Journal of Social & Organizational Matters

Vol 3 No 1 (2024): 176-196

Estimation of Quadratic Food Engel Curve in the Presence of Measurement Error and Endogeneity of Total Expenditure in Pakistan

Ghulam Abbas[1], Zahid Hussain Shaikh*[2]

[1]Department of Business Management and Economics, University of Baltistan, Skardu, Gilgit Baltistan, Pakistan.

[2*]Department of Mathematics & Social Sciences, Sukkur IBA University, Sukkur, Sindh, Pakistan.

Corresponding author: shaikhzahid@iba-suk.edu.pk

Keywords: *Engel Curve, Measurement Error, Endogeneity, Instrumental Variable, Turning Point*

DOI No:
https://doi.org/10.56976/jsom.v3i1.67

This study is conducted to compare the four different estimation techniques, Ordinary least square, Instrumental Variable, Lewbel (1996) approach and control function approach estimating the quadratic food Engel curve in the presence of the problem of measurement error and the problem of endogeneity of total expenditure. The authors have used the HIES data 2010-11 as it has been observed that the HIES data 2010-11 fits a quadratic shape of food Engel curve in Pakistani context. Moreover, the expenditure elasticity of food demand was observed between zero and one which indicates the fact that food is a necessary good for Pakistani households. The results also shows that endogeneity of total expenditure is more serious issue than the measurement error.

1. Introduction

The living standard and the welfare of households are measured via analysing the consumption patterns in economics (Sakketa & Gerber, 2024). Most of the literatures reported that households in developing countries spend major part of their income on food items (Zereyesus et al., 2017; Deaton 2018). In Pakistan, the share of food in the total consumption expenditure is reported about 48.91% in 2010-11 while 45.01% in 2011-12 which is largest than share of other commodities in total expenditure, as reported for housing, miscellaneous, and fuel & lighting 13.18%, 12.64% and 7.91% respectively (HIES 2011-12). These figures suggest that the share of food can be used as an indicator of household welfare. The relationship between households' food expenditure and their income is known as Engel's law in the literature of Economics. The empirical study on the relationship between food expenditure and income was performed by Engel - a German statistician. He found that the income elasticity of demand for food items was comparatively low. This observation, later on coined into the prominent Engel law. The Engel's law stipulates that as household income increases, the proportion of income spends on food decreases, revealing that the income elasticity of food is always positive but less than one (Zereyesus et al., 2018).

When households are able to sell their produce at a higher price, welfare gains are seen to be greater. It is important to remember that households generate and consume agricultural products (Shampio et al., 2024). Maintaining employment possibilities in the agricultural industry and promoting agricultural production depend on high agricultural prices, particularly for grain crops (Shahzad et al. 2023).

There are two arguments emerged about the Engel's law in the literatures. Firstly, the households increase the proportion of income on food with the increase in total income and this relationship specifies the linear relationship (Mbegalo & Yu, 2018).

Secondly, with the increase in income households change their preferences from food to luxurious goods which imply that food expenditure share will not increase in proportion to income but at lower spending levels, unit increases in total spending increase food budget share, above spending thresholds, unit increases in total spending reduce food budget share and specifies the quadratic relationship (Mbegalo & Yu 2018; Wirba 2021). As per the studies conducted by the researchers using the Pakistani data shows that the households with lower income spend a higher proportion of their income on food items as compared to the households with higher income (Moss et al., 2018)

Empirical evidences from the developing countries proved that the quadratic logarithmic food Engel curve is a feature of developing countries (Kedir & Girma 2017; Hassan 2019; Moss et al. 2018; Nsabimana et al., 2020). They also favoured and suggested a quadratic Engel curve in case of Pakistan using HIES 1987-88 data for rural households. The quadratic Engel curves have significant implication in designing tax policies, e.g a higher tax on food items as compare to non-food items implies that the tax burden is borne by low income households compared to high income households. This way, the formulation of government tax and transfer policies mainly depends upon the true specification of Engel's curve. In this connection, a misspecification of food Engel's curve leads misleading results and limits its effectiveness in formulation of policy advice.

Although, consumption basket of Pakistani population has changed significantly over the time. But the main objective of this study is to compare the four different estimation techniques, Ordinary least square, Instrumental Variable, this approach and control function approach to estimate the quadratic food Engel curve in the presence of the problem of measurement error and the problem of endogeneity of total expenditure to check whether endogeneity of total expenditure is more serious issue or the measurement error (Nsabimana et al., 2020).

In the estimation of Engel curve, when household's expenditure is used instead of household's income it may suffer from endogeneity problem. Endogeneity problem occurs when the regressors are correlated with the model error in the regression model. This problem arises because of three reasons namely measurement error, omitted variables and simultaneity (Selejio et al., 2020).

Measurement error was recognised in the demand literatures for the first time in early 1960s. He conducted a study using Israeli family budget survey and observed that ignorance of measurement error leads to a bias in ordinary least square (OLS) estimates of parameters of linear Engel curve. The empirical studies dealing with the measurement error prove that variation in the total expenditure data is because of the measurement errors. The study conducted by (Shitu, 2019) using the data of Norwegian budget survey found that the measurement error accounts about 27% variation in observed expenditure. Wirba (2023) use the non-linear error in variable model to investigate the Engel curve parameters and showed that about 42% variation in the measured total expenditure is because of measurement error. Using fuel data of United Kingdom, treated error of measurement (as the only the source of endogeneity) in both dependent and independent variables the demand equation and found that measurement error correction changes the estimates of the parameter of interest more than 15%.

Other source of endogeneity is simultaneity, which arises due to join determination of dependent and independent variables affecting each other. In case of food demand analysis, it is a common practice to choose food expenditure as dependent variable and total expenditure (instead of household's income) as independent simultaneously and these two variables are co-determined. It means that there is no strong justification to suppose the exogeneity of total expenditure in the demand equation. In the absence of measurement error, this problem is resolved through standard instrumental variable (IV) approach but the fact is that a higher proportion of variation in the total expenditure is due to the measurement error and it cannot be neglected (Kedir & Girma, 2018). The research conducted a study for treatment of both the total expenditure endogeneity and the measurement error in both sides of the demand equation via a control function approach. This study resulted out that the correction of measurement error and ignoring simultaneity problem leads to overestimation of the parameter. The endogeneity of total expenditure through simultaneity is more serious problem than measurement error.

