APPENDIX I. BASELINE REPORT – MILLENNIUM ENGEL CURVES

Millennium Engel Curves

EXTERNAL IMPACT EVALUATION OF THE MILLENNIUM VILLAGES PROJECT, NORTHERN GHANA

Date: **February 2014**

Submitted by Itad
In association with:

itad
Results in development

IDS Institute of Development Studies

London School of Hygiene & Tropical Medicine

Participatory Development Associates Ltd.

Report

Millennium Engel Curves

Table of Contents

Tables

Figures

Introduction

The estimation of Engel curves is a powerful tool of economic analysis. Engel curves describe the relationship between household expenditure on particular items and total household expenditure. Of particular interest are food Engel curves, which describe how food expenditure increases as total expenditure increases. Studies of food Engel curves have discovered an empirical regularity – also known as Engel's Law – according to which food consumption increases less than proportionally with income, and food expenditure as a share of total expenditure decreases as income increases.

The observation of this regularity has a number of useful practical implications. First, Engel curves can be used in welfare analysis. Engel's Law states that the food share decreases as living standards increase and high food shares can therefore be used as indicators of poverty. Second, Engel curves can be used to predict how food expenditure will change as income changes. These predictions can be used to simulate the impact of policies on consumers and the local economy. Third, the regularity of Engel curves can be exploited to assess the quality of the data collected and the extent of measurement error. Expenditure figures are collected with considerable error by household survey and the regularity of theoretical Engel curves offer a standard against which to test the accuracy of the data.

The analysis of Engel curves is not a simple task. First, it is difficult to choose the right functional form. We would like the Engel curve to adhere to basic requirements of demand theory such as adding-up. Adding-up requires that consumption not increase more than income, which implies that the change in expenditure of all goods as income change cannot exceed the increase in income. We would also like Engel curves to be able to capture a variety of demand behaviours such as elasticities that vary with income. In some contexts food can be a luxury for the poor – consumption increases more than proportionally with income – and a necessity for the rich – consumption increases less than proportionally with income. The number of theoretical and behavioural requirements that we would like to impose is such that the number of valid functional forms for the estimation of Engel curves is limited. This has sometimes led researchers to abandon the search for the ideal functional form and to leave the identification of the curves to the data themselves using non-parametric methods – an approach which we also follow in this paper.

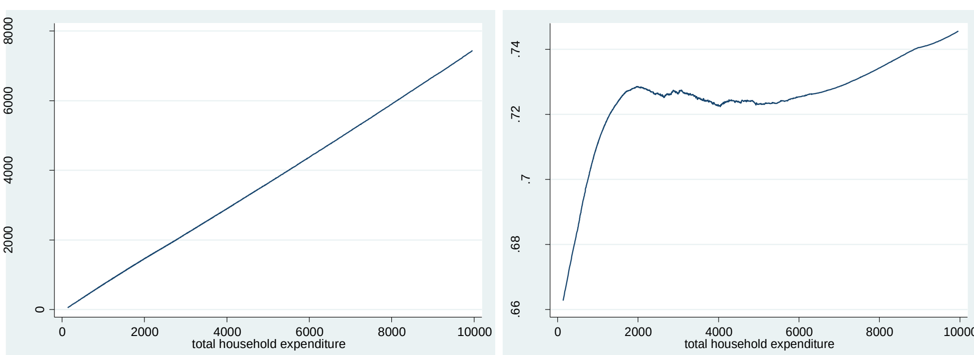
A second difficulty of Engel curves is that expenditures are often measured with error. The problem is particularly serious in datasets, such as ours, that collect information from very deprived households. A very large fraction of household expenditure (up to 80%) is devoted to the consumption of food and only one-third of food consumption consists of monetary expenditures. The fact that few households sell or purchase foods on the market implies that the quantities consumed have to be imputed using different sets of prices. In our survey we employ – in this order – prices at the household, community, and regional level. The widespread use of imputations in the calculation of expenditures adds an additional layer of error to standard reporting errors made by respondents and interviewers.

