Estimation of Quadratic Engel Curve in the presence of Measurement Error and Endogeneity of Total Expenditure using Pakistani Data

By

Ghulam Abbas

23/MPhil-ETS/PIDE/2012

Supervised by

Dr. Zahid Asghar

Department of Econometrics & Statistics

Pakistan Institute of Development Economics, Islamabad

2015

Estimation of Quadratic Engel Curve in the presence of Measurement Error and Endogeneity of Total Expenditure using Pakistani Data

By

Ghulam Abbas

23/MPhil-ETS/PIDE/2012

A Dissertation submitted to the Pakistan Institute of Development Economics, Islamabad, in partial fulfilment the requirements of the Degree of Master of Philosophy in Econometrics

Supervised by

Dr. Zahid Asghar

Department of Econometrics & Statistics

Pakistan Institute of Development Economics, Islamabad

In the Name of

I

The Most Beneficent
The Most Gracious

In the creation of the heavens and the earth, in the alteration of day and night, in the ships that sail and benefit the men, in the rain-------there are signs for those who think, understand and believe.

(AL-QURAN)

DEDICATION

Dedicated to My parents, siblings and my family

ACKNOWLEDGEMENT

All the deepest sense of gratitude to Allah Almighty the most compassionate and merciful and all my respect for His Holy Prophet (S.A.W.W). We are highly indebted to our teachers, who made us work and think hard to collaborate and improvise for the successful completion of the assigned task. Especially I would like to express my deepest gratitude to my Supervisor Dr. Zahid Asghar for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. His pain-staking effort in proof reading the drafts, are greatly appreciated. Indeed, without his guidance, I would not be able to put the topic together. I am also highly thankful to Dr. Hafsa Hina who not only help me but motivated me throughout my research and kept my moral high. Besides my Supervisor and teachers, I am also deeply thankful to my parents and brother Ghulam Nabi Noori, and cousins Sajjad Hussain, G.M. Rehmani, Ghulam Mustafa, and Hamidullah for their unconditional support, both financially and emotionally throughout my degree. I also place on rescored my deep sense of gratitude and indebtedness to my wife and son Arsalan Haider for sacrificing their time for my education. Sincere gratitude is also reserved for my siblings, Babi & nephews, and other family members their constant encouragement, endless prayers and continuous moral support. I would like to thank my best friends, Sheer Ali, Abdul Hameed, Ali Muhammad, and others. My greatest appreciation and friendship goes to my closest academic friends, Zahid Hussain Shaikh, Fawad Zaffar, Nasir Shabbir, Shujaat Ahmed and Muhammad Naveed who were always a great support in all my struggles and frustrations in my studies. Thanks them for helping me think rationally and even for hearing my problems.

I am also thankful to all other friends and fellows at PIDE who added fun to this journey.

Ghulam Abbas

Table of Contents

CHAPTER 1 1
INTRODUCTION 1
1.1 Objectives of the study 7
1.2 Significance of the Study 7
1.3 Planning of the Study 8
CHAPTER 2 9
REVIEW OF LITERATURE 9
2.1 Introduction 9
2.2 Review of Previous Studies 9
CHAPTER 3 17
THEORETCAL FRAMEWORK 17
3.1 Introduction 17
3.2 Theoretical Background of the study 17
3.2.1 Measurement error 17
3.2.2 Endogeneity of total Expenditure 18
3.2.3 Identification of Quadratic Engel curve 19
CHAPTER 4 22
DATA AND METHODOLOGY 22
4.1 Introduction 22
4.2 Source of Data and Characteristics 22
4.3 Description of variables 24
4.4 Construction of variables 25
4.4.1 Food share 25
4.4.2 Total Expenditure 25
4.4.3 Household size 25
4.4.4 Dummy Variables 26
4.5 Sample size 26
4.6 Methodology 28
4.7 Econometric models and estimation methods 28
CHAPTER 5 30
RESULTS AND DISCUSSIONS 30
5.1 Introduction 30
5.2 Descriptive statistics 30
5.3 Mean of food share at different expenditure quintiles 32
5.4 Preliminary Analysis 33
5.5 Quadratic Engel curve estimates 36
5.5.1 Non-linearity test 36

5.5.2 Standard Instrumental Variable Approach 36
5.5.3 Lewbel (1996) Approach 43
5.5.4 Control Function Approach 48
CHAPTER 6 55
MONTECARLO SIMULATION 55
6.1 Introduction 55
6.2 Data generating process 55
6.3 The MC exercise when extent of endogeneity is equal to zero 57
6.4 The MC exercise when extent of endogeneity is equal to 0.2 59
6.5 The MC exercise when extent of endogeneity is equal to 0.5 61
CHAPTER 7 64
CONCLUSIONAND RECOMMENDATIONS 64
7.1 Introduction 64
7.2 Summary and Conclusions 64
7.3 Policy Recommendations 67
7.4 Limitations of the study and Further Study 68
References 69

List of tables

Table	Title	Page
4.1	Description of variables	24
4.2	Distribution of sample size by provinces	26
5.1	Descriptive statistics of important variables	31
5.2	Mean of food share at different expenditure quintiles	32
5.3	First stage regression for testing for instrumental relevance	37
5.4	Instrumental variable estimation of quadratic food Engel curve	39
5.5	Lewbel (1996) approach of estimation of quadratic food Engel curve	45
5.6	First stage regression for estimation of quadratic food Engel curve via Control function approach	49
5.7	Control function approach of estimation of quadratic food Engel curve	51
6.1	Montecarlo simulation exercise when extent of endogeneity is equal to zero	58
6.2	Montecarlo simulation exercise when extent of endogeneity is equal to 0.2	60
6.3	Montecarlo simulation exercise when extent of endogeneity is equal to 0.5	62

List of figures

Figures	Title	Page
4.1	Household Demand for food and Non-food items	23
4.2	Distribution of sample size by Provinces	27
5.1	Percentile of Food Expenditure Share	32
5.2	Mean of food share at different expenditure quintiles	33
5.3	Non-parametric quadratic food Engel curve	34
5.4	Semiparametric quadratic food Engel curve	35

ABSTRACT

This study examines the suitable shape for food Engel curve for Pakistan using latest data of Household Income and Expenditure Survey (HIES) for the year 2010-11. We employed three estimation methods namely instrumental variable (IV), Lewbel (1996) and Control Function approach. Our results suggests that all the estimation methods employed in this study confirm that quadratic logarithmic Engel curve fits Pakistani data very well. Using this suitable quadratic shape we estimate the food Engel curve correcting measurement error and problem of endogeneity of total expenditure and found that the latter is more serious problem than the former one. The expenditure elasticity of food demand is also examined and obtain that each estimation method provide the elasticity between zero and one which indicate that food is a necessary good for Pakistani households. Furthermore in this study we investigate the relationship between food budget share and household size. We found that there is positive and statistically significant relationship between food budget share and household size.

CHAPTER 1

INTRODUCTION

The living standard and the welfare of household is measured via analysing consumption patterns in economics. Most of the literature reported that households spend major part of their income on food. The literature related to Pakistan observed that household with lower income spend a higher proportion of their income on food as compared to households with higher income (Haq and Arshad, 2009; Nisar *et al*., 2012). 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.[1] These figures suggest that the share of food is used as a reverse indicator of household welfare. For example, if household spend less on food items as compared to other consumption basket due to the result of income expansion it means that the household living standard has improved. Although welfare analysis can be checked by the help of absolute poverty-line or money metric approach as well but Kabubo-Mariara *et al.* (2010) indicate two drawbacks of this approach. Firstly this method presuppose market existence for all good and services and that prices reveal the utility weights all household with in a particular setting allocate to goods and services. But we cannot purchase some elements like public goods because the market is not existed and even if there is existed than it is not perfect. Therefore income as the only indicator of welfare is inadequate as it naturally fails to incorporate and reflect main dimensions of poverty associated to quality of life. Another flaw of income approach is that there is no surety that household with incomes at or above poverty line will truly assign their income so as to purchase the minimum basic

[1] As reported in HIES report 2011-12 of PBS.

needs bundles and therefor household may be rich regarding to their income but with some member dispossessed of some basic needs. A careful statistical analysis is necessary to make appropriate and suitable policies. A misspecification of Engel curve leads to limitation of its effectiveness and generation of misleading results. For example, if the true shape of Engel curve is quadratic but someone assumes a linear specification while estimating Engel curve then the welfare impacts on households with low income is underestimated.

Earnest Engel (1821-1896) observed that holding prices constant as households' income rises the proportion of income spend on food items declines. This theory is well-known as Engel's law in the literature of demand analysis later on. Although the earlier studies used the Working-Leser specification that specifies linear relationship between budget share of food and log total expenditure (Leser 1963; Deaton and Muellbauer 1980). But later on this specification fails to account the non-linear relationship in certain food budget share equation (Lewbel 1991; Hausman et al. 1995; Banks et al. 1997).

As mentioned above that literature related to Pakistan prove that, as income of low income households increase they spend almost all of their additional income on food (food is considered a necessary good). This is the one argument for quadratic Engel curve. Another argument is that as household income increases their preference shift to luxury goods.

Empirical evidences from the developing countries proved that the quadratic logarithmic food Engel curve is a feature of developing countries (Kedir and Girma 2007; Hassan 2012). Bhalotra and Attfield (1998) also favoured and suggested a quadratic Engel curve in case of Pakistan using data of rural households. The rural household might be best approximation of overall households in Pakistan because Pakistan is a developing country and the ratio of low income population are higher. But this study is fourteen years old and very old data set of HIES 1987-88 has been used. Now a lot of change with respect households' preferences and measurement technique have been made. Therefore we reinvestigate the functional form or shape of Engel curve in case of Pakistan using latest data set with the help of graphical analysis and specific tests of econometrics.

There are many other factors that affect the food demand such as own and cross prices, family size, education, taste, location, and weather etc. The household size is very important factor and many studies confirm the positive and significant impact of this variable on consumption patterns. In the case of Pakistan most of the studies prove that household size significantly and positively affect the demand for food, these studies include Siddiqui (1982); Burney and Khan (1991); and Malik and Sarwar (1993). In this study, we include log of household size and relax two assumptions relating to this, one is household data is correctly measured and other reported the linear relationship between household size and food demand. These assumptions are relaxed because we are interested in endogeneity of total expenditure and measurement error of expenditure data. This assumptions avoid us from more complexity and complications.

From the previous studies, it is observed that there are two modelling practice in selection of explanatory variables. Firstly in household surveys income variable is mismeasured, the households' expenditure is used as a proxy for the household income. Secondly, it is modelled such a manner that the budget share as the dependent variable while there is logarithm of total expenditure as the explanatory variable for example Banks et al. (1997); Bhallotra and Attfield (1998); Lewbel (1996). It should be noted that budget share equation is used in the demand analysis because of minimization of conditional heteroscedasticity of errors and the adding up condition of demand theory. In this study we follow these practices.

In the estimation of Engel curve, when household expenditure is used instead of household income it may suffer from endogeneity problem. So in this study, the main objective is to address the issue of endogeneity problem while estimating the Engel curve using food share equation. As mentioned above that a careful statistical and econometric analysis is required for making accurate and suitable economic policies with respect to indirect taxes, income distribution and subsidy etc. The inadequate policies come into existence due to misspecification of functional form and endogeneity problem. Our primary goal in this study is to get accurate estimates for parameters of interest of the demand equation taking into account the endogeneity problem and to check whether the endogeneity of total expenditure is more serious problem or the measurement error of the expenditure data.

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.

Measurement error was recognised in the demand literatures first time in early 1960s. Livitan (1961) 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 which is non-negligible. Afterward many studies have been conducted in the literature of demand analysis for treatment of measurement error. Some of the studies assumed that only total expenditure data are error ridden such as Hausman et al. (1995) and Hassan (2012). Later on some other studies control for error of measurement in both side of the demand equation such as Lewbel (1996), Kedir and Girma (2007), and Battistin and De Nadai (2013).

