International Journal of Basic and Applied Science 12 (1) (2023) 20-32

Published by: IOCSCIENCE

International Journal of Basic and Applied Science

Journal homepage: www.ijobas.pelnus.ac.id

Comparative analysis of linear and quantile regression models in predicting body mass index among students

Odoh, C. M.[1], Ugwu, N. D.[2], Aronu, C. O.[3]

[1]Department of Mathematics and Statistics, Delta State Polytechnic Ogwashi Uku, Delta State, Nigeria

[2,3]Department of Statistics, Chukwuemeka Odumegwu Ojukwu University, Uli Campus, Anambra, Nigeria

Article Info

Article history:

Received May 20, 2023
Revised Jun 16, 2023
Accepted Jun 30, 2023

Keywords:

Body Mass Index (BMI);
Linear regression;
Quantile regression;
Students.

ABSTRACT

This study applies quantile regression to model the Body Mass Index (BMI) of 152 students from Delta State Polytechnic in Delta State, Nigeria. The BMI serves as the response variable, while skin fold (SK), feeding habit (FH), feeding frequency (FF), and frequency of exercising (FE) are considered as explanatory variables. A comparison is made between the results obtained from the classical linear regression model and quantile regression models at different quantiles (p = 0.1, 0.25, 0.5, 0.75, and 0.9) to examine the impact of the variables on the students' BMI and assess if the relationships differ across quantiles. The findings reveal significant differences between the classical linear regression and quantile regression models, emphasizing the importance of quantile regression in capturing the nuances of the relationship between variables at different points of the BMI distribution. The study highlights the limitations of the classical linear regression model in providing a comprehensive understanding of the data and underscores the value of quantile regression in enhancing our insights into the relationship between BMI and the considered factors. This research contributes to the broader literature on quantile regression and its applications in exploring BMI and related health issues among students.

Corresponding Author:

Aronu, C. O.,
Department of Statistics,
Chukwuemeka Odumegwu Ojukwu University, Uli Campus, Anambra.
Ogwashi-Uku Azagba Road Delta State, Nigeria
Email: amaro4baya@yahoo.com

1. INTRODUCTION

A widely used approach for modelling the correlation between dependent and independent variables is the classical or ordinary regression model [1]. However, because it only considers the mean and falls short of capturing the whole image of the distribution, this model only offers a partial comprehension of the data set [2]. A different strategy termed quantile regression has been presented to get around this restriction. A more thorough investigation of the connection between variables is provided by quantile regression, which calculates the conditional median or other quantiles of the answer variables [3–4]. It expands the conventional regression model and is especially helpful when the linearity, homoscedasticity, and normalcy assumptions are broken [5]. Similar research was done by the authors [6] among university students in different areas, and they discovered that the quantile regression model gave useful insights into how variables affected BMI at various quantiles. Their findings highlighted how crucial it is for BMI prediction models to take into account quantile-specific connections. In a school-based environment, research by [7] investigated the use of linear and quantile regression models to predict BMI among teenagers. They found that, in contrast to linear regression, which offered a broad grasp of the typical association between factors and BMI, quantile regression allowed for a more in-depth analysis of the predictors' effects at various quantiles. These findings validated quantile regression as an effective approach for predicting BMI. The study by [8] stated that a cross-sectional study was carried out in Issele-uku, Delta State, Nigeria, to look at the prevalence of adult obesity.

The uses of quantile regression in many disciplines have been investigated in several papers. In this essay, we emphasize its use in simulating students' Body Mass Index (BMI). BMI, a measurement of nutritional status, is used to categorize people into various weight groups by calculating the ratio of weight to squared height [9]. The importance of BMI in comprehending health-related concerns has been stressed in earlier study [10–11]. Body mass index (BMI) is a measurement of a person's nutritional health based on their weight (kg) to height (m2) squared. Its value is used in classifying individuals to different weight classes, that is, underweight (<18.5), normal (18.5-24.99), overweight (25.00-29.99) and obese (≥30.00), based on the World Health Organization [9], standard classification in kg/m2. By utilizing quantile regression, the present study aim to gain a deeper insight into the relationship between BMI and other variables. The work by [1] asserts that ordinal variables in psychology are frequently mis-analyzed as metrics in statistical models, resulting in skewed effect-size estimates and exaggerated error rates. According to the study by [2], a Korean financial conditions index was developed as a result of the heightened interest in the impact of financial conditions on the actual economy in the wake of the global financial crisis. To investigate the effect of Korean and US financial circumstances on future Korean GDP growth, quantile regression models were used in the study. The results showed that the financial situation in Korea had an unbalanced impact on GDP growth across quantiles. The study by [3] examined how health-risky activities at home, school, and in the community affected Korean adolescents' body weight. Using quantile regression models, the complete distribution of the body mass index (BMI) relationship was examined. The findings showed that higher mother education was connected to higher BMI in boys, especially at upper quantiles, while higher paternal education was linked to lower BMI in girls.

With the use of quantile regression, the connection at various percentiles or quantiles was studied thereby providing a more detailed knowledge of the distribution of BMI among students [12]. This method contributes to the larger body of research on quantile regression and improves our understanding of the characteristics of BMI distribution among students. Additionally, many researchers have used linear and quantile regression in numerous domains, demonstrating its adaptability and efficiency in studying a range of phenomena. Since quantile regression methods may capture diverse connections, they are being utilized more often to predict BMI. The study by [13] looked at various investment levels and the applicability of the pecking order theory to industrial enterprises listed on the Borsa Istanbul. The panel quantile regression method they used led them to discover that the pecking order idea was not always true. Particularly, when internal funds were insufficient, high-leverage businesses showed a preference for equity financing at high investment levels, whereas low-leverage firms favoured borrowing as their top option.

