationship, goodness of fit test, AIC and BIC. Lee and Neocleous (2010), in their research "Bayesian Quantile Regression for Count Data with Application to Environmental Epidemiology," presented a Bayesian quantile regression model for count data and applied it in the field of environmental epidemiology. Their methods were applied to a new study of the relationship between long-term exposure to air pollution and respiratory hospital admissions in Scotland. They observed a decreasing relationship between pollution and the τth quantile of the response distribution, with a relative risk ranging between 1.023 and 1.070. Fuzi et al. (2010), in their paper "Bayesian Quantile Regression Model for Claim Count Data," applied Bayesian quantile regression model for the Malaysian motor insurance claim count data to study the effects of change in the estimates of regression parameters (or the rating factors) on the magnitude of the response variable (or the claim count). They also compared the results of quantile regression models from the Bayesian and frequentist approaches and the results of mean regression models from the Poisson and negative binomial.

Comparison from Poisson and Bayesian quantile regression models shows that the effects of vehicle year decrease as the quantile increases, suggesting that the rating factor has lower risk

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

for higher claim counts. On the other hand, the effects of vehicle type increase as the quantile increases, indicating that the rating factor has a higher risk for higher claim counts. Lv and Xu (2017), in their study "A Panel Data Quantile Regression Analysis of the Impact of Corruption on Tourism," adopted the quantile regression model to provide a broad description of the relationship between tourism demand and corruption across the demand distribution. Their empirical results indicate that the nonlinear relationship between corruption and tourism demand is only significant at the 50th and 75th quantiles. They also found a significant positive relationship between income and tourism demand across various quantiles, and the strength of the relationship is larger at lower demand levels. Grund et al. (2018), in their paper "Multiple Imputation of Missing Data for Multilevel Models: Simulations and Recommendations" based on theoretical arguments and computer simulations, provided guidance using Multiple Imputation (MI) in the context of several classes of multilevel models, including models with random intercepts, random slopes, cross-level interactions (CLIs), and missing data in categorical and group-level variables.

METHODOLOGY

This study adopted the method of estimating the quantile regression and negative binomial regression parameters. To do this, the study considered negative binomial estimates, quantile regression estimates, root mean square, and mean square error. Data used for this study was sourced from National Bureau of Statistics from Federal Road Safety Commission for 2021, on road accidents in the 36 states of the Nigeria plus FCT. The response is the number of people due to road accidents and the predictor variables are: number of cases, number of those involved in an accident and the number of people that got injured. The analysis began with a set of complete data; then, 30% of the values were made missing assuming that missingness is completely at random (MCAR). The missing values were later imputed using a multiple imputation method in the process and an outlier was injected. Analysis was conducted using the quantile regression and negative binomial regression technique on both the complete data and the data that has an outlier. Four methods of model comparison criteria, namely Root Mean Square Error (RMSE), Mean Square Error (MSE), Akaike Information Criteria (AIC) and Bayesian Information Criteria (BIC), were applied to compare the quantile regression and negative binomial regression. The preceding sections describe the various methods and then results and conclusion follow.

The Negative Binomial Regression Model

The introduction of negative binomial was due to overdispersion which violates the assumption of Poisson regression. In negative binomial regression, the mean of y is determined by the exposure time t and a set of k regressor variables (the *x's*). The expression relating these quantities is

$$\mu_i = exp\ exp\ (In(t_i) + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_k x_{ki} \qquad (5)$$

Often, $X_1 \equiv 1$, in which case β_1 is called the intercept. The regression coefficients $\beta_1, \beta_2, \ldots, \beta_k$ are unknown parameters that are estimated from a set of data. Their estimates are symbolized as $b_1, b_2, \ldots, b_k$.