A third difficulty in the use of Engel curves is that policy simulations based on Engel curves are based on the assumption that changes in expenditure patterns observed in a cross-section are a good prediction of how expenditure patterns change for the same person *over time*. But the prediction that a person will increase food consumption as their income increases in the same way as we observe different levels of consumption for two people with different incomes is a very strong one. Expectations change in the distribution of income and persistence of consumption habits may complicate the analysis considerably. In other words, elasticities and other relationships estimated via cross-sectional Engel curve should be interpreted as *long-term responses* to income changes and policy simulations should take this into account.

Empirical food Engel curves

The analysis of expenditure data from household surveys normally finds the following patterns: First, food consumption increases less than proportionally with income and, in a non-linear fashion, at a decreasing rate. Second, the share of food expenditure on food consumption decreases with total expenditure, though it is sometimes found that the food share increases with total expenditure for very poor households. The Millennium Village data however seem to defy Engel's Law and show a very different pattern (Figure 1). Food expenditure increases linearly, though not proportionally, with total expenditure. The share of expenditure on food *increases* sharply with total household expenditure for the first tercile of the expenditure distribution and the pattern becomes unclear afterwards, but never substantially decreases as we would expect.

Figure 1. Empirical food Engel curves in the study area



Figures were obtained applying a simple non-parametric smoother (lowess) to the data.

It is difficult to provide a reasonable explanation for the patterns of Figure 1 without further investigation of the data. If we accept the data as they are and the Engel curves they generate in Figure 1, then we must conclude that food is a luxury for a large fraction of the population in the study area. As income increases the consumption of food increases linearly and the food share increases. This is a plausible explanation. It has been found that in very poor areas, for example Pakistan (Bhalotra and Attfield 1998) and Ethiopia (Kedir and Girma 2007), the income elasticity of consumption can be larger than one and therefore the food share increases with income at least over the lower section of the income distribution.

However, this interpretation cannot be accepted before probing for the presence of measurement error. Expenditure data are reported with poor accuracy in household surveys and there are many opportunities for both respondents and interviewers to make mistakes. It is also possible that reported expenditures are 'correct' in the sense that respondents do really mean what they say and that interviewers report the information correctly, and yet do not capture actual household consumption. For example, respondents might be reporting not the actual expenditure on food but the expenditure that appears to be socially acceptable (Udry and Woo 2007) or expenditures incurred by individuals that are not part of their 'survey household' (Guyer 2004) as in the case of funerals and other celebrations.

Below we investigate the presence of measurement error and estimate Engel curves after adjusting for the presence of error. The error considered is standard error in the reporting of expenditure figures. This type of error has serious consequences in the estimation of Engel curves particularly when estimation employs share form equations (Lewbel 1996) (the most popular type of which is discussed below). When Engel curves are estimated using the food share as the independent variable and total expenditure as the dependent variable, the same measurement error may appear on both sides of the equation – because food share equals food expenditure divided by total expenditure. The correlation between the errors in the measurement of food expenditure and total expenditure may create a spurious correlation between total expenditure and the food share. This problem could be particularly serious in our data because much of food expenditure is obtained by imputation and, given that food is a large fraction of total expenditure and composed of a limited number of items (mostly cereals), errors in the reporting of food expenditures end up being strongly correlated with errors in total expenditure.

Engel curves adjusted by measurement error

We estimate Engel curves using the most common functional forms employed in the literature: the Working-Leser form popularised by Deaton and Muellbauer (1980) (the 'almost ideal demand system') and the quadratic version of this model introduced by Banks et al. (1997) (the 'quadratic almost ideal demand system'). Because we are focusing on food expenditure the 'system' reduces to just one equation. Following Attanasio et al. (2011) we introduce price variation in the data by including dummy variables for each locality in which data were collected. The Working-Leser equation has the form:

$$w_i = a + b ln x_i + cX + e$$

Where *w* is the share of expenditure on food for household *i*, *x* is total household expenditure and *X* is a vector of village dummies and other covariates. Our sample of households is relatively homogeneous in many social characteristics and in order to simplify we only include two demographic variables among the covariates: the age of the household head and household size.