In a study by [14], the quantile regression technique was used to examine the effects of several variables on the per capita income of participants in the microfinance program Amanah Ikhtiar Malaysia (AIM). The study concentrated on participants, primarily women, at various income distribution levels. The findings revealed that factors such as the value of assets, value of loans, household size, ratio of spending to income, and dummy state consistently influenced per capita income across various quantiles. The study by [15] made a valuable contribution by utilizing panel quantile regression to examine the dynamics of energy consumption in developing economies using data from 1990 to 2019. Their study served as a robustness check for previous findings, confirming the results obtained through other methodologies. Notably, the study discovered that the impact of remittances on energy consumption was more pronounced at upper quantiles. The study by [16] examined the impact of maternal pre-pregnancy BMI and gestational weight gain (GWG) on low birth weight (LBW) and macrosomia using the univariate and multivariate logistic regression analyses. It was found that insufficient GWG in the first and second trimesters increased the risk of LBW, while GWG <2kg in the second trimester to delivery also raised the risk. These associations were consistent across BMI categories, and LBW/macrosomia correlated with sustained low/high BMI percentiles in early childhood. The study by [17] examined COVID-19 deaths in public hospitals during three waves of the pandemic in Brazil. Quantile regression was used to analyze the relationship between time from diagnosis to death and socio-clinical variables. Findings revealed different associations in each wave, including factors such as immunodeficiency, obesity, neoplasia, education, O2 saturation, chronic neurological disease, race/colour, difficulty breathing, and chronic cardiovascular disease. Author [18] studied the complex association between obesity and fracture risk, considering variations in obesity definition, skeletal site, and gender. They used linear regression with waist circumference and BMI to determine fracture risk, adjusting for relevant covariates from a directed acyclic graph. These studies exemplify the applicability and benefits of using quantile regression in different research domains, providing valuable insights into the relationships between variables at various points in the distribution.

Also, relevant studies on BMI were reviewed to provide sufficient literature review depth on BMI. The study by [19] conducted a retrospective study on weight changes in patients receiving different biologic agents. Authors [20] examined the relationship between BMI, self-differentiation, and happiness in obese individuals. Coping styles were found to mediate this relationship. The study emphasizes the importance of specialized training to enhance happiness, self-differentiation, and coping styles in the obese population. Authors [21] investigated the causal associations between obesity, metabolic hormones, and female reproductive disorders. In a narrative review by [22], the association between adherence to the Mediterranean diet and overweight/obesity and age-related chronic diseases was examined. Authors [23] evaluated the validity of anthropometric measurements Mid Upper Arm Circumference (MUAC), Body Mass Index (BMI), and Calf Circumference (CC) for detecting malnutrition in older adults in Ethiopia. The study found significant correlations with the Mini-Nutritional Assessment (MNA) tool and high reliability (Cronbach's $\alpha = 0.847$). The area under the curve (AUC) indicated good overall accuracy (BMI: 0.98, MUAC: 0.94, CC: 0.96), with sensitivity and specificity ranging from 78-90% and 94-96%, respectively. Adjusted for age and sex, MUAC and CC had the best cut-off points with higher sensitivity and specificity. The study by [24] examined the prevalence and predictors of overweight and obesity in a multi-ethnic cohort from rural South Africa. Factors such as sex, age, education level, diabetes, and rural residence were associated with higher BMI. Women, older individuals, those with more education, and those in rural areas had higher odds of being overweight or obese. Conversely, smoking was associated with lower BMI and decreased odds of overweight or obesity. The study highlights the need for targeted community-based interventions to address obesity in this region, particularly among vulnerable groups. Authors [25] emphasized the significance of comprehending the factors contributing to overweight and obesity, as these conditions pose a considerable public health concern. The researchers noted that obesity is a risk factor for various diseases, including type 2 diabetes, which is becoming more prevalent in numerous low-income and middle-income countries (LMICs). Authors [26] investigated child malnutrition in Pakistan, analyzing data from the Pakistan Demographic and Health Survey. They found that 37.7% of children were stunted, 23% were underweight, and 8.0% suffered from wasting. Factors such as birth size, milk consumption, maternal education, socioeconomic status, and maternal BMI were associated with malnutrition. The study highlights the need for targeted public health policies and community-based education to address these issues.

Concisely, the work by [6] emphasized the necessity for capturing the effects of factors on BMI at different quantiles. The present study sought to address this requirement by contrasting the classical linear regression with quantile regression models. Consequently, the examination of quantile-specific impacts on BMI prediction in the context of Nigerian students' literature is still lacking. Quantile regression was necessary for the current study to acquire a deeper understanding of the link between BMI and other characteristics among students at Delta State Polytechnic, according to the study by [7], which validated the need for it. The research conducted by [8] took into account the same study area as the current study, but it did not pay particular attention to quantile regression; instead, it offered valuable data on the prevalence of obesity in the study area. However, the current study emphasized the need of researching BMI and obesity-related concerns in the Nigerian setting, reiterating the necessity of local population-focused research. In order to understand the distribution of BMI among students in greater depth, the study by [12] placed a strong emphasis on the use of quantile regression. The authors were able to use quantile regression to examine the relationship between BMI and other factors at various percentiles, revealing details about the peculiarities of the BMI distribution. As a result, the study provided evidence in favour of the objectives of the current, which was to use quantile regression to examine the association between BMI and explanatory factors among students at Delta State Polytechnic. The study by [19] looked at weight changes in individuals being treated for inflammatory bowel disease with various biologic drugs. Although unrelated to quantile regression or the student population in Nigeria, the study emphasized the significance of taking into account many parameters, including baseline BMI, in order to comprehend weight increases. It highlights the need of doing a thorough investigation of BMI prediction models and the impacts of explanatory factors, which is what the current study attempted to do by utilizing quantile regression. In a multi-ethnic sample from rural South Africa, the study by [24] looked at the prevalence and determinants of overweight and obesity. It included criteria including sex, age, education level, and living in a rural area that is linked to higher BMI while also emphasizing smoking's preventive qualities. Despite being conducted in a different setting from the one used in the current study, it highlights the need of comprehending the variables that contribute to overweight and obesity as well as the demand for focused treatments. This study, which examines BMI prediction models among students in Delta State, Nigeria, is thereby made even more pertinent by the research.

