Unfolding Factors Influencing Household Expenditure in Private Households in the Tanzania Mainland Using Quantile Regression

Damian Lewis[1]
Bahati Ilembo[2]

[1]masharichikolomo@gmail.com
[2]bmilembo@mzumbe.ac.tz

[1,2]Mzumbe University, Morogoro, United Republic of Tanzania

ABSTRACT

The purpose of this paper was to unfold factors that influence household expenditure using quantile regression methods. Secondary data were used with a probability sample of 9,552 households from the 2017-18 Household Budget Survey (HBS) covered the population residing in private households in Tanzania Mainland. The sample was designed to allow separate estimates for each of the 26 regions of the Tanzania Mainland, as well as urban and rural areas separately at the national level. The 2017-18 HBS adopted a two-stage cluster sample design. In the quantiles analysis, there has been consistency on the effect of household consumption, education level and household age on household expenditure. These showed a statistical significance at 5% level of significance. Except for the 25th quantile, all the three variables were statistically significant in the 50th and 75th quantiles. Further household age, household consumption and marital status of the household head had positive coefficients which implied a positive relationship with the household expenditure. It can thus be concluded that, of the seven independent variables, household age, household consumption and household marital status are the variables which influences positively household expenditure. By considering these findings, policymakers and researchers can develop targeted strategies to address economic disparities and enhance household financial well-being effectively to balance between household income and expenditure.

Keywords: Cluster, Household, Quantile Regression, Wellbeing

I. INTRODUCTION

Household consumption patterns are influenced by various factors such as food availability, social, culture, household size, number of people who are employed , level of education, income and the main job of head of household (Jayalakshmi & Indira, 2023; Maniriho et al., 2021; Siman et al. 2020; Mubarak et al., 2019). Demand led growth studies affirm that household consumption expenditure is a principal macroeconomic driving force (Alalo et al., 2020). According to İpek and Sekmen (2017), determination of the factors influencing consumption behavior has been the subject of many empirical studies in the microeconomic literature with the availability of micro level data. It is argued that, economists use two different approaches to analyze household decisions. The urinary approach, which ignores the differences between single- person and multi- person households; whereas the collective approach states that each person in the household must be characterized by specific preferences (Maniriho et al., 2021). It is therefore essential to understand the nature of household expenditure, a fundamental macroeconomic driver with immense significance for policy making (Alalo et al., 2020).

Studies on factors influencing household expenditure exists and mostly they use multiple linear regressions (Jayalakshmi & Indira, 2023; Maniriho et al., 2021; Mubarak et al., 2019; Rahim et al., 2018; Djibuti et al., 2017; Sekhampu & Niyimbanira, 2013) which examines the linear relationship between variables, mainly dependent and independent variables. This method is unable to identify the distribution of consumption expenditure for the household than the quantile methods (Cheah et al., 2021). According to İpek and Sekman (2017), the traditional least squares regression provides to approximate the conditional mean and median located at the centre of the distribution. Such mean- based estimators are lacking representing the whole distribution of the data. Thus, the quantile regression analysis which is proposed by Koenker and Bassert (1978), provides more comprehensive information of conditional quantile functions, whereby each function qualifies the behavior of a certain point in the conditional distribution and thereby it identifies the conditional distribution completely. Quantile regressions are more robust, able to perform the estimations when the error term is distributed non- normal and avoiding heteroscedasticity.

1.1 Statement of the Problem

Over three-fifths or 61.9% of Tanzanian Mainland poor people are living in households with 7 or more household members (URT, 2022) . Dar es Salaam region being the main commercial city in the country is the only place

its people suffer from the increase in the cost of living and whose household dwellers face higher expenditures because everything is sold at higher price, especially foodstuffs and non-food stuffs like clothes hence affect the household expenditure (URT, 2022). According to Kostakis (2014), consumers who live in urban places used to consume a lot of money on food. Many studies assessed more on food expenditure at household level than socioeconomic and demographic factors on household's expenditure, the way this study is doing. In general, there are no clear elaborating demographic, social, and economic factors influencing household expenditure to the best of our knowledge. This paper is an attempt to fill this gap by careful examining the demographic, social and economic factors that influence household expenditure by using quantile regression model.

1.2 Research Objectives

The broad objective of the paper is to determine the factors influencing household expenditure. Specifically, the paper is set to address the following two specific objectives: firstly, to examine the demographic factors affecting the household expenditure and secondly, to examine the social- economic factors affecting also the household expenditure. In line with these specific objectives, the paper attempts to answer one research question; is there any relationship between demographics and socio – economic factors against the total household expenditure?

The rest of the paper is organized as follows, section II presents the literature review, section III gives the methodology and section IV presents findings and discussion of the findings and section V concludes and presents the policy recommendations.

II. LITERATURE REVIEW

2.1 Theoretical Review

Traditional demand theory is based on the individual behaviour of a customer. Household purchasing power is determined by income, the higher the purchasing power, the more products that can be bought (Note that this may not be the case for goods). Previous research in economics looks at theories on food expenditure. As time passes, the empirical study of consumer behaviour regarding household food spending becomes more interesting. Numerous scholars have attempted to explain the features of the demand for food expenditures. The demand is mostly determined by four factors: social, cultural, psychological, and demographic. The growing distribution of food expenditures among household groups is the reason for the growing interest (Kinsey, 1994). According to Schiffman and Kanuk (2010), the theory of consumer behaviour investigates the how, when, and why people consume the products they consume. The theory focused on elements influencing decision-making at the individual and family levels in order to give marketers the abilities and information need to deal with consumer analyses, which are employed for both comprehending markets and developing marketing strategies (Omotoyinbo et al., 2017).

