Scientia Africana, Vol. 21 (No. 1), April, 2022. Pp 67-76 https://dx.doi.org/10.4314/sa.v21i1.6
 ISSN 1118 – 1931

MANLY TRANSFORMATION IN QUANTILE REGRESSION: A COMPARISON OF TWO TRANSFORMATION PARAMETER ESTIMATORS

[1]Nwakuya, M. T. and [2]Nkwocha, C. C.
[1&2]Department of Mathematics and Statistics, University of Port Harcourt
Email: [1]maureen.nwakuya@uniport.edu.ng, and [2]chibueze.nkwocha@uniport.edu.ng

Received: 05-02-2022
Accepted: 30-03-2022

ABSTRACT

This study implements the Manly transformations for normalization of variables in quantile regression analysis.The transformation parameter was estimated using two different methods namely; the maximum likelihood estimation (MLE) method and the two-step estimation method by Chamberlain and Buchinsky(CBTS).The transformation parameters obtained using the two different methods were used for the Manly transformation of data with outliers and data without outliers. The methods were applied to a quantile regression analysis at different quantiles (0.25, 0.50, 0.75, 0.95). Based on our findings, for data without outliers, the 25th quantile model was seen to be the best fit model compared to the other quantiles for the CBTS method with AIC=-43.46279, BIC=20.75212 and MSE=0.70956, while for the MLE the 50th quantile model was seen to be the best fit model with AIC=-348.3657, BIC=20.13548, and MSE=0.00864. Considering data with outliers the 25th quantile model was still seen to be the best fit model compared to the other quantiles for the CBTS method with AIC=-48.5671, BIC=21.8321 and MSE=0.92341, while for the MLE the 50th quantile model was still seen to be the best fit model with AIC=988.6763, BIC=710.09, and MSE=690.7965. Comparison of both methods for data without outliers the study concludes that the estimation of the transformation parameter using the MLE produced better results with lower AIC, BIC and MSE at all quantiles and for data with outliers the study concludes that the estimation of the transformation parameter using CBTS produced better results with lower AIC, BIC and MSE results as is shown in table (3.5) and table (3.6) respectively.

Keywords: Manly Transformation, Quantile Regression, Maximum Likelihood EstimationandChamberlain and Buchinsky Two Stage (CBTS) method.

INTRODUCTION

Conventional regressions assume that the covariates effects are constant across the population, which in most cases is not true . Quantile regression is an effective method to estimate not only the center, but also the lower or, upper tail of the conditional distribution of interest. An important advantage of quantile regression is its flexibility to describe the complete relationship between response and predictors. The Gaussiandistribution provides the foundation for most statistical methodological frameworks. As a result, statisticians and academics have noticed the widespread use of mappings that 'Gaussianize' data since Box and Cox (1964) released their seminal normalization transformation. Box and Cox (1964) transformation method is among the exponential family transformation, likewise the manly transformation method and Yeo-Johnson transformation amongst others. The Yeo-Johnson transformation was introduced by Yeo and John (2000) as an extension of the Box-Cox transformation to accommodate negative response values since the Box-Cox transformation is restricted to only positiveresponse values. Manly (1976) stated that an exponential transformation is quite effective at turning a skewed unimodal distribution into a nearly symmetric normal distribution.More recently, Watthanacheewakul (2020) proposed a modified BoxCox transformation as an appropriate method to transform right-skewed data to become normal.These methods mentioned above employ the use of maximum likelihood method in estimating their transformation parameter and it can be particularly sensitive to outliers.Chamberlain (1994) and Buchinsky (1994) proposed a two stage method(CBTS) for estimating the transformation parameter in quantile regression, this method incorporates the equi-variance property of quantiles. Quantile regression has been seen to be robust tooutliers (outliers are data points that differ significantly from other observations).QR models can detect heterogeneous effects of covariates at different quantiles of the outcome, but also offer more robust and complete estimates compared to the mean regression, when the normality assumption violated or outliers and long tails exist. These advantages make QR attractive and are extended to apply for different types of data, including independent data, time-to-event data and longitudinal data Huang, Q (2017). In this work we tried to implement the CBTS and maximum likelihood method in estimating the transformation parameter for manly transformation in a quantile regression analysis, using data without outliers and data with outlier in other investigate the robustness of the CBTS to outliers and compare its results to that of maximum likelihood estimation (MLE) method.