The paucity of research on Body Mass Index (BMI) prediction models relevant to Nigerian students, especially those enrolled at Delta State Polytechnic, is the issue that the present study seeks to address, as shown by the results of the review of related literature. There is a vacuum in our knowledge of the particular drivers of BMI among Nigerian students and the consequences of these factors on certain quantiles, despite the fact that current literature has examined the association between BMI and a variety of parameters. The creation of efficient treatments and policies to address the increased incidence of overweight and obesity among Nigerian students is hampered by this information gap. By contrasting classical linear regression with quantile regression models to predict BMI and offer insights into the various impacts of explanatory factors across quantiles, the study seeks to close this gap. By addressing this issue, the present study aims to advance quantile-specific research and policy, add to the body of knowledge, guide targeted health interventions, improve health education programs, and address BMI issues among students at Delta State Polytechnic in Nigeria.

2. RESEARCH METHOD

The section looks at the research method adopted in the study to examine and analyze the data obtained for the study between the variables considered for the study**.** Specifically, how the Linear Regression Model and Quantile Regression will be utilized to examine the impact of predictor variables on the response variable.

2.1. The General Linear Regression Model

A popular statistical framework for assessing the associations between a response variable and one or more predictor variables is the general linear regression model (GLM). In order to estimate the unknown parameters, it assumes a linear connection between the response variable and predictors. The GLM offers a reliable approach to parameter estimation by taking into account specific assumptions, such as the normal distribution of the response variable and the absence of correlation among predictors. This section looks at the theoretical foundations of the GLM, explores its assumptions, and discusses the estimation techniques used to obtain the parameter estimates.

The general linear regression model is given by the form:

$$Y = X\beta + \epsilon \quad \text{.........} \quad (1)$$

where $Y = (y_1, y_2, \dots, y_n)^T$ is an $n \times 1$ column vector of response variables, $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)^T$ is an $n \times 1$ column vector of errors, $\beta = (\beta_0, \beta_1, \dots, \beta_k)^T$ is an $(k+1) \times 1$ column vector of unknown parameters and

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

is an $n \times (k+1)$ matrix of predictor variables and k is the number of predictor variables.
The argument about the model in (1) borders on how to estimate the parameters vector β. Before this can be done, there are certain assumptions which are usually made concerning $Y, X$and ϵ and they are as follows:

a. Y is normally distributed,
b. the predictor variables X are stochastically fixed and uncorrelated,
c. ϵ is normally distributed with zero mean and a constant variance i.e. $\epsilon \sim N(0, \sigma^2 I_n)$, where 0 is an $n \times 1$ vector of zeros and I_n is the n-dimensional identity matrix and
d. The errors are uncorrelated i.e. $Cov(\epsilon_i, \epsilon_j) = 0$.

Given the above assumptions, a suitable estimator $\hat{\beta}$ is sought for β.

2.2. The ordinary least squares (OLS) estimator for $\boldsymbol{\beta}$

The OLS method is the most commonly used method for estimating the parameter β of the regression model in (1) given the aforementioned assumptions. The OLS method involves the minimization of the sum of square of the error ϵ given in the model in (1). Given

$$Y = X\beta + \epsilon,$$
$$\epsilon = Y - X\beta.$$

The sum of square of the error is thus expressed as:

$$\begin{aligned} \epsilon^T\epsilon &= (Y - X\beta)^T (Y - X\beta) \\ &= (Y^T - \beta^T X^T)^T (Y - X\beta) \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \\ &= Y^T Y - 2\beta^T X^T Y + \beta^T X^T X\beta. \end{aligned} \quad \text{.......} (2)$$

To obtain the OLS estimate $\hat{\beta}$ of β, we shall differentiate (2) w.r.t β and set the derivative to zero and thereafter solving for β. That is

$$\frac{\partial \epsilon^T \epsilon}{\partial \beta} = -2X^T Y + 2\beta X^T X = 0$$
$$2\beta X^T X = 2X^T Y$$

and thus

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y \quad \text{.......} (3)$$

whenever $(X^T X)$ is a non-singular matrix with an inverse. Thus $\hat{\beta}_{OLS}$ gives the OLS estimator of β. Also, since ϵ is normally distributed with $E(\epsilon) = 0$ and $Var(\epsilon) = E(\epsilon^T \epsilon) = \sigma^2 I_n$ the OLS estimator $\hat{\beta}_{OLS}$ is also normally distributed with mean and variance given by

$$E(\hat{\beta}_{OLS}) = \beta,$$
$$Var(\hat{\beta}_{OLS}) = \sigma^2 (X^T X)^{-1}.$$

2.3. General Arguments for Quantile Regression

Here we shall take a look at the theory and methodology behind quantile regression as distinct from the classical regression model specified in (1) under the aforementioned assumptions. We shall begin the section by looking at what quantiles mean.

2.4. Quantiles

In statistics and probability, quantiles are split points dividing the support of a probability distribution into continuous intervals each with equal probabilities, or dividing the observations in a sample in a similar way. Thus, given a probability threshold p, the quantile corresponding to p defined as $Q(p)$, that is the pth quantile is defined such that:

a. a proportion p of the observations of the distributions have a value lower or equal to p;
b. a proportion $1-p$ of the observations have a value greater or equal to p.

Mathematically, the quantile $Q_Y(p)$ of level $p\epsilon(0,1)$ of a random variable Y with cumulative distribution function (cdf) $F_Y(y)$ is defined in such a way that

$$P\{Y \leq Q_Y(p)\} \geq p \text{ and } P\{Y \geq Q_Y(p)\} \geq 1-p\,.$$

It can also be defined as

$$Q_Y(p) = F_Y^{-1}(p) = \inf\{y\epsilon\mathbb{R}: F_Y(y) \geq p\}.$$

Thus the quantile function $Q_Y(p)$ of the random variable Y is the inverse of its cdf. The most used quantiles in Statistics and Probability are:

a. $p = 0.5$: the median of the distribution
b. $p = \{0.1, 0.9\}$: the first and last deciles
c. $p = \{0.25, 0.75\}$: the first and last quartiles.

