QUADRATIC FOOD ENGEL CURVES WITH MEASUREMENT ERROR: EVIDENCE FROM A BUDGET SURVEY

Abbi M. Kedir[1] and Sourafel Girma

Abstract

We explore the issue of possible non-linear relationship between food share and log of real total household expenditure using parametric, non-parametric and semi-parametric techniques. There are almost no studies that estimate quadratic food Engel curves and simultaneously address the measurement error problem in the expenditure variable using budget surveys from developing countries. Our major contribution is the estimation of parametric quadratic food Engel curves controlling for measurement error in deflated total household expenditure using the 1994 Ethiopian Urban Household Budget Survey. Estimated non-parametric and semi-parametric regression results show the presence of a robust quadratic relationship between food share and real household expenditure unlike the findings from OECD countries. An analysis based on a restricted sample after excluding outliers showed that the quadratic relationship is robust. According to our IV estimates, measurement error significantly affects the precision of parameter estimates. Relative to the IV estimates estimators such as OLS and the outlier robust that are uncorrected for measurement error lead to the overestimation of the effect of income on food share in Addis Ababa but in other towns, OLS, outlier robust and LAD estimators lead to a significant underestimation of this effect. Almost an identical pattern is observed on the square term of real expenditure. Irrespective of the estimation techniques, we obtain negative and significant estimates for the quadratic logarithmic expenditure terms in the food budget share equations. Finally, we consider the potential policy distortions that result from a lack of careful handling of the statistical issues surrounding the estimation of the Engel relationship.

JEL Classification: C21, C23, D12, and I31.
Keywords: *quadratic food Engel curves, measurement error, deflated total household expenditure, instrumental variable estimator, outlier*.
Word Count: 7741

---

[1] Corresponding author. Department of Economics, University of Leicester, Tel: +44 (0) 116 252 2894. E-mail address: ak138@le.ac.uk

Acknowledgements
We would like to thank the editor and other members of the editorial board who gave us constructive comments to improve the earlier version of the paper. Thanks are also due to the Department of Economics of Addis Ababa University for allowing us to get access to the data collected by the Ethiopian Urban Household Survey. Kedir acknowledge financial support from the School of Economics University of Nottingham and Girma is grateful to the Leverhulme Trust under Programme Grant no. F114/BF.

1. INTRODUCTION

Engel curve analysis has attracted a considerable amount of attention and has been an important tool in understanding the dynamics of household welfare. It is useful to model income distribution and in the evaluation of indirect tax policy reform (You, 2003; Gibson, 2002; Banks et al 1997). The engel curve literature suggests that food share falls as household welfare improves. This empirical regularity has been confirmed in many applications especially on data sets from developed countries (Banks et al 1997; Blundell et al 1993). Many studies used the Working-Leser specification in which budget shares are assumed to be linear functions of the total expenditure (Deaton and Muellbauer, 1980; Leser, 1963). However, as suggested by a growing body of empirical evidence, a linear specification fails to account for non-linear relationships in certain budget share equations particularly for non-food items (Banks et al 1997; Hausman et al. 1995, and Lewbel, 1991). However, we argue that the non-linearity of the Engel relationship can also be observed for food items in food insecure countries. In the literature, a number of different functional forms and estimation methods for Engel curve, both parametric and non-parametric have been proposed and applied (You, 2003, Gozalo, 1997, Delgado and Miles, 1997)

We investigate whether our households devote a declining proportion of their total budget to food as their living standards rise. To check whether the non-linear relationship in our data is driven by outlying observation in real expenditure, we conducted a sensitivity analysis on the whole sample and a restricted sample which excludes the outliers. We also examine the point which we observe households devoting declining proportion of their budget for food (i.e. turning point analysis). These issues along with the treatment of the measurement error in real expenditure are the focus of the present study. As Banks et al. (1997) observe a quadratic specification allows for a variety of curvatures in Engel curves. Following the instrumental variable estimation approach for polynomial errors-in-variables due to Hausman et al. (1991,1995), we estimate quadratic Engel curves for food using household expenditure data for urban Ethiopia using parametric, non-parametric and semi-parametric regression techniques. The correct specification and estimation of food Engel curves is crucial to identify the reliable magnitude of income effects which can then be used as an input for distributional analysis, food and/or income transfer policies.

Our non-parametric and semi-parametric regression results strongly indicate the presence of a robust quadratic relationship between food share and real expenditure unlike the findings from Organisation for Economic Co-operation and Development (OECD) countries. According to the IV estimates, we found that measurement error significantly affects the precision of the parameter estimates. Relative to the IV estimates, estimators such as OLS and the outlier robust that did not control for measurement error lead to the overestimation of the effect of income on food share in Addis Ababa but in other towns, OLS, outlier robust and LAD estimators lead to a significant underestimation of this effect. Almost an identical pattern is observed on the square term of real expenditure. Irrespective of the estimation techniques, we obtain negative and significant estimates for the quadratic logarithmic expenditure terms in the food budget share equations. The threshold welfare level beyond which food share declines is found to lie between the 35^{th} and 47^{th} percentiles of the total expenditure distribution. This suggests that policy aimed at reducing food poverty should target nearly half of the households. In sharp contrast, estimates which erroneously neglect the problem of measurement error in the expenditure data imply that no more than a quarter of the households suffer from food insecurity. In our data, measurement error accounts for about 33 percent of the variability of the real household expenditure in the whole sample and about 28 percent in the restricted sample.

The paper is organised as follows. Section 2 reviews the relevant literature on Engel curve analysis. Section 3 introduces the quadratic errors-in-variables inference procedure. Section 4 discusses the data used in our empirical analysis. The main empirical results and their policy implications are presented in section 5. Finally section 6 concludes.