The empirical studies dealing measurement error prove that variation in the total expenditure data is because of measurement errors. Aasness et al. (1993) conducted a study modeling measurement error explicitly using the data of Norwegian budget survey. This study found that measurement error accounts about 27% variance of observed expenditure. Hausman et al. (1995) use the non-linear error in variable model to investigate the Engel curve parameters. Using the data of US consumer expenditure survey this study resulted out that about 42% variance in the measured total expenditure is because of measurement error. Lewbel (1996) conducted a study using fuel data of United Kingdom and considered measurement error as the only source of endogeneity. In the study the author treated error of measurement in both regressand and regressor of demand equation. This study employs a two-step approach. In the first step every elements of the original demand equation is multiplied by different powers of measured total expenditure and employed a generalized method of moment (GMM) technique to get consistent estimates of the transformed model. In the second step a functional relationship is developed between the parameter of transformed model and original demand model to recover the consistent estimates. This study found that measurement error correction changes the estimates of parameter of interest more than 15%.

In the empirical literature of quadratic Engel curve the treatment of measurement error on both side of the equation has an accurate estimate than that of only treatment in right hand side of estimating demand equation. The study of Lewbel (1996) has a higher turning point than that of Hausman et al. (1995). The study by Kedir and Girma (2007) implement both the approach of Hausman et al. (1995) and Lewbel (1996) using the same data of Ethiopian socioeconomic survey of urban households and resulted out that turning point for the Addis Ababa is lower in case of former approach than that of the later one.

Other source of endogeneity is simultaneity, sometime called endogeneity of total expenditure. Simultaneity arises due to jointly determined of two variables with each affecting the other. In case of food demand analysis it is common practice to choose both food expenditure and total expenditure simultaneously by individuals and these two variables are co-determined. It means that no strong justification is there to suppose exogeneity of total expenditure in the demand equation. Ignorance of total expenditure endogeneity leads to biased estimates while estimating Engel curve. In the absence of measurement error, this problem is resolved via standard instrumental variable approach but the fact is that a higher proportion of variation in total expenditure is due to measurement error so we cannot neglect it. Battistin and De Nadai (2013) conducted a study for treatment of both total expenditure endogeneity and measurement error in both side of the demand equation via a proposed control function approach. This study resulted out that the correction of measurement error only (ignoring simultaneity problem) leads to overestimation of the parameter of interest. This study concluded that the problem of endogeneity of total expenditure (simultaneity) is more serious issue than that of error of measurement.

Many studies shown 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 approach does not work when there is measurement error while Lewbel (1996) approach corrects 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 fail to provide inferential conclusion. This gap is filled up by the pioneer work of Battistin and De Nadai (2013). As in the literature of Engel curve in case of Pakistan as per best of our information there is no study that control for the unobserved expenditure endogeneity correcting measurement error. We followed the same setting of Battistin and De Nadai (2013) to fill-up this gap using most recent data and employing a quadratic Engel curve. We also evaluated the finite-sample properties (percentage of bias) of the control function approach employed in this study with a Monte Carlo simulation and compare the result with other alternative approaches such as ordinary least square, standard instrumental variable and the method employed by Lewbel (1996).

As mentioned above, in the previous literature some studies found that household size significantly and positively impact the food demand. There are few studies in case of Pakistan that favour this relationship. We also included the household size and check the whether this variable has a significant positive impact of this important variable on food demand using suitable specification of Engel curve. To control the regional variation, the dummy variable for all the four provinces of Pakistan namely Punjab, Sindh, Khyber Pakhtunkhwa and Baluchistan are also included in the estimating equation.

1.1 Objectives of the study

In this study, our objective is to identify an accurate and suitable specification for Engel curve and estimate this accurate Engel curve in the presence of endogeneity of total expenditure and measurement error of the expenditure data. The main objective of the study is to check whether the serious problem while estimating Engel curve is endogeneity of total expenditure or measurement error. A control function approach is used to resolve this problem and compared this approach with other estimators (used in this study). The impact of size of households on the food demand is also observed.

1.2 Significance of the Study

As per the best of our information, it is the first work in which both the two sources of endogeneity are considered and check whether endogeneity of total expenditure is more serious problem or measurement errors while estimation of Engle curve using latest Pakistani data set. Although Bhallotra and Attfield (1998) suggested a quadratic shape in case of rural Pakistan but there is no study yet that account the functional form of Engel curve in case of overall Pakistan.

1.3 Planning of the Study

Our study consists of seven chapters. First chapter includes introduction covering the, objective of the study and significance of the study. In second chapter previous relevant literatures are reviewed. In chapter three, theoretical frame work is discussed, Data and methodology are presented in chapter four. In chapter five, results and discussion are given. The Monte Carlo simulation has been performed in chapter six while conclusion and policy recommendation are given in the last chapter.

CHAPTER 2

REVIEW OF LITERATURE

2.1 Introduction

The Literature review provides not only base, but also both theoretical and empirical background of the research problem. This also gives efficient knowledge and valuable information about the problem of the research. So, the review of previous studies is very necessary and initial step in understanding, evaluating and solving the research problem. Therefore the review of previous studies related to shape of Engel curve and problem of endogeneity are presented in chronological order in the following.

2.2 Review of Previous Studies

Siddique (1982) conducted a study to confirm the Engel Law in the case of Pakistan. In the study the data of Households Income and Expenditure Survey (HIES) for the year 1971-72. The study confirms the Engel Law in the case of Pakistan. The food share expenditure has a negative relationship with household income. In other words the food share decreases as the household income increases. The study also resulted out that demand elasticities for necessities are lower while for luxury are higher.

Burney and Khan (1991) conducted a study to investigate the consumption patterns of both rural and urban households separately using Households Income and Expenditure Survey (HIES) data of 1984-85. This study confirmed Engel's law in case of Pakistan. This study also found that expenditure elasticity of food demand less than one in both urban and rural sector as well as households in various income group within each sector.

Burney and Khan (1992) studied the consumption patterns of households of Pakistan using functional forms of linear, double log and Working-Leser for six various income groups.

In this study, data of using Households Income and Expenditure Survey (HIES) data of 1984-85 was used. The study confirm the Engel's law in case of Pakistan and the income-specific estimates shown much richer consumption patterns compared to the conventional approach of using group data. This study also observed the effect of size and composition of on expenditure patterns. The expenditure elasticity of size of household size provide evidence of existence of economics of scale in consumption of some of the goods considered in the paper. The results, however, further shown that the degree of economies of scale are dissimilar across commodities and also varied across income groups. Furthermore the results presented in the study highlighted that the impact of household composition on expenditure patterns was not significant in Pakistan.

Bouis (1992) conducted a study to estimate both income and price elasticities for Pakistan. For this purpose the data on price and food consumption of Household Income and Expenditure Survey (HIES) was used. This study estimated the price and income elasticities by income groups by rural and urban households by income quartile for thirteen food groups. This study resulted out that wheat, vegetables, have lowest income elasticities while meat, dairy product, fruit have highest income elasticities. This study also concluded that income elasticities for food is higher for lower expenditure quartile than higher one and these elasticities for different income quartiles are lower for urban areas than that of rural areas.

Hausman *et al.* (1995) conducted a study to resolve the problem of errors in variables in non-linear regression model. Because in linear regression model solution to this problem is the instrumental variable technique. 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 finding suggests that Working-Leser specification must be widespread to higher order term in log of total expenditure. The finding also suggest that the error in variables (either in expenditure or reported income) must be accounted for. This study also observed the effect of family size and found that the effects of family size are small but statistically significant.

Lewbel (1996) proposed a simple GMM technique dealing with measurement error v_{ih} in expenditures for individual goods I imply measurement error V_h in total expenditures X_h. He considered the measurement error V_h as the only source of endogeneity. Consistent estimates of the parameters of original demand model are achieved appropriate transformation of the 2SLS, along with and efficient estimator of Generalized Method of Moment (GMM) because the ordinary 2SLS and GMM measurement error corrections fail in this context. The method is applied to U.K. fuel demand data from Family Expenditure Survey (FES) 1980-82 estimating standard linear logarithmic model. He also extend the linear logarithmic model into quadratic logarithmic model as well due to presence of strong empirical evidence that, for goods i. He resulted out that the effect of measurement error were 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 need 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 favouritism within families in the allocation of resources (income). In this study, the shape of Engel curve is also examined.

Using the data of Household Income and Expenditure Survey of Pakistan for the year 1987-88, argued that food Engel curve is quadratic in logarithmic. The study finding out that there is no discrimination of resource allocation among elderly over higher birth-order children.

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 share even after regulating for demographic difference and endogeneity but the quadratic logarithmic model seemed to be provide an acceptable parametric specification.

Ndanshau (1998) examined the impact income and household size on share of expenditure allocated to food and some other consumption items. This study demonstrates that there is positive and significant impact of income and household size on food and some other consumption items. This study is conducted on the basis of micro-survey data of 1989-90 of peasant households in Northern Tanzania.

Gong *et al. (*2000) examined the shape of Engle curve for china using semi-parametric approach and rejected the linear Engle curve for food. In this analysis, the data of Rural Household Income and Expenditure Survey (RHIES) of 1995. The data covers total sample of 7998 rural households 19 Chinese provinces.

Betti (2000) empirically analyzed the shape of Engle curve using the panel data of the Italian Household Budget Survey 1985-94. Empirical evidence indicates that the quadratic specification of Almost Ideal Demand System (AIDS) model fits the data well.

Hausman (2001) conducted a study for investigation of effect of error of measurement in both side of the estimating equation. This study found that the coefficient of linear model remains biased downward if both right- and left-hand side variables are measured with errors, compared with the coefficient derived from the model where only right-hand side is measured with error. Although the extent of the bias is not changed by mismeasurement of the left-hand side variable but the coefficient is measured with less accurately.

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 to flatten the curvature of the estimated functions.

Kedir and Girma (2003) conducted a study to investigate the shape of Engel curve using the Ethiopia Urban Households Survey (EUHS, 1994). This study confirm non-linear food Engel curves (quadratic) and this result was robust and statistically significant under different variety of specifications, non-parametric, endogeneity corrected semiparametric and outlier robust regression. Using measurement error corrected method of Lewbel (1996), this study found that at lower expenditure level the Engel curve is upward sloping. The threshold wellbeing level beyond which food Engel curve started downward sloping lies between the 35^{th} and 47^{th} percentiles of total expenditure distribution. For poverty reduction, this study suggested that about half of the urban population should be targeted while making policies. This is significantly dissimilar from the policy endorsement that emanates from the ordinary least squares estimator which infers that less than a quarter of population from the urban areas hurt from food insecurity.

Ahmed and Arshad (2007) investigated consumption patterns of households of Pakistan estimating Engel's equation for 22 commodities groups with specification of quadratic spline. This study resulted out that in case of rural households the expenditure elasticity of all the 22 groups were positive at all income level. Similar result found in case of urban except wheat which was inferior. The study also concluded that most of the Engel curves for food items were flatten than non-food items.

Kedir and Girma (2007) 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.

Caglayan and Astar (2012) conducted a study to investigate the consumption patterns of food and clothing using data of household consumption expenditure found from Turkish Statistical Institute. The study found that expenditure elasticities of both food and clothing were less than one which indicate that both commodities were necessities for Turkish households.

Hassan (2012) examined the shape of Engle curve using semi-parametric technique with base independence assumption and also control for endogeneity. In the study data of 12240 households has been drawn from the 2010 round of Bangladesh Household Income and Expenditure Survey (HIES) - a repeated cross-sectional household survey. The test for semi-parametric specification reveals that Engle curve for most of the categories have a quadratic shape. Quadratic model also estimated with equivalised household expenditure and check the significance of the quadratic term and observed that the quadratic term is highly significant 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. Comparing this model with other corresponding model it is confirmed that the control function model has better fit all categories. The result of the CF approach also support the quadratic shape of Engle curve for all categories under consideration except medical expenses. This study also confirmed the quadratic food Engle curve for developing countries.

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 take 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 suggest 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. The data set consist of food share in total expenditure, total expenditure, logged male wages as instrument, and regional dummies (North, centre and south) and covers total sample of 2311 . 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.

CHAPTER 3

THEORETICAL FRAMEWORK

3.1 Introduction

This chapter covered the theoretical background of the study which consist of functional relationship between budget share of food and total consumption expenditure. As discussed earlier that when we analysis these relationship with the help of real data, there rises the problem of endogeneity. So in this chapter we discussed the sources of endogeneity problem, mainly two sources those are measurement error and endogeneity of total expenditure, and the channels through which they work.

3.2 Theoretical Background of the study

In this segment, we set out the assumption about the measurement error model and the channel through which the measurement error and total expenditure endogeneity work and how these influence the different method of estimation employed in this study. For this reason same strategy already applied by Lewbel (1996) and Battistin and De Nadai (2013) is maintained throughout this analysis.