Now, the median $Q_Y(0.5) = m$ which is a quantile, of a real-valued random variable Y with cdf $F_Y(y) = P(Y \leq y)$ separates the distribution of Y into two halves where 50% of the observations of the distribution are below m and the other 50% are above m. Mathematically, the median exist in such a way that

$$P(Y \leq m) \geq \frac{1}{2} \text{ and } P(Y \geq m) \geq \frac{1}{2}.$$

If the probability distribution of the random variable Y is symmetric like for the normal distribution, the median m of Y and the mean (if it exists) are the same and the value is same as the value of the point where the symmetry occurs. The median can also be seen as the solution of a minimization problem. In fact, the median minimizes the absolute sum of deviations

$$\text{median}(Y) = \arg\min_m \mathbb{E}|Y-m|.$$

In general, quantiles can be viewed as particular centers of the distribution that minimizes the weighted absolute sum of deviations. For the pth quantile:

$$Q_p(Y) = \arg\min_c \mathbb{E}\left[\rho_p(Y-c)\right] \quad \text{.......(4)}$$

where $\rho_p(y)$ is a loss function defined as:

$$\begin{aligned}\rho_p(y) &= \left[p - \mathbb{I}_{(y<0)}\right]y \\ &= \left[(1-p)\mathbb{I}_{(y\leq 0)} + p\mathbb{I}_{(y>0)}\right]|y|.\end{aligned} \quad \text{.......(5)}$$

Thus, the loss function is an asymmetric absolute loss function; that is a weighted sum of absolute deviations, where a $(1-p)$ weight is assigned to the negative deviations and a p weight is used for the positive deviations. If the random variable Y is discrete with probability mass function $f_Y(y) = P(Y=y)$, (4) becomes

$$\begin{aligned}Q_p(Y) &= \arg\min_c \mathbb{E}\left[\rho_p(Y-c)\right], \\ &= \arg\min_c \left\{(1-p)\sum_{y\leq c}|y-c|f_Y(y) + p\sum_{y>c}|y-c|f_Y(y)\right\},\end{aligned}$$

and if Y is continuous with probability density function (pdf) $f_Y(y) = P(Y=y)$, (4) becomes

$$\begin{aligned}Q_p(Y) &= \arg\min_c \mathbb{E}\left[\rho_p(Y-c)\right], \\ &= \arg\min_c \left\{(1-p)\int_{-\infty}^{c}|y-c|f_Y(y)dy + p\int_{c}^{\infty}|y-c|f_Y(y)dy\right\},\end{aligned}$$

From definition, given a continuous random variable Y with pdf $f_Y(y)$, the cdf corresponding to the pdf is given by

$$F_Y(y) = P(Y \leq y) = \int_{-\infty}^{y} f_Y(t)dt,$$

and the corresponding survival function is given by

$$\overline{F_Y}(y) = P(Y > y) = \int_y^{\infty} f_Y(t)dt,$$

and consequently we have

$$c = F_Y^{-1}(0.5) = \text{medaian}.$$

Thus the median minimizes the expected value of the absolute difference. In general, and for a continuous random variable Y, the pth quantile of Y is given by

$$Q_p(Y) = \arg\min_c \mathbb{E}[\rho_p(Y - c)]$$

$$= \arg\min_c \left\{(1-p)\int_{-\infty}^{c} |y - c| f_Y(y)dy + p\int_{c}^{\infty} |y - c| f_Y(y)dy\right\}.$$

It follows that

$$Q_p(Y) = (1-p)\, F_Y(c) - p[1 - F_Y(c)] = 0,$$

and thus

$$F_Y(c) - p\, F_Y(c) - p + p\, F_Y(c) = 0$$

This implies that

$$F_Y(c) = p$$

and consequently we have

$$c = F_Y^{-1}(p) = \text{the pth quantile}.$$

2.5. Quantile regression

In the classical linear regression model in (1) the focus was on the conditional distribution of the response variable Y. That is, attention is only given to the mean effect of the predictor variable X on Y. Formally, given

$$Y = X\beta + \epsilon,$$
$$\mathbb{E}[Y|X] = X\beta.$$

Thus in many practical situations, we only look at the effects of a predictor on the conditional mean of the response variable. However, there might be some asymmetry in the effects across the quantiles of the response variable. In fact, it is not possible that the effect of a variable be the same for all observations. This is the rationale behind the use of quantile regression in place of the classical regression model which only accounts for the mean effect of the predictors on the response variable. In short, quantile regression offers a way out to account for these asymmetries which the classical linear regression model assumes away.

Given

$$\mathbb{E}[Y|X] = X\beta,$$

here β represents the marginal change in the mean of the response variable to a marginal change in the predictor variable X and the OLS estimator of β from (3) is

$$\hat{\beta}_{OLS} = (X^T X)^{-1} X^T Y.$$

Now, instead of looking at the mean effect, we can look at the effect at a given quantile. The condition quantile is thus defined as

$$Q_p[Y|X] = \inf\{y \epsilon \mathbb{R}: F_Y(y|x) \geq p\}.$$

The linear quantile regression model becomes:

$$Y = X\beta_p + \epsilon_p \quad \text{.......} \quad (6)$$

where $Y = (y_1, y_2, \dots, y_n)^T$ is an $n \times 1$ column vector of the response variable,

$$X = \begin{pmatrix} 1 & x_{11} & x_{21} & \cdots & x_{k1} \\ 1 & x_{12} & x_{22} & \cdots & x_{k2} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ 1 & x_{1n} & x_{2n} & \cdots & x_{kn} \end{pmatrix}$$

