Short Research Article

On the Comparison of Ordinary Least Squares and Quantile Regression with Nigerian Financial Data on Trade Balance, Foreign Inflow and Imports.

ABSTRACT

This research work compared the Ordinary Least Squares Regression and Quantile Regression models, as well as the differences between them thereby examining the compared models in terms of goodness of fit statistic and also recommend a suitable fit model for the data collected. These methods were applied to the Nigerian Financial Data on Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3) on the Nigerian Gross Domestic Product (Y) collected from Microtrends.net. The study found that the influence of Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3) vary on the Nigerian Gross Domestic Product (Y) depending on the quantile one is looking at. The study recommends the robustness and stability of the Quantile regression model considered as an alternative to the Ordinary Least Square Model.

1 Introduction

If a research(er) is concerned in the relationship between two or more variables, he or she will use the regression analysis method. The traditional regression analysis, the Ordinary Least Squares (OLS) focuses on the mean and sometimes called the ‘mean regression; specifically, the regression analysis assesses the association between the two variables (dependent and independent variable) by utilizing a function known as the conditional mean of the response to relate the mean of the dependent variables for each fixed value of the independent variable. Conditional-mean function for modeling and fitting is one of several regression-modeling techniques, which include the popular basic linear regression model, multiple regression, weighted least squares for models with “heteroscedastic” errors, and non-linear regression models. Conditional-mean models provide a number of appealing features.

The regression analysis has proved beyond a reasonable doubt to have been one of the most used and robust in the application for numerous kinds of research, especially when provisions are made to control for problems dealing with heteroskedasticity, due to the violation of OLS assumption. Ordinary least squares (OLS) estimators are nevertheless unbiased and consistent, as are regression predictions based on them. If the Ordinary least squares (OLS) estimators are considered to be no longer BLUE, that is, if these estimates are not efficient enough, the regression prediction will provide inefficient results. Because of the mismatch in the covariance matrix of the estimated regression coefficients, hypothesis tests such as t-tests and F-tests are no longer viable. It is on this background that this study proposes to introduce and compare the Ordinary Least Squares and Quantile Regression models when the stated assumptions fail to hold, as a resilient alternative to traditional least squares regression.

A common extension of the linear regression model is the quantile regression model. Changes in the conditional mean of the dependent variable in response to changes in the covariates are described by the linear regression model, whereas changes in the conditional quantile are described by the quantile regression model

Quantile regression is a statistical method for approximating and inferring conditional quantile functions. Quantile regression approaches, like the standard linear regression method, can be a reliable alternative for estimating models with the conditional median function and other supplementary range of conditional quantile functions. It combines techniques for approximating an entire group of conditional quantile functions with methods for estimating conditional mean functions, extremely effective for displaying changes in the conditional distribution of longitudinal data sets over time (Karlsson,2006). The quantile concept encompasses a wide range of words such as quintile, decile, quartile, and percentile. The n^{th} quantile represents the significance of a response whose fraction of the population is n. As a result, quantiles may be used to locate any point in a distribution. 50% of the population, for example, falls below the 0.50th quantile.

. Koenker and Bassett (1978) proposed the quantile regression approach, which aims to model and provide broad policy for implementing the regression analytical method (Koenker,2005).

The conditional median function $Q_q(y/x)$ may be used to think about the link between the regressors and the result, where the median is the 50^{th} percentile, or quantile q, of the empirical distribution, similar to the conditional mean function of linear regression.

In their paper "Quantile Regression Analysis as a Robust Alternative to Ordinary Least Squares," John and Nduka (2009) argued that the Quantile Regression approach was a good alternative to the OLS because, while the SSE is minimized by the OLS approach which calculates the mean, the errors are minimized by the median regression estimates of the

Quantile regression. Their work also established that Quantile regression estimates overcomes a number of issues that the OLS has, including heteroscedasticity because the OLS focuses on the information about a distribution's tails is lost when using the mean as a metric of position; since OLS is susceptible to severe outliers, so, findings can be greatly skewed. Fitzenberger etal (2002) compared the practical use of quantile regression to the least –squares regression technique. Their work underlined the importance of the correct distribution modelling by pointing out the critical multiple regression theory of constant and using pay distributions as an illustration.