QUANTILE REGRESSION

Linear regression analysis confines the covariates effects to be centered across the population of the response. This confinement might give an incomplete picture and thus can lead to possibly wrong conclusions as soon as all assumptions of the classical linear regression model are not met. Quantile regression (QR) has given a solution to; how to go further question, in regression analysis. This solution was proposed by Koenker and Bassett (1978). They introduced a new method labelled “quantile regression” that allows the estimation of the entire distribution of the response variable conditional on any set of linear covariates. In other words, the calculationof a single value, the conditional mean, is replaced by the computation of a whole set of numbers for the conditional quantileswhich are able to give a more complete picture of the underlying interrelations. QR framework that has pervaded the applied economics literature is based on the conditional quantile regression method. It is used to assess the impact of a covariate on a quantile of the outcome conditional on specific values of other covariates. In most cases, conditional quantile regression may generate results that are often not generalizable or interpretable in a policy or population context. In contrast, the unconditional quantile regression method provides more interpretable results as it marginalizes the effect over the distributions of other covariates in the model Borah and Basu (2013). QR

Scientia Africanა, Vol. 21 (No. 1), April, 2022. Pp 67-76 https://dx.doi.org/10.4314/sa.v21i1.6
© Faculty of Science, University of Port Harcourt, Printed in Nigeria ISSN 1118 – 1931

methods have the potential to deepen and expand the existing quantitative evidence from more common mean-based analyses Wei et al (2019). Chernozhukov et al (2022) describe several new methods for speeding up quantile regression computations when it is desirable to estimate a large number of distinct quantiles.Quantile regression is being in many areas of research, Nwakuya (2020) applied a Bayesian ordinal quantile regression approach in assessing the mental health of undergraduate students based on Age.

To present the mathematical notations, consider a classical linear regression model:

$$y_i = x_i'\beta + e_i \qquad i = 1, \dots, n \qquad (1)$$

Where $y_i$is the response variable, $x_i$is the covariates, $\beta \sim N(\beta, (X^T X)^{-1}\sigma^2)$ is the covariates effect and $e_i \sim N(0, \sigma_e^2)$ is the error term. Assume that the expected value of the error term conditional on the covariates is zero $(E(e_i|x_i) = 0)$, then the conditional mean ofy_i with respect to $x_i$is

$$E(y_i|x_i) = x_i'\beta \qquad (2)$$

The covariates effect β can be estimated by the well-known method of least squares:

$$\hat{\beta} = \underbrace{argmin}_{\beta \in \mathbb{R}^k} \sum_i (y_i - x_i'\beta)^2 \qquad (3)$$

A solution to equation (3) is given by$\hat{\beta} = (X'X)^{-1}X'y$.

Assume that $y_i = x_i'\beta_\tau + e_{i,\tau}$and that not the expected value, but the τ-th quantile of theerrorterm conditional on thecovariatess is zero$(Q_\tau(e_{i,\tau}|x_i) = 0)$. Then it is ready to see that the τ-th conditional quantile of y_i with respect to $x_i$can be written as:

$$Q_\tau(y_i|x_i) = x_i'\beta_\tau \qquad (4)$$

hence for any τ in the interval (0, 1), the parameter vector β_τ can be estimated by:

$$\widehat{\beta_\tau} = \underbrace{argmin}_{\beta_\tau \in \mathbb{R}^k} \left\{ \sum_{i \in \{i | y_i \geq x_i'\beta_\tau\}} \tau |y_i - x_i'\beta_\tau| + \sum_{i \in \{i | y_i < x_i'\beta_\tau\}} (1-\tau)|y_i - x_i'\beta_\tau| \right\} \qquad (5)$$

$$\therefore \widehat{\beta_\tau} = \underbrace{argmin}_{\beta_\tau \in \mathbb{R}^k} \sum_i \rho_\tau (y_i - x_i'\beta_\tau) \qquad (6)$$