is an $n \times (k+1)$ matrix of predictor variables and k is the number of predictor variables, $\epsilon_p = \left(\epsilon_{1,p}, \epsilon_{2,p}, \dots, \epsilon_{n,p}\right)^T$ is an $n \times 1$ column vector of quantile errors and $\beta_p = \left(\beta_{0,p}, \beta_{1,p}, \dots, \beta_{k,p}\right)^T$ is a $(k+1) \times 1$ column vector of unknown quantile coefficient which corresponds to the marginal change in the pth quantile following a marginal change in X.
If we assume that $Q_p[\epsilon|X] = 0$, then (6) becomes:

$$Q_p[Y|X] = X\beta_p \quad \text{.......(7)}$$

Observe that in the classical linear regression model in (1), the interest is on predicting the average value of the response variable Y given the explanatory variable(s) X and hence, there is always one regression model. In the case of the quantile regression model in (7), the interest is not in predicting the mean or average value of the response variable Y, but predicting its value for a given quantile of its observations. Thus there are many regression models in (7) with each corresponding to a given quantile of interest. For example $p = 0.5$ gives us the median regression with model

$$Q_{0.5}[Y|X] = X\beta_{0.5},$$

$p = 0.1$ gives us the first decile regression with model

$$Q_{0.1}[Y|X] = X\beta_{0.1},$$

$p = 0.9$ gives us the last decile regression with model

$$Q_{0.9}[Y|X] = X\beta_{0.9},$$

$p = 0.25$ gives us the first quartile regression with model

$$Q_{0.25}[Y|X] = X\beta_{0.25},$$

and $p = 0.25$ gives us the last quartile regression with model

$$Q_{0.75}[Y|X] = X\beta_{0.75}.$$

The idea or rationale behind the use of quantile regression is well woven within the context of application. In the analysis of Body Mass Index (BMI) it is usually the lower or upper quantile that is of interest other than the average of the data. In this context, people with obesity or people with malnutrition are of interest and using the classical linear regression model will not allow us capture these two extremes in our analysis as all the observations on the weights of all the individuals are averaged out given some set of explanatory variables. This may lead one to make an erroneous predictive conclusion on how the explanatory variables impact the response variable which in this case is the weight of the individuals. However, quantile regression gives us the allowance to observe the effects of the explanatory variables on each quantile or specific quantile of the weight of the individuals.

2.6. Estimating the quantile regression model

As shown in sub-section 2.1, the unknown parameter vector β in classical regression linear regression model in (1) is estimated using the OLS method by minimizing the sum of squared errors. That is

$$\hat{\beta} = \arg \min_{\beta}\{(Y - X\beta)^T(Y - X\beta)\}.$$

For quantile regression, the approach is different. Given that the pth quantile minimizes the risk associated with the asymmetric absolute loss function

$$Q_p(Y) = \arg \min_{c} \mathbb{E}\left[\rho_p(Y - c)\right],$$

it follows that the pth quantile of Y solves

$$\min_{c} \sum_{i=1}^{n} \rho_p(y_i - c).$$

Given that

$$Q_p[Y|X] = X\beta_p,$$

it follows that the quantile estimator of the parameter vector β_p is given by:

$$\beta_p = \arg \min_{\beta} \frac{1}{n}\sum_{i=1}^{n} \rho_p\left(y_i - X_i\beta_p\right) \quad \text{.......(8)}$$

were the minimization problem in (8) is solved using linear programming which are well implemented in many software packages and in this study, we shall use the R software to perform the exercise.

Quantile regression will be applied to model BMI (kg/m2) of 152 students at Delta State Polytechnic, Nigeria. BMI is calculated as weight divided by height squared. Explanatory variables include skin fold, feeding habit, feeding frequency, and exercise frequency. Quantile and classical regression models will be compared to assess variable impact on BMI.

The classical regression model for these variables is given by:

$$BMI = \beta_0 + \beta_1 SK + \beta_2 FH + \beta_3 FF + \beta_4 FE + \epsilon \quad \text{.......} (9)$$

We shall formulate the quantile regression model for the quantiles $p = 0.1, 0.25, 0.5, 0.75, 0.9$ in the following order:

$$BMI = \beta_{0,0.1} + \beta_{1,0.1} SK + \beta_{2,0.1} FH + \beta_{3,0.1} FF + \beta_{4,0.1} FE + \epsilon_{0.1} \quad \text{.......} (10)$$

$$BMI = \beta_{0,0.25} + \beta_{1,0.25} SK + \beta_{2,0.25} FH + \beta_{3,0.25} FF + \beta_{4,0.25} FE + \epsilon_{0.25} \quad \text{.......} (11)$$

$$BMI = \beta_{0,0.5} + \beta_{1,0.5} SK + \beta_{2,0.5} FH + \beta_{3,0.5} FF + \beta_{4,0.5} FE + \epsilon_{0.5} \quad \text{.......} (12)$$

$$BMI = \beta_{0,0.75} + \beta_{1,0.75} SK + \beta_{2,0.75} FH + \beta_{3,0.75} FF + \beta_{4,0.75} FE + \epsilon_{0.75} \quad \text{.......} (13)$$

$$BMI = \beta_{0,0.9} + \beta_{1,0.9} SK + \beta_{2,0.9} FH + \beta_{3,0.9} FF + \beta_{4,0.9} FE + \epsilon_{0.9} \quad \text{.......} (14)$$

We shall use the classical linear regression model in (9) and the quantile linear regression models in (10) – (14) to fit the data on the variables using the 'quantreg' package in the R programming software. The results are well tabulated in Tables in the next section. We shall compare the results of the classical linear regression model to those obtained from the quantile regression model at respective quantile value in order to determine whether the value of each parameter in the models are either overstated or understated by the linear regression model for a given quantile. The would also use the comparison to determine if the nature of relationship between respective explanatory variables and the BMI will change from that of the classical linear regression model for a given quantile.