When making purchases, consumers must weigh the trade-off between having limitless demands and limited income. Customers need to find a balance between their preferences and budgets to buy the goods they desire. It is assumed that the buyer is reasonable and seeks to attain a maximum of satisfaction within the constraints of his or her money and the going rate. The customer allocates their income in a way that maximizes their level of enjoyment (Colander & Su, 2018).

Engel's law is an economics observation which relates amount spent on food and income. Precisely, the law states that the rise in household's income reduces the contribution of that income spent on food, even if the entire amount spent on food rises. In other words, while people will buy more food in total as they get richer, the portion of their total spending that goes towards food will fall. Put otherwise, the range of 0 to 1 shows the income elasticity of food demand. Engel's law says customers raise their expenditures for food products in percentage terms less than their income increases, not that food expenditure stays constant as income rises. The consumption level -per person is seen as a key indicator of an economy's production success since Neoclassical economists (broad) usually believe that consumption is the final objective of economic activity (Nuru et al., 2015).

Economists like Keynes (1936), Duesenberry (1949), Friedman (1957) and Modigliani (1963) have been influenced by the factors that determine consumption expenditure to examine variables that are qualitative as well as quantitative and can affect consumption, such as wealth, income, interest rates, capital appreciation, and liquid assets. The reason these components are studied is that consumer expenditures, in any economy, are a crucial part in the process of economic growth (Branson, 1989). Keynes made this connection in a conceptual breakthrough that confirmed the correlation between income and consumer expenditure in 1936.

2.1.1 Inter Temporal Choice Theory (ICT)

People's income limits their consumption that is why they don't consume as much as they would like to. People choose to consume less than or equal to their available budget as a result of this financial limitation. They encounter an

intertemporal budget constraint when determining how much to save for the future versus how much to consume today. A consumer has two life periods of time: period one corresponding to their youth, and period two to their old age. Consumers generate revenue in period one (say, Y_1) and spend C_1, then in period two (Y_2) they generate income and spend C_2. Because consumers have the option to borrow and save, their consumption in a given period may be higher or lower than their income during that same period(Syariah & Ilmu, n.d.).

In period one, income less consumption equals saving and defined by the equation, $S = Y_1 - C_1$. In the period two, the accumulated savings plus the interest rate earned on the saving, and the second period income gives the consumption denoted by $C_2 = (1 + r)S + Y_2$. In the period one, if consumption is less than income, consumer is saving and saving is greater than zero. In addition, if consumption is greater than income, the consumer borrowing and saving is less than zero with the same rate of interest. To derive the consumer's budget constraint, combine the two equations above to obtain$C_2 = (1 + r)(Y_1 - C_1) + Y_2$. When this equation is arranged for easier interpretations, it becomes:

$$(1 + r)C_1 + C_2 = (1 + r)\, Y_1 + Y_2 \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(*)}$$

Dividing both sides of equation $(*)\ by\ (1 + r)$ gives the following:

$$\frac{(1 + r)C_1 + C_2}{1 + r} = \frac{(1 + r)\, Y_1 + Y_2}{1 + r} \to C_1 + C_2 = Y_1 + Y_2$$

The traditional way of corresponding to the intertemporal budget constraint is by an equation that relates the income and consumption for the two periods. The price of C_2 expressed in terms of C_1 is indicated by $\frac{1}{1+r}$ and the factor $(1 + r)$ represents the future income and consumption discount rate. The conceptual framework is given hereunder:

2.2 Empirical Review

Studies on the relationship between socio- economic and demographic factors versus total household expenditure are well documented in the literature. Mignouna et al. (2020) while using quantile regression model in the analysis, attempted to explain the consumption patterns of 1700 producers of yams, for a randomly selected sample. It was revealed that, the primary determinant of the disparities in household consumption expenditures and quantiles is educational attainment. They also found that there is insignificant relationship between total income and consumption expenditure, which goes against theoretical expectations.

Çağlayan (2012) performed a micro econometric analysis to examine the factors that influence household consumption spending in Turkey's rural and urban districts. The research looked into Turkey's household consumption spending determinants. An attempt to assess the regional disparities for the total expenditure distribution in that study was done and the data analysis employed was quantile regression. The results showed that whereas consumption expenditures in rural estimations fall with age, they grow in general and in urban estimations. Only income, age, marital status, insurance, and household size were significant in rural estimations.

Zin and Nabilah (2015) used both Ordinary Least Squares (OLS) and Quantile Regression (QR) models to explain the determinants of households' consumption behaviour through socioeconomic variables such as, gender, age, ethnicity, education levels, marital status, employment status, and household size for the years 1999, 2005, and 2010. They found that education level has a positive and significant effect on consumption expenditures for urban areas, but household size, and work status of family heads are most relevant variables determining consumption expenditures for the rural areas.

Quantile regression has gained reputation in applied research and is very useful in providing estimated parameters of a regression model by minimizing the weighted sum of absolute residuals (Xu, 2023). Accordingly, since Koenker and Bassett (1978) introduced the theory of linear quantile regression, the quantile regression has gained popularity and extensively applied regression model method. Due to its advantages, the quantile regression offers a novel regression method that addresses many shortcomings of least squares (OLS) method. To this background, the choice of quantile regression as a method of data analysis was made expected to yield robust results.