2. LITERATURE

The presence of measurement error has been duly recognised in the literature of linear Engel curves estimation since the early 1960's but the treatment of measurement error in quadratic Engel curves is quite a recent phenomenon and is often undertaken in the context of OECD data sets and on non-food items for which the quadratic form is well established. Based on evidence from an Israeli family budget data Liviatan (1961) documents that neglected measurement error induces non-negligible bias in OLS estimates of linear Engel curve parameters. Econometric advances in the early 1990's by Hausman et al (1991) developed a root *n* consistent estimator for polynomial specifications with measurement error. This helped the estimation of quadratic Engel curves for data sets from different countries. Aasness et al. (1993) explicitly model measurement error in the estimation of a system of consumer functions from Norwegian household budget data, and conclude that measurement error accounts for about 27% of the variability of the observed total consumption expenditure. Hausman et al (1995) employ non-linear error-in-variables models in examining the parameters of some Engel curves using US consumer expenditure survey data, and estimate that about 42% of the total variance of measured expenditure is due to measurement error. Within a framework of Generalised Method of Moments (GMM) estimation of household demand for fuel in the United Kingdom, Lewbel (1996) finds that correction for measurement error changes parameter estimates by more than 15%. Using data from the British Family Expenditure Survey, Hikaru & Kozumi (2001) consider Engel curves estimation with measurement error from Bayesian perspectives, and report that the observed mean household total expenditure over-estimates the mean of the `true' total expenditure. In fact, under a classical measurement error with a normally distributed regressor, Kuha and Temple (1999) have shown that OLS-type `naïve' estimators tend to generate estimates where the quadratic model appears to curve less steeply than it actually does. Another evidence from the UK is provided by Banks et al (1997) which found the presence of a quadratic relationship between expenditure on non-food items and total household expenditure. However, the study could not reject the Working-Leser linear specification for food. In a recent application on budget survey data from Canada, You (2003) showed that robust estimators point to lower income elasticities and have better performance than the standard least squares and Tobit estimators when there are obvious outliers in the data. Betti (2000) showed evidence of non-linearities in Engel curves for seven commodities including food using Italian budget data but did not control for measurement error.

In the following few paragraphs, we discuss some of the scanty empirical literature on Engel curves from developing countries. Moffit (1989) investigates the magnitude of the effect of in-kind transfers on the consumption of subsidised goods by evaluating the experience of an actual conversion from food stamps to cash in Puerto Rico. A linear budget share equation gives –0.89 as the coefficient of the household welfare measure. This statistically significant term suggests that as households' welfare improves the share of the household budget devoted to food will diminish. A panel data version of the Working-Leser Engel function is proposed by Wan (1996). Among the commodities considered by this study using data from China since the early 1980s, pork, vegetables, poultry, beef and lamb are found to be luxuries while cereals and other foods as necessities. In developing countries, most studies estimate income elasticities of calorie intake/availability without a focus on the food share-expenditure relationship (Subramanian and Deaton 1996). Kebede (2003) uses a budget survey from rural Ethiopia to estimate a Quadratic Almost Ideal Demand System (QUAIDS) in an attempt to unveil the intra-household allocation of resources. His findings indicate that the budget devoted to food increases as household incomes increased suggesting that many rural households live near subsistence levels. It is also observed that at very low levels of income where households are on the verge of starvation, additional income is most likely to be used to increase food purchases.

As the review above demonstrates, there is an acute lack of studies of Engel Curves for developing countries that employ non-linear errors-in-variables estimators. In addition, the issue is not investigated using household information from urban areas and none of the studies in developing countries analysed the potential curvature in the Engel relationship for food in conjunction with the treatment of measurement error in the expenditure variable. Therefore, this study is aimed at contributing to the literature by estimating quadratic Engel Curves allowing for measurement error in the total expenditure variable.

2. THE ECONOMETRIC FRAMEWORK

We argue that with error-ridden expenditure data, standard econometric approaches do not provide the appropriate framework for answering the central question of this study: *When does food share starts to decline as income increases once we control for measurement error in reported total expenditure?* We apply quadratic errors-in-variables techniques so as to recover consistent estimates of the Engel curve relationship and also conducted a sensitivity analysis to investigate the impact of outliers on parameter estimates and turning points. To our knowledge, this is the first paper to do so using developing country data where measurement error problems are arguably more severe (Gibson, 2002; Deaton, 1997). We also evaluate the effects of neglecting measurement error and identify the magnitude of measurement error in the total expenditure data.

In many of the applications of Engel curve analysis, expenditure data for many countries have been analysed as a basis of quantifying total expenditure (or income) eslasticities of demand for different types of commodities but it is well known that this variable is measured with error (Hausman et al, 1995; You, 2003). The measurement error in expenditure data is more likely to be severe in recall surveys than diary surveys (Gibson, 2002). The presence of errors in variables induces non-zero correlation between the contaminated regressors and the equation disturbance, so that OLS estimates are biased and inconsistent (Aigner et al. 1984; Fuller, 1987). One way of dealing with errors in variables is to assume that they have a non-normal distribution, and exploit the information contained in higher order moments of the data to produce identifying restrictions on the parameters of interest. This approach has not thus far gained much currency in the applied economic literature. Another approach is to assume some knowledge of the measurement error variance, in which case the model can be identified by purging the contaminating effect of the measurement error from the covariance matrix of the data. But by far the most popular method of getting round the identification problem caused by errors in variables appears to be instrumental variables (IV) estimation. Unfortunately, for non-linear models standard IV techniques lead to inconsistent estimates because the true variables are not additively separable from the measurement error (Amemiya, 1985). Hence the estimation problem is not trivial.

In this paper we employ the IV estimation approach for polynomial errors-in-variables due to Hausman et al. (1991, 1995), in the context of the following quadratic Engel curve model:

$$y_i = \beta_o + \beta_1 z_i + \beta_2 z_i^2 + \phi r_i + \varepsilon_i, \quad \text{i =1,...N} \qquad (1)$$

where i indexes households, y is the budget share of food; z represents the log of `true' total household expenditure, which is not observed; r denotes household size, which is assumed to be correctly measured, and □ is a stochastic disturbance term.

There are both conceptual and pragmatic reasons why expenditures available from household surveys might be preferred to an indicator such as household income in developing countries. If the income stream accruing to the consumer and his needs were steady over time, it could be argued that the conditions of the static theory of consumer demand were satisfied and that income is an appropriate measure of welfare. However, this supposition is unrealistic because both the income of a household and its needs change over time, and the income received in a particular period may be a very poor indicator of its standard of living. This is attributed to the permanent income hypothesis (PIH). In the light of the PIH, it is often argued that expenditures reflect not only what a household is able to command based on its current income, but also whether that household can access credit markets or household savings at times when current incomes are low. In this way, expenditure is thought to provide a better picture of a household's long run standard of living than a measure of current income. Further, calculating consumption expenditure is often easier than calculating household incomes, particularly for the poor[2].

As stated in the introduction, we explicitly recognise that errors are present in our expenditure data. Instead of the `true' expenditure variable z_i, we observe the expenditure variable x_i, which is related to z_i via the following measurement model:

$$x_i = z_i + \eta_i \qquad \text{i = 1,..,N} \qquad (2)$$

where η_i is a mean zero measurement error with unknown variance σ_η^2.

[2] See Deaton and Grosh, 2000 and Hentschel and Lanjouw, 1996 for a detailed argument about the appropriate welfare measure in the context of developing countries.