Ibrahim (2015), proposed quantile regression (QR) model as an alternative to ordinary least squares (OLS) regression. The work established the quality of fit statistic of the Quantile regression coefficient as well as heteroskedasticity test statistic. Marasinghe (2014) used quantile regression to model climate data and the work established that quantile regression is seen as an emerging statistical technique for explaining the link between response and a predictor or a replacement of the ordinary least squares regression. He established the fact that while quantile regression provides a more thorough view in estimating the derivative when dealing with non-normality and non-constant variance assumption, least square regression is preferred to modelling temperature research when discussing residuals with normality and constant variance assumption

2.0 Methodology

The OLS model

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + e_i \qquad i = 1,2,\dots,k \qquad 2.1$$

$Y = \ Gross\ Domestic\ Product$

$x_1 = Trade\ Balance$

$x_2 = Foreign\ Inflow$

$x_3 = Imports$

Several approaches for creating models have been developed to build a simple regression model by finding a mean to place Y as a function of X. Here we look at the least square method for evaluating a regression model.

$$y_i = \beta_0 + \beta_1 x_i + e_i \qquad i = 1,2,\dots,n \qquad 2.2$$

Multiple Regression

In matrix terminology, a general linear regression model

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + e_i \qquad 1,2,\ldots,n \qquad 2.3$$

Thus we have:

$$y_1 = \beta_0 + \beta_1 x_{11} + \cdots + \beta_{p-1} x_{1,p-1} + e_1$$

$$y_2 = \beta_0 + \beta_1 x_{21} + \cdots + \beta_{p-1} x_{2,p-1} + e_2$$

$$y_3 = \beta_0 + \beta_1 x_{31} + \cdots + \beta_{p-1} x_{3,p-1} + e_3$$

$$\vdots$$

$$y_n = \beta_0 + \beta_1 x_{n1} + \cdots + \beta_{p-1} x_{n,p-1} + e_n$$

Then $Y = X\beta + \varepsilon$, which is

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1,p-1} \\ 1 & X_{21} & X_{22} & \cdots & X_{2,p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & X_{n2} & \cdots & X_{n,p-1} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Recall that the error has zero (0) mean and zero (0) covariance matrix

$$\begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 I$$

since the ε_i are independent.

Note that $E(Y) = E(X\beta + \varepsilon) = E(X\beta) = X\beta$ since $E(\varepsilon) = 0$.

Goodness-of-Fit

Calculating a value that illustrates how well the OLS regression line fits the data is frequently useful.. The coefficient of determination, often known as the R-squared (R^2) of the regression, can be defined by the formula below

$$R^2 = SSE/SST = 1 - SSR/SST$$

R^2 is the fraction of the sample variation in y that is explained by x which is the ratio of the explained variation to the overall variance.

The squared correlation coefficient between the real y_i with fitted values $\hat{y_i}$ may also be calculated using R-squared. That is,

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i-\bar{y})(\hat{y_i}-\bar{y})\right)^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2\sum_{i=1}^{n}(\hat{y_i}-\bar{y})^2} \qquad 2.4$$

The loss function is given by

$$\rho_\tau(z) = (\tau - Iz) * z$$

where $I(z)=1$ if $z<0$ and $I(z)=0$ if $z\geq 0$. The loss function's minimal value is zero, which happens when z = 0. The loss function is linear to the right of z = 0 and slopes upward with slope τ, whereas the loss function is linear to the left of z = 0 and slopes downward with slope ($\tau-1$).

If the cdf of the random variable Y is continuous, the theoretical expected loss function may be expressed as follows.

$$E[\rho_\tau(z)] = (\tau-1)\int_{-\infty}^{z}(x-z)dF_Y(x) + \tau\int_{z}^{\infty}(x-z)dF_Y(x)$$

The empirical analogy of the above-mentioned theoretical anticipated loss is as follows: (letting $Y_i - u = z_i$)

$$E[\rho_\tau(y_t-u)] = \frac{(\tau-1)}{N}\sum_{y_t<u}(y_t-u) + \frac{\tau}{N}\sum_{y_t\geq u}(y_t-u)$$

where the empirical density assigns each observation an equal probability (say 1/N). The τ-th quantile will be generated by minimizing this. That is, we are looking for the estimated quantile u*.