III. METHODOLOGY

This study used secondary data with a probability sample of 9,552 households from the 2017-18 HBS covered the population residing in private households in Tanzania Mainland. The sample was designed to allow separate estimates for each of the 26 regions of the Tanzania Mainland, as well as urban and rural areas separately at the national level. The HBS adopted a two-stage cluster sample design. The first stage involved the selection of enumeration areas (primary sampling units - PSUs) from the 2012 Population and Housing Census (2012 PHC) Frame. A total of 796 PSUs (69 from Dar es Salaam, 167 from other urban areas, and 560 from rural areas) were selected. The National Bureau of Statistics (NBS) carried out a listing exercise in which households residing in selected PSUs were freshly listed to update the 2012 PHC list before selecting households. The second stage of sampling involved systematic sampling of

households from the updated PSUs list. A sample of 12 households was selected from each selected PSU. All household members regardless of their age, who were usual members of the selected households, and all visitors who were present in the household on the night before the survey interview, were eligible for the survey.

3.1 Theoretical Model

In configuring the basic principle of quantile regression, it is ideal to set the general linear regression model as:

$$y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p + \vartheta \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 1$$

Under the premise of satisfying the Gauss- Markov hypothesis, equation 1 can be expressed in terms of expectation as:

$$\mathrm{E}(y/x) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

Where the $b_i's; i = 0,1,2, \ldots\ldots, p$ are the coefficients to be estimated

Model 2 is the expression of the mean reversion model, which is the result of taking the mathematical expectation on both sides of equation 2. Similarly, the median regression model can also be set as follows:

$$\mathrm{M}(y/x) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p + M(\vartheta) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 3$$

Where:

$\mathrm{M}(y/x)$ is the conditional median about x, and $M(\vartheta)$ is the median of the random disturbance term. The quantile regression model is thus;

$$Q_y(\psi/x) = b_0 + b_1x_1 + b_2x_2 + \cdots + b_px_p + Q_\vartheta(\psi) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 4$$

For the mean regression model, the ordinary least squares method (OLS),can be used to estimate the unknown parameters. For the median regression model, the least square method (LAD) can be used while for quantile regression model, the linear programming method (LP) can be used to estimate the minimum weighted absolute deviation, and the regression coefficient of explanatory variables can be obtained. The following expressions follow;

Ordinary Least Squares approach gives;

$\min \mathrm{E}\left(y - b_0 - b_1x_1 - b_2x_2 - \cdots - b_px_p\right)^2$ forms the following solution:

$\widehat{\mathrm{E}}\ (y/x) = \widehat{b_0} + \widehat{b_1}x_1 + \widehat{b_2}x_2 + \cdots + \widehat{b_p}\ x_p$

The LAD approach gives;

$\min \mathrm{E}\left|y - b_0 - b_1x_1 - b_2x_2 - \cdots - b_px_p\right|$ whose solution is;

$$\widehat{\mathrm{M}}\ (y/x) = \widehat{b_0} + \widehat{b_1}x_1 + \widehat{b_2}x_2 + \cdots + \widehat{b_p}\ x_p$$

Quantile regression approach gives;

$\min \mathrm{E}\lambda_\psi\left(y - b_0 - b_1x_1 - b_2x_2 - \cdots - b_px_p\right)$ and the solution is;

$$\widehat{Q_y}\ (\psi/x) = \widehat{b_0} + \widehat{b_1}x_1 + \widehat{b_2}x_2 + \cdots + \widehat{b_p}\ x_p$$

$$Where, \lambda_\psi(t) = t(\lambda - I(t < 0)), \psi\epsilon(0,1)$$

The quantile regression model introduced by Koenker and Bassert (1978) is defined solution to minimise the equation below for the αth regression quantile, for $0 < \alpha < 1$.

$$min_{z\in R^P}[\sum_{t\in(t:y_t>x_tb)} \alpha/y_t - x_tb| + \sum_{t\in(t:y_t<x_tb} (1-\alpha)|y_t - x_tb \ldots\ldots\ldots\ldots 5$$

Where $\{x_t: 1, \ldots\ldots\}$a sequence of p is – vectors of a known design matrix, and $\{y_t: 1, \ldots\ldots, T\}$ is a random sample on the regression process$\psi_t = y_t - x_t\phi$ having distribution function F.

Relevant to this study, the Kroenker and Bassert quantile regression model is applied on the Household Budget Survey micro data set for the reference year and for this matter 2017. The dependent variable is household expenditure defined as total value of all expenditures on individual and collective consumption goods and services incurred by residents' households.

IV. FINDINGS & DISCUSSIONS

4.1 Descriptive Statistics

Table 1 shows descriptive statistics of the sample. They include sex, marital status, level of education, household size, household age, total household expenditure and total consumption. These variables form the socio-demographics and have been used in the conceptual framework to examine their influence on household expenditure (HE).

Table 1
Descriptive Statistics

Variable	Category	Count (%)	Statistic
Sex of HH head	Male	556(74.53)	
	Female	190 (25.47)	
Marital status of the HH head	Widowed	64 (8.58)	
	Divorced	41 (5.50)	
	Married	490 (65.68)	
	Separated	36 (4.83)	
	Single	115 (15.42)	
Education level of the HH head	High education	127 (17.02)	
	Primary education	193 (25.87)	
	Secondary	426 (57.10)	
Household size	Small size	563 (75.47)	
	Medium size	169 (22.65)	
	Larger size	14 (1.88)	
Employment status of the HH head	Employed	195 (26.14)	
	Self employed	462 (61.93)	
	Unemployed	89 (11.93)	
Age of HH head			Mean: 43.14
			Min. 19
			Max. 83
Total HH expenditure (TZS)			Mean: 809347.2
			Min. 93522.92
			Max. 10,100,000
Total HH consumption (TZS)			Mean: 626,933.3
			Min. 51,874.75
			Max. 9,701,477

The results show that most of the households, 556 (74.53%) were male – headed, while only 190 (25.4%) households were female – headed. A total of 490 (65.68%) household heads were married, followed by those who were single (15.42%) and widowed (8.58%). The total number of divorced families was 41 (5.50%) while the separated families were 36 (4.83%). With regard to education attainment, the results show that the majority of household heads 426 (57.10%) obtained Secondary school education while only 193 (25.87%) household heads obtained primary school education. There were 127 (17.02%) household heads that attained a high education level. The dominance size of the household was recorded to be those between 0-5 members which accounted for 563 (75.47%), and was termed as small size, followed by medium size household between 6 – 10 members which represented 169 (22.65%) followed by large sized households and contributed 14 (1.88%) and sized between 11 to 19 members.