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 1, 2023 (pp. 1-11)

Using this notation, the fundamental negative binomial regression model for an observation *i* is written as:

$$\Pr (Y = y_i/\mu_i, \alpha) = \frac{\Gamma(y_i+\alpha^{-1})}{\Gamma\alpha^{-1}\Gamma(y_i+1)} (\frac{1}{1+\alpha\mu_i})^{\alpha^{-1}} (\frac{\alpha\mu_i}{1+\alpha\mu_i})^{y_i} \quad (6)$$

where

$\mu_i = t_i$

$\alpha = \frac{1}{v}$

Comparison Criteria

Akaike's Information Criteria (AIC)

One of the most commonly used information criteria is the Akaike's Information Criteria (AIC). The idea of AIC (Akaike, 1973) is to select the model that minimizes the negative likelihood penalized by the number of parameters as specified in the equation:

$$AIC = -2 \log (L) + 2p \quad (7)$$

where L refers to the likelihood under the fitted model and p is the number of parameters in the model. Specifically, AIC is aimed at finding the best approximating model to the unknown true data generating process and its applications.

Mean Squared Error (MSE)

The MSE measures the average of the square deviation between the fitted values with the actual data observation. The mean-squared error is determined by the residual sum of squares resulting from comparing the predictions $\hat{y}$ with the observed outcomes y:

$$MSE = \frac{1}{n}\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \quad (8)$$

Bayesian Information Criteria (BIC)

Another widely used information criterion is the BIC. BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models: Kass and Rafftery (1995). BIC is presented thus:

$$BIC = -2 \log (L) + p \log (n) \quad \text{where } n>p \quad (9)$$

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 6, Issue 1, 2023 (pp. 1-11)

RESULTS/FINDINGS

Table 1: Deviance result

Overdispersion		**DF**
Null deviance	3495.11	36
Residual deviance	509.46	33

The results show the presence of overdispersion.

Quantile Estimates on a Specified Quantile for the Count Data Model with and without Multiple Imputation of Missing Data

Table 2: Quantile estimates on a specified quantile for the count data model with complete data

	25%	**50%**	**75%**	**95%**
$\boldsymbol{\beta_0}$	3.7892e+00***	3.9707e+00***	4.28600128***	4.6733e+00***
$\boldsymbol{\beta_1}$	-1.2461e-03***	-1.7657e-03***	-0.00222374***	-1.9677e-03***
$\boldsymbol{\beta_2}$	2.7372e-04***	3.1949e-04***	0.00051783***	3.6893e-04***
$\boldsymbol{\beta_3}$	6.6060e-04***	7.2476e-04***	0.00043977***	4.9369e-04***

*** p-value significant at 0.05 showing all the values are significant.

The results show a significant effect of all variables at all quantiles.

Table 3: Quantile estimates on a specified quantile for the count data model with 30% missing data imputed using multiple imputations

	25%	**50%**	**75%**	**95%**
$\boldsymbol{\beta_0}$	3.8970e+00***	4.2714e+00***	4.4050e+00***	4.6500e+00***
$\boldsymbol{\beta_1}$	-1.3081e-03***	-1.6506e-03***	-2.0447e-03***	-3.0335e-03***
$\boldsymbol{\beta_2}$	2.8615e-04***	3.1593e-04***	5.6021e-04***	7.4902e-04***
$\boldsymbol{\beta_3}$	6.1071e-04***	5.3809e-04***	2.5991e-04***	8.1016e-05***

*** p-value significant at 0.05 showing all the values are significant.

The results show a significant effect of all variables at all quantiles.

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

Negative Binomial Estimates for the Count Data Model with Complete Data and 30% Missing Data with Multiple Imputation

Table 4: Negative binomial estimates for the count data model with complete data

ESTIMATE	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$
	4.0553655***	-0.0016260***	0.0003775***	0.0005396***

*** p-value significant at 0.05 showing all the values are significant.

The results show a significant effect of all variables. It can be noted that the results are similar to results of the 50th quantile.

Table 5: Negative binomial estimates for the count data model with 30% missing data with multiple imputation

ESTIMATE	$\boldsymbol{\beta_0}$	$\boldsymbol{\beta_1}$	$\boldsymbol{\beta_2}$	$\boldsymbol{\beta_3}$
	4.2286220***	-0.0018254***	0.0004527***	0.0003711***

*** p-value significant at 0.05 showing all the values are significant.

The results show a significant effect of all variables. It can also be noted that the results are similar to results of the 50th quantile; hence, we can say that the median regression is similar to the mean regression.