3.2.1 Measurement error

The measurement error is an important source of endogeneity that arises in the estimation of demand model when the expenditure data are mismeasured. In this study, same setting is replicated that is already considered in Lewbel (1996) and Battistin and De Nadai (2013).

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

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

Let C^*_{food} and C_{food} equal to the true and observed (i.e. measured) consumption expenditure on food by an economic agent that may be a household or entire economy respectively. The measurement error therefore is $C_{food} - C^*_{food}$ and may denote by v_i. And if we used C_{food} instead of C^*_{food} then the equation can define as

$$C_{food} = C^*_{food} + v_iY^* \quad (2)$$

Throughout this study, variables with a star indicate that these are measured without any error.

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

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

As shown in Lewbel (1996) measurement error v_i and V must be related by

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

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

So we conclude that the measurement error non-linearly enters into left hand side of the demand model while estimating Engel curve. See for detail Lewbel (1996).

3.2.2 Endogeneity of total Expenditure

As commonly Y^* and $C^*{}_{food}$ 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 model. In the availability of a set of valid instruments (Q) the parameters of interest (b_i) in the above equation (1) can estimate using instrumental variable approach as mostly employed in empirical application (Blundell *et al.* 2007; Attanasio *et al.* 2012) that is the method of two stage least square (2SLS) or by a method of standard control function as employed by Battistin and De Nadai (2013). In the first stage of later approach, it is necessary to relate the endogenous variable ($logY^*$) to the exogenous variable (Q) and error term η^* that derives endogeneity.

3.2.3 Identification of Quadratic Engel curve

We considered the following standard assumptions usually employed by previous studies. These assumptions offer the base for the identification results derived below. For this purpose we use the same setting considered by Lewbel (1996) and Battistin and De Nadai (1913).

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 (iid) random variables such that

(i) $E[Y/Q] \neq 0;$

(ii) $E[\mu_i/Q] = 0;$

(iii) $E[v_i] = 0\ and v_i \perp (Y^*, Q, \mu_i)$

Where (i) and (ii) of assumption 1 are the standard assumption of the validity of instrumental variable, whereas (iii) shows that measurement errors are independent of 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) there is

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

When$E[\mu_i /Y^*] = 0$, the last part of right hand side of the equation (5) removes. This implies that b_{i1} and γ_{i1} identified via two stage least square regression of C_{food} on $YlogY$ and $Y(logY)^2$ without a constant and using Q as instruments. When$E[\mu_i/Y^*] \neq 0$, the 2SLS approach yields incorrect inference for b_{i1} due to an omitted variable.

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

$$logY^* = g(Q) + \eta^* \qquad (6)$$

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) $E[\mu_i/Q] = 0$

Define η as residual, obtained 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 pursue from the structure of measurement error in equation (3) that,

$$\eta = \eta^* + logV - E[logV] \qquad (7)$$

The following theorem used both 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).

Theroem (Identification of Quadratic Engel curves)

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

$$S_{food} = \frac{b_{i0} + b_{i1} logY^* + b_{i2}(logY^*)^2 + \mu_i + v_i}{V} \qquad (8)$$

Under assumption 2 we can write

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

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

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

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

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

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

Where

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

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

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

This implies that through a two stage least square (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}).

CHAPTER 4

DATA AND METHODOLOGY

4.1 Introduction

This chapter consists of six sections. In first section, the source of data and their characteristics are discussed, while section two and three cover the description of variable and construction of variable respectively. Section four consists of sample size, section five includes methodology and in the last section econometric models and estimation methods are presented.

4.2 Source of Data and Characteristics

This quantitative analysis is based on the Household Income and Expenditure survey (HIES) of Pakistan for the year 2010-11. The survey covers total households of 16341. The 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 only 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 bloch) 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. The food and non-food (household total demand) item are shown in figure 4.1 in detail. Data on food items were in biweekly and monthly basis while data on non-food items were monthly and annual basis. The data of household income from main occupations were also on the basis of monthly basis. In the study, the biweekly and monthly data of all the expenditure items converted into annual multiplying 26 and 12 respectively. The income and expenditures are measured in Pakistani currency (Pak rupee). Dummies include the numerical value 0, 1, 2, 3, and 4 and described briefly in table 1and in section 4.4.4 as well. In this study the prices of all goods and services assumed constant. The source of the data is Pakistan Bureau of Statistics (PBS).

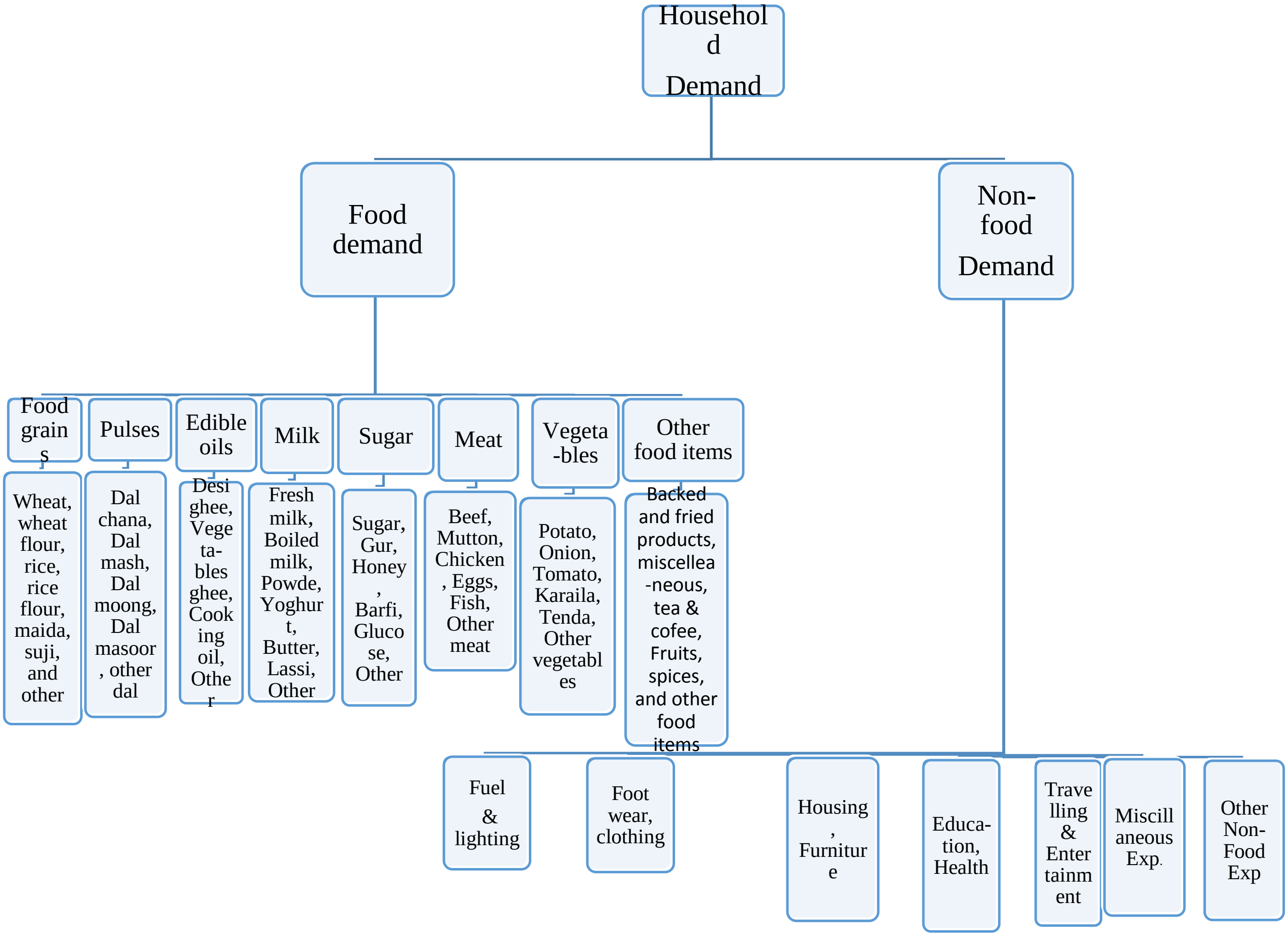

Figure 4.1 Household Demand for food and Non-food items

4.3 Description of variables

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

Table 4.1 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

4.4 Construction of variables

4.4.1 Food share

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

4.4.2 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

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

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

4.4.3 Household size

Household size is the number of family members who normally live and eat together and consider the living space occupied by them as their usual place of residence. Household members include head of household, his/her wife(s)/husband(s), unmarried sons and daughters and other direct dependents such as parents, unmarried brothers and sisters, and divorced sisters or daughters. In the study the household size is included as an important factor as employed by previous literatures such as Siddiqui (1982); Burney and Ashfaq (1991); Malik and Sarwar (1993) and Castaldo (2007).

4.4.4 Dummy Variables

In this study, dummy variables are included to control the regional dummies for the four provinces, Punjab, Sindh, Khyber Pakhtunkhwa and Baluchistan. The dummy variable include numerical values zero (0) to four (4). The value ‘’1’’ represents the households belonging to Punjab, “2” if the households from Sindh, “3” if households belong to Khyber Pakhtunkhwa and “4” for households belong Baluchistan. The numerical value “0” for all provinces represents that if the households is not belonging to the respective provinces.

4.5 Sample size

The survey included total sample size of 16341 households. We drop 1755 observations due to unavailability of data on households’ food expenditure and income from households’ main occupation (instrumental variable in this study). We further deleted 50 observations of household size that is greater than 20, because household size is consists of number of family member up to 38 and 99.67% observation consists of less than 21 members. The final sample size for estimation purpose in this study remains 14233 after deleting further 853 observation that were outliers in the consumption expenditure. The distribution of sample sizes by different provinces is given table 4.2.

Table 4.2 Distribution of sample size by Provinces

Provinces	**Sample size**	**Percentage**
Punjab	5757	40.45
Sindh	3843	27.00
KPK	2373	16.6
Baluchistan	2258	15.86
Total	**14233**	**100%**

Source: calculated by the author

The distribution of sample size by province are depicted in the following figure 4.2

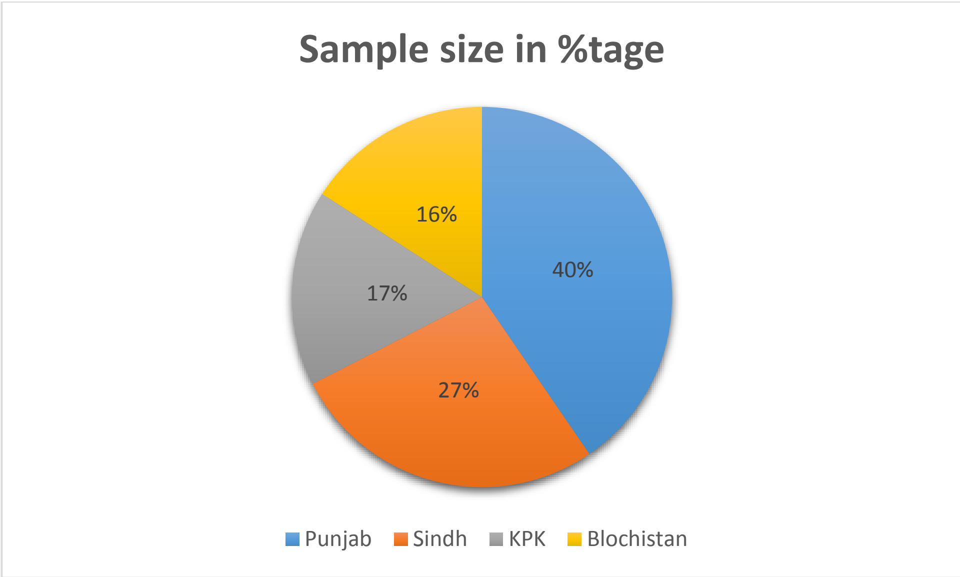

Figure 4.2 Distribution of sample size by Provinces

4.6 Methodology

In the estimation of quadratic Engel curve, the conventional Ordinary Least Square (OLS) method is not applicable due to its linearity assumption. It is discussed in first chapter when we use consumption expenditure data as the proxy of income variable there arises the problem of endogeneity. There are two main sources of endogeneity problem, measurement errors and simultaneity some time called endogeneity of total expenditures. The standard instrumental approach resolve the endogeneity problem in the absence of measurement error while the estimation approach proposed by Lewbel (1996) corrects the measurement error only and does not considers simultaneity as a source of endogeneity. To address both sources in the same time, Battistin and De Nadai (2013) proposed a control function approach. In case of Pakistan, as per our best information there is no study on this important issue. This study bridge the gap using a quadratic Engel curve which is more appropriate in case of Pakistan as discussed in chapter 5. In this study, we employed all the three methodology mentioned above and compare the final one with other two. These estimations have been done by using the recent statistical software STATA 12 version.