The researchers observed in various studies that treating total expenditure endogenous and its measurement is error ridden the ordinary least square (OLS) yields a biased result of the parameter of interest. The instrumental variable (IV) approach does not work when there is measurement error while this approach corrected the measurement error treating it as the only source of endogeneity. When someone wishes to estimate the Engel curve both allowing for endogenous expenditure and correcting error of measurement, the above approaches failed to provide inferential conclusion. This gap is filled up by the pioneer work of researchers. As in the literature of Engel curve in case of Pakistan as per best of our information there is no study that controls for the un-observed expenditure endogeneity correcting measurement error. We followed the same setting of authors to fill-up this gap using HIES (2010-11) data and employing a quadratic Engel curve.

Based on the previous debate, the research questions that emerge are:

1. **What range of income defines the turning points of the food Engel's curves, if any?**
2. **Which one is more serious problem in estimation of food Engel's curve: endogeneity of total expenditure or the measurement error?**

The conforming objectives of the study are:

3. **To explore the income-range defining the turning points of the food Engel curves, if any.**
4. **To check whether the endogeneity of the total expenditure is more serious problem or the measurement error of the expenditure data.**

The rest of manuscript is organized as follows. Section two consisted of literature review on estimation and policy recommendation of food Engel's curve in the presence of endogeneity and total expenditure. Section three comprises on theoretical frame work. Section four dwells on the description of data and variables. Section five consisted of estimation process and econometrics modelling strategy adopted for this study. Section six presents the empirical results and discussion. In the last section, summery, conclusion and policy recommendations have been discussed.

2. Literature Review

As discussed earlier that the relationship between households' food expenditure and their income is known as Engel's law in the literature of Economics. Most of the literatures on Engel's curve based on the Working-Leser log linear model in which household food budget is expected to be a linear function of household total expenditure. Leser (1963) conducted a study and found that food Engel's curve has a linear specification.

Hausman et al. (1995) conducted a study to resolve the problem of errors in variables in non-linear regression model. In this study the authors proposed a new technique for non-linear error in variable model to investigate the Engel curve parameters. Using the data of US consumer expenditure survey (CES) for the year 1982, this study resulted out that about 42% variance in the measured total expenditure is because of measurement error. The findings suggest that Working-Leser specification must be widespread to higher order term in log of total expenditure.

Lewbel (1996) proposed a simple GMM technique applied to U.K. fuel demand data from Family Expenditure Survey (FES) 1980-82 estimating standard linear logarithmic model. He also extended the linear logarithmic model into quadratic logarithmic model as well. He resulted out that the effect of measurement error was found to be statistically significant, and the estimated measurement error correction changed parameter estimates by more than fifteen percent.

Banks et al. (1997) conducted a study and propose a new model to analyse the consumption patterns of household non-parametrically. The study, proposes Engel curve needs quadratic specification in the logarithm of total expenditure. In this study data of U.K Family Expenditure Survey (FES) was used and the results strongly rejected the functional form of Working-Leser for some goods, while for other commodities; particularly food vindicates the Working-Leser specification.

Bhalotra and Attfield (1998) investigate the shape of Engel's curve using the data of Household Income and Expenditure Survey of Pakistan for the year 1987-88 and resulted out that food Engel curve is quadratic in logarithmic.

Blundell et al. (1998) provided a new method to investigate the shape of Engel curves. The method is a semi-parametric one which accounts endogeneity problem also. This semi-parametric specification compared to the Working-Leser and quadratic logarithmic parametric specification through a recently developed specification test of Sahalia et al. (1994). This study strongly rejected the Working-Leser specification for few budget shares even after regulating for demographic difference and endogeneity but the quadratic logarithmic model seemed to be provided an acceptable parametric specification.

Kuha and Temple (2003) conducted a study to see the effect of measurement error. In this study, quadratic regression model was used where the explanatory variable was measured with error.

This study resulted out that the consequence of error of measurement was flatten the curvature of the estimated functions.

Kedir and Girma (2018) examined the consumption pattern based on the Ethiopia Urban Households Survey (EUHS, 1994). The study corrected error of measurement in both side of the budget share equation. This study found robust and statistically significant quadratic relationship between food budget share and total expenditure for Ethiopian households. This study also found that at low level of total expenditure, the Engel curve was sloping upward and started to downward sloping at 63rd percentile of total expenditure distribution. In this study the estimator approach of Hausman et al. (1995) which demonstrated the importance of accounting measurement errors in both side of the budget share equation.

Attaniaso et al. (2009) studied food Engel curve among the low income population targeted by a conditional cash transfer programme in Colombia. This study proposed a log-linear specification after accounting for the endogeneity of total expenditure and controlling for variability of prices across villages. This study found that a 10% increase in total expenditure leads to 1% declines in share of food.

Hassan (2012) examined the shape of Engle curve using semi-parametric technique with base independence assumption and also control for endogeneity. The test for semi-parametric specification reveals a quadratic shape of Engle curve for food, clothing, transport, education and other items. Due to OLS estimates suffer from endogeneity, the control function approach also used by taking the log income, square of log income and household land owning status as instruments. The result of the CF approach also supports the quadratic shape of Engle curve for all categories under consideration except medical expenses.