Fitting the Quantile Regression Model with 30% Multiple Imputations of Missing Data

25th Percentile of $\hat{y}$ = 3.8970 - 0.0013081 β_1+ 0.00028615β_2 + 0.0006107 β_3

50th Percentile of $\hat{y}$ = 4.2714 - 0.0016506 β_1 + 0.000031593β_2+ 0.00053809 β_3

75th Percentile of $\hat{y}$ = 4.4050 - 0.0020447 β_1 + 0.00056012 β_2+ 0.00025991 β_3

95th Percentile of $\hat{y}$ = 4.4050 - 0.0020447β_1 + .00056012 β_2+ 0.00025991 β_3

Fitting the Negative Binomial Regression Model with 30% Multiple Imputations of Missing Data

$\hat{y}$ = Exp (4.2286220 - 0.0018254 β_1 + 0.0004527 β_2+ 0.0003711 β_3)

Comparison of the AIC's, BIC, MSE and RMSE of Both Models

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

Table 6: Quantile estimates on a specified quantile for the count data model with complete data

QUANTILE REGRESSION MODEL					NEGATIVE BINOMIAL MODEL
	25%	50%	75%	95%	
AIC	324.6557	352.391	398.1594	373.4109	390.65
BIC	5224.061	11038.84	37995.7	19471.6	31003.35
MSE	5209.617	11024.4	37981.26	19457.15	31001.79
RMSE	72.17768	104.9971	194.8878	139.4889	176.071

The result in Table 6 shows clearly that the quantile regression presented lower values for AIC, BIC, MSE and RMSE, hence a better model.

Table 7: Quantile estimates on a specified quantile for the count data model with imputed values and an outlier

QUANTILE REGRESSION MODEL					NEGATIVE BINOMIAL MODEL
	25%	50%	75%	95%	
AIC	329.2749	321.2383	398.3156	411.659	1943
BIC	5916.79	4764.442	38156.4	54718.95	5.17×10^{22}
MSE	5902.346	4749.998	38141.96	54704.51	5.17×10^{22}
RMSE	76.8267	68.92	195.300	233.990	2262×10^{8}

The result in Table 6 shows clearly that the quantile regression presented lower values for AIC, BIC, MSE and RMSE, hence a better model.

DISCUSSION

The ratio of the deviance to the DF is 15.438, hence over dispersion warranting the use of negative binomial regression. Table 2 showed the quantile regression estimates of count data with complete data. The quantile parameter estimates obtained for each independent variable is given as -1.2461e-03 β_1, 2.7372e-04β_2 and 6.6060e-04β_3 at the 25th Quantile, -1.7657e-03 $\beta_1$3.1949e-04 β_2 and 7.2476e-04 β_3 at the 50th Quantile, -0.00222374 β_1, 0.00051783 β_2 and 0.00043977 β_3 at the 75th Quantile and -1.9677e-03 β_1, 3.6893e-04 β_2 and 4.9369e-04 β_3 at the 95th quantile. Table 3 shows the quantile regression estimates of count data with 30% multiple imputations of missing data. The parameter estimates obtained for each independent variable is given as -1.3081e-03β_1, 2.8615e-04β_2 and 6.1071e-04β_3 at the 25th quantile, -1.7657e-03β_1, 3.1949e-04β_2 and 7.2476e-04β_3 at the 50th quantile, -0.00222374β_1, 0.00051783β_2 and 0.00043977β_3 at the 75th quantile, and -1.9677e-03β_1, 3.6893e-04β_2 and 4.9369e-04β_3 at the 95th quantile. Table 4 shows the negative binomial estimates for the count data model with complete data. The estimates for each independent variable is given as -0.0016260 β_1, 0.0003775 β_2 and 0.0005396β_3**.** Table 5 shows the negative binomial estimates for the count data model with multiple imputation and an outlier. The estimates for each independent variable are given as -0.0018254β_1, 0.0004527β_2 and 0.0003711β_3. Table 6

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

revealed the comparison between the two models; from the results, we can see that AIC, BIC, RMSE, and MSE of quantile regression were smaller than those of negative binomial values and were lower than those of the negative binomial regression at all quantile levels except for the 75th quantile where the negative binomial presented a better result. Table 7 revealed the comparison between the two models. The quantile regression presented better results with both complete data and multiple imputed data than the negative binomial regression in terms of AIC, BIC, RMSE, and MSE.