4.7 Econometric models and estimation methods

In this study we initially estimate the quadratic Engel Curve using three models namely conventional standard instrumental variable (IV) method, Lewbel (1996) Method, and control function (CF) approach. We test whether linear Engel curve fits the data very well or quadratic one in case of Pakistan in each of the above estimation methods. We then compare 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 suitable shape of Engel curve.

The conventional Ordinary Least Square (OLS) is not a good estimator for estimation of non-linear models due to its linearity assumption as discussed earlier so we employ the three methods mentioned above. The regression model of standard instrumental variable (IV) method in case of quadratic function can be written as follows

$$S_{food} = b_0 + b_1 logY + b_2(logY^2) + \theta logHHS + \varphi' D + \mu \qquad (13)$$

Where $logHHS$ is the log of household size and D is represents the vector of dummy variables, use to control the effect of regional variation. These dummies include Punjab, Sindh, KPK and Baluchistan that is for households who live in the four provinces. While estimation of instrumental variable approach household income from main occupation and it squared have been used as instruments by takin log.

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

$$YS_{food} = b_0 Y + b_1 YlogY + b_1 Y(logY)^2 + \theta YlogHHS + \varphi' DY + \varepsilon \qquad (14)$$

This equation is estimated via two stage least square method where the endogenous variables Y, $(logY)$, $Y(logY^2)$, and DY are instrumented with Qs and their interaction term with Ds. The Control Function approach, proposed by Battistin and De Nadai (2013), can be written as

$$YS_{food} = b_0 Y + b_1 YlogY + b_1 Y(logY)^2 + \theta YlogHHS + \varphi' DY + \rho Y\hat{\eta} + \omega \qquad (15)$$

In this approach we firstly run 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 then plug into the main regression.

CHAPTER 5

RESULTS AND DISCUSSIONS

5.1 Introduction

In this chapter, the results and interpretations of the three methods employed in this study are given. First of all descriptive statistics of the dependent and some important independent variables are presented. After that the preliminary graphical analysis is reported. Finally the results of all the methods are discussed briefly so that the reader can understand easily.

5.2 Descriptive statistics

The descriptive statistics involves the mean, standard deviation, minimum, maximum and some percentiles. Table 2 summarise the descriptive statistics of some important variables used in the study. Mean of food budget share is 0.57 which represents the average share of expenditure allocated to food items in total expenditure of the households. Average log of households' total expenditure and income from main occupation are 11.95 and 11.61 respectively. The mean value of household size is 6.81 which explains that there is 6.81 members in a household on average. Standard deviations, minimum, maximum, and percentiles are also shown in the following table 5.1.

Table 5.1 Descriptive statistics of important variables

	Food share	Log Expenditure	Log HH income (from main occupations)	HH size
Mean	0 .57	11.95	11.61	6.81
Std.Dev.	0.11	0.47	0.67	2.81
Min	0.13	10.96	5.70	1
Max.	0.88	13.13	15.68	20
Percentile of Food Expenditure Share				
1	0.26			
5	0.37			
10	0.42			
25	0.50			
50	0.57			
75	0.64			
90	0.70			
99	0.78			
Sample size	14233			

Source: calculated by Author

The expenditure share allocated to food is also depicted in the following figure 5.1

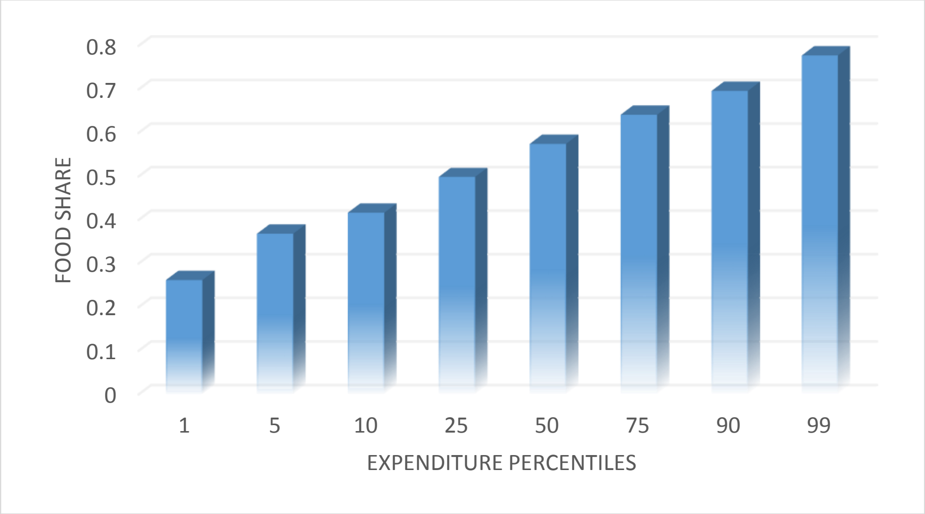

Figure 5.1 Percentile of Food Expenditure Share

5.3 Mean of food share at different expenditure quintiles

In the following table 5.2 we shown the mean of food share (dependent variable) at different expenditure quintiles. The results clearly shows that food expenditure share varies substantially over its own quintiles and as we move towards higher expenditure quintiles the food share declines. The food share at different expenditure quintiles is also drawn in figure 5.2.

Table 5.2 Mean of food share at different expenditure quintiles

	1st quintile	2nd quintile	3rd quintile	4th quintile	5th quintile
Mean	0.608	0.59	0.578	0.551	0.511

Source: calculated by Author

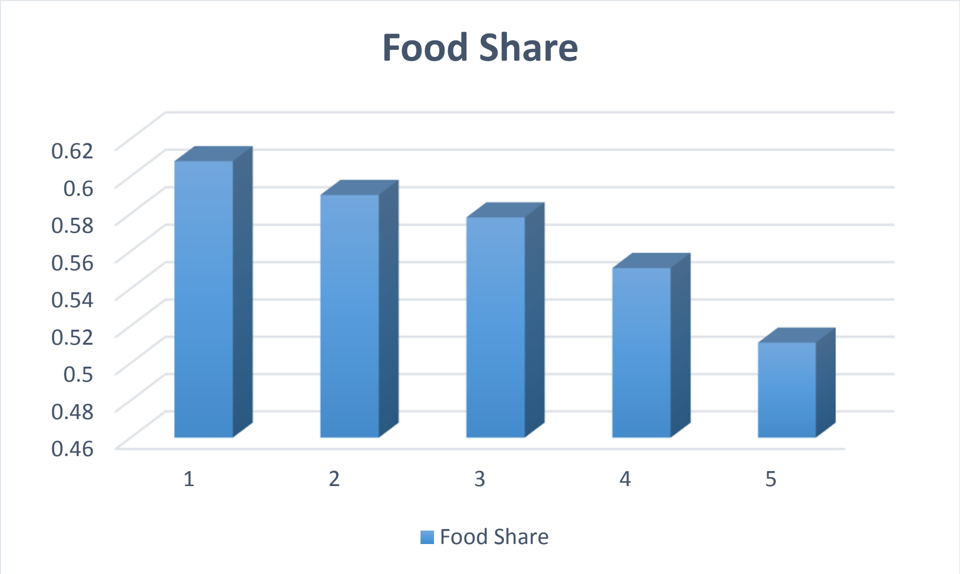

Figure 5.2 Mean of food share at different expenditure quintiles

5.4 Preliminary Analysis

As mentioned in chapter 1 that in case of developing countries the data confirms the quadratic Engel curve. In case of Pakistan, Bhalotra and Attfield (1998) clearly suggested quadratic Engel curve using rural household data but we also mentioned that this study is too old and very old data set of HIES 1987-88 had been used. So in this section, we revisit the shape of Engel curve. Therefore non- and semiprametric graphical analysis is presented in the following figures 5.3 and 5.4 respectively using recent data from all provinces of Pakistan including both rural and urban households.

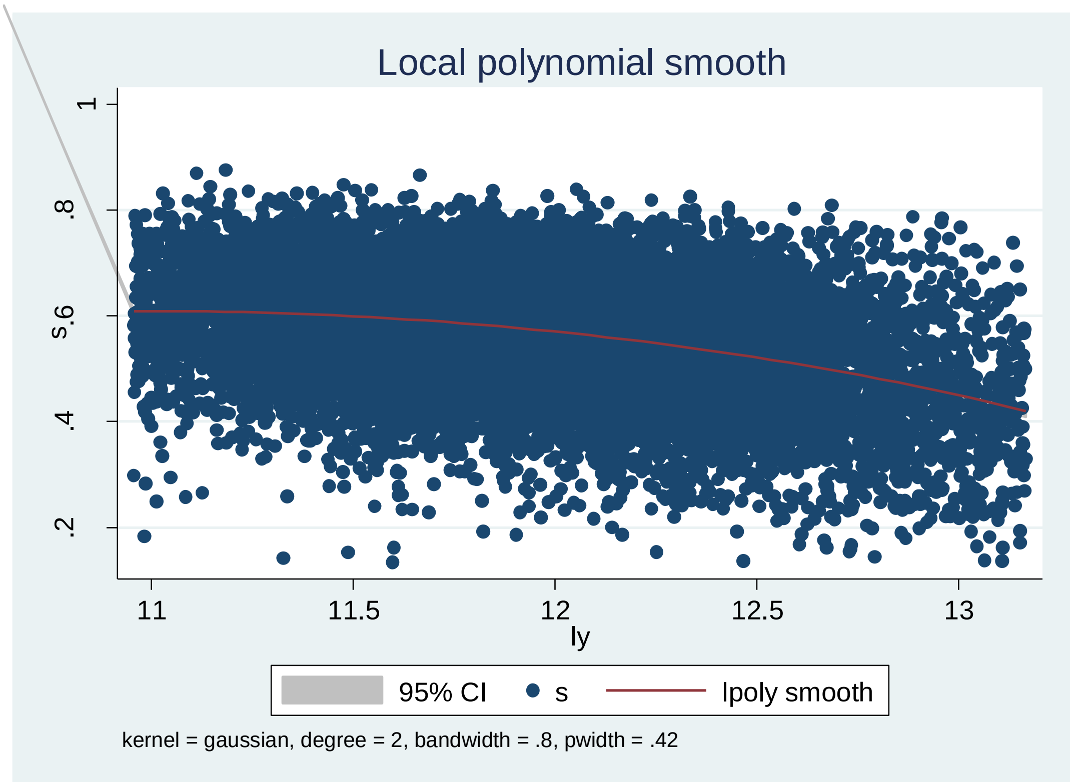

Figure 5.3 Non-parametric quadratic food Engel curve

The above non-parametric graphical result shown in figure 5.3 clearly indicate a quadratic relationship between food share and log of total expenditures. To obtain the above result firstly we run a non-parametric regression of food share on log of total expenditures. Than the smoothed food share value (bandwidth=0.8) along with 95% confidence interval are drawn against the log of total expenditure. We also experimented with different bandwidths to draw the graph and found the shape of relationship unchanged. The result via semi parametric approach is presented in figure 5.4 which is quite consistent with non-parmetric result presented in above figure 5.3.