Further, descriptive statistics showed that the average age among household members was 43 years, with range of 64. The average household total expenditures were 809,347.2 TZS monthly while the average household consumption was 626,933.3 TZS monthly. The minimum monthly expenditure by households was 93522.92 TZS. The maximum monthly expenditure by a household was 10,100,000 TZS. The minimum level of consumption used by a household was 51,874.75 TZS per month and the maximum consumption used by a household was 9,701,477 TZS per month.

4.2 Classical Linearly Regression Models' Assumptions and Diagnostic Test

4.2.1 Normality Test

The Shapiro –Wilks was used to test for normality and the null hypothesis was that the data are normally distributed as shown below.

Table 2
Shapiro–Wilks Test Results

Variable	Observation	W	V	Z	Prob>z
resid	746	0.591	197.462	12.930	0.000

The results shows the Z- statistic and its associated p-value (Prob > z). If the p-value is greater than 0.05 (commonly chosen significance level), it suggests that the null hypothesis is accepted. Therefore, we conclude that the

data are approximately normal. On the other hand, if the p-value is less than or equal to 0.05, it indicates that the null hypothesis is not accepted. In this case, we conclude that the data are skewed on either side, left or right. The p-value was found to be 0.000, which is less than 0.05 (5%) and therefore there is enough evidence not to accept the null hypothesis under 5% level of significant and conclude that the data do not follow a normal distribution. This is supported by observing the value of the mean on HH expenditure which is not zero as one of the requirements for the normal distribution. Similar diagnosis results are also shown in Figure 2 below.

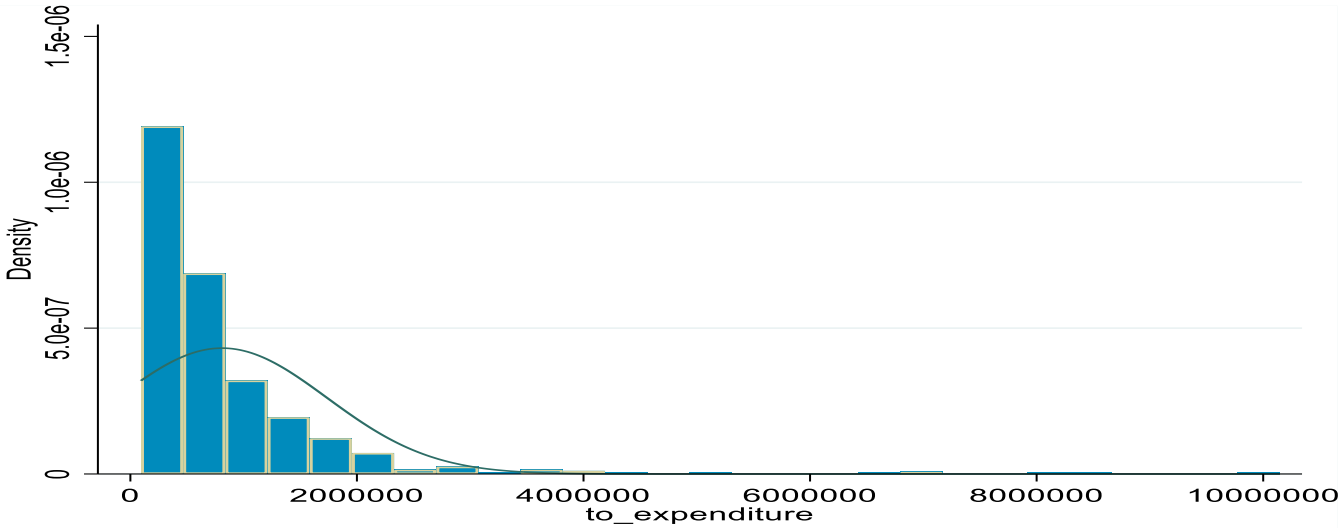

Figure 2
A Histogram Plot Showing Normality Test

The curve is skewed to the right side which confirms the data to be non-normal distributed as data deviate from the centre. Right-skewness often indicates the presence of outliers or extreme values on the high end of the expenditure scale. These outliers could be due to unusual or one-time expenses, luxury purchases, or other factors driving up total expenditure for certain cases.

Quantile regression allows researcher to examine the relationship between variables at different points of the conditional distribution, providing insights into how the relationship may vary across different quantiles. This flexibility is particularly useful when the relationship between variables is not constant across the entire distribution. Quantile regression helped to show if there is an effect of the predictor's variable varying across different levels of household expenditure. When the data distribution is skewed, traditional OLS regression is particularly vulnerable to outliers (Koenker & Hallock, 2001). Given that, quantile regression takes into account the whole distribution, it is less impacted by extreme values and produces reliable results. For this case, the 25th, 50th, and 75th quantiles were used to assess the relationship between the quantiles.