Suppose that in equation (1) $\beta_2 < 0$ and consider the turning point of the true model for $E\{Y \mid z\}$, $\beta^* = -\frac{\beta_1}{2\beta_2}$. Kuha and Temple (1999) show that the degree and direction of bias in the estimator for the turning point depend on the variance of and the location of the turning point relative to the population mean of *z,* say μ_z. The true turning point always lies between μ_z and the OLS estimates of the turning point. In general the effect of measurement error is to "flatten" the observed relationship between y and x, where the quadratic model curves less steeply and has a smaller maximum value for y[3]. The identifying information for our model comes in the shape of a p-dimensional vector of instrumental variables q_i which contains sufficient independent variation that helps predict the unobserved regressor z via the equation:

$$z_i = q_i^{'}\alpha + v_i \equiv w_i + v_i, \quad \text{i = ,..,N} \tag{3}$$

The instruments used in this study include the log and log square of income, the gender of the head of household and regional dummies. In equation (3) $w_i \equiv q_i' \alpha$ can be interpreted as the part of z which is linearly related to the instrumental variables (Hausman et al, 1991). The two most crucial assumptions imposed on the data are:

Assumption 1: $E(\varepsilon_i \mid q_i, r_i) = E(\eta_i \mid q_i, r_i) = 0$ *and* $E(\varepsilon_i.\eta_i \mid q_i, r_i) = \sigma_{\varepsilon\eta}$

Assumption 2: υ_i is independent of q_i and r_i with $E(\upsilon_i) = 0$ and $E(\upsilon_i.\varepsilon_i \mid q_i, r_i) = \sigma_{\varepsilon\upsilon}$.

Substituting equation (3) into equation (1), gives;

$$y_i = \gamma_0 + \gamma_1 w_i + \gamma_2 w_i^2 + \gamma_\phi r_i + e_i \tag{4}$$

where the composite error term e_i can be shown to be orthogonal to the regressors of the reduced form model (4). The parameters of the reduced form model are related to the structural parameters as follows: $\gamma_0 = \beta_0 + \beta_2 \sigma_\eta^2$; $\gamma_1 = \beta_1$; $\gamma_2 = \beta_2$ and $\gamma_\phi = \phi$.

---

[3] This is equivalent to the measurement-error-induced bias towards zero of the slope estimators in linear regressions.

Hence equation (4) can be used to identify β_1 and β_2, but not β_0 and σ_η^2. To identify the remaining coefficients of interest we follow Hausman et al. (1991) in multiplying equation (1) by the mismeasured regressor x_i and substituting $w_i+\upsilon_i$ for z_i, to obtain a second reduced form equation[4]

$$x_i.y_i = \delta_0 + \delta_1 w_i + \delta_2 w_i^2 + \delta_3 w_i^3 + \delta_\phi w_i r_i + error \qquad (5)$$

It can easily be shown that there exists a one-to-one mapping between the coefficients of the above model and the structural parameters. That is, $\delta_o = \beta_1\sigma_\eta^2 + \beta_2\upsilon_3 + \sigma_{\varepsilon\eta} + \sigma_{\varepsilon\upsilon}$ where $\upsilon_3 = E(\upsilon^3)$; $\delta_1 = \beta_0 + 3\beta_2\sigma_\eta^2$; $\delta_2 = \beta_1$; $\delta_3 = \beta_2$ and $\delta_\phi = \phi$.

The intercept and the measurement error variance can thus be estimated from the following recursive formulae: $\sigma_\eta^2 = \dfrac{\delta_1 - \gamma_0}{2\beta_2}$ and $\beta_0 = \delta_1 - 3\beta_2\sigma_\eta^2$. But there is more than one solution for β_2, β_1 *and* ϕ using the reduced form parameters $\gamma_1, \gamma_2, \gamma_\phi, \delta_3, \delta_2$ *and* δ_ϕ which results in over-identification. Given the estimated reduced form parameters, a method of obtaining efficient estimators for the structural parameters is the optimal minimum distance (or minimum chi-square) technique. Define the 5×1 vector of structural parameters as $\theta = (\beta_0\ \beta_1\ \beta_2\ \phi'\ \sigma_\eta^2)'$ and let $\hat{\pi} = (\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_\phi : \hat{\delta}_3, \hat{\delta}_2, \hat{\delta}_\phi : \gamma_0\hat{\delta}_1)' \equiv (\hat{\pi}_1 : \hat{\pi}_2 : \hat{\pi}_3)'$ be the vector of the reduced form parameters[5]. Using the fact that $\pi_1 = \pi_2$ and letting $\hat{\pi}_r \equiv (\hat{\pi}_2 : \hat{\pi}_3)$, the solution to the minimum chi-square problem is

$$Q = \arg\min_\theta \left[\hat{\pi}_r - h(\theta)\right]' \hat{\Omega}^{-1} \left[\hat{\pi}_r - h(\theta)\right] \qquad (6)$$

[4] In a similar vein, Lewbel (1996) proposes a GMM by multiplying the original demand equation by different power of the observed total consumption expenditure.

[5] Note that he estimate of δ_o is not useful in the identification of the structural coefficients.

where Ω is the asymptotic covariance matrix of the restricted reduced form parameters and $h(\theta)$ is the 5×1 vector function mapping $\pi_r = \begin{bmatrix} \delta_3 \\ \delta_2 \\ \delta_\phi \\ \gamma_0 \\ \delta_1 \end{bmatrix}$ to θ,

which takes the following form: $h(\theta) \equiv \begin{bmatrix} \beta_2 \\ \beta_1 \\ \phi \\ \beta_1 + \beta_2\sigma_\eta^2 \\ \beta_1 + 3\beta_2\sigma_\eta^2 \end{bmatrix}$.

We solved equation (6) by using the general method described in Section 3.2 of Hausman et al. (1991). The resulting minimum chi-square estimator of θ, $\hat{\theta}$, is asymptotically distributed as $\sqrt{N}(\theta-\hat{\theta}) \sim N\left[\left(0, \hat{H}'\hat{\Omega}^{-1}\hat{H}\right)^{-1}\right]$, where $\hat{H}$ is a consistent estimator of the Jacobian matrix $\frac{\partial h(\theta)}{\partial \theta'} = \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & \sigma_\eta^2 & 0 & \beta_2 \\ 1 & 0 & 3\sigma_\eta^2 & 0 & 3\beta_2 \end{bmatrix}$.

Finally, one can make an appeal to the general theory of minimum chi-square estimation to establish that, under the null hypothesis of correctly specified model,

$$\sqrt{N}\,(\hat{\pi}-\hat{\pi}_{21})'\left[Var((\hat{\pi}-\hat{\pi}_{21})\right]^{-1}(\hat{\pi}-\hat{\pi}_{21}) \xrightarrow{d} \chi^2(3). \qquad (7)$$

This result is used to test the validity of the overidentifying restrictions implied by the assumptions of the model.