Lewbel (2012) examined that the endogeneity problem arises due dependence between instrumental variables and higher moments of expenditure distribution. He provided a new technique to obtain identification in mismeasured regressor model, triangular system and simultaneous equation system. This approach can be used in the application, in the absence of other sources of identification for instance instrumental variable or repeated measurements. Associated estimator takes the form of two stage least square (2SLS) or generalized method of moments (GMM) as employed by Lewbel (1996). He concluded that this methodology yields estimates that are close to estimates that are obtained by using an ordinary external instrument. De Nadai and Lewbel (2013) conducted a study to observe the effect of measurement error. They proposed a nonparametric estimator based on sieve approximation of the conditional moment which takes a form of conditional Generalized Moment Method (GMM). The study suggests that in case of food Engel curve Working-Laser specification is suitable. The study found that about one-third variance of total expenditure is because of error of measurement while three and seven percent of standard deviation of measurement errors in total expenditure are accounted for by measurement errors in clothing and food expenditure respectively. This study verified via simulation study that the proposed estimator greatly reduces mean squared error relative to other alternative estimators.

Battistin and De Nadai (2013) proposed control function approach for the estimation of Engel curve dealing both sources of endogeneity namely, error of measurement and the endogeneity of total expenditure. This method used to estimate the Engle curve for food using the data obtained from the Bank of Italy's Survey on Households Income and Wealth (SHIW) for the year 2010. This study resulted out that ignorance of endogeneity of total expenditure may result in harshly biased estimates. Particularly, the degree of error of measurement would be significantly overestimated. This investigation is uncommon because it deals both sources of endogeneity problem namely endogeneity of total expenditure and measurement error in the left- and right-hand side of the demand equation.

Moss et al. (2018) conducted a study to evaluate the distribution of food budget shares. The author used the Rwandan data and employed the quantile estimation technique. The study resulted out that the food expenditure shares statistically different between provinces involved in coffee cultivation and those involved in not growing coffee.

Mbegalo and Yu (2018) used Tanzanian household budget data from 2008/2009 and 2010/2011 to quantitatively examine the welfare implications of rising food prices in rural Tanzania. To attain this objective scrutinized the food Engel curve using a semi-parametric approach. The result of their study showed that the data fit a quadratic parametric food Engel curve.

Almas et al. (2019) observed the influence of large income changes on food budget shares among rural poor households in Kenya using data from a randomized controlled trial delivering unconditional cash transfers to poor households. Their study found expenditure and calorie elasticities of 0.78 and 0.60, respectively.

Nsabimana et al. (2020) investigated a study using the household expenditures and demographic survey of Rwanda to estimate quadratic food Engel curves. The study results show that the food Engel curve has a quadratic shape in rural setting, while in the urban setting empirical findings supported a linear specification of food Engel curve.

Wirba (2023) conducted a study to estimate the quadratic food Engel curves using the 2001, 2007 and 2014 Cameroon household consumption surveys. The author used an augmented heteroscedasticity-based identification strategy to estimate mean and quantile regressions to address potential mismeasurement of regressors. Exploratory non-parametric analyses proposed a quadratic shape for the food Engel curves. The findings of the study recommended that the reduction of taxes on food items would be more beneficial to households below the poverty line.

3. Theoretical Framework

3.1 Measurement Error

The measurement error is an important source of endogeneity that arises in the estimation of demand model when the expenditure data are not correctly measured. In this study, we follow Lewbel (1996) and Battistin and De Nadai (2013) approach to deal with measurement error issue.

To understand easily, as used in various papers, the identification results for the following quadratic demand model is derived:

$$\boldsymbol{S^*}_{\boldsymbol{food}} = \boldsymbol{C^*_{food}}/\boldsymbol{Y^*} = \boldsymbol{b_{i0}} + \boldsymbol{b_{i1}} \boldsymbol{logY^*} + \boldsymbol{b_{i2}}(\boldsymbol{logY^*})^2 + \mu_i \quad (1)$$

Where $\boldsymbol{C^*_{food}}$ and $\boldsymbol{C_{food}}$ represent the true and the observed (i.e. measured) consumption expenditure on food respectively by a household or an entire economy. While, $\boldsymbol{Y^*}$ denotes the true total consumption expenditure and $\boldsymbol{S^*}_{\boldsymbol{food}}$ denotes share of food expenditure in the total consumption expenditure. The measurement error is $\boldsymbol{C_{food}} - \boldsymbol{C^*_{food}}$ and denote by $\boldsymbol{v_i}$. And if we used $\boldsymbol{C_{food}}$ instead of $\boldsymbol{C^*_{food}}$ then the equation can be defined as

$$\boldsymbol{C_{food}} = \boldsymbol{C^*_{food}} + \boldsymbol{v_i}\boldsymbol{Y^*} \quad (2)$$ [1]

When over all goods are summing up, this specification implies classical measurement error in $\boldsymbol{logY^*}$.

$$\boldsymbol{logY} = \boldsymbol{logY^*} + \boldsymbol{logV} \quad (3)$$

As shown in Lewbel (1996) method of correcting measurement error on both side of the equation, measurement error $\boldsymbol{v_i}$ and $\boldsymbol{V}$ must be related as

$\boldsymbol{V} = 1 + \sum_{i=1}^{I} \boldsymbol{v_i}$ and this implies that

$$\boldsymbol{S_{food}} = \frac{\boldsymbol{S^*_{food}} + \boldsymbol{v_i}}{\boldsymbol{V}} \quad (4)$$

Equation (4) indicates that the measurement error non-linearly enters into the left hand side of the demand equation while estimating Engel curve (see Lewbel (1996) for detail).

3.2 Endogeneity of Total Expenditure

As commonly $\boldsymbol{Y^*}$ and $\boldsymbol{C^*}_{\boldsymbol{food}}$ may be chosen by the individuals simultaneously, it means that there is no clear economic justification to suppose exogeneity of the total expenditure in the right-hand side of the demand equation.Having a set of valid instruments ($\boldsymbol{Q}$) the parameters ($\boldsymbol{b_i}$) in equation (1) can be estimated using IV approach with two stage least square (2SLS) which is mostly employed in empirical applications (Blundell et al., 2007; Attanasio et al., 2012). Another method of standard control function was employed by Battistin and De Nadai (2013) to deal with the problem of endogeneity. In the first stage of later approach, it is necessary to relate the endogenous variable ($\boldsymbol{logY^*}$) with the exogenous variable (Q) and the error term $\boldsymbol{\eta^*}$ that derives endogeneity.