Table 3
Socio-Economic and Demographic Factors on Total Household Expenditures (25th Quantile level)

Expenditure	Coefficient	Std. errs.	t	P>|t|
hhhage	497.3693	368.256	1.35	0.177
consumption	1.103898	0.0062922	175.44	0.000
Sex	-6847.527	9356.05	-0.73	0.464
household__size	-14110.62	8893.299	-1.59	0.113
education level	-12505.17	5811.773	-2.15	0.032
employment	-242.0163	6756.306	-0.04	0.971
marital status	1129.337	3478.97	0.32	0.746
constant	4939.027	30839.72	0.16	0.873

Pseudo R^2= 0.7325

From Table 3 it was found that the coefficient of the variable consumption is positive and significantly influenced household expenditure while education level is negative but also significantly influencing household expenditure. The negative sign for education means the more educated (attaining high level of education) household head is, the less amount of household expenditure it becomes and vice versa. Similarly, it implies that in the 25th quantile,

a one-unit increase in total household consumption leads to an increase in expenditure by approximately 1.103898 units. Regarding the level of education, the findings from Table 3 shows that the negative signed coefficient indicates that, at the 25th quantile, higher education levels are associated with lower to expenditure by approximately 12505.17 units, this effect is significant at 5% level (p-value = 0.032 < 0.05). Furthermore in 25th quantile, the variables age and marital status were positively related to household expenditure but sex, household size and employment status were altogether negatively related to the household expenditure. They were all insignificant. The pseudo R-squared value of 0.73 indicates that approximately 73% of the variability in household expenditure at the 25th quantile is explained by the variables used in the model. The choice of the variables are in agreement with Bagarani et al. (2009), Nguyen et al. (2007) and İpek and Sekmen (2017) when attempting to analyze the effects of household heterogeneity on consumption expenditure, especially when using household budget survey data.

Table 4

Socio-Economic and Demographic Factors on Total Household Expenditures (50th Quantile level)

expenditure	**Coefficient**	**Std. error.**	**t**	**P>|t|**
hhhage	1373.398	548.0292	2.51	0.012
consumption	1.191756	0.0093639	127.27	0.000
Sex	-11307.54	13923.44	-0.81	0.417
household_size	-11746.52	13234.78	-0.89	0.375
education_level	-33016.95	8648.933	-3.82	0.000
employment	-7708.665	10054.56	-0.77	0.444
marital_status	986.0487	5177.314	0.19	0.849
constant	25294.01	45894.88	0.55	0.582

Pseudo R^2= 0.7488

Table 4 shows that, the coefficient for household age, consumption and marital status are positive and thus positively related to household consumption while sex, household size, education level, and employment status are negatively related to household expenditure. Unfortunately, it is only household age, consumption and education level are statistically significant ($p - value < 0.05$). Generally, this implies that household expenditure at the 50th quantile is strongly positively impacted by total household consumption, household age and education level and negatively affected by sex, household size, education level and employment status. Generally, the correlation between education level of the head of the household and household expenditure at the 50th quantile is highly significant (p-value < 0.000). On the other hand, there is insufficient evidence to conclude that the insignificant variables have a significant effect on household expenditure at the 50th quantile. The pseudo R-squared value of 0.7488 indicates that approximately 74.88% of the variability in household expenditure at the 50th quantile is explained by the variables used. This suggests that the model is good- fit for the data at the 50th quantile.

The median household expenditure at the 50th quantile has been determined so that, for each household in the dataset, half of the expenditures fall below this amount and the other half fall beyond it. Within this framework, the coefficients offer valuable perspectives on the variables linked to median household spending. This further suggests that it is essential to understand the nature of household expenditure which is vital macroeconomic driver with immense significance for policy making. This is in agreement with Tian et al. (2016) when argued that household consumption expenditure is a principal macroeconomic driving force and that all economic activities are affected by the level of private household expenditure. Thus, understanding the portion of the distribution of household expenditure through median becomes important.

Table 5

Socio-Economic and Demographic Factors on Total Household Expenditures (75th Quantile level)

expenditure	**Coefficient**	**Std. error**	**t**	**P>|t|**
hhhage	3215.405	957.2903	3.36	0.001
consumption	1.34454	0.0163567	82.20	0.000
Sex	-16928.24	24321.28	-0.70	0.487
household_size	-7376.964	23118.34	-0.32	0.750
education_level	-50246.32	15107.84	-3.33	0.001
employment	-2047.186	17563.18	-0.12	0.907
marital_status	4240.84	9043.665	0.47	0.639
constant	-15532.26	80168.59	-0.19	0.846

Pseudo R^2= 0.7679

Table 5 shows that total consumption, household age and marital status have got positive coefficients implying that, they all positively related to household expenditure. Meaning that, household expenditure increases with household consumption, household age as well as marital status. Except for marital status, household age and household consumption are significantly associated to household expenditure at 5% level of significance. The variables sex, household size, education level and employment are negatively related to household expenditure with education level being statistically significant. This could be that people with greater levels of education typically have different spending priorities or are more frugal with their money.

The insignificance of the variable household size means that, it has no discernible impact on total household spending. At the 75th quantile level, neither job status nor marital status significantly affects total household spending. In conclusion, a detailed knowledge of the many levels at which socioeconomic and demographic characteristics impact total household spending is made possible by the comparison across quantile as also shown in Figure 3 and as suggested by Xu (2023). According to Xu (2023), quantile regression offers a novel regression method that addresses many shortcomings of the least squares method.