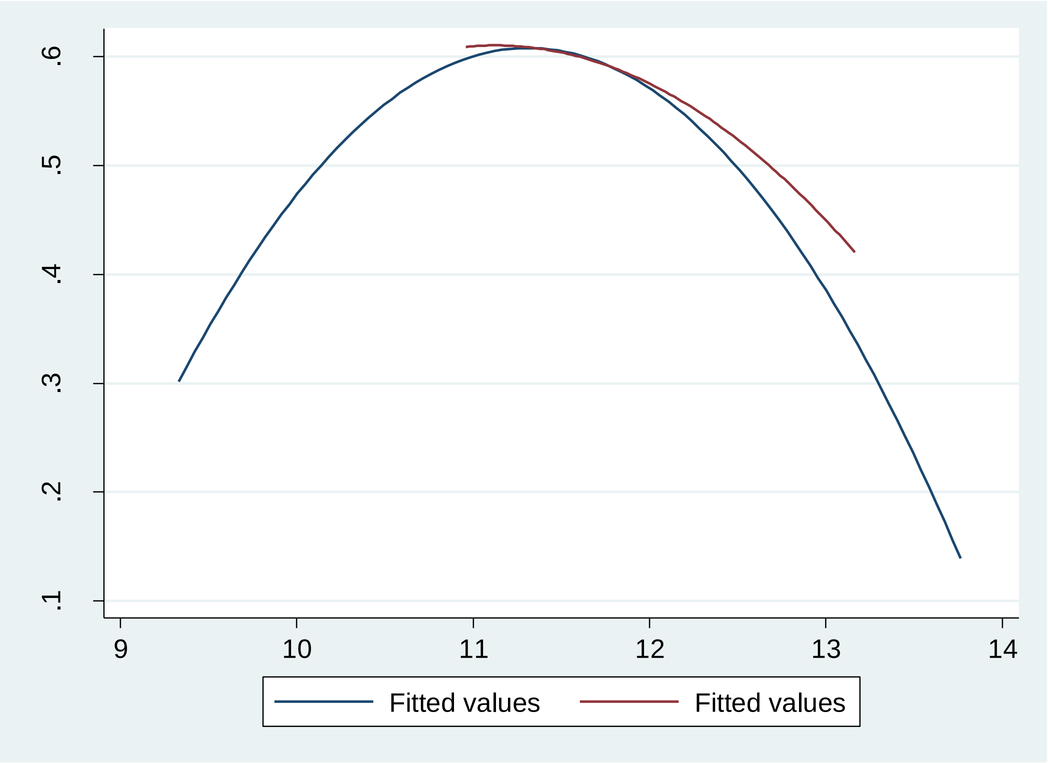

Figure 5.4 Semiparametric quadratic food Engel curve

Both of the above graphs show that initially there is more proportionate change in budget share of food in response to income increases. This is because the economic condition of these households are very poor and food is luxury good for them. As income of households increases the proportionate change in food budget share decreases and reach a point where the changes is zero than start to decreases. This nature of fluctuation in data drawn in the above graphs clearly indicate a quadratic Engel curve and validate the result of Bhalotra and Attfield (1998).

5.5 Quadratic Engel curve estimates

In this section we firsrtly perform a non-linearity test to check whethere linear Engel curve fits the data very well or the quadratic one. After confirmation of suitable quadratic shape we presented the estimation results of quadratic Engel curve using standard instrumental approach (IV), Lewbel (1996) method, and Control Function approach in table 5.3 to 5.7.

5.5.1 Non-linearity test

Null and Alternative Hypotheses

$H_0: A\ quadratic\ equation\ does\ not\ fit\ the\ data\ significantly\ better\ than\ a\ linear\ equation$

$H_A: A\ quadratic\ equation\ fit\ the\ data\ significantly\ better\ than\ a\ linear\ equation$

```
F(1, 14226) =       19.35
   Prob > F =        0.0000
```

The critical values of F-statistics is 3.84 at 5% significance level and is less than that of calculated value. The P-value is also less than 5% so we reject the null and conclude that a quadratic equation fits our data significantly better than linear one. So we will use the quadratic Engel curve in this study.

5.5.2 Standard Instrumental Variable Approach

5.5.2.1 Testing of instrumental relevance

Before estimation we test whether the instrumental variable is relevant with the endogenous variable. The procedure of testing is to regress the endogenous variable on instrumental variable as in equation (6) and test whether parameter of this regression model is significant or not. The estimating equation (6) has can be written as

$$logY = g(Q) + \eta^* = \pi_0 + \pi_1 \log(Q) + \eta^*$$

The results from this first stage regression is presented in the following table 5.3.

Table 5.3 First stage regression for testing instrumental relevance

Source	SS	df	MS	
Model	1219.8972	1	1219.8972	
Residual	1871.0693	14231	.131478413	
Total	3090.96649	14232	.217184267	

Number of obs	= 14233
F(1, 14231)	= 9278.31
Prob > F	= 0.0000
R-squared	= 0.3947
Adj R-squared	= 0.3946
Root MSE	= .3626

ly	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lq	.4394274	.004562	96.32	0.000	.4304854	.4483695
_cons	6.850359	.0530692	129.08	0.000	6.746337	6.954382

Null and Alternative Hypotheses

$H_0: \pi_1 = 0$ (instrumental variable is irrelvent with endogenous variable)

$H_A: \pi_1 \neq 0$ (instrumental variable is relvent with endogenous variable)

$$Test\ statistics: \hat{t}_\pi = \frac{\widehat{\pi_1} - \pi}{\mathrm{SE}(\widehat{\pi_1})} = \frac{0.4 - 0}{0.0046} = 96.32$$

$t -$ Critical value $= \pm 1.96$

As the absolute value of t-calculated is greater than that of absolute value of critical one, so we reject the null hypothesis and conclude that the instrumental variable is highly relevant with endogenous variable.

5.5.2.2 Testing for endogeneity

Null and Alternative Hypotheses

H_0: $Explanatory$ varibles are exogenous

H_A: $Explanatory\ variables\ are\ not\ exogenous$

```
Durbin (score) chi2(2)          =  183.718  (p = 0.0000)
Wu-Hausman F(2,14224)           =  93.0014  (p = 0.0000)
```

The critical values of Chi-squared is 5.99 and the critical value of F-statistics is 3.00 respectively at 5% significance level. In case of both Durbin and Wu-Hausman test the computed values are greater than that of critical values so we reject the null and conclude that explanatory variables are not exogenous or in other words explanatory variables are endogenous.

The estimation of standard instrumental variable for quadratic Engel curve is different with that of linear one. In case of instrumental variable approach, the quadratic Engel curve is estimated in two step procedure. In the first step, the endogenous variable that is total expenditure in our case, is regress on the instrumental variable and obtain the fitted values. In step two, we take the square of fitted variable and used in the regression as the instrumental variable for endogenous variable $(logY)^2$ [2]. The result of Instrumental approach is reported in the following table 5.4.

---

[2] This is the method of Wooldridge (2002) and the log squared fitted variable $(logYF)^2$ is identical to squared of instrumental variable $(logQ)^2$ and provides same result.

Table 5.4 IV estimation of quadratic food Engel curve

```
Instrumental variables (2SLS) regression          Number of obs =    14233
                                                  Wald chi2(6)  = 4655.67
                                                  Prob > chi2   =  0.0000
                                                  R-squared     =  0.2711
                                                  Root MSE      =  .09293
```

s	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ly	1.391133	.344458	4.04	0.000	.7160081	2.066259
ly2	-.0634667	.0142923	-4.44	0.000	-.0914791	-.0354544
lhhs	.0707874	.0028463	24.87	0.000	.0652087	.0763661
punjab	-.1009801	.002462	-41.02	0.000	-.1058055	-.0961547
sindh	-.024714	.001257	-19.66	0.000	-.0271777	-.0222504
kpk	-.0202438	.0009226	-21.94	0.000	-.0220519	-.0184356
_cons	-7.046557	2.069539	-3.40	0.001	-11.10278	-2.990335

Methods for testing the significance of quadratic term, calculating turning points and expenditure elasticity are presented as under with their respective results.

5.5.2.3 Testing for the significance of Quadratic Model

After getting the result of Instrumental variable method we have tested whether there is a significant overall relationship between, food share (S) and total expenditure (Y). The null and alternative hypothesis are as follows

$$H_0: \beta = \gamma = 0 \ (there\ is\ no\ overall\ reslationship\ between\ S\ and\ Y)$$

$$H_A: \beta\ and/or\ \gamma \neq 0 \ (there\ is\ an\ overall\ reslationship\ between\ S\ and\ Y)$$

$$F = \frac{R^2}{1-R^2}\left(\frac{n-k}{k-1}\right) = \frac{0.2711}{1\text{-}0.2711}\left(\frac{14233-7}{6}\right) = 881.85$$

$$F_{critical\ 5\%}\,(6, \infty) = 2.10$$

The computed value of F statistic is 881.85 and at the level of significance of 5% the critical value of the F distribution with 6 and ∞ degrees of freedom is 2.10. As the computed value is greater than that of critical one, so we reject the null hypothesis and conclude that there is significant overall relationship between food share and log of total expenditure.

5.5.2.4 Testing the Quadratic Model

In using a regression model to examine a relationship between two variables, we need to find out the most accurate and simplest model that express that relationship. Therefore we need to examine whether there is a significant difference between the quadratic model and the linear model. To test the significance of the contribution of the quadratic effect, we used the following null and alternative hypothesis.

$$H_0: \beta_2 = 0 \text{ (the quadratic term is insignificant)}$$

$$H_A: \beta_2 \neq 0 \text{ (the quadratic term is significant)}$$

$$Test\ statistics: \hat{t}_{\beta_2} = \frac{\widehat{\beta_2} - \beta_2}{\mathrm{SE}(\widehat{\beta_2})} = \frac{-0.063 - 0}{0.014} = -4.44$$

$$t - \text{Critical value} = \pm 1.96$$

As the absolute value of t-calculated is greater than that of absolute value of critical one, so we reject the null hypothesis and conclude that quadratic term is significant. It means that using instrumental variable method, the quadratic model is significantly better than the linear one for representing the relationship between food share and total expenditure of households.

5.5.2.5 Testing the Household size effect

In case of IV estimation method we test the third objective of this study which is to examine the effect of household size on food demand. To test the effect of household size on food demand, we used the following null and alternative hypothesis.

$H_0: \theta = 0$ (household size effect is insignificant)

$H_A: \theta \neq 0$ (household size effect is significant)

$$Test\ statistics: \hat{t}_\theta = \frac{\hat{\theta} - \theta}{\mathrm{SE}(\hat{\theta})} = \frac{0.070 - 0}{0.00285} = 24.87$$

$t - \text{Critical value} = \pm 1.96$

Again the value of t-calculated is greater than that of critical value, so we reject the null hypothesis and conclude that household effect is significant. As the sign of coefficient of $loghhs$ is positive again it means that high significant and positive relationship exist between households size food demand.

5.5.2.6 Calculation of Turning Point

The turning point tells us the point where function stops to increase or change its direction. The turning point provide us the value of log total expenditure where slope of the food Engel curve is zero. In other words, where the share of food expenditure stops to increase and this value of food share is maximum. After this point the share of food share is decreased as total expenditure is increased. In this study the turning points and their 95% confidence intervals are calculated via the Delta method. The formula is as under:

$$\tau = -\hat{\beta}/(2\hat{\gamma}) - 1/(2\hat{\gamma})\hat{\beta} + \frac{1}{2}\hat{\beta}/(\hat{\gamma})^2\,\hat{\gamma}$$

The turning point and associated 95% confidence interval obtain through instrumental variable approach are as under.

s	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_nl_1	10.95955	.2465246	44.46	0.000	10.47637	11.44273

The Instrumental variable method gives the turning point in log value of 10.96 which corresponds to total expenditure of Rs.57526 while the lower and upper 95% confidence interval of turning points are 10.48 and 11.44 which correspond to Rs.35596 and Rs.93221 respectively. This result shows that below only 1.5th percentile food is luxury good.

5.5.2.7 Calculation of Expenditure Elasticity

The elasticity is the measurement the responsiveness of change in one economic variable due to change in another variable. The Expenditure elasticity tells us the responsiveness of change in demand for food (food budget share) to a change in total expenditure. Food is consider as luxury good if elasticity is more than unity, necessary if elasticity lies between zero and one and inferior if it is less than zero. In this study the expenditure elasticity of food budget share (food demand) and their 95% confidence intervals are calculated. The formula for calculation of expenditure elasticity of food budget share in case of quadratic functional form is as under:

$$\text{e} = 1 + (\widehat{\beta_1} + 2\beta_2 logY/S)$$

The expenditure elasticity and their 95% confidence intervals obtained via instrumental variable (IV) approach are presented in the following.

s	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
esely	.7776029	.0069891	111.26	0.000	.7639045	.7913013

Above table shows that instrumental variable approach gives the expenditure elasticity equal to 0.777 which indicates that the food is necessary good and this results is consistent with the result provided by turning point. The 95% lower and upper confidence interval of expenditure elasticity are 0.763 and 0.79 respectively. It means that we are 95% confident that the expenditure elasticity lies between 0.763 and 0.79.