3. RESULTS AND DISCUSSIONS

In this section, we shall carry out the fitting of the regression model in (9) and the quantile regression models in (10) – (15). The results are contained in Tables (1-5).

Table 1. Comparison of the performance of the methods setting $p = 0.1$

Coefficients	Classical Linear Regression	Quantile Linear regression $p = 0.1$
β_0	23.0286	18.9917
β_1	1.1210	0.4250
β_2	-0.3516	0.3250
β_3	0.4401	-0.0833
β_4	-0.5349	0.3250

The result presented in Table 1 shows coefficients of linear and quantile regression models (p=0.1). Linear regression had higher β_0 than quantile regression. β_0 represents autonomous BMI (independent of factors). First decile autonomous BMI is 18.9917 kg/m2, average is 23.0286 kg/m2. $\beta_1, \beta_2, \beta_3, \beta_4$ differ between linear and quantile regression, measuring marginal effects of variables on BMI. $\beta_1 = 1.1210$ implies 1 unit increase in skin fold leads to BMI increase. $\beta_1 = 0.4250$ in quantile regression for the first decile, showing smaller increase. Negative coefficients indicate BMI decrease with variable increase. By contrasting the coefficients of classical linear regression and quantile regression models for BMI prediction, the present study's findings add to the conclusion reached by [13] and [24] and further the results of those studies.

Table 2. Comparison of the performance of the methods setting $p = 0.25$

Coefficients	Classical Linear Regression	Quantile Linear regression $p = 0.25$
β_0	23.0286	21.5750
β_1	1.1210	-0.7500
β_2	-0.3516	-0.5250

β_3	0.4401	0.0250
β_4	-0.5349	0.9250

The result presented in Table 2 shows the individual model coefficients of the two methods setting *p*=0.25. It was found that the classical linear regression clearly gave a value of β_0 higher than the one obtained for the quantile regression even though both are positive. The result clearly reveals the autonomous BMI of the first quartile of the population is 21.5750 kg/m² while the average autonomous BMI is 23.0286 kg/m². Observe also the difference in the value of the $\beta_1, \beta_2, \beta_3, \beta_4$ for the case of the classical linear regression model and the quantile linear regression. For example, $\beta_1 = 1.1210$ implies that for every one unit increase in skin fold, there would be a corresponding increase in the BMI and the rest parameters are interpreted similarly. Observe that for the quantile regression, $\beta_1 = -0.7500$ this implies that one unit increase in skin fold will only lead to a 0.7500 kg/m² decrease in the BMI only in the first quartile of the population. Here the classical linear regression model is stating an increase while the quantile regression is specifying a decrease. The outcome was consistent with the findings obtained by [6–8], which supported the use of quantile regression for a more thorough examination of the impact of variables at various quantiles.

Table 3. Comparison of the performance of the methods setting $p = 0.5$

Coefficients	Classical Linear Regression	Quantile Linear regression $p = 0.5$
β_0	23.0286	22.0500
β_1	1.1210	0.0000
β_2	-0.3516	-0.3500
β_3	0.4401	0.6000
β_4	-0.5349	-0.0500

The result presented in Table 3 shows the individual model coefficients of the two methods setting *p*=0.5. It was found that the classical linear regression clearly gave a value of β_0 higher than the one obtained for the quantile regression even though both are positive. The result clearly reveals the autonomous BMI of the median of the population is 22.0500 kg/m² while the average autonomous BMI is 23.0286 kg/m². Observe also the difference in the value of the $\beta_1, \beta_2, \beta_3, \beta_4$ for the case of the classical linear regression model and the quantile linear regression. For example, $\beta_1 = 1.1210$ implies that for every one unit increase in skin fold, there would be a corresponding increase in the BMI and the rest parameters are interpreted similarly. Observe that for the quantile regression, $\beta_1 = 0.0000$ this implies that one unit increase in skin fold will have no effect on the BMI only in the first median of the population. Here the classical linear regression model is stating an increase while the quantile regression is specifying no effect. This finding conflicts with those of [6] and [14], who found that the quantile regression model gave useful insights into the ways in which various factors influenced BMI and per capita income at different quantiles.

Table 4. Comparison of the performance of the methods setting $p = 0.75$

Coefficients	Classical Linear Regression	Quantile Linear regression $p = 0.75$
β_0	23.0286	25.4000
β_1	1.1210	1.3250
β_2	-0.3516	-0.6500
β_3	0.4401	0.5250
β_4	-0.5349	-0.6250

The result presented in Table 4 shows the individual model coefficients of the two methods setting *p*=0.75. It was found that the classical linear regression clearly gave a value of β_0 lower than the one obtained for the quantile regression even though both are positive. The result clearly reveals the autonomous BMI of the third quarter of the population is 25.4000 kg/m² while the average autonomous BMI is 23.0286 kg/m². Observe also the difference in the value of the $\beta_1, \beta_2, \beta_3, \beta_4$ for the case of the classical linear regression model and the quantile linear regression. For example, $\beta_1 = 1.1210$ implies that for every one unit increase in skin fold, there would be a corresponding increase in the BMI and the rest parameters are interpreted similarly. Observe that for the quantile regression, $\beta_1 = 1.3250$ this implies that one unit increase in skin fold will result in an increase in the BMI only in the third quartile of the population. This result was consistent with the findings of the study by [6–9], which advocated quantile regression and highlighted its advantages in examining the relationships between factors and BMI at various quantiles.