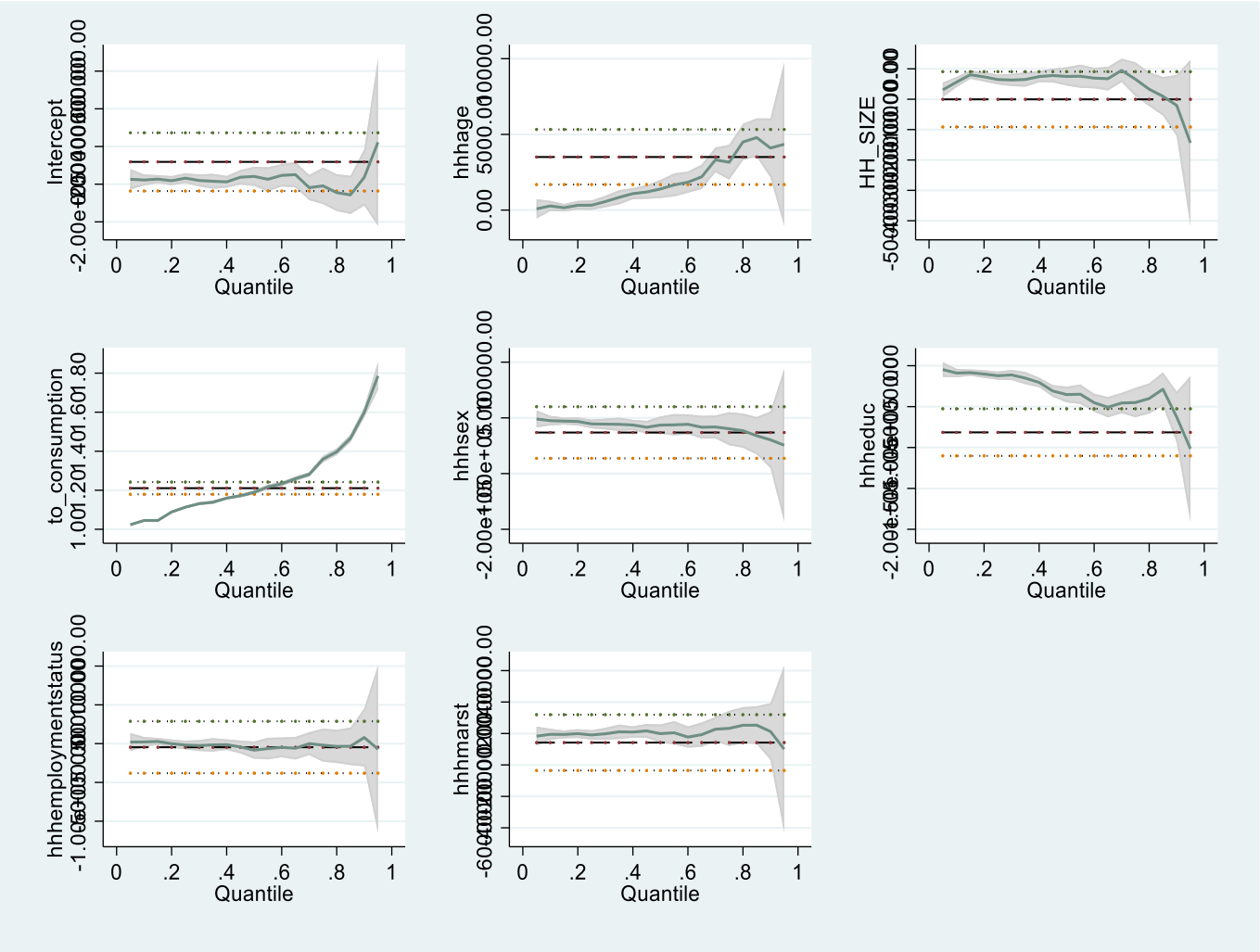

Figure 3

Quantile Plots for the Estimated Parameters with their 95% Confidence Band

Figure 3 provides another way of looking at the quantile and their interpretation to improve on the results. These considerations include consistency of coefficients and coefficient magnitude as well as determining influential observations. The consistency of coefficients evaluate whether the coefficients' direction and significance hold true for different quantiles. There is a strong correlation with the predictor and the result throughout a range of conditional quantiles if coefficients rise or fall with quantiles and continue to be statistically significant. From the plot it seems that the coefficient like total consumption and household head age has got a strong correlation.

Also, the magnitudes of the coefficients were compared across quantiles to inform on the correlations. A higher correlation between the predictor and the result for that quantile is indicated by a larger coefficient magnitude. From the

graph above it seems that many coefficients have the almost same size of magnitude; for example employment status and household marital status seem to have larger magnitude.

Examining potential relationships that might occur between quantiles and predictors would be another dimension that can be learned from the quantile regression. It becomes more important when the interaction effect is realized which calls for an investigation of any patterns of interaction by looking at the variations in the relationship's slope between quantiles. From the graph above it can be revealed that, household age and household size as they cross the quantiles, shows the interaction effect among the quantiles. The findings are in agreement with Mubarak et al. (2019) Cheah et al. (2021) and Simatupang and Junaidi (2020) but contrary to Siman et al. (2020) who concluded that the number of household members did not affect the household consumption.