5.5.3 Lewbel (1996) Approach

As mentioned in chapter one, that this method assumes measurement errors only source of endogeneity and does not consider endogeneity of total expenditure as a source of endogeneity. This method also consider measurement errors in both left- and right hand side of the food budget share equation as pointed out by Lewbel (1996). In this study we employ this method and we follow the same setting as employed by Lewbel (1996) to correct the measurement errors. This method is a two stage least square method involves two steps, in first step we multiply both side of the food share equation by total expenditure. In second step, the endogenous variables (Y, $(logY)$, $Y(logY^2)$) are instrumented with the exogenous variable (Qs) and their interaction term with dummies. From the first stage regression of the two stage we are able to test for endogeneity and over identifying restrictions.

5.5.3.1 Test of endogeneity

Null and Alternative Hypotheses

H_0: $Explanatory$ varibles are exogenous

H_A: $Explanatory\ variables\ are\ not\ exogenous$

```
Tests of endogeneity
Ho: variables are exogenous

Durbin (score) chi2(6)           =  325.166  (p = 0.0000)
Wu-Hausman F(6,14220)            =  55.4108  (p = 0.0000)
```

The critical values of Chi-squared is 12.59 and the critical value of F-statistics is 2.10 respectively at 5% significance level. In case of both Durbin and Wu-Hausman test the computed values are greater than that of critical values so we reject the null and conclude that all the explanatory variables are endogenous.

5.5.3.2 Test of overidentifying restrictions

Null and Alternative Hypotheses

H_0: $All\ instrumental$ varibles are exogenous

H_A: $Some\ of\ the\ instrumental\ variables\ are\ not\ exogenous$

```
Tests of overidentifying restrictions:

Sargan (score) chi2(10)=  34.3964  (p = 0.0002)
Basmann chi2(10)       =  34.4386  (p = 0.0002)
```

The critical values of Chi-squared is 18.31 at 5% significance level. In case of both Sargan and Basmann test the computed values are greater than that of critical values so we reject the null and conclude that some of the instrumental variables are not exogenous. It means that instrumental variables are correlated with the model error and may conclude that our instruments are not valid. Although this test concludes that our instruments are week but we observed other possible instruments and we found the instruments used in this study are the best instruments compare to other possible instruments because computed statistics in case of both Sargan and Basmann test of overidetifying restrictions are lower using this instrument.

The estimation result applying procedure of Lewbel (1996) are presented in the following table 5.5.

Table 5.5 Lewbel (1996) approach of estimation of quadratic food Engel curve

f	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
y	-3.516978	1.578075	-2.23	0.026	-6.609948	-.4240082
yly	.7869512	.2597409	3.03	0.002	.2778683	1.296034
yly2	-.0383596	.0107007	-3.58	0.000	-.0593327	-.0173866
ypunjab	-.1014112	.0031081	-32.63	0.000	-.107503	-.0953194
ysindh	-.0276965	.001463	-18.93	0.000	-.0305639	-.0248291
ykpk	-.0234229	.0011151	-21.01	0.000	-.0256085	-.0212374
ylhhs	.1272133	.0056927	22.35	0.000	.1160559	.1383707

The result of quadratic estimation via Lewbel (1996) approach is presented in table 5.5. We used the same methods for testing the significant of quadratic term and calculating turning point and expenditure elasticity as above instrumental variable methods. The estimates of parameter of interest via this method is lower than that of instrumental variable approach which is consistent with the result obtained by Battistin and De Nadai (2013).

5.5.3.3 Testing for the significance of Quadratic Model

The table shows that the F statistic is not computed stata itself. This may be because F statistics is not important in this case but all the estimates of the parameter is significant. So it is concluded that there is significant overall relationship between food share and log of total expenditure.

5.5.3.4 Testing the Quadratic Model

To test the significance of the contribution of the quadratic effect, we again use the following null and alternative hypothesis as specified above in instrumental variable methods.

$$H_0: \beta_2 = 0 \text{ (the quadratic term is insignificant)}$$

$$H_A: \beta_2 \neq 0 \text{ (the quadratic term is significant)}$$

$$Test\ statistics: \hat{t}_{\beta_2} = \frac{\widehat{\beta_2} - \beta_2}{\mathrm{SE}(\widehat{\beta_2})} = \frac{-0.0384 - 0}{0.0107} = -3.58$$

$$t - \text{Critical value} = \pm 1.96$$

As again the absolute value of t-calculated is greater than that of absolute value of critical value, so we reject the null hypothesis and conclude that quadratic term is significant. It means that using Lewbel (1996) method, the quadratic model is significantly better than the linear one for representing the relationship between food share and total expenditure of households.

5.5.3.5 Testing the Household size effect

To test the effect of household size on food demand, we used the same null and alternative hypothesis as we used before.

$$H_0: \theta = 0 \text{ (household size effect is insignificant)}$$

$$H_A: \theta \neq 0 \text{ (household size effect is significant)}$$

$$Test\ statistics: \hat{t}_{\theta} = \frac{\hat{\theta} - \theta}{\mathrm{SE}(\hat{\theta})} = \frac{0.1272 - 0}{0.0057} = 22.35$$

$$t - \text{Critical value} = \pm 1.96$$

Again the value of t-calculated is greater than that of critical value, so we reject the null hypothesis and conclude that household effect is significant. As the sign of coefficient of $loghhs$ is positive it again confirm a significant and positive relationship between households size and food budget share.

5.5.3.6 Calculation of Turning point

The turning point and its 95% confidence interval obtained via this measurement error corrected method are presented in the following.

f	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_nl_1	10.25754	.5257712	19.51	0.000	9.227049	11.28803

The Lewbel (1996) approach gives the turning point of log total expenditure equal to 10.2575 which corresponds to total expenditure of Rs.28567. The lower and upper 95% confidence interval of turning points are 9.227 and 11.288 which correspond to total expenditure of Rs.10199 and Rs.79860. The confidence intervals associated with this estimator is wider than that of IV which support the well-known “bias versus variance trade-off” problem. Which states that correcting bias makes the corrected estimator more variable than biased estimator (Carroll et al. 1995).

5.5.3.7 Calculation of Expenditure Elasticity

The expenditure elasticity and its 95% confidence interval obtained via this measurement error corrected method are presented in the following.

f	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
esely	.7703093	.0092639	83.15	0.000	.7521524	.7884662

The Lewbel (1996) approach gives the value of expenditure elasticity equal to 0.77 as presented in above table and again indicate the food is necessary good. The expenditure elasticity obtained from Lewbel (1996) approach is similar to that of IV. The 95% lower and upper confidence interval of expenditure elasticity are 0.763 and 0.79 respectively. It means that we are 95% confident that the expenditure elasticity lies between 0.75 and 0.778.

5.5.4 Control Function Approach

We discussed in the first chapter that the problem of measurement error (both side of the equation) and endogeneity of total expenditure is handled in the pioneer work of Battistin and De Nadai (2013). Unfortunately there is no such study in case of developing countries including our country as well. This study aims to bridge this gap using the same setting employed by Battistin and De Nadai (2013) that is using of control function approach but a quadratic shape of Engel curve. The difference between the above mentioned study and this study is that in case of former a linear specification is used while in this study a quadratic specification is used due substantial empirical evidence in support of each of these studies. This methodology of control function is the extended work of Lewbel (1996) having two steps involved. In first step, a linear non-parametric regression of the endogenous independent variable ($logY$) is run on the set of exogenous variables or instrumental variables ($logQs$) and their interaction term with dummies as and save the residuals that derive endogeneity. In second step, we plugged this residual as an independent variable in the main food share equation after than we multiply both side of the equation by endogenous variable in level (Y) to correct the measurement errors as shows in equation (15). The result of control function approach for both first and second regression are reported in following table 5.6 and 5.7.

Table 5.6 Result of first stage regression for estimation of quadratic Engel curve via Control Function approach

R-squared = 0.3581
Root MSE = .40694

ly	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
q	.0001497	.0000772	1.94	0.053	-1.63e-06	.000301
lq	-.5822597	.0993436	-5.86	0.000	-.7769888	-.3875306
qlq	-.0000181	.000011	-1.64	0.100	-.0000396	3.46e-06
qlq2	5.38e-07	3.93e-07	1.37	0.171	-2.32e-07	1.31e-06
qpunjab	.0002727	.0000741	3.68	0.000	.0001275	.000418
qsindh	.0001027	.0000368	2.79	0.005	.0000306	.0001747
qkpk	.0000567	.0000266	2.14	0.033	4.69e-06	.0001088
lqpunjab	-.1416556	.0201694	-7.02	0.000	-.1811909	-.1021204
lqsindh	-.0423509	.0101503	-4.17	0.000	-.0622471	-.0224546
lqkpk	-.0083038	.0072471	-1.15	0.252	-.0225094	.0059017
qlqpunjab	-.0000371	.0000107	-3.45	0.001	-.0000581	-.000016
qlqsindh	-.0000144	5.32e-06	-2.71	0.007	-.0000249	-4.01e-06
qlqkpk	-8.55e-06	3.85e-06	-2.22	0.026	-.0000161	-1.01e-06
qlq2punjab	1.27e-06	3.92e-07	3.23	0.001	4.98e-07	2.03e-06
qlq2sindh	5.11e-07	1.93e-07	2.64	0.008	1.32e-07	8.90e-07
qlq2kpk	3.22e-07	1.40e-07	2.30	0.021	4.80e-08	5.96e-07
_cons	17.31946	.9064073	19.11	0.000	15.54276	19.09616

5.5.4.1 Test of endogeneity

Null and Alternative Hypotheses

H_0: $Explanatory$ varibles are exogenous

H_A: $Explanatory\ variables\ are\ not\ exogenous$

```
Tests of endogeneity
Ho: variables are exogenous

Durbin (score) chi2(7)          =  331.873  (p = 0.0000)
Wu-Hausman F(7,14218)           =  48.4912  (p = 0.0000)
```

The critical values of Chi-squared is 14.07 and the critical value of F-statistics is 2.01 respectively at 5% significance level. In case of both Durbin and Wu-Hausman test the computed values are greater than that of critical values so we reject the null and conclude that all the explanatory variables are endogenous.

5.5.4.2 Test of overidentifying restrictions

Null and Alternative Hypotheses

H_0: $All\ the\ instrumental$ varibles are exogenous

H_A: $Some\ of\ the\ instrumental\ variables\ are\ not\ exogenous$

```
Tests of overidentifying restrictions:

Sargan (score) chi2(10)=  69.2473  (p = 0.0000)
Basmann chi2(10)       =  69.4978  (p = 0.0000)
```

The critical values of Chi-squared is 18.31 at 5% significance level. In case of both Sargan and Basmann test the computed values are greater than that of critical values so we reject the null and conclude that some of the instrumental variables are not exogenous. Although some of instruments are not exogenous but we use this instruments because we observed other possible instruments but computed value of both Sargan and Basmann were higher in those cases.

Table 5.7 Control Function approach for estimation of quadratic Engel curve

f	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
y	-6.523918	1.804449	-3.62	0.000	-10.06057	-2.987262
yly	1.289513	.297634	4.33	0.000	.7061609	1.872865
yly2	-.0585441	.0122373	-4.78	0.000	-.0825287	-.0345595
yeta	.0526844	.006476	8.14	0.000	.0399917	.065377
ypunjab	-.1118383	.0030018	-37.26	0.000	-.1177216	-.1059549
ysindh	-.0318178	.0014532	-21.89	0.000	-.0346661	-.0289696
ykpk	-.0237457	.0011244	-21.12	0.000	-.0259494	-.0215419
ylhhs	.0672274	.0060661	11.08	0.000	.0553381	.0791166

The result of quadratic estimation via Control Function approach is reported in the above table 5.7. This result again is consistent with the result of Battistin and De Nadai (2013) which shows that if we correct measurement error and control for 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 most serious problem than that of measurement error. We used the same methods for testing the significance of overall model and quadratic term and calculate turning point as employed in different estimation methods above. The results of tests and calculation are presented briefly in the following.

5.5.4.3 Testing for the significance of Quadratic Model

The table shows that the F statistic is not computed stata itself. This may be because F statistics is not important in this case but all the parameter estimates are highly statistically significant. So it is concluded that there is significant overall relationship between food share and log of total expenditure.