3. DATA

Our empirical analysis is based on the 1994 socio-economic survey of urban households in Ethiopia (EUHS, 1994). The survey questionnaire includes modules on household demographics including education, rural-urban migration, employment and income, consumption, ownership of durables, housing, health, welfare and welfare change indicators. A sample of 1500 households was selected from seven major urban centres of the country. These are Mekele and Dessie in the north, Bahir Dar in the north west; Addis Ababa in the centre, Dire Dawa in the east, Awassa in the south and Jimma in the south west were selected. Mekele and Dessie were selected to represent areas often affected by drought and the socio-economic groups in the north. Bahir Dar was included as a representative town in the main cereal producing areas of the country. Addis Ababa is by far the largest city and the capital, and represents the diversity of the country's population. Dire Dawa is mainly a trading centre, while Awassa is the administrative centre of the south, and was chosen to represent the large Enset culture[6]. Finally, Jimma was selected to represent the urban characteristics of the main coffee growing regions of the country. The total sample size was distributed over the selected urban centres proportional to their populations, based on the Central Statistical Authority's population figure projections. Accordingly, the sample included 900 households in Addis Ababa, 125 in Dire Dawa, 75 in Awassa, and 100 in each of the other four towns.

Traditional Engel curve analysis assumes that all households in a given survey face the same prices. But because transportation and distribution networks tend to develop along with economic growth, there is much greater scope for spatial price variation in less developed than more developed countries. We relax the usual constant price assumption in the present paper and employ carefully constructed spatial price deflator on the raw expenditure data[7]. The spatial price indices are reported in the Table 1. Table 2 gives summary statistics of the variables used in the study for households where complete information is available[8]. On average food share for other urban areas is higher by about 3.5 percentage points compared to Addis Ababa, and this proves to be significant at 1% level using a t-test for the equality of means. By contrast no appreciable difference in the means of deflated total expenditure is found at conventional significance levels.

4. RESULTS

The estimation results of the quadratic food Engel curves are reported in Tables 3, 4 and 5. The findings at this stage of the study suggest that the incorporation of the square of the log of household expenditure in our Engel curve equation is amply

[6] This is one of the major food cultures in southern Ethiopia. Enset is often referred to as false banana.
[7] Details on how they are constructed under different price assumptions can be found in Kedir et al (2003).
[8] Apart from missing observations, we also dropped households with reported 0 or 100% food share values.

justified. First we report the results from our preliminary investigation of the data using non-parametric and semi-parametric regression techniques. Second, we discuss the sensitivity analyses we conducted on the estimates using different estimators to examine the robustness of the quadratic relationship. The different estimators we considered are ordinary least squares (OLS), least absolute deviation (LAD) and outlier robust estimators. We generated the estimators both for the whole sample and for a restricted sample after removing the outlying observations[9]. Finally, we present and discuss the impact of measurement error on the estimated parameters.

A. Non-parametric and semi-parametric analysis

Non-parametric regression analysis provides a strong alternative to linear regression because it allows the data to determine the 'local' shape of the conditional mean relationship. In applied microeconometric literature, it is believed that it is useful to consider a parametric specification against a non-parametric alternative (Blundell and Duncan, 1998). Our preliminary non-parametric and semi-parametric analysis showed a clear evidence of the quadratic relationship between food share and deflated total expenditure. As clearly shown in Figures 1 and 2, the quadratic relationship is not driven by outliers because the exclusion of outliers did not result in significant changes of the shape of the Engel curves. The figures show that the quadratic relationship is stronger in the case of other urban areas both for the whole and restricted sample relative to the sample drawn from Addis Ababa.

The semi-parametric regression results shown in Figures 3 and 4 are quite consistent with the non-parametric/Kernel fits reported in Figures 1 and 2. In the semi-parametric framework, we considered the case of endogenous real total household expenditure and the Figures 3 and 4 shows that such an approach has a slight impact on the curvature of the Engel curve relationship. Both the non-parametric and the semi-parametric approaches show that there is a distinct nonlinear behaviour in food share which is stronger for the data from other urban areas than Addis Ababa.

The nonparametric and semi-parametric regressions reveal an approximate inverted U-shaped relation between budget shares for food and log of total expenditure. This makes it apparent that the Leser-Working linear curve formulation is not an accurate

[9] We thank the referees of the original version of the paper who pointed out that the quadratic relationship in our data might have been driven by outlying observations. Therefore, we have undertaken a robustness check with and without the outliers to establish whether that is the case. During our initial exploratory analysis, we also investigated if there is a need for a further (i.e. third order) polynomial term, but this notion is decidedly rejected by the Ramsey RESET test.

approximation to our data. The inverted U-shaped relationship has important welfare implications. For instance, let us look at turning points. From Figure 1 for the whole sample that the budget shares allocated to food starts to decline at about 400-492 *Birr*[10], which roughly corresponds to the median expenditure value in the raw data. Taken at face value this implies that for 50 percent of the households in the sample are food insecure. Beyond this turning point, the share of food declines. This is more rapid for households in Addis Ababa than for households outside the capital.

B. Estimators without controlling for measurement error

Tables 3 and 4 report the robustness results of our parameters across estimation techniques for the whole sample and the restricted sample respectively. We found that irrespective of the method, the quadratic relationship is robust and statistically significant under all estimation techniques. This is also true whether or not we removed outlying observations of the regressor - real total household expenditure. According to estimates of Tables 3 and 4, turning points have been affected by estimation techniques and outliers. Their effect is more significant for the sample of households from other urban areas than the estimates for the capital city- Addis Ababa. When we compare the results reported in these tables with results in Table 5, it is evident that the coefficient of the expenditure term increased as we controlled for measurement error. This is more pronounced as outliers were excluded from the analysis.

When survey data contains some extreme-value observations which cause estimation bias, the use of outlier robust estimators which have various desirable properties is appropriate. Our robust estimates point to smaller parameters of the expenditure term for 'other urban areas' but not for 'Addis Ababa'. The impact of our regressors on the food share are stronger for the restricted sample. Therefore, outliers have led to underestimation of the parameter estimates. Trimming reduced the implied share of the measurement error variance supporting the argument that avoiding outliers diffuses the impact of measurement error on parameter estimates.

C. Measurement Error Corrected IV Estimates

We present our measurement error corrected estimates in Table 5. We consider the estimates of Tables 3 and 4 relative to the measurement error corrected IV estimates reported in Table 5. The comparison reveals that the OLS and the outlier robust

[10] The *Birr* is the name of the Ethiopian currency. The interval for the turning point is calculated based on the exponential of 6 and 6.2, after `eyeballing' the graphs from the kernel regressions.

estimators lead to the overestimation of the effect of income on food share in Addis Ababa but in other towns, all the measurement error uncorrected estimators (i.e. OLS, outlier robust and LAD) severely underestimate it. Almost an identical pattern is observed on the square term of real expenditure. However all estimators in Table 4 underestimate the effect of income on food share as the IV estimates are significantly larger.