The quadratic form simply adds a quadratic term for the log of total household expenditure to the Working-Leser form above:

$$w_i = a + b ln x_i + c(ln x_i)^2 cX + e$$

We estimate these equations with and without adjusting for measurement error. We adjust for measurement error using the standard two-step simultaneous equation procedure. First, we find an exogenous variable that is well correlated with total household expenditure and is not correlated with measurement error in the dependent variable (food expenditure). Next, we use this variable as an instrument to correct the estimates of food share equations by two-stage least squares (using the *ivregress* command in stata). In the case of equation (1) we also experiment with the use of the Lewbel (1996) estimator (using the *ivreg2* command in stata). In the case of equation (2) the instrumentation of total household expenditure using two-stage least squares is more complicated because of the presence of the quadratic term in the log of total household expenditure. In this case we employ the standard two-step procedure of simultaneous equation models. We first run a regression of total expenditure on the selected instruments and then we employ the residuals and, following Attanasio et al. (2011), their square and cubes in the estimation of equation (2).

Table 1. Estimation of parametric Engel curves

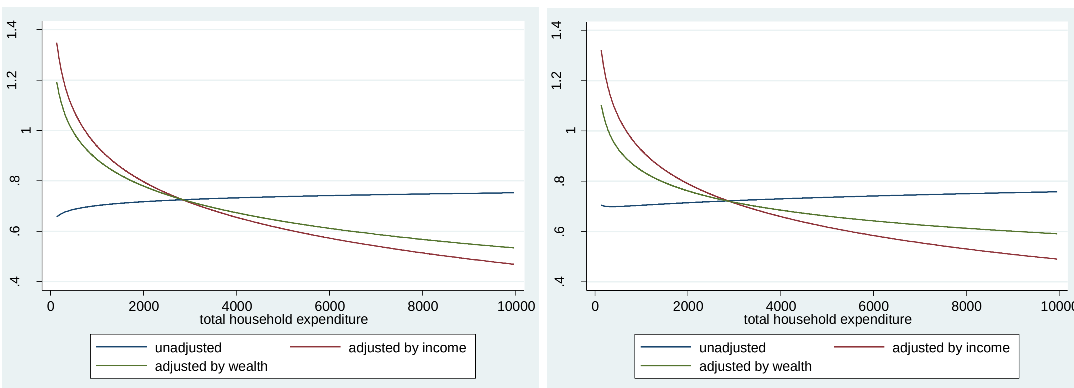
	Standard share form			Quadratic form		
	Unadjusted	Adjusted by income	Adjusted by wealth	Unadjusted	Adjusted by income	Adjusted by wealth
Log of expenditure	0.022**	-0.204**	-0.152***	-0.064	-0.222**	-0.183**
	0.010	0.086	0.028	0.081	0.106	0.077
Log of expenditure squares				0.005	0.002	0.005
				0.005	0.005	0.005
Elasticities						
1st quintile	1.031	0.79	0.83	0.89	0.76	0.77
2nd quintile	1.031	0.75	0.81	0.90	0.72	0.76
3rd quintile	1.030	0.72	0.79	0.91	0.69	0.75
4th quintile	1.030	0.69	0.77	0.92	0.66	0.74
5th quintile	1.029	0.61	0.74	0.93	0.60	0.72

Note: 1) All regressions include the following covariates: age of head of household, household size, and dummy variables for 103 localities. 2) Instruments of models 'adjusted by income' are income and the square of income. Instruments of models 'adjusted by wealth' are the value of the stock of wealth and its square. 3) The estimation of the adjusted standard share forms was performed using the ivregress command in stata. The estimation of the adjusted quadratic forms was performed running regressions of total household expenditure on the instruments, calculating the residuals and including the residuals, their squares and cubes in the second stage. 4) The calculation of elasticities was performed at the mean values of the *estimated* expenditure share for each quintile of the expenditure distribution. 5) The use of the Lewbel method to adjust for measurement error in the standard share form (ivreg2 command in stata) produced similar results: a coefficient of -0.130 with the income instrument and of -0.121 with the wealth instrument.