5.5.4.4 Testing the Quadratic Model

To test the significance of the contribution of the quadratic effect, we again use the following null and alternative hypothesis as specified above.

$$H_0: \beta_2 = 0 \text{ (the quadratic term is insignificant)}$$

$$H_A: \beta_2 \neq 0 \text{ (the quadratic term is significant)}$$

$$Test\ statistics: \hat{t}_{\beta_2} = \frac{\widehat{\beta_2} - \beta_2}{\mathrm{SE}(\widehat{\beta_2})} = \frac{-0.059 - 0}{0.0122} = -4.78$$

$$t - \text{Critical value} = \pm 1.96$$

As again the absolute value of t-calculated is greater than that of absolute value of critical value, so we reject the null hypothesis and conclude that quadratic term is significant. It means that in case of control function approach also, the quadratic model is significantly better than the linear one for representing the relationship between food share and total expenditure of households.

5.5.4.5 Testing the Household size effect

To test the effect of household size on food demand, we used the same null and alternative hypothesis as specified above in the three methods.

$$H_0: \theta = 0 \text{ (household size effect is insignificant)}$$

$$H_A: \theta \neq 0 \text{ (household size effect is significant)}$$

$$Test\ statistics: \hat{t}_{\theta} = \frac{\hat{\theta} - \theta}{\mathrm{SE}(\hat{\theta})} = \frac{0.067 - 0}{0.0060} = 11.08$$

$$t - \text{Critical value} = \pm 1.96$$

Again the calculated t-value is greater than that of critical value, so we reject the null hypothesis and conclude that household effect is significant. As the sign of coefficient of $loghhs$ is again positive, so we again conclude that control function approach also confirm a significant and positive relationship between households size and food demand.

5.5.4.6 Calculation of Turning Point

The turning point estimated through this estimator of control function approach and its 95% confidence interval are presented in the following.

f	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
_nl_1	11.01317	.2413905	45.62	0.000	10.54005	11.48629

In case of Control function approach the dalta method provide turning point equal to 11.013 which correspond to total expenditure Rs.59874 while the 95% confidence interval of turning points are 10.54 and 11.49 which correspond to total expenditure of Rs.37798 and Rs.97734 respectively. This turning point is greater than one provide by Lewbel (1996) while very closer to the one estimated via standard IV approach. This result shows more accurate picture of the status of households in developing countries as well as Pakistan.

5.5.4.7 Calculation of Expenditure Elasticity

The expenditure elasticity obtained through this control function approach and its 95% confidence interval are presented in the following.

f	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
esely	.8059585	.010789	74.70	0.000	.7848124	.8271046

The control function approach gives the value of expenditure elasticity equal to 0.805 and this again indicates that the food is necessary good for Pakistani households. The 95% lower and upper confidence interval of expenditure elasticity are 0.784 and 0.826 respectively. It means that we are 95% confident that the expenditure elasticity lies between 0.784 and 0.826.

The expenditure elasticities obtained via all three estimation methods indicate that food is a necessary good in case of Pakistan and are consistent with the result of Siddiqui (1982) and Burney and Khan (1991)

CHAPTER 6

MONTECARLO SIMULATION

6.1 Introduction

In this chapter we presented Montecarlo simulation study to investigate the finite sample property (percentage of bias) of the estimators employed in this study. The aim of this simulation exercise is to compare the endogeneity (both measurement error and simultaneity) corrected method employed by Battistin and De Nadai (2013) with other estimators, standard instrumental variable (IV) and Lewbel (1996). This chapter consists of four sections. In section one the data generating process has been given, while in section two, three and four the results of simulation study for different scenarios are presented for all the three approach.

6.2 Data generating process

The data generating process for the exercise of the monte carlo simulation is as under.

$$S^{*}_{food} = \frac{C^{*}_{food}}{Y^{*}} = b_{i0} + b_{i1} logY^{*} + b_{i2}(logY^{*})^{2} + \rho_i \eta^{*} + \xi_i$$

$$logY = g(Q) + \eta^{*} = \pi_0 + \pi_1 \log Q + \eta^{*}$$

We consider that

$$Var\ (logY) = \pi_1{}^2 \text{var}\ (\log Q) + E\{\ Var\ (\eta/Q = z)\}$$

In this exercise we consider different scenarios for heteroscedasticity to explore the sensitivity of our results. We set

$$Var(\eta/Q = q) = \tau_0 + \tau_1 (logQ)^2$$

So that there is

$$Var\ (logY) = \tau_0 + ({\pi_1}^2 + \tau_1)\text{var}\ (\log Q)$$

Steps of generating samples

1. Values of $\log(q)$ are generated from a random normal distribution with mean zero and variance $\sigma^2_{\log q} = 1$.
2. Values of q are drawn randomly from uniform distribution times 5.
3. The value of η are drawn from a normal distribution with mean equal to zero and variance equal to $\tau_0 + \tau_1(logq)^2$ conditional on each random draw $\log(q)$, where we set $\tau_0 = 0.1$.
4. The values of Y are drawn from a random normal distribution with mean equal to zero and variance $\left(\sigma^2_x\right) = 0.1$.
5. The values of $logY$ are drawn randomly, setting $\pi_0 = 3.6$ and $\pi_1 = 0.3$
6. Values of $logY$ are drawn randomly setting $b_{i0} = -2,\ b_{i1} = 0.75\ and\ b_{i2} = -0.05$ and by drawing values of ξ_i from a normal variable centred at zero and variance equal to $\sigma^2_{\xi_i}$.
7. Measurement error is added in the model by drawing the values of $\log(V)$ from random normal variable with mean zero and variance $\sigma^2_{\log V}$.

The simulation exercise is done considering different scenarios for extent of heteroscedasticity regulated by τ_1, of endogeneity regulated by ρ and measurement error regulated by $\sigma^2_{\log V}$.

We set three scenarios for the parameter τ_1 corresponding none, moderate and large heteroscedasticity setting 0, 0.3, and 0.5 respectively.

We consider three scenarios corresponding to none, moderate and large measurement error setting a value for $\sigma^2_{\log V}$ which gives the noise to signal ratio $\mathrm{var}(\log V)\ /var(logY^*)$ equal to 0, 0.25 and 0.5

We set three scenarios for the parameter ρ corresponding none, moderate and large endogeneity setting ρ eual to 0, 0.2, and 0.5 respectively.

The Montecarlo simulation exercise has been performed for sample size of 1000 and 5000 for each of the estimation methods. The results obtained from thus simulation study are presented in the following tables considering different scenarios.

6.3 The MC exercise when extent of endogeneity is equal to zero

The simulation result for Montecarlo simulation when extent of endogeneity of total expenditure equal to zero is presented in the following table 6.1.

Table 6.1 The MC exercise when extent of endogeneity is equal to zero

Percentage bias for $\widehat{\beta}_2$ defined as $(\widehat{\beta}_2 - b_2)/\lvert b_2\rvert$								
Extent of Endogeneity (ρ_i)=0								
			Extent of Heteroscedasticity (τ_1)					
Sample size	**N/S Ratio**	**Estimators**	**0**		**0.3**		**0.5**	
			PB (I)	**S.E (II)**	**PB (III)**	**S.E (IV)**	**PB (V)**	**S.E (VI)**
N=1000	**0**	**IV**	.000871	.012686	.000154	.003271	.000592	.015546
		Lewbel	.000783	.015863	.0001768	.004249	.002197	-.01866
		CF	. .000779	.0158427	.000174	.0042358	.0000489	.0021949
	0.25	**IV**	.000362	.012213	.000094	.003148	.000045	.00406
		Lewbel	.000044	.015533	-.00003	.004137	-.00001	.00246
		CF	.000085	.015491	-.000027	.004134	-.00009	.002506
	0.5	**IV**	.000362	.012213	.000094	.003148	.000045	.00406
		Lewbel	.000044	.015533	-.00003	.004137	-.0001	.00246
		CF	.000085	.015491	-.000027	.004134	-.00009	.002506
N=5000	**0**	**IV**	.0001358	.0120568	.0000402	.0032039	-.000101	.01999
		Lewbel	.0002604	.0152578	.0000755	.004067	.0000126	.0024998
		CF	.0002569	.0152417	.0000741	.0040641	.0000127	.0024808
	0.25	**IV**	-.000017	.0120937	.0000007	.0031674	.0000146	.0061039
		Lewbel	-.000004	.015176	.0000148	.004111	-.000053	.004969
		CF	-.000004	.015154	.0000124	.004104	-.000022	.003232
	0.5	**IV**	-.000017	.0120937	.0000007	.0031674	.0000146	.0061039
		Lewbel	-.000004	.015176	.0000148	.004111	-.000053	.004969
		CF	-.000004	.015154	.0000124	.004104	-.000022	.003232

Note: The simulation results presented in the above table are obtained from the data generating process as discussed in section 6.2 of this chapter. The performance of the estimators employed in this study for the shape parameter of quadratic Engel curve is evaluated, the true value of the parameter being -0.05. The label IV, Lewbel, CF refer to simulation results obtained using instrumental variable approach, method of Lewbel (1996) and control function approach respectively. PB and S.E refers to percentage of bias and its standard error respectively. Different combination of noise to signal ratio and extent of heteroscedasticity are considered by row and column.

The presentation of the results of Montecarlo simulation exercise is organized according to the extent of endogeneity of total expenditure. We start our discussion from the results presented in table 1 where we considered that total expenditure is exogenous as considered in Lewbel (1996). The setting of simulation exercise is almost same as considered in Battistin and De Nadai (2013). When we consider the case of no endogeneity problem. In table 6.1, the results shows that in the absence of measurement error the CF approach outperforms instrumental variable and Lewbel (1996) at every extent of heteroscedasticity. When we move downward error through column I with extent of measurement error as we consider mild (0.25) and large (0.5) in each panel, the Lewbel (1996) outperforms instrumental variable approach, while control function approach provide the similar result as provided by Lewbel (1996). As we continue to move downward to large sample size the result shows that there is need for large sample size to implement the measurement error correction. A point is to be noted that in case of both mild and large extent of measurement error the result is same but over the sample size these are different, which indicates that the measurement error problem is not much serious. Overall the method of control function approach is very best than other two estimators. One thing is common in all three methods that as we consider large sample size (5000) the percentage of bias reduces.

6.4 The MC exercise when extent of endogeneity is equal to 0.2

The simulation result for Montecarlo simulation when extent of endogeneity of total expenditure equal to zero is presented in the following table 6.2.

Table 6.2 The MC exercise when extent of endogeneity is equal to 0.2

Percentage bias for $\widehat{\beta}_2$ defined as $(\widehat{\beta}_2 - b_2)/\lvert b_2 \rvert$								
Extent of Endogeneity (ρ_i)=0.3								
			Extent of Heteroscedasticity (τ_1)					
Sample size	N/S Ratio	Estimators	0		0.3		0.5	
			PB (I)	S.E (II)	PB (III)	S.E (IV)	PB (V)	S.E (VI)
N=1000	0	IV	-.002538	.241019	-.3143	1.155488	.0170575	8.177095
		Lewbel	.0014644	.298178	-.1882	1.07338	-.097885	.9518097
		CF	.001519	.297837	-.1886	1.073487	-.099674	.9698364
	0.25	IV	-.002509	.238275	-.2823	1.08436	-.127096	3.23053
		Lewbel	-.000619	.307234	-.1837	1.15256	-.111715	1.32380
		CF	-.000114	.307130	-.1838	1.15214	-.110219	1.25934
	0.5	IV	-.002509	.238275	-.2823	1.08436	-.127096	3.23053
		Lewbel	-.000619	.307234	-.1837	1.15256	-.111715	1.32380
		CF	-.000114	.307130	-.1838	1.15214	-.110219	1.25934
N=5000	0	IV	-.006581	.238323	-.2879	1.30399	.554178	47.0675
		Lewbel	.000405	.298865	-.1816	1.08286	-.086305	.918534
		CF	.0001632	.298508	-.1823	1.082535	-.087408	.913807
	0.25	IV	-.007684	.239504	-.3016	1.08682	-.182582	4.67315
		Lewbel	-.003885	.305656	-.1937	1.11755	-.08838	1.45744
		CF	-.003949	.305352	-.1941	1.11653	-.087617	1.50613
	0.5	IV	-.007684	.239504	-.3016	1.08682	-.182582	4.67315
		Lewbel	-.003885	.305656	-.1937	1.11755	-.08838	1.45744
		CF	-.003949	.305352	-.19411	1.11653	-.087617	1.50613

Note: The simulation results presented in the above table are obtained from the data generating process as discussed in section 6.2 of this chapter. The performance of the estimators employed in this study for the shape parameter of quadratic Engel curve is evaluated, the true value of the parameter being -0.05. The label IV, Lewbel, CF refer to simulation results obtained using instrumental variable approach, method of Lewbel (1996) and control function approach respectively. PB and S.E refers to percentage of bias and its standard error respectively. Different combination of noise to signal ratio and extent of heteroscedasticity are considered by row and column.