Assume that the quantile u in the previous anticipated loss is linearly related to certain explanatory factors. Using the linear function, we can create the conditional expected loss function

$$u_t = \beta_0 + \beta_1 x_t^1 + \beta_2 x_t^2 + \ldots + \beta_{K-1} x_t^{K-1}$$

and write the expected loss as

$$E[\rho_\tau(y_t-\beta_0-\sum_{j=1}^{K-1}\beta_j x_t^j)]=\frac{(\tau-1)}{N}\sum_{y_t<\beta_0+\sum_{j=1}^{K-1}\beta_j x_t^j}(y_t-\beta_0-\sum_{j=1}^{K-1}\beta_j x_t^j)$$

$$+\frac{\tau}{N}\sum_{y_t\geq\beta_0+\sum_{j=1}^{K-1}\beta_j x_t^j}(y_t-\beta_0-\sum_{j=1}^{K-1}\beta_j x_t^j)$$

By choosing the proper values for the β's, the empirical expected loss is now minimized.

3.0 Results and Discussion

4.1.1 Resulting model

$Y_i = 18.14 + 0.88x_1 - 2.62x_2 + 6.06x_3$,

Looking at the result considering the independent variables, only Imports (x_3) had a statistical significance on the dependent variables while Trade Balance (x_1) and Foreign Inflow (x_2) was statistically insignificant with the dependent variable(Y). From the regression coefficients, Trade Balance (x_1) ($\beta_1 = 0.88$) and Imports (x_2) ($\beta_3 = 6.06$) were seen to positively improve the Nigerian Gross Domestic Product (Y), while Foreign Inflows ($\beta_2 = -2.62$) had a negative decrease in the Nigerian Gross Domestic Product (Y).

4.2 The quantile Regression

The test was carried out at the 10%, 25%, 50%, 75% and 90% quantities.

N:B- Full tables of each quantile is found at the appendix of this work.

Table 1 Quantile Regression Estimate

	10%	25%	50%	75%	90%	OLS
β_0	4.547	2.091	4.329	30.039*	59.619**	18.136
β_1	0.846**	1.218**	1.487**	0.312	-0.267	0.877

β_2	5.950*	7.222*	0.992	-10.458	-11.869	-2.622
β_3	3.576**	4.398**	5.383**	7.997**	8.493**	6.059**

N.B: Sig at 99% (**); Sig at 95(*).

At the 10% quantile, the regression coefficients signify a positive effect between Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3) with the dependent variable Nigerian Gross Domestic Product (Y). The probability values suggest a significant relationship amongst Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3).

At the 25% quantile, Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3) had a positive relationship or effect on the dependent variable Nigerian Gross Domestic Product (Y). The probability values suggests a significant relationship amongst Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3).

At the 50% quantile, Trade Balance (x_1), Foreign Inflow (x_2) and Imports (x_3) had a positive relationship or effect on the dependent variable Nigerian Gross Domestic Product (Y), the probability values suggest significance in Trade Balance (x_1) and Imports (x_3), while Foreign Inflow (x_2) had a statistical non-significant relationship with the Nigerian Gross Domestic Product (Y).

At the 75% quantile, the regression coefficients signifies that Trade Balance (x_1) $(\beta_1 = 0.312)$ and Imports (x_2) $(\beta_3 = 7.997)$ had a positively influence on the Nigerian Gross Domestic Product (Y), while Foreign Inflows $(\beta_2 = -10.45)$ had a negative influence which will bring about a decrease in the Nigerian Gross Domestic Product (Y). The probability values suggest significance in only Imports (x_2).