3.3 Identification of Quadratic Engel Curve

We considered the following standard assumptions.

[1] Throughout in this article, variables with a star indicate that these are measured without any error.

Assumption.1 Validity of the instrumental variable and conditions on measurement error

Let $Y, Y^{*}, C_{food}, Q, \mu_i, v_i$ be vectors of identically and independently distributed random variables such that

$$E[Y/Q] \neq 0;$$
$$E[\mu_i/Q] = 0;$$
$$E[v_i] = 0 \text{ and } v_i \perp (Y^{*}, Q, \mu_i),$$

Where (i) and (ii) of Assumption 1 are the standard conditions of the validity of the IV, whereas (iii) shows that measurement errors are independent of the total consumption expenditure. It should be noted that (iii) also implies that $E[V] = 1$.

By using the same result as derived by Lewbel (1996), we can write

$$E[C_{food}/Q] = \tilde{a}_{i1}E[Y/Q] + b_{i1}E[YlogY/Q] + \gamma_{i1}E[Y(logY)^2/Q] + E[Y^{*}\mu_i/Q] \quad (5)$$

When$E[\mu_i/Y^{*}] = 0$, the last part of right-hand side of Equation (5) vanishes. This implies that b_{i1} and $\gamma_{i1}$are identified via 2SLS regression of C_{food}on $YlogY$ and $Y(logY)^2$ without a constant using Q as instrument. When$E[\mu_i/Y^{*}] \neq 0$, the 2SLS approach yields incorrect inference aboutb_{i1} due to an omitted variable.

To solve this problem, we use control function approach as suggested by Battistin and De Nadai (2013). We relate the error-free variables in the first-stage as:

$$logY^{*} = g(Q) + \eta^{*} \quad (6)$$

Assumption 2.Control function restrictions

Suppose η^{*} ,Q and $\mu_i$be such that

$$E[\mu_i/Q, \eta^{*}] = E[\mu_i/Q] + E[\mu_i/\eta^{*}] = E[\mu_i/\eta^{*}] = \rho_i\eta^{*}$$

Because from the assumption 1 (ii) $E[\mu_i/Q] = 0$,

Where η is residuals obtained by regressing endogenous variable ($logY$) on the set of valid instrumental variable (Q). It means that η is the analogue of η^{*} when$logY$ is replaced to $logY^{*}$ in equation (6). It pursues from the structure of measurement error in Equation (3) that

$$\eta = \eta^{*} + logV - E[logV]. \quad (7)$$

The following theorem used in Lewbel (1996) and Battistin and De Nadai (2013) permits to separate endogeneity, arises due to correlation between μ_i and$logY^{*}$, from measurement error incoming via Equation (1).

Theorem

Let suppose equations (1) and (3) hold, equation (4) can be written as

$$S_{food} = \frac{1}{V} b_{i0} + b_{i1} logY^{*} + b_{i2}(logY^{*})^{2} + \mu_{i} + v_{i} \tag{8}$$

Under assumption 2 we can write

$$S_{food} = \frac{1}{V} b_{i0} + b_{i1} logY^{*} + b_{i2}(logY^{*})^{2} + \rho_{i}\eta^{*} + \xi_{i} + v_{i} \tag{9}$$

With$E[\xi_i / Y^{*}, \eta^{*}] = 0$.

Now multiply Y either side and take conditional expectation w.r.t Q

$$E[YS_{food}/Q] = E[(Y^{*}V)\frac{b_{i0}+b_{i1}logY^{*}+b_{i2}(logY^{*})^{2}+\rho_{i}\eta^{*}+\xi_{i}+v_{i}}{V}/Q]. \tag{10}$$

Solving Equation (10) we get the following estimating equation.

$$E\left[\frac{YS_{food}}{Q}\right] = b_{i0} = \alpha_{il} E\left[\frac{Y}{Q}\right] + \beta_{il} E\left[\frac{YlogY}{Q}\right] + \gamma_{il}\left[\frac{Y(logY)^{2}}{Q}\right] + \tilde{\rho}_{il} E\left[\frac{Y\eta}{Q}\right], \tag{11}$$

Where,

$$\alpha_{i1} = b_{i0} - b_{i1}E[VlogV] - b_{i2}\{E[V(logV)^{2}] - 2E[VlogV]^{2}\} - \rho_{i}cov(VlogV),$$

$$\beta_{i1} = b_{i1} - 2b_{i2}E[VlogV],$$

$\gamma_{i1} = b_{i2}$, $\tilde{\rho}_{i1} = \rho_{i}$

This implies that through a 2SLS regression of C_{food}on$Y$, $YlogY, Y(logY)^{2}$and$Y\hat{\eta}$, using instrumental variable Q would consistently estimate the coefficient of quadratic Engel curve (b_{i2}).

3.4 Description of Data and Variables

The analysis is based on the Household Income and Expenditure survey (HIES) data of Pakistan for the year 2010-11, conducted by Pakistan Bureau of Statistics (PBS). The solid reason behind the use of HIES 2020-11 data instead of latest one is that, the main objective of this study is to compare the four different estimation techniques, Ordinary least square, Instrumental Variable, Lewbel (1996) approach and control function approach estimating the quadratic food Engel curve in the presence of the problem of measurement error and the problem of endogeneity of total expenditure to check whether endogeneity of total expenditure is more serious issue or the measurement error. As we know that, over the time consumption basket of Pakistani population has changed significantly that is why it is needed to re-investigate the shape of Engel curve using latest data. To obtain this objective, the authors are working on another paper on estimation of food Engel curve using separate household Income and Expenditure surveys; HIES 2004-05, HIES 2010-11 and HIES 2018-19, employing quantile regression technique and the approach of Hausman et al. (1995) to re-investigate the suitable functional form of Engel curve and to compare the results of these two estimator techniques. The HIES 2010-11 covers a sample of 16341 households.