IMPLICATION TO RESEARCH AND PRACTICE

This work has brought to light the robustness of quantile regression over the negative binomial regression. Even in the presence of outliers, the quantile regression still proved to be a better method. The implication of this research is that, researchers working with count data in the presence of overdispersion and outlier should consider employing the quantile regression as an alternative, especially if their interest is in having a picture of the whole distribution and not just the mean effect.

CONCLUSION

The work started by checking for overdispersion, which was found to be present, as shown in Table 1. The data was analyzed using both quantile regression and negative binomial. Then 30% values were made missing in the data, assuming missing at random. The incomplete data was completed using multiple imputation method while injecting an outlier. Both data were analysed and from the results, the quantile regression in general presented better results based on the model comparison criteria used for both sets of data. In line with the findings, we can conclude that Quantile Regression is robust and a better alternative to the negative binomial regression when the researcher is interested in the effect of the independent variable on different points of the distribution of the response variable and when there is overdispersion and presence of an outlier.

Future Research

This research is limited to count data with overdispersion and an outlier. It can be extended to a situation where there is more than one outlier. Also, quantile regression can be compared to Poisson regression in situations of no overdispersion.

REFERENCES

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In: Petrov, B.N. and Csaki, F., Eds., International Symposium on Information Theory, pp 267-281.

Allison P., (2003). Handling Missing Data by Maximum Likelihood. SAS Global Forum 2012. Development (Version 9.0).

Allison, P. D. (2000). Multiple imputation for missing data: A cautionary tale. Sociological methods & research, *28*(3), 301-309.

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ

Borah, B, J. and Basu, A. (2013). Highlighting Differences Between Conditional and Unconditional Quantile Regression Approaches Through an Application to Assess Medication adherence.

Chernozhukov, V., Fernández-Val, I. &Melly, B. (2022).Fast algorithms for the quantile regression process. Empir Econ 62**,** 7–33. https://doi.org/10.1007/s00181-020-01898-0

Fuzi, M. F. M., Jemain, A. A., & Ismail, N. (2016). Bayesian quantile regression model for claim count data. *Insurance: Mathematics and Economics*, *66*, 124-137.

Graham JW. (2009). Missing data analysis: making it work in the real world. Annu Rev Psychol; 60: 549-76.

Grund, S., Lüdtke, O., & Robitzsch, A. (2018). Multiple imputation of missing data for multilevel models: Simulations and recommendations. *Organizational Research Methods*, *21*(1), 111-149.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. Journal of the American statistical association, 90(430), 773-795.

Koenker R. and Bassett G. (1978). Regression Quantiles. Econometrica 46(1), 33-50.

Lee, D., & Neocleous, T. (2010). Bayesian quantile regression for count data with application to environmental epidemiology. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(5), 905-920.

Lv, Z., & Xu, T. (2017). A panel data quantile regression analysis of the impact of corruption on tourism. *Current Issues in Tourism*, *20*(6), 603-616.

Machado José A. F. and J. M. C. Santos Silva (2005). Quantiles for Counts. Journal of the American Statistical Association, 100,1226-1237.

Nwakuya, M. T. (2020). Assessment of Mental Health of Undergraduate Students Based on Age: A Bayesian Ordinal Quantile Regression Approach. Quarterly Journal of Econometrics Research6(1), 12-17.

Nwakuya M. T. and Nwabueze C. J. (2022). A Negative Binomial Regression on Road Accident Fatalities During COVID-19 Hit Era in Nigeria. International Journal of Statistical Distributions and Applications; 8(3): 40-46.

Schafer J. L. (1997), “Analysis of Incomplete Multivariate Data”, Chapman & Hall, London, pp 87-95

Schafer, J. L., & Olsen, M. K. (1998). Multiple imputation for multivariate missing-data problems: A data analyst's perspective. Multivariate behavioral research, *33*(4), 545-571.

Wei, Y., Kehm, R.D., Goldberg, M. and Terry, M. B. (2019). Applications for Quantile Regression in Epidemiology. Curr Epidemiol Rep **6,** 191–199. https://doi.org/10.1007/s40471-019-00204-6

Article DOI: 10.52589/AJMSS-CLQ73EUZ
DOI URL: https://doi.org/10.52589/AJMSS-CLQ73EUZ