Where the check function is $\rho_\tau(e) = e(\tau - I(e < 0))$, $0 < \tau < 1$

METHODOLOGY

This section describes the Manly transformation and the two methods for estimating the transformation parameter, namely the Chamberlain and Buchinsky two step**(CBTS)** method and the maximum likelihood Method (MLE). The analysis was done using the Iris data from R that comprises of 150 data points with four factors namely sepal length, sepal width, petal width and petal length of plants. In the analysis the data was transformed using the manly transformation. But the transformation parameter lambda was estimated using the CBTS and MLE. The estimated lambdas were then used in the manly transformationfor data without outliers and for data with an outlier. Three methods of model comparison criteria (MSE, AIC AND BIC) were used in this work. These methods were also discussed in this section. Finally the next section presented the results and conclusion.

Manly Transformation

The family of transformation applied over a long period can be used for data transformation for any population so that the transformed data can be normally distributed. Let y be a random variable distributed as non-normal and Y^* be the transformed value of y and λ the transformation parameter. Box-Cox(1964) gave a simple modified form of the power transformation to avoid discontinuity at $\lambda = 0$. He considered;

$$Y^* = \begin{cases} \frac{y^\lambda - 1}{\lambda} & ,\lambda \neq 0 \\ lny & ,\lambda = 0 \end{cases} \quad for\ y > 0 \quad (7)$$

This equation () is known as the Box-Cox transformation. Manly (1976), proposed a one parameter exponential transformation as an alternative to Box Cox transformation because it allows negative y values.The transformation by Manly is given as:

$$Y^* = \begin{cases} \frac{\exp(\lambda y) - 1}{\lambda} & ,\lambda \neq 0 \\ y & ,\lambda = 0 \end{cases} \qquad (8)$$

Where λ is the transformation parameter, Y* is the transformed response variable and *y* is the observed response variable and it is restricted to be positive. It has been found that this transformation is quite effective in turning skewed unimodal distribution into nearly symmetric distributions but is not quite useful for bimodal or U-shaped distribution, WatthanacheewakulL.(2014).Traditional remedies for deviation from normality include employing a more appropriate distribution as well as transforming data to near-normality. Zhu, and Melnykov (2018), merged both approaches by introducing a mixture model with components derived from the multivariate Manly transformation and this mixture models show good performance in modeling skewness and have excellent interpretability.

Maximum Likelihood Estimation of transformation parameter

Maximum Likelihood Estimation (MLE) approach involves forming an assumption about the underlying probability distribution function (pdf) that generates the observed data set, and then estimating parameters of the assumed distribution. The steps in the estimation process typically involve two steps;

-Specification of a probability distribution for u_i .

-Computation and maximisation of the likelihood function.

Thepdfof each transformed observation takes the following form:

$$f(y^*|\mu, \delta^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} e^{-\frac{1}{2\sigma^2}(y^* - \mu)^2} \quad (9)$$

Where y^* is the transformed observation, μ is the mean and σ^2 is the variance. The Manly likelihood function in relation to the original observations is given by

$$L(\mu, \sigma^2, \lambda / y) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} \exp\left\{-\frac{1}{2\sigma^2}\left[\frac{e^{-\lambda y} - 1}{\lambda} - \mu\right]^2\right\} \cdot J(y^*; y) \qquad (10)$$

Given that,$J(y^*; y) = \Pi \left|\frac{\partial y^*}{\partial y}\right|$

Given that$\lambda$is the transformation parameter.The maximum likelihood estimate of transformation parameter is obtained by solving the likelihood equation below;

$$\frac{d}{d\lambda} \ln L(\lambda) = \frac{-n\left[\Sigma e^{2\lambda y_i} y_i - \Sigma \frac{1}{n}(e^{\lambda y_i})(e^{\lambda y_i} y_i)\right]}{\Sigma(e^{2\lambda y_i}) - \Sigma \frac{1}{n}(e^{\lambda y_i})^2 + \frac{n}{\lambda} + \Sigma y_i} = 0 \qquad (11)$$

Lakhana Watthanacheewakul (2020).