Both measurement error corrected IV estimates and measurement error uncorrected estimators yield negative and significant estimates for the quadratic logarithmic expenditure terms in the food budget share equations. The results strongly suggest that a small reduction in income or increase in food prices may lead to severe food shortages. However as discussed below, the qualitative similarity between the two estimators masks some important statistical and interpretation differences.

For instance, Hausman-type tests reveal significant (joint) differences between the quadratic Engel curves coefficient estimates obtained from the IV and OLS estimates. In other words, these tests reject the null that the measurement errors present in the expenditure data are not serious enough to invalidate OLS based inferences. This conclusion is reinforced by the magnitude of the relative importance of the measurement error variances implied from the IV estimates. Looking at the results for the whole sample from Addis Ababa, our results indicate that about a third of the total variance in real household expenditure is due to measurement error. The corresponding figure for other urban areas is 37%. The measurement error variances are smaller for the restricted sample for both locations. As indicated above this shows the role played by the exclusion of outliers in reducing the potential measurement error problem in real household expenditure. However, it is obvious from the results that trimming does not get rid of the measurement error problem completely. This is because a non-negligible proportion of the variance of real household expenditure (i.e. about 30%) is still due to measurement error in the restricted sample. Thus a sizeable proportion of `noise' is found to exist in the expenditure variable, justifying the use of errors-in-variables technique which seeks to extract the true `signal' from the data. As Table 5 shows, the validity of the IV technique we adopted in this paper is confirmed by the test of the overidentifying restrictions derived from the model (see reported p-values).

To briefly highlight the impact of household size on food share, we restrict our discussion only to the first columns of Tables 3 and 5. The IV estimates of the coefficient on household size falls short of significance for the Addis Ababa sample, while a negative size-food share relationship is established for households outside the capital. By contrast OLS estimates show that household size is not a significant determinant of food share for households outside Addis Ababa, while it attracts positive coefficients in the Engel curve specification for households in the capital. It appears that measurement error in the expenditure data has imparted an upward bias in the OLS coefficient of the correctly measured household size variable. Since household size and total expenditure are positively correlated in the data, the direction of this bias is consistent with predictions from econometric theory (Caroll et al, 1995).

D. Turning points

We now turn our attention to the discussion of the turning points of our quadratic Engel curves. Turning points differ between Addis and Other urban areas. These differences are caused, inter alia, due to differences tastes; preferences in consumption and regional food price variation. As expected, the results suggest that households in the capital, Addis Ababa, have more choices for which they can allocate their budget relative to households in other cities. In Table 4, turning points tend to increase for 'other urban areas' as the estimators change but for Addis the reverse is true. In Table 3 we cannot be conclusive about the trend as the turning point has dramatically declined for the LAD estimators in Addis Ababa. Under the same estimator, the turning point increased for the sample of households from other urban areas.

If we compare the first columns of Tables 3 and Table 5, in spite of the apparent closeness of the OLS and IV coefficients of the expenditure terms for Addis Ababa households, a marked difference is observed in both the point estimate and the confidence interval[11] of the Engel curves' turning point. According to the OLS estimates for Addis Ababa, food share increases with household income for households with less than 182 *Birr,* and this corresponds to the 15th percentile of the data. By contrast, the measurement error corrected turning point estimate points to the conclusion that food does not display the characteristics of a necessity until total expenditure reaches 351 *Birr.* The raw data shows that 35% of the households in the capital city spend less than this amount for food. Thus OLS overstates the welfare of about 20% of the households. The story is similar when one considers the sample from the other urban areas. The OLS and IV based turning points are 310 *Birr* and 501 *Birr* respectively, expenditure levels which approximately correspond to the 26th and 47th percentiles. This demonstrates that inferences that neglect the measurement error in the expenditure data would erroneously classify 21% of the households as having enough to eat.

[11] We used the `delta' method to compute the standard error of the turning point.

Notice that the confidence intervals of the turning points associated with the measurement error corrected estimators are wider. This is a manifestation of the well known “bias versus variance trade off” problem, where the very process of correcting for bias makes the corrected estimator more variable than the biased estimator (cf. Carroll et al, 1995, Section 2.4). But this does not alter the fact that inferences using OLS-based confidence intervals would severely underestimate the “satiety” level of food. Incidentally, compared to OLS, the measurement error corrected IV estimator yields turning points closer to the non-parametric regressions. This is consistent with the notion that measurement error flattens the curvature of OLS-estimated quadratic functions (Kuha and Temple, 1999).

Overall our findings suggest that a significant number of households may have just enough to eat but at the same time can be malnourished, even if they may be deemed non-food poor according to some poverty lines (Dercon and Taddesse, 1997). Thus the approach of assessing food poverty by studying the curvature of Engel curves can be a useful complement to the various analysis based on poverty lines.

E. Policy Implications

Both quantitative and qualitative studies put food security as the top priority to be addressed in Ethiopia. For many decades, Ethiopia has suffered from the extreme form of food deprivation typified by famines. This has been mainly because the poor do not have adequate purchasing power to secure access to food. The failure of “entitlement” or effective demand in the short and long run can occur in a variety of ways which are linked to the system of production and distribution of income in an economy (Sen, 1981), including a decline in employment income; or a rise in the price of food. It arises not only within the agricultural sector but also outside it as evidenced by an increasing number of households residing in urban areas failing to get enough food on their tables (Kedir et al, 2003). Policy needs to focus on enhancing households' entitlement to prevent aggravated hunger and malnutrition. The geographical and income profiles of the poor can be a basis to target them so that they benefit from policy interventions. Direct food transfers, food subsidies, employment generation and/or income transfers are possible policy options open to the Ethiopian government to improve the food situation of urban households.

Access to food can also be enhanced for urban households through an appropriate mix of market and transfer mechanisms. A major policy goal in Ethiopia should be how to design food and income transfer programmes in ways that do not adversely affect the development of a food marketing system that stimulates production incentives, income growth and more affordable food over the long run (Jayne and Molla, 1995). To guide future food policy, especially with respect to potential commodity price stabilisation and food aid monetisation[12], it is also vital to understand the factors affecting food market prices in Ethiopia and the behaviour of food markets more generally. Policies need to be based on accurate food poverty studies and robust statistical analyses such as the one provided by this paper.