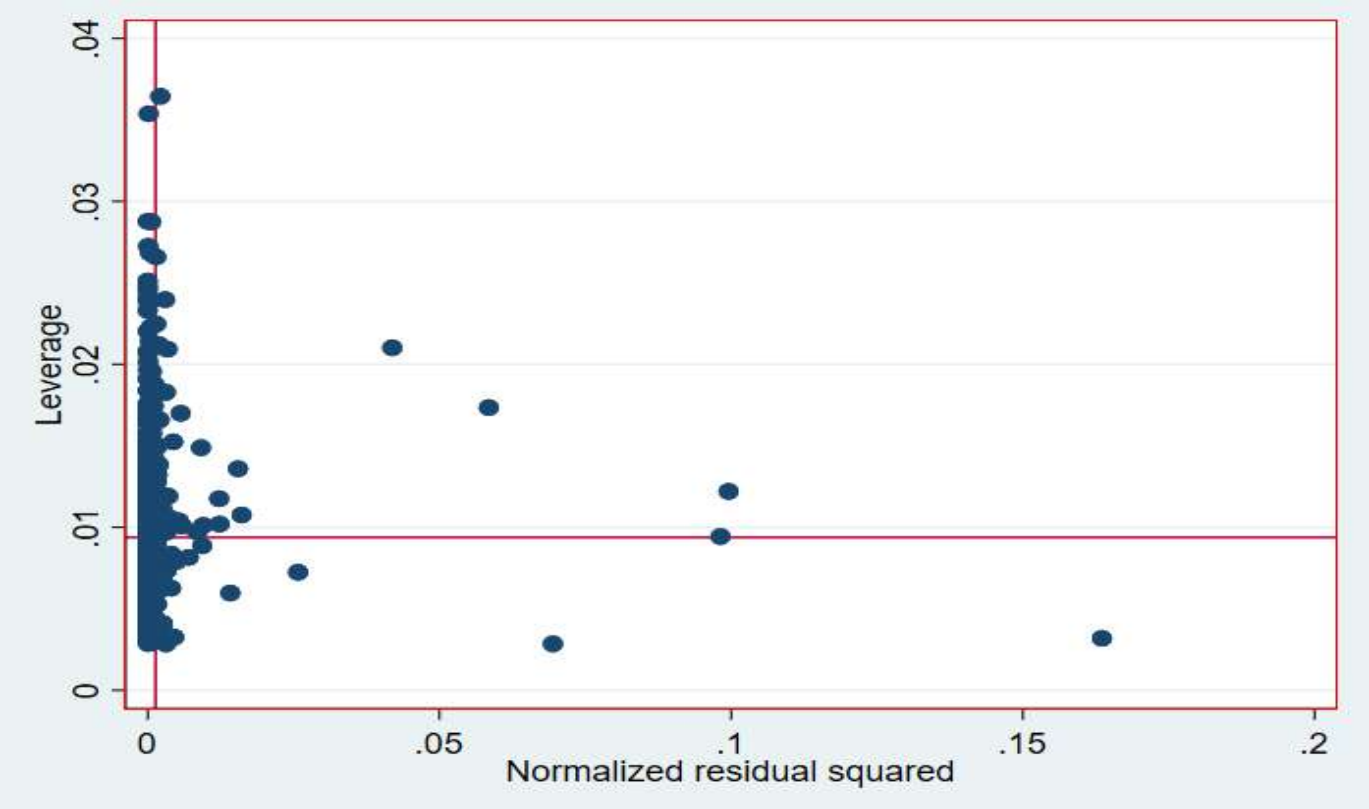

Figure 4
Normalized Leverage Against Residual Squared

Figure 4 shows the model capturing the variability in the data when the residuals, which represent the differences between the observed and predicted values of the dependent variable, are randomly distributed around zero for most data points. In other words, if the model accurately predicts the dependent variable across different levels of the predictors, the residuals should be small.

Influential observations, refers to those observations that significantly affect the model parameter estimation. Influential observations in the context of quantile regression are frequently recognized by the combination of big residuals and high leverage. The predicted quantile regression coefficients can be strongly impacted by these findings, especially for quantiles when the observations diverge more from the general trend. It is therefore possible to determine whether the model sufficiently explains these points or whether they are possible outliers that need more research by finding the influential observations. Finally, leverage in quantile regression quantifies each data point's impact on the quantile regression coefficients' estimation. Points with high leverage have a greater impact on the projected quantile regression line, just like in linear regression. The disparities between the values expected by the quantile regression model at a specific quantile level and the observed values are represented by residuals in quantile regression. Usually, they are standardized to make cross-observation comparisons easier. This means that the changing structure of the consumption expenditures in different points (quantiles) with household socio- economic factors such as age, sex, marital status, among others, can be observed as also documented by İpek and Sekmen (2017).

V. CONCLUSIONS & RECOMMENDATIONS

5.1 Conclusions

This paper was set to use the strengths of quantile regression to unfolding factors influencing household expenditure. The variables used included household age, sex, consumption, household size, education level, employment status and marital status. In the quantiles analysis, there has been consistency on the effect of household consumption, education level and household age on household expenditure. These showed a statistical significance at 5% level of significance. Except for the 25th quantile, all the three variables were significant in the 50th and 75th quantiles. Further household age, household consumption and marital status of the household head had positive coefficients which implied a positive relationship with the household expenditure. It can thus be concluded that, of the seven independent variables, household age, household consumption and household marital status are the variables which influences positively household expenditure.

5.2 Recommendations

The summary of the conclusion above suggests that in order to balance between household expenditure and the factors that influence it, the following is recommended. The government should enhance income support programs such as minimum wage policy in order to ensure a stable income for the households. The significance relationship between level of education and household expenditure suggests that there should be an investment in education together with vocational training in order to enhance skills of household members as well as their ability to generate income through self-employment. Finally, there should be sustainable financial literacy programs that will help households be able to make budgets as well as control their day-to-day spending.

REFERENCES

Alalo, M., Aljabber, A., & Naseeb, A. (2020). Household expenditure patterns in Kuwait. *Journal of Research in Emerging Markets, 2*(2), 1.