When we consider the case of mild endogeneity setting equal to 0.2 in the above table 6.2, the results shows that in case of both presence and absence of measurement error the CF approach and procedure of Lewbel (1996) outperform instrumental variable regardless of extent of heteroscedasticity. When we move downward error through column I with extent of measurement error as we consider mild (0.25) and large (0.5) in each panel, the Lewbel (1996) outperforms instrumental variable approach, while control function approach provide the similar result as provided by Lewbel (1996). One thing is common in all three methods that as we move to small sample size to large one the percentage of bias reduces.). A point is to be noted that in case of both mild and large extent of measurement error the result is same but over the sample size these are different indicates that the measurement error problem is not much serious. Overall the method of control function approach is very best than other two estimators.

6.5 The MC exercise when extent of endogeneity is equal to 0.5

The simulation result for Montecarlo simulation when extent of endogeneity of total expenditure equal to zero is presented in the following table 6.3.

Table 6.3 The MC exercise when extent of endogeneity is equal to 0.5

Percentage bias for $\widehat{\beta}_2$ defined as $(\widehat{\beta}_2 - b_2)/\lvert b_2\rvert$								
Extent of Endogeneity (ρ_i)=0.5								
			Extent of Heteroscedasticity (τ_1)					
Sample size	N/S Ratio	Estimators	0		0.3		0.5	
			PB (I)	S.E (II)	PB (III)	S.E (IV)	PB (V)	S.E (VI)
N=1000	0	IV	-.007651	.602163	-.78602	2.88891	.041755	20.4203
		Lewbel	-.001308	.75362	-.47097	2.70752	-.206889	2.51693
		CF	-.001425	.752733	-.47196	2.709647	-.213693	2.364225
	0.25	IV	-.006816	.595057	-.70584	2.71146	-.31781	8.07218
		Lewbel	-.007208	.777218	-.47376	3.04587	-.17150	2.1726
		CF	-.006063	.777249	-.47235	3.02475	-.16989	2.2037
	0.5	IV	-.006816	.595057	-.70584	2.71146	-.31781	8.07218
		Lewbel	-.007208	.777218	-.47376	3.04587	-.17150	2.1726
		CF	-.006063	.777249	- .4724	3.02475	-.16989	2.2037
N=5000	0	IV	-.016655	.595147	-.71992	3.25982	1.385595	117.695
		Lewbel	.0010325	.755661	-.44722	2.75353	-.208587	2.502112
		CF	.0006492	.754588	-.44847	2.75124	-.222257	2.310071
	0.25	IV	-.019184	.597745	-.75410	2.7170	-.456476	11.6849
		Lewbel	-.011618	.769804	-.48327	2.86066	-.221070	2.80242
		CF	-.011702	.768912	-.48394	2.85344	-.260034	2.434285
	0.5	IV	-.019184	.597745	-.75410	2.7170	-.456476	11.6849
		Lewbel	-.011618	.769804	-.48327	2.86066	-.221070	2.80242
		CF	-.011702	.768912	-.48394	2.85344	-.260034	2.434285

Note: The simulation results presented in the above table are obtained from the data generating process as discussed in section 6.2 of this chapter. The performance of the estimators employed in this study for the shape parameter of quadratic Engel curve is evaluated, the true value of the parameter being -0.05. The label IV, Lewbel, CF refer to simulation results obtained using instrumental variable approach, method of Lewbel (1996) and control function approach respectively. PB and S.E refers to percentage of bias and its standard error respectively. Different combination of noise to signal ratio and extent of heteroscedasticity are considered by row and column.

In above table 6.3 we consider the case of large extent of endogeneity (0.5). The results shows that in the presence of measurement error with both mild and large extent Lewbel (1996) approach outperforms instrumental variable and at every extent of heteroscedasticity, while control function approach provide the similar result as provided by Lewbel (1996). This shows that correction of measurement error is very important. As we continue to move downward to large sample size the result shows that there is need for large sample size to implement the measurement error correction. Overall the method of control function approach is very best than other two estimators. One thing is common in all three methods that as we move to small sample size to large one the percentage of bias reduces.

CHAPTER 7

CONCLUSIONAND RECOMMENDATIONS

7.1 Introduction

This chapter consists of three sections, first section covers the summary of results of all three methods and concluding remarks. In section two policy recommendations are discussed while in the last section, limitations of the study and recommendations for further study are presented.

7.2 Summary and Conclusions

The objective of the study is to estimate the Engel curve in the presence of measurement error and endogeneity of total expenditure using Pakistani data. In this study there are two main objectives. First objective is to confirm whether linear specification fits the data well or the quadratic one in case of Pakistan. Second objective is to investigate whether measurement error is serious problem or endogeneity of total expenditure. As mentioned above that in the presence of endogeneity problem, few studies shown 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 control function approach which corrects measurement error and controls for endogeneity of total expenditure. In case Pakistan as per best of our information there is no study that bridge this gap. Through this study we tried to fill up this gap using a quadratic specification due to empirical reasoning and nature of the data.

The quadratic food Engel curve is estimated using three different methods namely instrumental variable method (IV), Lewbel (1996) approach and the control function approach employed by Battistin and De Nadai (2013) using recent data of Household Income and Expenditure survey (HIES) of Pakistan for the year 2010-11. The results of these four estimation methods are presented in tables 5.3-5.7 and make us able to derive the following conclusions.

- The descriptive statistics as shown in table 5.1 shows that food expenditure is the largest expenditure in a household budget which constitutes about 56% of total expenditure. Food budget share at different quintiles shown in table 5.2, indicate that as household income increases their budget share allocated to food declines. These results provide us the base to analysis the household welfare with the help of food Engel curve because other welfare indicators, like absolute poverty line and such other has some draw backs.
- All the three methods employed in this study confirms 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 provide by Bhalotra and Attfield (1998) and Kedir and Girma (2007) and verifies that the quadratic logarithmic food Engel curve is a feature of developing countries.
- We 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 similar to that presented by Battistin and De Nadai (2013).
- Another conclusion is that turning points of IV is greater than that of Lewbel (1996) and closer to CF approaches. This suggest that ignoring the problem of endogeneity of total expenditure leads to overestimate the welfare of households. This prove that endogeneity of total expenditure is more serious problem than that of measurement error.

- We have calculated and presented the expenditure elasticity food demand for each estimators. We found that all estimation methods employed in this study provide the expenditure elasticity of food demand in between zero and one. The expenditure elasticity for each methods are closer to each other and the range of 0.77 to 0.81. These elasticity results confirm the result provide by Siddiqui (1982) and Burney and Khan (1991) that food is a necessary good for Pakistani households. This results are consistent with those provide by Siddiqui
- In this study we observed the effect of household size on food demand. The results obtained through all the methods employed in this study confirm that the households size has a positive and statistical significant impact on food demand (food budget share)
- We evaluate the finite sample property (percentage of bias) of the each estimators employed in this study via a Montecarlo simulation. The simulation results confirms that control function approach suggested by Battistin and De Nadai (2013), which account the endogeneity of total expenditure correcting measurement errors, has a lower bias.

7.3 Policy Recommendations

The basic results provided by this study is that the food expenditure is the largest expenditure within the household budget and the share of this expenditure in the budget decreases as the income increases. On the basis of results I want to provide some suggestion and policy recommendations.

- As mentioned in introduction in first chapter that careful statistical analysis is necessary to make appropriate and suitable policies. A misspecification of Engel curve leads to limitation of its effectiveness and generation of misleading results. For example, if the true shape of Engel curve is quadratic but someone assumes a linear specification while estimating Engel curve then the welfare impacts on households with low income is underestimated. So in case of Pakistan as we confirm that a quadratic shape for food Engel curve is suitable therefore it is suggested to researchers that they must be careful while doing statistical analysis of food Engel curve.
- We observed that as income of households rise the expenditure share allocated food items declines and food is the necessary items for Pakistani households. This indicates that it is possible for low income groups to be affected more from the tax policies applied to food expenditure. 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 higher authority must also give subsidies in food items for households so that they can easily purchase foods, to maintain their required calories level, and other basic needs of life. Government must also launch some other targeted base financial assistance program as BISP for low income households so poverty can be reduced.

7.4 Limitations of the study and Further Study

In this study our main objectives were to analyse the shape of Engel curve and identify the more serious problem of endogeneity while estimation of Engel curve. These objectives are achieved limiting our focus and using Pakistani data for households across overall Pakistan. We also limited our focus considering measurement error in the both left-and right side of the budget share equation and compare with the endogeneity of total expenditure. So it can be a gateway for new researchers to revisit these objectives estimating Engel curve for both urban and rural households separately. In the future research, presence of measurement error can be assumed only in the total expenditure as considered by Hausman *et al* (1995) and must compare with both results obtained via Lewbel (1996) approach as done by Kedir and Girma (2007) in case of Ethiopia and control functional approach. Furthermore other important socio-economic variables such as age, education, etc. can also be included in their studies to check their impact on the consumption patterns of households.

References

Ahmed, E. and Arshad, M. (2007). "Household Budget Analysis for Pakistan under Varying the Parameter Approach". PIDE Working Paper *41: 1-19.*

Attanasio, O., Battistin, E. and Mesnard, A. (2012). "Food and Cash Transfers: Evidence from Colombia". Economic Journal *122:92–124.*

Banks, J., Blundell, R. and Lewbel, A. (1997). "Quadratic Engel Curves and Consumer Demand". Review of Economics and Statistics *79: 527–539.*

Battistin, E. and De Nadai, M. (2013). "Identification and Estimation of Engel Curves with Endogenous and Unobserved Expenditures". Journal of Applied Econometrics.

Betti, G. and 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. and Attfield, C. (1998). "Intrahousehold Resource Allocation in Rural Pakistan: A Semiparametric Analysis". Journal of Applied Econometrics *13: 463–480*.

Blundell, R., Duncan, A. and Pendakur, K. (1998). "Semiparametric Estimation and Consumer Demand". Journal of Applied Econometrics *13: 435–461*.

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. and 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. and 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*

Caglayan, E. and Astar, M. (2012). "An Econometric Analysis of Engel's Curve: Household Food and Clothing Consumption in Turkey". Annals of the Alexandru Loan Cuza University-Economic *59 (1):313–319*

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). "Measurement Error in Nonlinear Models". London: Chapman and Hall.

Castaldo, A. and Reilly, B. (2007). "Do Migrant Remittances Affect the Consumption Patterns of Albanian Households?". South-Eastern Europe Journal of Economics *1:25–54.*

Deaton, A. and Muellbauer.J. (1980). "An Almost Ideal Demand System". American Economic Review *70:312–326.*

De Nadai, M. and Lewbel, A. (2012). "Nonparametric Errors in Variables Models with Measurement Errors on Both Side of the Equation". Boston College Working Papers in Economics *No.790.*

Gong, X. Soest, V. A., and 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. and 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., Newey, W. K, and Powell, J. L. (1995). "Nonlinear Errors in Variables Estimation of Some Engel Curves". Journal of Econometrics *65:205–233.*

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.*

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

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

Kedir, A. and Girma, S. (2007). "Quadratic Engel Curves with Measurement Error. Evidence from a Budget Survey". Oxford Bulletin of Economics and Statistics *69: 123–138*.

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

Laser, CEV. (1963). "Forms of Engel Functions". Econometrica *31: 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.*

Lewbel, A. (2012). "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models". Journal of Business and Economic Statistics *30: 67–80.*

Malik, S. J. and 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*.

Ndanshau, M.O.A. (1998). "An econometric analysis of Engel's curve:The case of peasant households in Northern Tanzania". UTAFITI [New Series] Special Issue, 1998-2001, *4: 57–70*.

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

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

Gujarati, N. D. and Porter, C. D. (2008). “Basic Econometric 5th Edition”.

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