To highlight the potential policy distortion induced by measurement error, we examine the rate at which a rise in real household expenditure leads to a decline in the budget share of food. To this end, we report in Table 6 elasticity estimates at selected points of the total expenditure distribution for two estimators – OLS and IV. According to the OLS estimator for Addis Ababa, food share starts to decline at about the 25th percentile while this does not occur (i.e. a negative elasticity is not observed) until we reach the 50th percentile for the measurement error corrected IV estimator. In other words, a food poverty reduction policy based on OLS or other naïve estimates would overestimate the welfare of up to 25% percent of the population.

The findings of our paper show the need for a careful analysis of household consumption behaviour. With the aid of our quadratic Engel curve specifications, we uncovered the implied extent of urban food insecurity. For sound food/income transfer policies targeted at alleviating food shortage problems in Ethiopia, we showed the income level beyond which households start to reduce the food share of their budget. In connection to that, we also identified the implied percentage of households that need to be classified as potential beneficiaries of such interventions.

5. CONCLUSION

This paper was motivated by the acute lack of estimated food Engel relationships that account for non-linearity in the relationship and measurement error using data from a developing country. We focus particularly on food due to the crucial aspect of food security issues in the Ethiopian context and our findings have relevance to policy making. We analysed the Ethiopian urban household budget survey and our results pointed to more curvature in the food Engel curve than is permitted by the standard Working-Leser form. We showed that under a variety of estimators we can reject the linear Working-Leser form for food. This a direct contrast to findings on OECD data sets. Important contributions of this study include the introduction non-linearity in the engel relationship, a detailed estimation of the engel curves using different functional

[12] Food aid monetisation refers to the sales of food aid by the state onto the market to influence market prices.

forms to test the robustness of our findings and more importantly the explicit recognition of the measurement error problem in deflated total household expenditure. Our careful investigation highlighted the severity of the measurement error problem and its impact in the precision of the income elasticities. The impact of outliers have also been examined under different functional forms. Through a robustness check we found that the quadratic form for the food Engel curve is not driven by outlying observations in real household expenditure.

In this study, we recognise and deal with the fact that observed household expenditure is an imperfect measure of actual outlays. The finding that about a third of the total variance in expenditure is due to measurement error vindicates our approach. We also explicitly focus on identifying the proportion of urban households in Ethiopia that are more likely to suffer from food insecurity in urban Ethiopia. We have also highlighted the possible food and income policy implications that emerge from our results.

According to the estimator which has taken due account of measurement error in expenditure data, the threshold welfare level beyond which food ceases to be a luxury lies between the 35^{th} and 47^{th} percentiles of the total expenditure distribution. Economic policy that aims at reducing food poverty will thus have to target nearly half of the urban households. This is markedly different from the policy recommendation that emanates from the naïve ordinary squares estimator. The latter implies that no more than a quarter of the urban population suffer from food insecurity. In this respect, the present study has demonstrated the potential for serious policy distortions resulting from a lack of careful statistical analysis. From a technical point of view, this study hopefully give some guidance to future researchers wishing to use our data set and similar data sets from other developing countries.

References

Aasness, J., Biorn, E., and Skjerpen. T. (1993). 'Engel functions, Panel Data and Latent Variables', *Econometrica* Vol. 61, pp. 139-1422.

Aigner, D.J., Hsiao, C., Kapteyn, A., and Wansbeek, T. (1984). 'Latent Variable Models in Econometrics', in Griliches, Z., Intriligator, M.D. (eds.), 'Handbook of Econometrics', Vol. II, North-Holland, Amsterdam, pp.1321-1393.

Amemiya, Y. (1985). 'Instrumental Variable Estimator for the Non-linear Errors-in-Variables Model', *Journal of Econometrics,* Vol. 28, pp. 273-289.

Ayalew, T. (2000). 'Liquidity Constraint and the Demand for Food: Income Elasticity of Calorie in Rural Ethiopia', mimeo, Centre for Economic Studies, Katholieke Universitiet Leuven, Belgium.

Banks, J., Blundell, R., and Lewbel, A. (1997). 'Quadratic Engel curves and consumer demand', *The Review of Economics and Statistics,* Vol. 79, pp. 527-539.

Betti, G. (2000). 'Quadratic Engel Curves and Household Equivalence Scales: the Case of Italy 1985-1994', mimeo, Department of Statistics, London School of Economics.

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

Deaton, A., (1997). *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy,* John Hopkins University Press, Baltimore and London.

Deaton, A. and Grosh, M. (2000). 'Consumption', in *Designing Household Survey Questionnaires for Developing Countries: Lesson from the 15 years of the Living Standards Measurement Study*, by M. Grosh and Glewwe, P. (eds), Chapter 17.

Deaton, A., and Muellbauer, J. (1980). *Economics and Consumer Behaviour,* Cambridge University Press, Cambridge.

Delgado, M. and Miles, D. (1997) Household Characteristics and Consumption Behaviour: a non-parametric approach. *Empirical Economics* 22:409-429.

Dercon, S., and Tadesse, M., (1997). 'A Comparison of poverty in rural and urban Ethiopia', mimeo, Centre for the Study of African Economies, University of Oxford and Department of Economics, Addis Ababa University.

Fuller, W.A. (1987). *Measurement Error Models*. New York: Wiley.

Gozalo, P. (1997) Nonparametric Bootstrap Analysis with Applications to Demographic Effects in Demand Functions. *Journal of Econometrics* 81:357-393.

Hasegawa, H. and Kozumi, H. (2001). 'Bayesian Analysis on Engel Curves Estimation with Measurement Error and an Instrumental Variable'*, Journal of Business and Economic Statistics,* Vol. 19, pp. 292-298.

Hausman, J., Newey, W., Ichimura, H., and Powell, J. (1991). 'Measurement Error in Polynomial Regression Models'. *Journal of Econometrics*, Vol. 50, pp. 273-295.

Hausman, J., Newey, W., and Powell, J. (1995) 'Nonlinear Errors in Variables Estimation of Some Engel Curves', *Journal of Econometrics,* Vol. 65, pp. 205-233.

Hentschel, J. and Lanjouw, P. (1996). 'Constructing an Indicator of Consumption for the Analysis of Poverty: Principles and Illustrations with Reference to Ecuador', World Bank Living Standards Measurement Study, Working Paper No. 124.

Islam, N. (1989). 'Undernutrition and Poverty: Magnitude, Pattern and Measures for Alleviation', in: John, W. H., Stanley R. (eds.), A reprint from 1988 World Food Conference Proceedings, Volume II, Iowa State University Press, 1989.

Jayne, T.S., and Molla, D., (1995). 'Toward a Research Agenda to Promote Household Access to Food in Ethiopia'. Working Paper 2, Food Security Research Project, and Ministry of Economic Development and Cooperation, Addis Ababa.