At the 90% quantile, the regression coefficients signifies that Trade Balance (x_1) $(\beta_1 = -0.267)$ and Foreign Inflows $(\beta_2 = -11.869)$ had a negative influence on the Nigerian Gross Domestic Product (Y), while Imports (x_2) $(\beta_3 = 8.493)$ had a positive influence which will bring about an increase in the Nigerian Gross Domestic Product (Y). The probability values suggest significance in only Imports (x_2).

4.3 Discussion and Comparison of Results

Looking at the OLS results, an increase in Trade Balance (x_1) will bring about an increase in the Nigerian Gross Domestic Product (Y). This is also suggested by the quantile regression model at the 10%, 25%, 50%, and 75% quantile while at the 90% quantile, Trade Balance (x_1) will cause a decrease in the Nigerian Gross Domestic Product (Y). Foreign Inflows (x_2) affects Nigerian Gross Domestic Product $\beta_2 = -2.622$ by looking at the OLS model. The quantile regression shares same view with the OLS at just the 75% and the 90% quantile with coefficients $\beta_2 = -10.458$ and $\beta_2 = -11.869$ respectively. While at the 10%, 25% and 50% quantile Foreign Inflows had a positive influence $\beta_2 = 5.95$, $\beta_2 = 7.222$ and $\beta_2 = 0.992$ on the Nigerian Gross Domestic Product (Y). Looking at the Import (x_3) on the influence of Nigerian Gross Domestic Product (Y). The OLS model suggest that Imports will bring about an increase in the Nigerian Gross Domestic Product (Y). This is also suggested by the quantile regression model at the 10%, 25%, 50%, 75% and 90% quantile, Imports (x_3) will cause an increase in the Nigerian Gross Domestic Product (Y). This thus implies that the influence of these financial inflows on one quantile alone or also a generalisation of the OLS model only will be misleading as there are slight changes which was made known to us at different quantiles giving us a broad understanding of the nature of the effect of financial inflows on the Nigerian Gross Domestic Product at different positions.

5 Conclusion and Recommendation

This work investigated the robustness of quantile regression as a good alternative to the ordinary least squares regression while comparing the models. This work provided a fundamental model and its key characteristics, as well as a brief analysis of a major application employing Nigerian financial data.

At each quantile of the conditional distribution function, quantile regression allows you to ask questions about the relationship between the response variable and the covariate, going beyond the primary purpose of establishing merely the conditional mean.

The comparison of the OLS and Quantile regression has proven to be very important, because if looking at the OLS model alone, a lot of information would be lost. By looking at the Quantile Regression model at different quantiles, more information about the variables were brought to limelight. The performance is stable, and robust against common deviations from the model assumptions.

References

Fitzenberger, B, Koenker R. and J.M Editors ,(2002). Economic Applications of Quantile Regression: Physica-Verlag Heidelberg, New York.

Ibrahim, A. (2015).Analysis of Quantile Regression as Alternative to Ordinary Least SquaresRegression. Unpublished thesis of the Department of Mathematics Ahmadu Bello University, Zaria Nigeria.

John O.and Nduka, (2009) Quantile Regression as a robust alternative to the ordinary least squares. Scientia Africana. 2(8), 61-65.

Karlsson, S. (2006) Estimation and inference for Quantile Regression of Longitudinal Data, With Applications in Biostatistics. International journal of social science, 8,.8-36.

Koenker, R. (2005). Quantile Regression. New York, NY: Cambridge University Press.

Koenker, R. and G. Bassett, Jr. (1978.). Regression quantiles. Journal of Econometrica, 46(1), 33-50.

Koenker, R., and Hallock, K. (2001). Quantile regression. Journal of Economic Perspective, 15(4), 143-156.

Koenker, R., and Machado, J. (1999). Goodness of fit and related inference processes for quantile regression. Journal of the American Statistical Association, 94, 1296-1310.

Marasinghe,D.(2014) "Quantile Regression for Climate Data". All Theses. 1909.https://tigerprints.clemson.edu/all theses/1909.