As usual, the main problem in instrumental variable estimation is finding a valid instrument. A standard choice in similar studies has been household income. We follow this practice by using income and income squared as instruments (because about 10% of households report negative incomes, the use of logs is not possible). In addition, we report estimates obtained using as instrument the stock of wealth held by each household. We calculate the stock of wealth as the value of all durables and production assets, including animals, owned by the households (but excluding land whose value was not reported by respondents). The stock of wealth has two advantages over income as an instrument for total expenditure. First, reporting of animal holdings and other assets occurs with much less error than reporting of income. Indeed income is normally reported with more error than expenditure and the relationship between expenditure and income is often a very weak one, thus leading to the potential problem of the 'weak instrument' (Angrist and Pischke 2009). The explanatory power of the stock of wealth is much higher. The R-square is 0.02 for a regression of total household expenditure on income and income squared while it is 0.12 for a regression of total household expenditure on the log of the stock of wealth. Second, wealth is a good approximation of long-term household income and well represents permanent income.

Table 1 reports the results of estimating the standard and the quadratic share form Engel equations with and without adjustment for measurement error. The adjustment for measurement error has the remarkable effect of reversing the sign of the relationship between food share and total expenditure. As appears in Figure 2, after adjusting for measurement error, food expenditure decreases as income increases (left chart). The quadratic form does not improve goodness of fit (right chart). In fact, the standard form appears to represent the data better than the quadratic form. The quadratic term in the log of household expenditure is never statistically significant and the shape described by the quadratic form is the same as the one described by the standard form. Income and wealth instruments appear to produce very similar results. Elasticities of the adjusted models are decreasing in total household expenditure: poorer households spend more on food as income increases than richer ones.

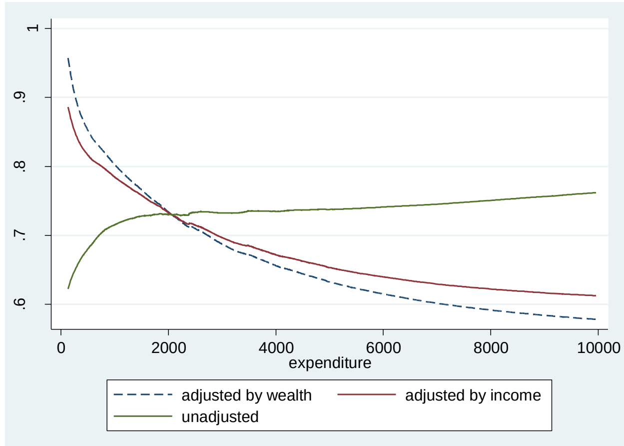
Figure 2. Engel curves corrected for measurement error (standard form on the left and quadratic form on the right)



The results of adjusting for measurement error are truly remarkable as they produce food Engel curves of the standard type that are decreasing in total expenditure. On the other hand, the use of a quadratic form fails to find non-linearities in the relationship between food expenditure and total expenditure. Poorer households have higher food expenditure elasticity but the value of the elasticity is always below one: food is a necessity for all households in the study.

Because we are considering a very deprived population we further investigate whether the food share increases as total expenditure increases, at least for some households, using non-parametric methods. Non-parametric methods obtain curves generated by the data without imposing an assumption or particular form. In order to control for measurement error we need to employ a semi-parametric model. Semi-parametric regressions of food share over total food expenditure adjusted for measurement error produce curves that are very similar to those produced by the standard share form (Figure 3).