Bagarani, M., Forleo, M., & Zampino, S. (2009). Households food expenditures behaviours and socioeconomic welfare in Italy: A micro econometric analysis. In *113th EAAE Seminar*, Crete, Greece.

Branson, T. P. (1989). Conformal transformation, conformal change, and conformal covariant. In *Proceedings of the Winter School "Geometry and Physics"* (pp. 115-134). Circolo Matematico di Palermo.

Çağlayan, E., & Astar, M. (2012). A micro econometric analysis of household consumption expenditure determinants for both rural and urban areas in Turkey.

Cheah, Y. K., Abdul Adzis, A., Abu Bakar, J., & Applanaidu, S. D. (2021). Factors associated with household expenditure on oil and fat products in Malaysia: Application of quantile regression. *Food Research, 5*(3), 112-120.

Colander, D. C., & Su, H. C. (2018). *How economics should be done: Essays on the art and craft of economics*. Edward Elgar Publishing.

Djibuti, M., Gotsadze, G., Mataradze, G., & Zoidze, A. (2007). Influence of household demographic and socio-economic factors on household expenditure on tobacco in six new independent states. *BMC Public Health, 7*(1), 1-6.

Duesenberry, J. S. (1949). *Income, saving and the theory of consumer behaviour*. Harvard University Press, Cambridge.

Friedman, M. (1957). The permanent income hypothesis. In *A theory of the consumption function* (pp. 20-37). Princeton University Press.

İpek, E., & Sekmen, Ö. (2017). Effect of household heterogeneity on consumption expenditure: A simultaneous quantile regression analysis. *The Empirical Economics Letters, 16*(12), 1329-1336.

Jayalakshmi, N., & Indira, M. (2023). Factors influencing household expenditure on school education in Karnataka. *International Journal of Management and Development Studies, 12*(11), 26-38.

Keynes, J. M. (1936). *A treatise on money*. Macmillan and Co. Limited.

Kinsey, J. D. (1994). Food and families' socioeconomic status. *The Journal of Nutrition, 124*(12), 1878S-1885S.

Koenker, R., & Bassett Jr, G. (1978). Regression quantiles. *Econometrica: Journal of the Econometric Society, 46*(1), 33-50.

Koenker, R., & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives, 15*(4), 143-156.

Kostakis, I. (2014). The determinants of households' food consumption in Greece. *International Journal of Food and Agricultural Economics (IJFAEC), 2*(2), 17-28.

Maniriho, A., Musabanganji, E., Nkikabahizi, F., Ruranga, C., & Lebailly, P. (2021). Analysis of the determinants of households' expenditures in Rwanda. *UKH Journal of Social Sciences, 5*(1), 8-17.

Mignouna, D. B., Akinola, A. A., Abdoulaye, T., Alene, A. D., Manyong, V., Maroya, N. G., & Asiedu, R. (2020). Potential returns to yam research investment in sub-Saharan Africa and beyond. *Outlook on Agriculture, 49*(3), 215-224.

Modigliani, F. (1965). The monetary mechanism and its interaction with real phenomena. In *The state of monetary economics* (pp. 79-107). NBER.

Mubarak, M. S., Yunus, A. K., Mandela, A., & Yunus, A. (2019). Analysis of factors affecting expenditure of poor households in Luwu District, Indonesia. *IOSR Journal of Economics and Finance (IOSR-JEF), 10*(5), 54-58.

Nguyen, B. T., Albrecht, J. W., Vroman, S. B., & Westbrook, M. D. (2007). A quantile regression decomposition of urban-rural inequality in Vietnam. *Journal of Development Economics, 83*(2), 466-490.

Nuru, F., Msalilwa, U., Mbwambo, L., & Balama, C. (2015). Assessment of household charcoal consumption in urban areas: The case of Dar es Salaam City, Tanzania. *Tanzania Journal of Forestry and Nature Conservation, 84*(2), 22-32.

Omotoyinbo, C., Worlu, R., & Ogunnaike, O. (2017). Consumer behaviour modelling: A myth or heuristic device? *Perspectives of Innovations, Economics & Business, 17*(2), 101-119.

Rahim, A., Hastuti, D. R. D., & Bustanul, N. (2018). Estimation of household consumption expenditure of small-scale fishermen in Indonesia. *Russian Journal of Agricultural and Socio-Economic Sciences, 83*(11), 375-383.

Sekhampu, T. J., & Niyimbanira, F. (2013). Analysis of the factors influencing household expenditure in a South African township. *International Business & Economics Research Journal, 12*(3), 279-284.

Schiffman, G., & Kanuk, L. (2010). *Consumer behavior* (10th ed.). Prentice Hall.

Siman, S., Tawakal, M. A., Risamasu, P. I. M., & Kadir, R. (2020). Effect of household size, working hours, health and income on consumption expenditure of poor households. *Enfermería Clínica, 30*(1), 512-515.

Simatupang, J., & Junaidi, J. (2020). Poor households expenditure for preventive and curative health needs. *International Journal of New Economics and Social Sciences, 11*(1), 197-208.

Tian, D., Hu, N., Wang, X., & Huang, L. (2016). Trade margins, quality upgrading, and China's agri-food export growth. *China Agricultural Economic Review, 8*(2). https://doi.org/10.1108/CAER-12-2013-0156

URT. (2022). *Tanzania panel survey*. United Republic of Tanzania.

Xu, M. (2023). Quantile regression model and its application research. *Academic Journal of Science and Technology, 8*(3), 172-176.

Zin, W. Z. W., & Nabilah, S. F. (2015). *Malaysian household consumption expenditure: Rural vs urban*. In Department of Statistics, Malaysia, MyStats 2015 Conference Papers.