Scientia Africana, Vol. 21 (No. 1), April, 2022. Pp 67-76 https://dx.doi.org/10.4314/sa.v21i1.6
© Faculty of Science, University of Port Harcourt, Printed in Nigeria ISSN 1118 – 1931

Chamberlain and Buchinsky Two Step (CBTS) Estimation

The two-step estimation was proposed by Chamberlain (1994) and Buchinsky (1994). They suggested the following numerically attractive simplification in form of a two-step procedure (CBTS) which exploits the equivariance property of quantiles inother to estimate the parameters. The procedure is as follows:

First estimate $\beta_\tau(\lambda)$ conditional on λ by solving the minimization problem;

$$\hat{\beta}_\tau(\lambda) = arg\min_{\beta} n^{-1} \sum_{i=1}^{n} \rho_\tau(y_{\lambda i}\bar{g}_{y_\tau} - x'\beta) \qquad (12)$$

Secondly estimate λ_τ by solving the minimization problem;

$$\hat{\lambda}_\tau = \min_{\lambda \in \mathbb{R}} n^{-1} \sum_{i=1}^{n} \rho_\tau \left(y_i - (\lambda x'\hat{\beta}_\tau(\lambda) + 1)^{-1/\lambda}\right) \qquad (13)$$

Conditioned on the assumption that, $\lambda x'\hat{\beta}_\tau(\lambda) + 1 > 0$

In implementing this procedure, for this paper we assume that $\lambda x'\hat{\beta}_\tau(\lambda) + 1$ is strictly positive.

Comparison Criteria

Akaike's Information Criteria (AIC)

One of the most commonly used information criteria is AIC. The idea of AIC (Akaike, 1973) is to select the model that minimises the negative likelihood penalised by the number of parameters as specified in the equation.

$$\text{AIC} = -2\log(L) + 2p \qquad (14)$$

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**:

$$MSE = \sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad (15)$$

Bayesian information criteria (BIC)

Another widely used information criteria is the BIC. BIC is derived within a Bayesian framework as an estimate of the Bayes factor for two competing models; Schwarz (1978), Kass and Rafftery (1995). BIC is defined as:

$$\text{BIC} = -2\log(L) + p\log(n) \qquad (16)$$

RESULTS

Table 1: Results Summary from CBTS Method Without Outlier

	25%	50%	75%	95%
Intercept	**-0.4262978**	-1.1070109	-1.0729753	-1.4473967
Sepal.Width	**0.4825532**	0.6367298	0.6305471	0.8335235
Petal.Length	**0.5252625**	0.6922661	0.7355744	0.8442709
Petal.Width	**-0.4123057**	-0.6922661	-0.6829649	-0.5766708

AIC	**-43.46279**	-5.406937	66.48484	347.9554
BIC	**20.75212**	20.95704	21.51937	29.68686
MSE	**0.70956**	0.91450	1.47683	9.64432

From table 3.1the model of best fit using the model selection criteria is the 25% quantile model, because it has the smallest AIC, MSE and BIC values compared to other quantile results.

Table 2: Results Summary fromMLEMethod without Outlier

	25%	50%	75%	95%
Intercept	1.63385	**1.84461**	2.01079	2.42655
Sepal.Width	0.65597	**0.65916**	0.63587	0.63186
Petal.Length	0.69066	**0.71928**	0.77466	0.63337
Petal.Width	-0.51315	**-0.59353**	-0.67459	0.31060
AIC	-286.0308	**-348.3657**	-284.3705	-138.5484
BIC	20.18337	**20.13548**	20.18494	20.41898
MSE	0.01983	**0.00864**	0.02027	0.14169

From table 3.2 the model of best fit using the model selection criteria is the50% quantile model, because it has the smallest AIC, MSE and BIC values compared to other quantile results.