Figure 3. Semi-parametric Engel curves (adjusted and unadjusted)



Note. The curves were obtained by semi-parametric regression using the simple method of differencing described by Johnston and DiNardo (1997) and Yatchew (2003). First, the observations in the sample are sorted in ascending order of the logarithm of per capita expenditure. Second, all variables are first differenced using optimum weights. Third, differenced food shares are regressed on all differenced covariates (age of household head, household size, locality dummies and first, second and third order residuals from a regression of the log of expenditure on the instruments). Fourth, the estimated coefficients obtained from this regression are then multiplied by the original values of the explanatory variables, and their product is subtracted from the non-differenced food share. Finally, the food share so adjusted is regressed non-parametrically on total household expenditure using locally weighted regressions (lowess command in stata).

Conclusions

The empirical analysis of this paper finds that food Engel curves in the study area have the usual shape, whereby the share of household food expenditure decreases as total household expenditure increases. The opposite relationship observed in the data – an increasing food share as expenditure increases – appears to be the result of measurement error. The reasons for such large errors in the data are not explored but are likely due to a combination of reporting errors by respondents and the imputation of quantities of food produced by households for their own consumption. One concrete possibility is that errors in reporting and imputing some expenditure categories like cereals have the effect of generating a spurious correlation between food expenditure and total expenditure in the case of positive errors (overestimation of expenditures). In the case of negative errors (underestimation of expenditures) the underreporting of any specific expenditure (for example maize) does not generate the same spurious correlation between food expenditure and total expenditure, which could be the result of generalised underreporting of food expenditure by some households. This and other issues would require further investigation.

The main conclusions of this paper are the following:

- Food Engel curves in the study area have the usual shape: the share of food expenditure decreases as total household expenditure increases.
- Food is a necessity for all households (elasticity <1) and there is no sign that food could be a luxury for very poor households.

- Elasticities are decreasing with total household expenditure: poor households will spend more on food as their income increases.
- The estimated food shares and elasticities can be used to perform simulations of the impact of income changes on expenditure patterns and prices, but it should be kept in mind that the elasticities obtained in this way should be interpreted as long-term elasticities and are probably inadequate to simulate changes occurring in the short run.
- The widespread measurement error in expenditure data casts some doubts on the validity of poverty estimates based on consumption data, but the size and the direction of the bias are not investigated.

References

Angrist, J. D., and J-S. Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

Attanasio, O., E. Battistin, and A. Mesnard. 2011. Food and Cash Transfers: Evidence from Colombia. *The Economic Journal* 122 (March):92-124.

Banks, J., R. Blundell, and A. Lewbel. 1997. Quadratic engel curves and consumer demand. *Review of Economics and Statistics* 79 (4):527-539.

Bhalotra, S., and C. Attfield. 1998. Intrahousehold resource allocation in rural Pakistan: A semiparametric analysis. *Journal of Applied Econometrics* 13 (5):463-480.

Deaton, A., and J. Muellbauer. 1980. *Economics and consumer behavior*. Cambridge: Cambridge Univerity Press.

Guyer, J. I. 2004. *Marginal Gains: Monetary transactions in Atlantic Africa*. Chicago: The University of Chicago Press.

Johnston, J., and J. Dinardo. 1997. *Econometric Methods*. New York, NY: McGraw-Hill.

Kedir, A., and S. Girma. 2007. Quadratic engel curves with measurement error: Evidence from a budget survey. *Oxford Bulletin of Economics and Statistics* 69 (1):123-138.

Lewbel, A. 1996. Demand Estimation with Expenditure Measurement Errors on the Left and the Right Hand Side. *The Review of Economics and Statistics* 78 (4):718-25.

Udry, C., and H. Woo. 2007. Households and the Social Organisation of Consumption in Southern Ghana. *African Studies Review* 50 (2):139-153.

Yatchew, A. 2003. *Semiparametric Regression for the Applied Econometrician*. Cambridge: Cambridge University Press.