This survey provides significant information on income, consumption expenditure, savings, liabilities, employment, education, health and other economic and social indicators at both national and provincial level with rural urban breakdown. In this study we focus on the food consumption by taking the budget share of food (in total expenditure) as dependent variable. Independent variables include households log total consumption expenditure (contained all food and non-food items), squared of log total expenditure, log of households' size, and macro areas dummies (Punjab, Sindh, KPK, And Blochistan) for those who living in different provinces of Pakistan to control the household regional variation.

The instrumental variable used in the study is households' income from main occupations. Data on food items were given on fortnightly and monthly basis while data on non-food items were available on monthly and annual basis. The income and expenditures are measured in Pakistani currency (Pak rupee). The prices of almost all goods and services were assumed constant in the article. After doing some preliminary cleaning, the final sample size for estimation purpose remains 14233. The distribution of sample size by provinces is given in Appendix-A.

3.5 Description of Variables

Variables used in the study and their definitions are presented in the following table

Table No 1: Description of Variables

Variables	Definition of variables
S_{food}	The budget share of total expenditure allocated to food which is the ratio of annual total food expenditure to annual total expenditure on all goods and services by households.
Y	Total expenditure by the households on all goods and services including food, non-food, durable and non-durable items (annual)
$logY$	Log total expenditure
Q	Households' income from the main occupation (Instrumental variable).
$logQ$	Log of households' income from main occupations
hhs	Number of family members (individuals) living under a roof.
$loghhs$	Log of Household size
Punjab	regional dummy for Punjab (who live in Punjab) that is =1, Punjab, =0 otherwise
Sindh	regional dummy for Sindh, =2, Sindh, =0 otherwise

KPK	regional dummy for KPK, =3, KPK, =0 otherwise
Bloch	regional dummy for Baluchistan, =4, Bloch, =0 otherwise

3.6 Construction of Variables

3.6.1 Food Share

$$food\ budget\ \text{share} = \frac{Food\ Expenditure}{Total\ Expenditure}$$

3.6.2 Total Expenditure

The total expenditure is calculated summing all the four expenditure categories, food, non-food, durables and non-durables. The expenditure of these four categories is calculated multiplying its quantities with prices as

$$Total\ Expenditure = \sum_{i=0}^{I} p_i\, q_i$$

Where

$$p_i = price\ of\ \text{ith commodity}$$

$$q_i = quantity\ of\ \text{ith commodity}$$

3.6.3 Estimation Procedure

We firstly started the estimation of quadratic Engel Curve using ordinary least square (OLS) method, conventional standard IV method, Lewbel (1996) method, and control function (CF) approach, then compared the results of each of the three methods to investigate whether measurement error problem is more serious or the endogeneity of total expenditure using the quadratic shape of Engel curve as the data support this functional form for Pakistan.

The regression model of OLS and standard IV method in case of quadratic function can be written as follows:

$$S_{food} = b_0 + b_1 logY + b_2(logY^2) + \theta logHHS + \varphi' D + \mu, \qquad (12)$$

Where, $logHHS$ is the log of household size and D represents the vector of dummy variables, use to control the effect of regional variation. While estimating through instrumental variable approach we used household income from main occupation and its squared as instruments both in log form.

The empirical model proposed by Lewbel (1996) for the quadratic functional form is given by:

$$YS_{food} = b_0 Y + b_1 YlogY + b_1 Y(logY)^2 + \theta YlogHHS + \varphi' DY + \varepsilon. \qquad (13)$$

Equation (13) was estimated via two stage least square method where the endogenous variables Y, $(logY)$, $Y(logY^2)$, and DY wereinstrumented with Qs and their interaction term with Ds. The CF approach, proposed by Battistin and De Nadai (2013), can be written as:

$$YS_{food} = b_0Y + b_1YlogY + b_1Y(logY)^2 + \theta YlogHHS + \varphi' DY + \rho Y\hat{\eta} + \omega \quad (14)$$

In this approach, we run a non-parametric regression of endogenous variable $(logY)$ on set of instruments (Q,$logQ$,$QlogQ$, $Q(logQ)^2$) and their interaction with D then save the residual ($\hat{\eta}$) and finally plug into the main regression.

4. Results and Discussion

Tables 2 summaries the results on key variables based on HIES data 2010-11. The average share of expenditure allocated to food items in total expenditure is observed 57% in our sample. This statistic shows that households spend more than 50% of their income on food than non-food items. Average log of total expenditure and income are observed 11.95 and 11.61(respectively. Overall, the average household size in of the sample was 6.8 persons.

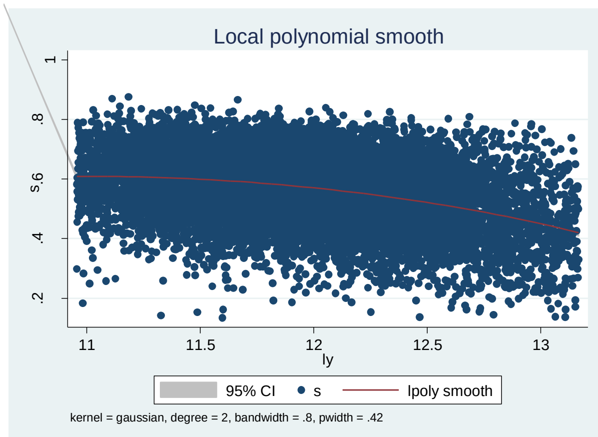
Table No 2: Summery Statistics of Study Variables