Table 5. Comparison of the performance of the methods setting $p = 0.9$

Coefficients	Classical Linear Regression	Quantile Linear regression $p = 0.9$
β_0	23.0286	27.0750
β_1	1.1210	2.3500
β_2	-0.3516	-0.2500
β_3	0.4401	0.9750
β_4	-0.5349	-2.1500

The result presented in Table 5 shows the individual model coefficients of the two methods setting *p*=0.9. It was found that the classical linear regression clearly gave a value of β_0 lower than the one obtained for the quantile regression even though both are positive. The result clearly reveals the autonomous BMI of the last decile of the population is 27.0750 kg/m^2 while the average autonomous BMI is 23.0286 kg/m^2. Observe also the difference in the value of the $\beta_1, \beta_2, \beta_3, \beta_4$ for the case of the classical linear regression model and the quantile linear regression. For example, $\beta_1 = 1.1210$ implies that for every one unit increase in skin fold, there would be a corresponding increase in the BMI and the rest parameters are interpreted similarly. Observe that for the quantile regression, $\beta_1 = 2.3500$ this implies that one unit increase in skin fold will result in a increase in the BMI only in the last decile of the population. The findings show that the effects of the factors on BMI vary across different quantiles of the population, according to the quantile regression technique. It emphasizes the need of taking into account the quantile-specific effects for a more thorough understanding of the link between factors and BMI by pointing out that the associations between the variables and BMI are more significant in the higher BMI range.

By contrasting classical linear regression with quantile regression models for BMI prediction, the outcome of the present study extends previous findings from [13] and [24]. Additionally, it was discovered that the results agreed with [6–8], supporting the use of quantile regression for a thorough examination of variable influences across quantiles. The results of [6] and [14], which stressed the use of quantile regression in evaluating the effect of variables on BMI and per capita income at various quantiles, were found to be contradicted by one of the findings. Further results were found to be consistent with [6–9], which supports the use of quantile regression to study associations between variables and BMI at different quantiles.

4. CONCLUSION

The study compared classical linear regression with quantile regression models for predicting BMI in 152 students from Delta State Polytechnic, Nigeria. The findings revealed significant differences between the models, indicating varying effects of explanatory variables across quantiles. Quantile regression offered a more nuanced understanding of the data, highlighting the limitations of the classical linear regression model. The study contributes to the body of knowledge by emphasizing the importance of tailored health interventions, enhanced health education programs, and quantile-specific research and policies to address BMI concerns. The research implications include the need for personalized interventions, informed lifestyle decisions, and policy strategies promoting healthy BMI ranges. The research limitations suggest the potential for sample size expansion, improved generalizability, longitudinal design, integration of more variables, comparison analyses, and intervention studies. Future research should address these limitations to enhance understanding and practical actions for maintaining a healthy BMI range among students of Delta State Polytechnic, Nigeria.

REFERENCES

[1] P. C. Bürkner and M. Vuorre, "Ordinal Regression Models in Psychology: A Tutorial," Advances in Methods and Practices in Psychological Science, vol. 2, no. 1, 2019. [Online]. Available: https://doi.org/10.1177/2515245918823199

[2] N. Kwark and C. Lee, "Asymmetric effects of financial conditions on GDP growth in Korea: A quantile regression analysis," Economic Modelling, vol. 94, pp. 351-369, 2021. [Online]. Available: https://doi.org/10.1016/j.econmod.2020.10.014

[3] T. H. Kim, E. K. Lee, and E. Han, "Quantile regression analyses of associated factors for body mass index in Korean adolescents," Public Health, vol. 129, no. 5, pp. 424-435, 2015. [Online]. Available: https://doi.org/10.1016/j.puhe.2015.01.021

[4] A. Nicolucci and C. Maffeis, "The adolescent with obesity: what perspectives for treatment?," Ital J Pediatr, vol. 48, no. 1, p. 9, 2022. [Online]. Available: https://doi.org/10.1186/s13052-022-01205-w

[5] Q. Huang, H. Zhang, J. Chen, and M. He, "Quantile Regression Models and Their Applications: A Review," Journal of Biometrics & Biostatistics, vol. 8, no. 3, 2017

[6] D. A. Amugsi, Z. T. Dimbuene, G. Asiki, and C. Kyobutungi, "Quantile regression analysis of modifiable and non-modifiable drivers' of blood pressure among urban and rural women in Ghana," Scientific Reports, vol. 8, no. 1, 2018. https://doi.org/10.1038/s41598-018-26991-4

[7] J. A. Mitchell, M. Dowda, R. R. Pate, K. Kordas, K. Froberg, L. B. Sardinha, E. Kolle, and A. Page, "Physical Activity and Pediatric Obesity: A Quantile Regression Analysis," Medicine and Science in Sports and Exercise, vol. 49, no. 3, 2017. [Online]. Available: https://doi.org/10.1249/MSS.0000000000001129

[8] O. Agofure, "Prevalence of obesity among adults in Issele-Uku, Delta State Nigeria," Alexandria Journal of Medicine, vol. 54, no. 4, pp. 463-468, 2018. [Online]. Available: https://doi.org/10.1016/j.ajme.2017.10.005

[9] C. Tukuitonga and I. Keller, "Implementing the World Health Organization Global Strategy on Diet, Physical Activity and Health," Scandinavian Journal of Nutrition/Naringsforskning, vol. 49, no. 3, 2005

[10] I. Janssen, P. T. Katzmarzyk, and R. Ross, "Waist circumference and not body mass index explains obesity-related health risk," The American Journal of Clinical Nutrition, vol. 79, no. 3, pp. 379-384, 2004. [Online]. Available: https://doi.org/10.1093/ajcn/79.3.379

[11] W. Maziak, S. Rastam, T. Eissenberg, T. Asfar, F. Hammal, M. E. Bachir, M. F. Fouad, and K. D. Ward, "Gender and smoking status-based analysis of views regarding waterpipe and cigarette smoking in Aleppo, Syria," Preventive Medicine, vol. 38, no. 4, 2004.

[12] H. Yun, D. W. Kim, E. J. Lee, J. Jung, and S. Yoo, "Analysis of the Effects of Nutrient Intake and Dietary Habits on Depression in Korean Adults," Nutrients, vol. 13, no. 4, p. 1360, Apr. 2021.