Table 3: Results Summary fromCBTSMethod With an Outlier

	25%	50%	75%	95%
Intercept	**-10.6789**	-22.7312	-26.6612	-31.4693
Sepal.Width	**7.33456**	8.01567	8.92617	22.67813
Petal.Length	**7.89651**	8.82614	9.67813	22.98165
Petal.Width	**-7.58671**	-9.68120	-9.87110	-8.07623
AIC	**-48.56710**	-6.78234	78.23412	505.6624
BIC	**21.8321**	22.03489	24.89412	33.6712
MSE	**0.92341**	2.3867	3.69812	8.77631

From table 3.3the model of best fit using the model selection criteria is the 25% quantile model, because it has the smallest AIC, MSE and BIC values compared to other quantile results.

Table 4: Results Summary fromMLEMethod With an Outlier

	25%	50%	75%	95%
Intercept	-24.83032	**-44.21738**	-49.53094	-40.6526
Sepal.Width	9.39032	**14.15994**	15.41346	13.90437
Petal.Length	10.28397	**15.00064**	17.52720	18.89159
Petal.Width	-8.33953	**-14.57947**	-16.82453	-16.4337
AIC	1001.26	**988.6703**	992.0208	1026.736
BIC	771.2872	**710.809**	726.412	910.355
MSE	751.2447	**690.7965**	706.3695	890.3125

Scientia Africana, Vol. 21 (No. 1), April, 2022. Pp 67-76 https://dx.doi.org/10.4314/sa.v21i1.6
© Faculty of Science, University of Port Harcourt, Printed in Nigeria ISSN 1118 – 1931

From table 3.4The model of best fit using the model selection criteria is the50% quantile model, because it has the smallest AIC, MSE and BIC values compared to other quantile results.

Table 5: Model comparison using AIC, BIC and MSE at different quantile for CBTS and MLE Method Without Outlier.

	MLE with Manly Transformation			TwoStage(CBTS)Estimation		
T	AIC	MSE	BIC	AIC	MSE	BIC
25%	**-286.0308**	**0.3752726**	**20.18337**	-43.46279	0.8423644	20.75212
50%	**-348.3657**	**0.3048652**	**20.13548**	-5.406937	0.9562941	20.95704
75%	**-284.3705**	**0.3773553**	**20.18494**	66.48484	1.21525	21.51937
95%	**-138.5484**	**0.6135492**	**20.41898**	347.9554	3.105531	29.68686

From table (3.5) above based on the AIC, MSE and BIC the maximum likelihood technique gave lower values at all quantiles, making it the preferred method.

Table 6: Model comparison using AIC, BIC and MSE at different quantile for CBTS and MLE Method With An Outlier.

	MLE with Manly Transformation			**Two Stage (CBTS)Estimation**		
Quantile	**AIC**	**MSE**	**BIC**	**AIC**	**MSE**	**BIC**
25%	1001.26	751.2447	771.2872	**-48.5671**	**0.92341**	**21.8321**
50%	988.6703	690.7665	710.809	**-6.78312**	**2.3867**	**22.03489**
75%	992.0208	706.3695	726.412	**78.23412**	**3.69812**	**24.89412**
95%	1026.736	890.3125	910.355	**505.61123**	**8.77631**	**33.66712**

From table (3.6) above based on the AIC, MSE and BIC the CBTS gave lower values at all quantiles, making it the preferred method.