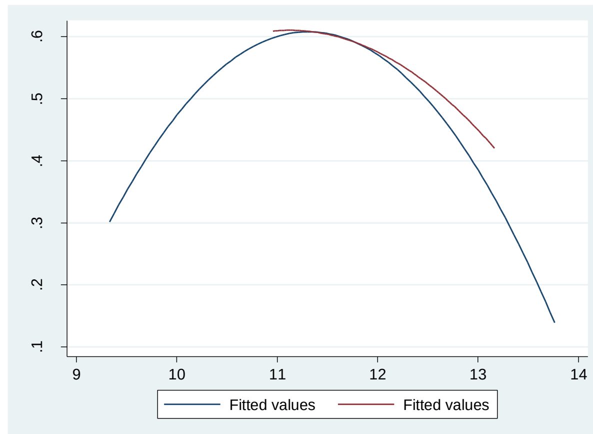
Variable	Mean	Std.Dev.	Min.	Max.
Food share	0.57	0.11	0.13	0.88
Log Expenditure	11.95	0.47	10.96	13.13
Log HH income	11.61	0.67	5.70	15.68
HH size	6.81	2.81	1	20

Most empirical studies that analysed the income and food demand relationship assume the linear specification. But later on, this assumption had been challenged and there was a major concern about the shape and non-linearity of this relationship, we used non-parametric and semi-parametric estimation methods to obtain the Engel curves. These are shown in Figure 1 and 2. The nonparametric and semi-parametric regression analyses presents attractive alternative to linear regression because they allow the data to determine the shape of the conditional mean relationship (Kedir & Girma (2018).

Figure No 1: Non-Parametric QFEC **Figure No 2: Semi-Parametric QFEC**

Vol 3 No 1 (2024): 176-196





Irrespective of the bandwidth used, non-parametric curve in Figure 1 clearly indicates a quadratic relationship between food share and log of total expenditures. We used different bandwidths to observe this relationship and found that the shape of relationship is unchanged. The result by using semi-parametric approach is presented in Figure 2 which is quite consistent with non-parametric result presented in Figure 1.Both Figures indicate that, initially, there is more proportionate change in budget share of food in response to increase in income up to some extent. This is because the economic conditions of these households as they are very poor, and food is luxury item for them. Later on the food share decreases as a result of further increases in households' income. Figure 1 and Figure 2 also demonstrates the fluctuating behaviour of the data that leads to the indication of Quadratic Food Engel Curve (QFEC). Our results are consistent with the study of Bhalotra and Attfield (1998).

4.1 Quadratic Engel Curve Estimates

Before estimation, non-linearity test, instrumental relevance test and test for endogeneity were performed. The results non-linearity test and instrumental relevance tests are presented in Appendix-B section i and ii respectively. While the result of the test for endogeneity are presented in Appendix-C for three approaches, IV, Lewbel (1996) and CF approaches, employed in this study. The results found from the four estimation methods are summarised in Table 3. After estimation of quadratic food Engel curve, the test of quadratic model is also performed in case of all three approaches (IV, Lewbel 1996 and CF) whose result is presented in Appendix-D.

Table No 3: Estimates of Quadratic Engel Curve Parameters, Hies 2010-11 Data

Variable	OLS	IV	Lewbel (1996)	CF
Log total expenditure	0.31	1.39	0.79	1.29
	(4.34)	(4.04)	(3.03)	(4.33)
Square log of total expenditure	-0.017	-0.063	-0.04	-0.06
	(-5.9)	(-4.44)	(-3.58)	(-4.78)
Log of household size	0.06	0.07	0.13	0.07
	(32.21)	(24.87)	(22.35)	(11.08)
Constant	-0.55	-7.05	-3.52	-6.52
	(-1.60)	(-3.4)	(-2.23)	(-3.62)

Note: the dependent variable is the share of expenditure on food and t-statistics are provided in parentheses.

The results of all four employed estimation methods, shown in Table 3 indicate that the quadratic model is significantly better than the linear one for representing the relationship between food share and total expenditure of households. Furthermore, these methods demonstrate highly significant and positive relationship between households' size and food demand.

The turning point provides us the point where the share of food expenditure stops to increase and this value of food share is considered the maximum. After this point the share of food share is decreased as total expenditure is increased. The turning points and expenditure elasticity with their respective confidence interval were calculated and are given in the following Table 3.

Table: 4 Estimates of Turning Points and Elasticities, HIES 2010-11 Data

Variable	OLS	IV	Lewbel (1996)	CF
Turning point	8.83	10.96	10.26	11
[95% confidence interval]	[7.78-9.89]	[10.48-11.44]	[9.23-11.29]	[10.54-11.49]
Expenditure elasticity	0.80	0.78	0.77	0.81
95% confidence interval]	[0.80-0.81]	[0.76-0.79]	[0.75-0.79]	[0.75-0.79]
Number of observations	14233	14233	14233	14233

Note: Turning points and confidence intervals are estimated using the delta method.

Table 4 shows that he OLS estimation method gives the turning point 8.83 and the lower and upper 95% confidence intervals of turning points are 7.78 and 9.89 respectively. The IV method gives the turning point of log total expenditure equals to 10.96 corresponding to total expenditure of Rs.57526 while the lower and upper 95% confidence interval of turning points are10.48 and 11.44 respectively. These results show that food is luxurious good only for those households which are below the 1.5th percentile. The Lewbel (1996) approach gives the turning point of log total expenditure equal to 10.26 which corresponds to total expenditure of Rs.28567. The lower and upper 95% confidence intervals of turning points are 9.23 and 11.29. The confidence interval associated with this estimator is wider than that of IV method which supports the well-known "bias versus variance trade-off" problem. This states that correcting bias makes the corrected estimator more variable than biased estimator (Carroll et al. 1995).The turning point log total expenditure, provided by CF approach is equal to 11.01 which corresponds to the total expenditure of Rs.59874 while the 95% confidence interval of turning points are10.54 and 11.49 respectively.

This turning point is greater than that provided by Lewbel (1996) while very closer to the one estimated via standard IV approach. The result of quadratic estimation via Control Function approach is consistent with the result of Battistin and De Nadai (2013). The result shows that if we correct measurement error problem and control the endogeneity of total expenditure than the estimate of the parameter of interest is very closer to IV estimate in magnitude. This result shows that endogeneity of total expenditure is the most serious problem than that of measurement error of total expenditure.