Kebede, B. (2003). 'Intra-household Distribution of Expenditures in Rural Ethiopia: A Demand Systems Approach'. CSAE WPS/2003-08, Centre for the Study of African Economies and St. Anthony's College, Oxford University.

Kedir, A., Disney, R., and McKay, A. (2003). 'Price Deflators and Food Poverty in Urban Ethiopia'. Mimeo, School of Economics, University of Nottingham.

Kuha, J., and Temple, J. (2003) Covariate measurement error in quadratic regression. *International Statistical Review*, vol. 71(1), pp. 131-150.

Leser, C.E.V. (1963). 'Forms of Engel Functions', *Econometrica,* Vol. 31, pp. 694-703.

Lewbel, A. (1991). 'The Rank of Demand Systems: Theory and Non-Parametric Estimation', *Econometrica* Vol. 94, pp. 979-1000.

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

Liviatan, N. (1961). 'Errors in Variables and Engel Curve Analysis', *Econometrica* Vol. 29, pp. 336-362.

Moffit, R. (1989). 'Estimating the Value of an In-kind Transfer: the Case of Food Stamps', *Econometrica* Vol. 57, pp. 385-409.

Sen, A.K. (1981). *Poverty and Famine: An Essay on Entitlement and Deprivation,* Clarendon Press, Oxford.

Subramanian, S. and Deaton, A. (1996). 'The Demand for Food and Calories', *Journal of Political Economy,* Vol. 104, pp. 133-162.

Wan, G. H. (1996) Using Panel Data to Estimate Engel Functions: food consumption in China. Applied Economics Letters, vol. 3, pp. 621-624.

You, J. (2003) Robust Estimation of Models of Engel Curves, *Empirical Economics*, 28:61-73.

Table 1: Regional Cost of Living Indices based on Corrected Unit Values

Urban area	Fishers' Ideal index
Addis	100.00
Awassa	81.0
Bahardar	99.0
Dessie	105.0
Dire Dawa	113.0
Jimma	101.0
Melee	99.0

Table 2: Descriptive statistics for the food share data

Variable	Addis Ababa		Other Urban Areas	
	Mean	Std. Dev	Mean	Std. dev
Food share	0.674	.178	0.710	.174
Log of real total expenditure	6.213	0.919	6.216	.903
Log of real income	6.307	1.283	6.270	1.158
Log household size	1.746	0.506	1.643	0.565
Sample size	879		558	

Table 3: **OLS, outlier robust and least absolute deviations (LAD) estimates of the quadratic Engel Curve parameters: Whole Sample**

	Addis Ababa			Other Urban Areas		
	OLS with robust s.e	Outlier robust	LAD	OLS with robust s.e	Outlier robust	LAD
Log of real total expenditure	**0.294**	**0.343**	**0.267**	**0.283**	**0.240**	**0.348**
	(4.48)**	**(6.43)****	**(3.96)****	**(3.09)****	(3.52)**	(4.72)**
Square of log of real total expenditure	-0.028	-0.032	-0.026	-0.025	-0.021	-0.030
	(5.39)**	(7.61)**	(4.95)**	(3.34)**	(3.75)**	(4.97)**
Log of household size	0.038	0.034	0.053	-0.004	-0.008	0.000
	(2.97)**	(2.92)**	(3.67)**	(0.22)	(0.54)	(0.00)
Constant	-0.111	-0.246	-0.021	-0.071	0.076	-0.226
	(0.55)	(1.47)	(0.10)	(0.25)	(0.37)	(1.02)
Turning point (95% confidence interval)	186.4 (118.4, 293.9)	215.3 (157.6, 294.1)	162.6 (92.0 , 287.4)	310.1 (188.1, 511.3)	307.0 (195.8, 482.0)	322.8 (231.8, 449.4)
Observations	879	879	879	558	558	558

Notes:

(i) Absolute value of t-statistics in parentheses

(ii) * significant at 5%; ** significant at 1%

(iii) Confidence intervals for turning points are computed using the delta method

Table 4: OLS, outlier robust and least absolute deviations(LAD) estimates: Restricted sample

	Addis Ababa			Other Urban Areas		
	OLS with robust s.e	**Outlier robust**	**LAD**	**OLS with robust s.e**	**Outlier robust**	**LAD**
Log of real total expenditure	0.384	0.383	0.324	0.450	0.472	0.613
	(6.61)**	(7.80)**	(5.23)**	(6.16)**	(6.47)**	(7.46)**
Square of log of real total expenditure	-0.036	-0.036	-0.031	-0.038	-0.040	-0.051
	(7.72)**	(9.26)**	(6.45)**	(6.29)**	(6.70)**	(7.68)**
Log of household size	0.039	0.040	0.052	-0.004	-0.004	0.003
	(3.93)**	(3.97)**	(4.05)**	(0.32)	(0.28)	(0.22)
Constant	-0.359	-0.352	-0.168	-0.580	-0.640	-1.057
	(2.03)*	(2.31)*	(0.87)	(2.63)**	(2.88)**	(4.23)**
Turning point [95% confidence interval)	211.5 (160.8 , 278.1)	209.8 (162.4 , 270.9)	171.2 (112.9 , 260.3)	382.4 (308.0 , 474.8)	389.5 (313.1 , 484.4)	406.7 (336.9 , 491.3)
Observations	789	789	789	502	502	502

Notes:

(i) The top and bottom 5% observations in terms of standardised residuals are omitted in the restricted sample.

(ii) Absolute value of t-statistics in parentheses

(ii) * significant at 5%; ** significant at 1%

(iii) Confidence intervals for turning points are computed using the delta method.

Table 5: Measurement error corrected IV estimates of the quadratic Engel Curve parameters

	Addis Ababa		Other Urban Areas	
	Whole sample	**Restricted sample**	**Whole sample**	**Restricted sample**
Log of real total expenditure	0.281	0.509	0.435	0.780
	(2.36)*	(6.79)**	(4.59)**	(8.74)**
Square of log of real total expenditure	-0.024	-0.046	-0.035	-0.063
	(2.56)*	(7.45)**	(4.35)**	(8.52)**
Log of household size	0.019	0.028	-0.040	-0.063
	(1.64)	(2.81)**	(2.70)**	(8.52)**
Constant	-0.179	-.701	-0.600	-0.018
	(0.48)	(3.24)**	(2.11)*	(1.51)
Turning point [95% confidence interval]	351 [207,592]	248.1 [200.6 , 306.8]	501 [352,713]	468.3 [402.7 , 544.6]
p-value for overidentifying restrictions [$\chi^2(3)$]	.107	.198	.132	.211
Implied share of measurement error variance	31%	26%	37%	30%
Number of observations	879	789	558	502

Notes:

(i) The top and bottom 5% observations in terms of standardised residuals are omitted in the restricted sample.