DISCUSSIONS

Given the results above, the summary result of the CBTS for data without outliers showed that the 25th quantile model was of best fit with an AIC of -43.46279, BIC of 20.75212 and MSE of 0.70956. While the 50th quantile model was the best fit for the MLE method with AIC of -348.3657, BIC of 20.13548 and MSE of 0.00864. The results from data with outliers still showed that the 25th quantile model was the model of best fit for CBTS method with AIC of -48.56710, BIC of 21.321 and MSE of 0.92341 and for the MLE method, the 50thquantile still appeared the best fit with AIC of 988.6763, BIC of 710.809 and MSE of 690.7965. Now considering the MLE method alone for both data with outliers and data without outliers, it was observed that the AIC, BIC and MSE for data without outliers were very small compared to that of data with outliers, exposing the outlier sensitivity of MLE. The comparison of the CBTS and MLE methods at both cases showed that MLE produced smaller AIC, BIC and MSE at all quantiles for data without outliers. While the CBTS produced smaller AIC, BIC and MSE at all quantiles for data with outliers.

CONCLUSIONS

Data transformation plays a vital role when it comes to normalization of variables in statistical analysis;the study was performed using Manly data transformation for normalizing of the response variables in quantile regression analysis.The Manly transformation parameter was estimated using two different methods namely; the maximum likelihood estimation (MLE) method and the two-step estimation method by Chamberlain and Buchinsky(CBTS).The methods were applied to a quantile regression analysis at different quantiles (0.25, 0.50, 0.75, 0.95). Based on our findings, the CBTS method showed that the 25th quantile model was the best fit for the data with outliers and without outliers, while the MLE method showed the 50th quantile as the model of best fit at both scenarios. The comparison of the two estimation methods revealed that for data without outliers the MLE performed better but for data with outliers the CBTS performed better. These observations can lead us to conclude that in agreement to literature the MLE is very sensitive to outliers while the CBTS is robust to outliers.

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.

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

Buchinsky, M. (1994). Changes in the US wage structure 1963-1987: Application of quantile regression. Econometrica: Journal of the Econometric Society, 62(2), pp 405-458.

Box G. E. P. and Cox D. R. (1964). An Analysis of Transformation (with Discussions). Journal of the Royal Statistical Society, Ser. B; 26(2), 211-252

Chamberlain, G. (1994). Quantile Regression, Censoring, and the Structure of Wages.Advances in Econometrics: Sixth World CongressVol. 1, pp 171-209. Cambridge University Press.

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

Huang, Q., Zhang, H., Chen, J. and He, M. (2017). Quantile Regression Models and Their Applications: A Review. Journal of Biometrics & Biostatistics 08(03). DOI: 10.4172/2155-6180.1000354

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

Koenker, R., & Bassett, G. (1978). Regression Quantiles. Econometrica, 46(1), 33–50. https://doi.org/10.2307/1913643

Manly, B. F. (1976). Exponential data transformations. Journal of the Royal Statistical Society: Series D,The Statistician, 25(1), 37-42.

Nwakuya, M. T. (2020). Assessment of Mental Health of Undergraduate

Scientia Africana, Vol. 21 (No. 1), April, 2022. Pp 67-76 *https://dx.doi.org/10.4314/sa.v21i1.6*
© Faculty of Science, University of Port Harcourt, Printed in Nigeria *ISSN 1118 – 1931*

Students Based on Age: A Bayesian Ordinal Quantile Regression Approach. Quarterly Journal of Econometrics Research6(1), 12-17.

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

Yeo, I. K., & Johnson, R. A. (2000). A new family of power transformations to improve normality or symmetry. Biometrika, 87(4), 954-959.

Zhu, X. and Melnykov, V. (2018).Manly transformation in finite mixture modeling.Computational Statistics & Data Analysis; 121: 190-208. DOI:10.1016/j.csda.2016.01.015

WatthanacheewakulL.(2014). A New Family of Transformations for Lifetime Data.Proceedings of the World Congress on Engineering, Vol1 WCE 2014, July 2-4, London UK.

WatthanacheewakulL.(2020). Modified Box-Cox Transformation and Manly Transformation with Failure Time Data; https://pdfroom.com/books/modified-box-cox-transformation-and-manly-transformation-with failure-time-data/8Pe5xkNZ2nN