Appendix A

Quantile Regression Output

10% Quantile

QuantReg Regression Results

Dep. Variable:	y	Pseudo R-squared:	0.6279
Model:	QuantReg	Bandwidth:	78.13
Method:	Least Squares	Sparsity:	156.9
No. Observations:	50		
Df Residuals:	46		
Df Model:	3		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.5467	11.900	0.382	0.704	-19.408	28.501
TB	0.8457	0.254	3.326	0.002	0.334	1.357
FI	5.9497	2.689	2.213	0.032	0.537	11.363
IMP	3.5764	0.275	13.017	0.000	3.023	4.129

25% Quantile

QuantReg Regression Results

Dep. Variable:	y	Pseudo R-squared:	0.6650
Model:	QuantReg	Bandwidth:	40.17
Method:	Least Squares	Sparsity:	84.15
No. Observations:	50		
Df Residuals:	46		
Df Model:	3		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	2.0910	7.475	0.280	0.781	-12.955	17.137

	coef	std err	t	P>|t|	[0.025	0.975]
TB	1.2182	0.307	3.974	0.000	0.601	1.835
FI	7.2220	3.348	2.157	0.036	0.482	13.962
IMP	4.3980	0.258	17.019	0.000	3.878	4.918

50% Quantile

QuantReg Regression Results

Dep. Variable:	y	Pseudo R-squared:	0.7175
Model:	QuantReg	Bandwidth:	39.82
Method:	Least Squares	Sparsity:	81.76
No. Observations:	50		
Df Residuals:	46		
Df Model:	3		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	4.3289	7.845	0.552	0.584	-11.462	20.120
TB	1.4872	0.470	3.164	0.003	0.541	2.433
FI	0.9916	4.879	0.203	0.840	-8.829	10.812
IMP	5.3833	0.411	13.101	0.000	4.556	6.210

75% Quantile

QuantReg Regression Results

Dep. Variable:	y	Pseudo R-squared:	0.7693
Model:	QuantReg	Bandwidth:	55.51
Method:	Least Squares	Sparsity:	140.1
No. Observations:	50		
Df Residuals:	46		
Df Model:	3		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	30.0386	11.595	2.591	0.013	6.699	53.379

```
TB           0.3117      0.957      0.326      0.746      -1.615      2.238
FI         -10.4582      7.301     -1.432      0.159     -25.155      4.238
IMP          7.9973      0.632     12.656      0.000       6.725      9.269
==============================================================================
```

90% Quantile

```
                         QuantReg Regression Results
==============================================================================
Dep. Variable:                  y   Pseudo R-squared:             0.7789
Model:                   QuantReg   Bandwidth:                     79.64
Method:             Least Squares   Sparsity:                      221.4
 No. Observations:                50
 Df Residuals:                    46
                         Df Model:                         3
==============================================================================
                coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept    59.6194     12.641      4.716      0.000      34.175      85.064
TB           -0.2668      1.653     -0.161      0.872      -3.594      3.060
FI          -11.8692     10.400     -1.141      0.260     -32.803      9.065
IMP           8.4929      0.742     11.453      0.000       7.000      9.986
```

Appendix B: OLS Regression Output

```
 OLS Regression Results
==============================================================================
Dep. Variable:                    GDP   R-squared:                     0.899
Model:                            OLS   Adj. R-squared:                0.892
Method:                 Least Squares   F-statistic:                   135.9
Prob (F-statistic):          7.21e-23
 Log-Likelihood:              -267.63
No. Observations:                  50   AIC:                           543.3
Df Residuals:                      46   BIC:                           550.9
Df Model:                           3
Covariance Type:            nonrobust
```

==

+

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	18.1356	10.223	1.774	0.083	-2.442	38.713
TB	0.8774	0.612	1.432	0.159	-0.356	2.110
FI	-2.6221	6.357	-0.412	0.682	-15.419	10.175
IMP	6.0589	0.535	11.315	0.000	4.981	7.137

==

Omnibus:	9.506	Durbin-Watson:	0.701
Prob(Omnibus):	0.009	Jarque-Bera (JB):	13.571
Skew:	0.550	Prob(JB):	0.00113
Kurtosis:	5.303	Cond. No.	48.1

==

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.