The expenditure elasticities obtained using OLS, IV, Lewbel (1996) and CF estimation methods are 0.80, 0.78, 0.77, and 0.81 respectively which indicates that food is a necessary good in case of Pakistan and is consistent with the result of Siddiqui (1982) and Burney and Khan (1991).

5. Conclusion and Recommendations

The main objective of the study is to compare the results of various estimation techniques employed in the study estimating the quadratic food Engel curve in the presence of measurement error and endogeneity of the total expenditure in Pakistan. Beside this, the study also investigates whether the measurement error or the endogeneity of the total expenditure is serious problem. A misspecification of Engel curve leads to a limitation of its effectiveness and generation of misleading results. Hence, a careful statistical analysis is necessary to make appropriate and suitable policies to enhance households' welfare. Therefore, we also tried to investigate the true shape (quadratic) of this relationship for Pakistan. In the presence of endogeneity problem, few studies demonstrated that Ordinary Least Squared (OLS) method provide biased estimate while the standard instrumental variable (IV) approach works only in the absence of measurement error. The approach employed by Lewbel (1996) assumes that the measurement error is the only source of endogeneity. Battistin and De Nadai (2013) proposed a new method of CF approach which corrects measurement error and controls the endogeneity of total expenditure. In case of Pakistan, as per best of our information there is no such study that incorporates all these estimation techniques for the same data set.

Through this study we tried to fill up this gap using a quadratic specification due to empirical reasoning and nature of the data. We estimated the quadratic food Engel curve using four different methods namely Ordinary Least Square method (OLS), instrumental variable method (IV), Lewbel (1996) approach and the control function approach employed by Battistin and De Nadai (2013), using HIES data (2010-11) to check whether the endogeneity of total expenditure is more serious problem or measurement error and concluded that the former is more serious issue than the later one.

The major findings of the study are as follows. First we find that food expenditure has largest share in a household budget which constitutes about 57% of total expenditure. The average household size is 6.8 persons. Therefore, we suggest that policy interventions should also include measures to boost the income of the very poor households so that they can manage their food and non-food consumptions. Secondly, both non-parametric and semi-parametric estimation results show a quadratic relationship between food share and log total expenditure.

Third, all the four estimation methods employed in this study also confirm that the quadratic term is statistically significant which prove that in case of Pakistan the quadratic functional form fits the data well. This result is consistent with those provided by Bhalotra and Attfield (1998) and Kedir & Girma (2018) and verifies that the quadratic logarithmic food Engel curve is a feature of developing countries. Although investigation of suitable functional form (linear or quadratic) is not our primary objective but to meet our primary objective as mentioned earlier we confirm that the HIES data fits quadratic shape of Engel curve.

It is observed that the point estimates of the parameter of interest (quadratic term) in absolute term in case of IV is larger than Lewbel (1996). The point estimate of control function is also larger than that of Lewbel (1996) and much closer in magnitude to those obtained via IV method. These results are like that presented by Battistin and De Nadai (2013). Further, turning points of IV is greater than that of Lewbel (1996) and closer to CF approaches. This suggests that ignoring the problem of endogeneity of total expenditure leads to overestimate the welfare of households. This also lead that endogeneity of total expenditure is more serious problem than that of measurement error.

The expenditure elasticity of food demand is computed for all four estimators. We found that all estimation methods provide the expenditure elasticity of food demand between zero and one which are closer to each other and gives the range of 0.77 to 0.81. These elasticity results are in line provided by Siddiqui (1982) and Burney and Khan (1991) that food is a necessary good for Pakistani households. In this study we observed the effect of household size on food demand. The results demonstrate that the household size has a positive and statistically significant impact on food demand (food budget share) in all estimation procedures.

5.1 Policy Recommendations

Turning to policy implication, we observed that as households' income increases the expenditure share allocated to food items declines and food is the necessary items for Pakistani households because expenditure elasticity was between zero and one. This indicates that if government impose more commodity tax or increase tax rate, it affects the low-income groups more as compare to their rich counterparts. Thus, the policy makers should make tax policies in the way that its impact on low-income group should be less for compulsory expenditures. The government should take affective steps to give subsidies on food items so that poor household may easily purchase foods to maintain their required calories level, and other basic needs of life. The government must also launch some other targeted base financial assistance program as BISP for low-income households so that poverty can be reduced.

6. References

Almås, I., Haushofer, J. & Shapiro, P. (2019). The income elasticity for nutrition: Evidence from unconditional cash transfers in Kenya. NBER Working Paper No.25711

Attanasio, O., Battistin, E. & Mesnard, A. (2012). Food and Cash Transfers: Evidence from Colombia. *Economic Journal*, 1(2), 92–124.

Banks, J., Blundell, R. & Lewbel, A. (1997). Quadratic Engel Curves and Consumer Demand. *Review of Economics and Statistics,* 7(9), 527–539.

Battistin, E. & De Nadai, M. (2013). Identification and Estimation of Engel Curves with Endogenous and Unobserved Expenditures. *Journal of Applied Econometrics,* 2(2), 12-34.

Betti, G. & Street, H. (2000). Quadratic Engel Curves and Household Equivalence Scales: The Case of Italy 1985 – 1994. *Department of Statistics, London School of Economics.*

Bhalotra, S. & Attfield, C. (1998). Intrahousehold Resource Allocation in Rural Pakistan: A Semiparametric Analysis. *Journal of Applied Econometrics,* 1(3), 463–480.

Bouis, E., H. (1992). Food Demand Elasticities by Income Group by Urban and Rural Populations for Pakistan. *The Pakistan Development Review,* 31(4), 997–1017.

Burney, A., N. & Khan, H., A. (1991). Household Consumption Patterns in Pakistan: An Urban-Rural Comparison using Micro Data. *The Pakistan Development Review,* 30(2), 145–171.