[13] D. Yıldırım and A. K. Çelik, "Testing the pecking order theory of capital structure: Evidence from Turkey using panel quantile regression approach," Borsa Istanbul Review, vol. 2, no. 2, pp. 460-479, Jun. 2020.

[14] J. Duasa and N. H. Zainal, "Determinants of household income of micro-finance women participants: a quantile regression approach," Ecofeminism and Climate Change, vol. 1, no. 1, pp. 63-74, 2020.

[15] Y. Chen, K. Raza, and M. Alharthi, "The nexus between remittances, education, and energy consumption: Evidence from developing countries," Energy Strategy Reviews, vol. 46, p. 101069, 2023. [Online]. Available: https://doi.org/10.1016/j.esr.2023.101069

[16] H. Uchinuma, K. Tsuchiya, T. Sekine, S. Horiuchi, M. Kushima, S. Otawa, H. Yokomichi, K. Miyake, Y. Akiyama, T. Ooka, R. Kojima, R. Shinohara, S. Hirata, Z. Yamagata, M. Kamijima, S. Yamazaki, Y. Ohya, R. Kishi, N. Yaegashi, K. Hashimoto, C. Mori, S. Ito, Z. Yamagata, H. Inadera, T. Nsakayama, H. Iso, M. Shima, Y. Kurozawa, N. Suganuma, K. Kusuhara, T. Katoh, "Correction: Gestational body weight gain and risk of low birth weight or macrosomia in women of Japan: a nationwide cohort study (International Journal of Obesity, (2021), 45, 12, (2666-2674), 10.1038/s41366-021-00947-7)," International Journal of Obesity, vol. 46, no. 6, 2022.

[17] F. M. Costa Leite, J. R. Tovar Garbin, M. Junior, and L. C. Lopes-Júnior, "COVID-19 Mortality in Public Hospitals in a Brazilian State: An Analysis of the Three Waves of the Pandemic," International Journal of Environmental Research and Public Health, vol. 19, no. 21, 2022.

[18] A.-F. Turcotte, S. Jean, S. N. Morin, F. Mac-Way, and C. Gagnon, "Relationships between Obesity and Incidence of Fractures in a Middle-Aged Population: A Study from the CARTaGENE Cohort," JBMR Plus, vol. 5, no. 1, 2023.

[19] P. Kaazan, Z. Tan, P. Maiyani, M. Mickenbecker, S. Edwards, C. McIvor, J. M. Andrews, "Weight and BMI Patterns in a Biologicals-Treated IBD Cohort," Digestive Diseases and Sciences, vol. 67, no. 12, pp. 5628-5636, 2022. [Online]. Available: https://doi.org/10.1007/s10620-022-07488-7

[20] S. Vahidi, A. Aghausefi, and H. Namvar, "Predicting happiness based on self-differentiation mediated coping styles in obese men and women," Journal of Applied Family Therapy, vol. 2, no. 2, pp. 460-479.

[21] S. Venkatesh, T. Ferreira, S. Benonisdottir, N. Rahmioglu, C. M. Becker, I. Granne, K. T. Zondervan, M. V. Holmes, C. M. Lindgren, and L. B. Wittemans, "Obesity and risk of female reproductive conditions: A Mendelian randomisation study," PLoS Medicine, vol. 19, no. 2, 2022. [Online]. Available: https://doi.org/10.1371/journal.pmed.1003679.

[22] L. J. Dominguez, N. Veronese, G. Di Bella, C. Cusumano, A. Parisi, F. Tagliaferri, S. Ciriminna, and M. Barbagallo, "Mediterranean diet in the management and prevention of obesity," Experimental Gerontology, vol. 174, p. 112121, 2023. [Online]. Available: https://doi.org/10.1016/j.exger.2023.112121.

[23] A. Dawed, T. C. Mekonnen, M. Genetu, S. E. Tadesse, R. Dewau, A. Muche, A. A. Zerga, F. Y. Ayele, and T. K. Gill, "Comparing the validity of anthropometric measurements in identifying malnutrition status of older age people in Borena district, North Central Ethiopia: A cross-sectional study," BMC Geriatrics, vol. 22, no. 1, p. 776, 2022.

[24] M. H. Smith, J. W. Myrick, O. Oyageshio, C. Uren, J. Saayman, S. Boolay, C. Werely, M. Möller, B. M. Henn, and A. W. Reynolds, "Epidemiological correlates of overweight and obesity in the Northern Cape Province, South Africa," PeerJ, vol. 11, 2023. [Online]. Available: https://doi.org/10.7717/peerj.14723.

[25] F. Teufel, J. A. Seiglie, P. Geldsetzer, M. Theilmann, M. E. Marcus, C. Ebert, W. A. L. Arboleda, K. Agoudavi, G. Andall-Brereton, K. K. Aryal, B. W. Bicaba, G. Brian, P. Bovet, M. Dorobantu, M. S. Gurung, D. Guwatudde, C. Houehanou, D. Houinato, J. M. A. Jorgensen, G. B. Kagaruki, K. B. Karki, D. Labadarios, J. S. Martins, M. T. Mayige, R. W. McClure, J. K. Mwangi, O. Mwalim, B. Norov, S. Crooks, F. Farzadfar, S. S. Moghaddam, B. K. Silver, L. Sturua, C. S. Wesseh, A. C. Stokes, U. R. Essien, J. W. De Neve, R. Atun, J. I. Davies, S. Vollmer, T. W. Bärnighausen, M. K. Ali, J. B. Meigs, D. J. Wexler, and J. Manne-Goehler, "Body-mass index and diabetes risk in 57 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 685,616 adults," The Lancet, vol. 398, no. 10296, pp. 238–248, 2021

[26] M. Siddiqa, G. H. Shah, T. L. Mayo-Gamble, and A. Zubair, "Determinants of Child Stunting, Wasting, and Underweight: Evidence from 2017 to 2018 Pakistan Demographic and Health Survey," Journal of Nutrition and Metabolism, vol. 2023, Article ID 2845133, pp. 1-12, 2023.