(iii) Absolute value of t-statistics in parentheses

(ii) * significant at 5%; ** significant at 1%

(iii) Confidence intervals for turning points are computed using the delta method.

Table 6: Some Percentile Elasticities

Percentile	Addis Ababa		Other Urban Areas	
	OLS	**IV**	**OLS**	**IV**
5^{th}	.033 (.016)	.058 (.033)	.057 (.019)	.115 (.023)
10^{th}	.065 (.012	.036 (.025)	035 (.015)	.084 (.013
25^{th}	-.021 (.089)	.012 (.016)	.001 (.009)	.035 (.011)
50^{th}	-.052 (.007)	-.014 (.010)	-.026 (.009)	-.003 (.013)
75^{th}	-.089 (.074)	-.045 (.013)	-.052 (.012)	-.040 (.019)
95^{th}	-.138 (.013)	-.087 (.027)	-.091 (.019)	-.095 (.030)

N.B. Standard Errors are given in parentheses.

Figure 1: Nonparametric kernel fits with 95% pointwise confidence interval: Whole sample

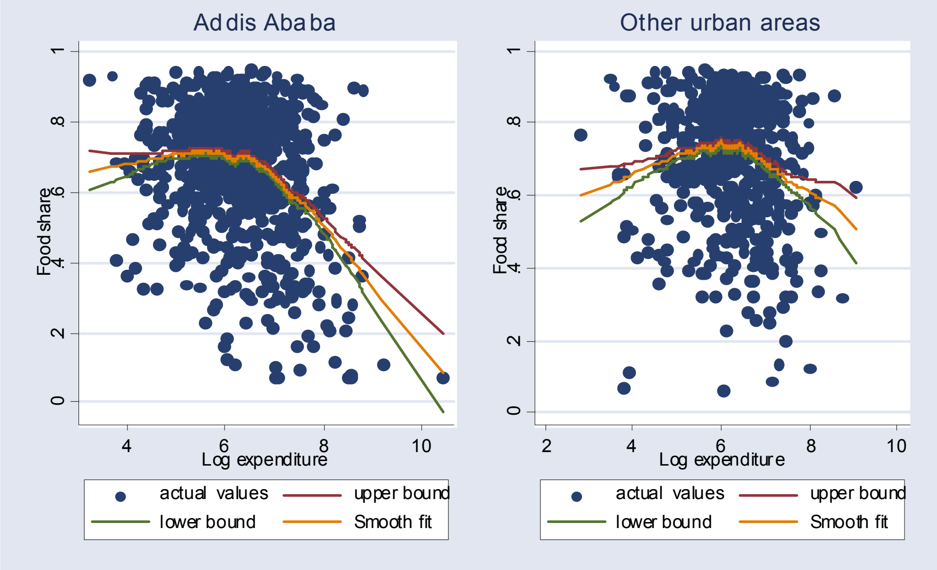

Figure 2: Nonparametric kernel fits with 95% pointwise confidence interval: Restricted sample

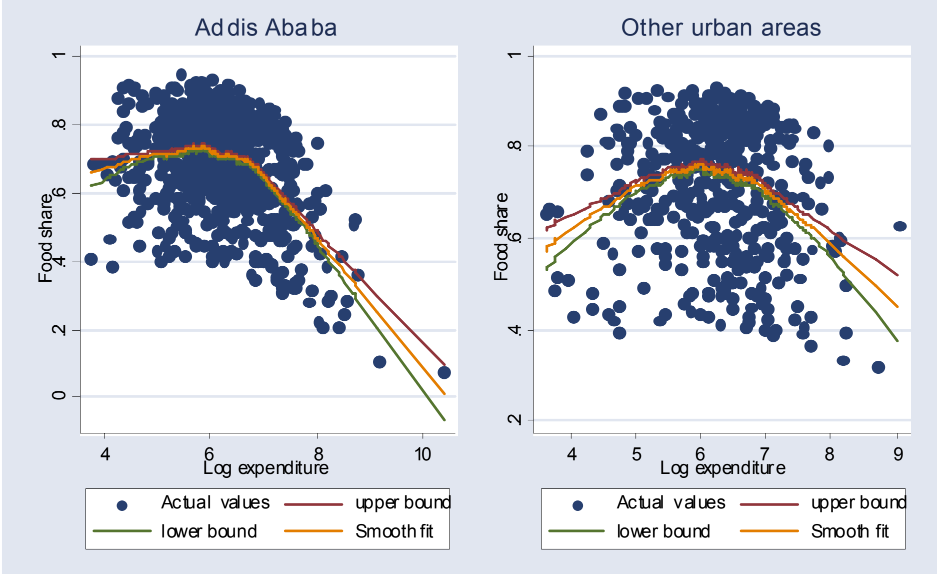

Figure 3: Semiparametric models of food share

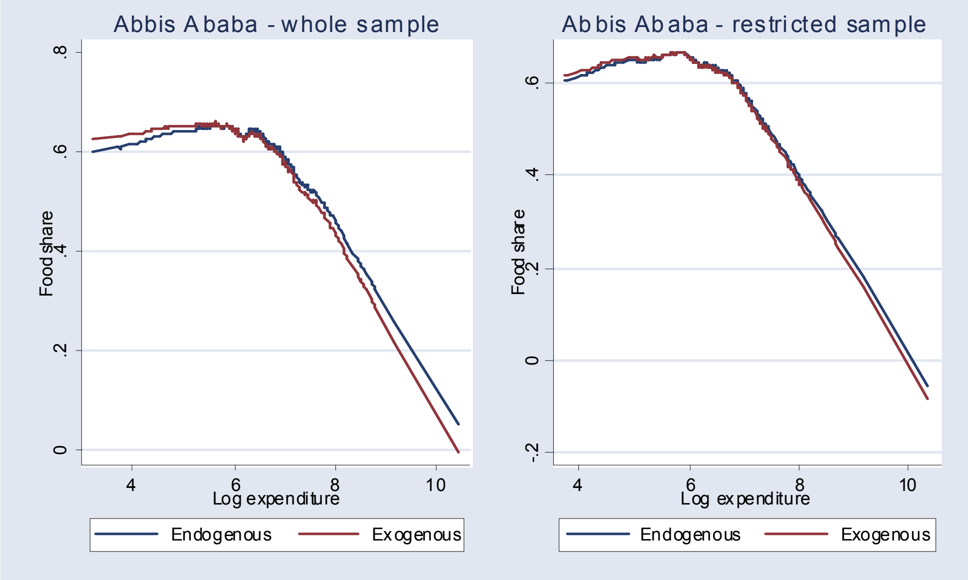

Figure 4: Semiparametric models of food share