Burney, A., N. & Khan, H., A. (1992). Household size, its Composition and Consumption Patterns in Pakistan: An Empirical Analysis using Micro Data. *Indian Economic Review,* 27(1), 57–72.

Deaton, A. & Muellbauer.J. (1980). An Almost Ideal Demand System. *American Economic Review,* 7(3), 312–326.

Engel, Ernst (1857). Die Productions- und Consumtionsverhältnisse des Königreichs Sachsen. *Zeitschrift des statistischen Bureaus des Königlich Sächsischen Ministerium des Inneren.* 8(9), 28–29.

Gong, X., Soest, V., A., & Zhang, P. (2000). Sexual Bias and Household Consumption: A Semiparametric Analysis of Engel Curves in Rural China. *Institute for the Study of Labor (IZA) Discussion Paper No.212.*

Govt. of. Pakistan (GoP) (2010-11). Household Income and Expenditure Survey" Report by Pakistan Bureau of Statistics (PBS), Islamabad.

Haq, R. & Arshad, N. (2009). Inequality and Welfare by Food Expenditure Components. *The Pakistan Development Review*, 48(4), 755–768.

Hassan, A., S. (2012). Engel Curve and Equivalence Scales for Bangladesh. *ASARC Working Paper No. 15.*

Hausman, J. (2001). Mismeasured Variables in Econometrics Analysis: Problem from the Right and Problems from the Left. *The Journal of Economic Perspectives,* 5(4), 57–67.

Hausman, J., Newey, W. K, & Powell, J. L. (1995). Nonlinear Errors in Variables Estimation of Some Engel Curves. *Journal of Econometrics* 6(5), 205–233.

Kabubo-Mariara, J.,Wambugu, A. & Musau, S. (2011). Multidimensional Poverty in Kenya: Analysis of Maternal and Child Wellbeing. *Cahiers De Recherche PMMA, 12*.

Kedir, A. & Girma, S. (2003). Quadratic Engel Curves with Measurement Error. Evidencefrom a Budget Survey. Discussion Paper 03/17, Department of Economics, University of Leicester, UK.

Kedir, A. & Girma, S. (2018). Quadratic Engel Curves with Measurement Error. Evidence from a Budget Survey. *Oxford Bulletin of Economics and Statistics* 6(9), 123–138.

Kuha, J. & Temple, J. (2003). Covariate Measurement Error in Quadratic Regression.*International Statistical Review,*71 (1), 131–150

Laser, CEV. (1963). Forms of Engel Functions. *Econometrica,* 3(1), 694–703.

Lewbel, A. (1991). The Rank of Demand System: Theory and Nonparametric Estimation. *Econometrica*, 59, 711–730.

Lewbel, A. (1996). Demand Estimation with Expenditure Measurement Error on the Left and Right Hand Side. *Review of Economics and Statistics,* 78, 718–725.

Malik, S., J. & Sarwar, N. (1993). Some Tests for Differences in Consumption Patterns: The Impact of Remittances using Household Integrated Economic Survey Data of Pakistan. *The Pakistan Development Review,* 32(4), 699–711.

Mbegalo, T., & Yu, X. (2018). The impact of food prices on household welfare and poverty in rural Tanzania. Discussion papers, No. 216, Georg-August-Universität Göttingen, Courant Research Centre-Poverty, Equity and Growth (CRC-PEG), Göttingen.

Moss, C., Oehmke, F., Lyambabaje, A. & Schmitz, A. (2018). Distribution of budget shares for food: An application of quantile regression to food security. *Econometrics,* 4(2), 1-12.

Nisar, A., Chima, A. & Saleem, A. (2012). Food Consumption Analysis in Pakistan: Expenditure Elasticities Approach using HIES Data. *Interdisciplinary of Contemporary Research in Business* 4(4), 466–475.

Nsabimana, A., Swain, R., Surry, Y., & Ngabitsinze, J. (2020). Income and food Engel curves in Rwanda: A household microdata analysis. *Agricultural and Food Economics,* 1(1), 1-20.

Sakketa, T. G., & Gerber, N. (2024). Rural Shadow Wages and Youth Agricultural Labor Supply in Ethiopia: Evidence from Farm Panel Data. In S. W. Polachek & K. Tatsiramos (Eds.), Change at Home, in the Labor Market, and On the Job, 4(8), 61–105.

Sampaio Morais, G. A., Silva, F. F., de Freitas, C. O., & Braga, M. J. (2022). Irrigation, technical efficiency, and farm size: The case of brazil. *Sustainability (Switzerland),* 13(3), 1–21.

Selejio, O., Lokina, R. B., & Mduma, J. K. (2020). Smallholder Agricultural Production Efficiency of Adopters and Nonadopters of Land Conservation Technologies in Tanzania. *Journal of Environment and Development,* 27(3), 323–349.

Shahzad, A., Ullah, S., Dar, A. A., Sardar, M. F., Mehmood, T., Tufail, M. A., … Haris, M. (2021). Nexus on climate change: Agriculture and possible solution to cope future climate change stresses. *Environmental Science and Pollution Research,* 28(12), 14211–14232.

Shittu, A. M. (2019). Off-farm labour supply and production efficiency of farm household in rural Southwest Nigeria . *Agricultural and Food Economics* . Berlin/Heidelberg : Springer Berlin Heidelberg . https://doi.org/10.1186/s40100-014-0008-z

Siddique, R. (1982). An Analysis of Consumption Pattern in Pakistan. *The Pakistan Development Review* 4(2), 275–296.

Wirba, L, E., (2023), Identification of quadratic food Engel curves: Evidence from Cameroon. American Economics Research Consortium (AERC) final report (Paper No.523.

Wooldridge, J. M. (2002). Econometrics Analysis of Cross Section and Panel Data. The MIT Press, Cambridge, Massachusetts London, England.

Zereyesus, Y., Embaye, W., Tsiboe, F., and Amanor-Boadu, V. (2018). Implications of non-farm work to vulnerability to food poverty-recent evidence from Northern Ghana. *World Development*, 91(3), 113-124