KIEL WORKING PAPER

The nutrient-income elasticity in ultra-poor households: Evidence from Kenya

No. 2114 October 2018

Hamidou Jawara and Rainer Thiele

Kiel Institute for the World Economy
ISSN 2195–7525

The nutrient-income elasticity in ultra-poor households: Evidence from Kenya

By

Hamidou Jawara[1] and Rainer Thiele[2]

Abstract

The relationship between nutrient intake and wealth of poor households continues to be an issue of huge policy relevance. In this paper, we contribute to the ongoing debate on the nutrient-income elasticity using a sample of ultra-poor households with orphans and vulnerable children (OVC) in Kenya. To estimate the nutrient-income elasticity for these households, we employ panel data techniques that enable us to tackle measurement error and simultaneity bias. In addition, we use semi-parametric panel data models to address nonlinearities. For most of the nutrients considered, we find that income elasticities are significantly different from zero but below unity. Caloric intakes turn out to be less income-inelastic than macro and micro nutrient intakes.

Keywords: Nutrient-income elasticity; OVC households; Kenya

JEL *classification:* C14, D12, Q11

[1] PhD candidate in Quantitative Economics, University of Kiel
Email: hamido.jawara@economics.uni-kiel.de
[2] Head of Research Area "Poverty Reduction , Equity, and Development", Kiel Institute for the World Economy, and Adjunct Professor, University of Kiel
Email: rainer.thiele@ifw-kiel.de

1. Introduction

Motivated by the assertion that hunger and poverty tend to be synonymous (Banerjee and Duflo, 2011), i.e. the prevalence of hunger and malnutrition is associated with poor economic status, a significant amount of mostly empirical literature on the relationship between nutritional status and income or total expenditure has emerged over the last decades (Ogundari and Abdulai, 2013).

Theoretically, two explanations can be given on why nutrition might be related to income or expenditure. The first is the efficiency-wage hypothesis (Leibenstein, 1957; Mirrlees, 1957; Stiglitz, 1976), according to which employers reward labor based on productivity and the latter is determined by nutritional status. Thus, unemployment and therefore poverty exists because some people do not have enough to eat (Subramanian and Deaton, 1996). The second explanation, which has dominated much of the academic literature, is that nutrition status is determined by income and food demand (Subramanian and Deaton, 1996). Accordingly, nutritional problems such as malnutrition, which are assumed to be associated with underdevelopment, would be attenuated by economic prosperity (Abdulai and Aubert, 2004). This latter argument has motivated a long-standing debate on whether what Subramanian and Deaton (1996) called a "calorie Engel curve" exists or not for the poor. A calorie Engel curve would imply that the poor switch from poor-quality food but less expensive calories to more expensive calories but higher-quality food as income increases (Skoufias et al., 2011). The main behavioral parameter of interest in these debates is the calorie-income elasticity (Gibson and Rozelle, 2002).

A number of empirical studies have estimated the calorie-income elasticity, but the evidence is still mixed (Santeramo and Shabnam, 2015). While some authors (such as Behrman and Deolalikar, 1987; Bouis and Haddad, 1992; Skoufias et al., 2009) find that the responsiveness of calories to income changes is not significantly different from zero, others (e.g. Subramanian and Deaton, 1996; Abdulai and Aubert, 2004; Aromolaran, 2004; Skoufias et al., 2011) obtain a positive and statistically significant calorie-income elasticity. This inconclusiveness has important policy implications: If it is true that nutrient intake does not respond to changes in income, then common interventions such as cash transfer programs will not suffice to eliminate malnutrition and food insecurity.

Against this background, the objective of this paper is to provide new evidence on the nutrient-income elasticity in a setting of very poor households in Sub-Saharan Africa. Most of the evidence so far is from Asian and Latin American countries and does not cover the very poorest households. By contrast, we consider ultra-poor households with orphans and vulnerable children (OVCs) in Kenya. The size of this group is non-negligible: It is estimated that there about 2.6 million OVCs in Kenya and 12% of all households have at least one OVC (Lee et al., 2014). Lee et al. (2014) estimate that more than half of OVC households are in the lowest two quintiles of the wealth distribution in Kenya and 22% of them have recently experienced moderate to severe food shortages. Studying whether income gains may help extremely poor households in Kenya improve their nutritional status is thus highly policy relevant, especially at a time when social programs to tackle food and nutrition problems are ubiquitous in the country.

Another notable feature of our study is that in addition to calorie elasticities it also provides evidence on the elasticity of other (macro and micro) nutrient intakes to changes in income. Most of the previous studies on the relationship between dietary behavior and changes in income have focused on caloric intake, but calories are not the only important component of human diets. As highlighted by Skoufias et al. (2009), a positive relationship between energy intake and income does not necessarily translate into a positive relationship between nutrient intake and income and vice versa. Households may for example use additional income to switch to more nutrient-rich food with the same calorie content. It is therefore important to also investigate the impact of income on the intake of macro and micro nutrients. The macro nutrients we study are protein, fat, and fiber and the micronutrients are vitamins (A, D, and Folate) and minerals (zinc, iron, and calcium). There is evidence that higher intake of these nutrients is associated with better health via a reduction in malnutrition and other nutrient-deficiency-related health problems such as anemia.

Furthermore, we contribute to the methodological debate on the issue of non-linearities that arise if nutrition elasticities differ between income groups. In particular, the calorie-income elasticity may be higher for poorer household, who have insufficient food to eat, than for richer households. This issue was first raised by Strauss and Thomas (1990), who found for the case of Brazil that indeed the calorie-expenditure elasticity is higher for poorer households. They concluded that this non-linearity cannot be adequately captured in a fully linear parametric model. Following their study, nonparametric techniques were used to estimate the calorie-income elasticity (e.g. Subramanian and Deaton, 1996; Skoufias, 2002; Abdulai and Aubert, 2004; Skoufias et al., 2011). Nonparametric models are desirable because they allow for curvature without imposing any functional form on the data (DiNardo and Tobias, 2001). Yet, they have a major drawback in that they prevent the inclusion of a large set of control variables. Partial-linear models that allow for the inclusion of variables in both parametric and nonparametric fashion can be considered as a good compromise between fully parametric and nonparametric models (Rodriguez-Poo and Soberon, 2017). In our case, this is desirable because it allows for a flexible characterization of the nature of the relationship between wealth and nutrients intake while at the same time allowing for other covariates to be controlled for in a parametric fashion. Specifically, we employ the partial-linear panel data model suggested by Baltagi and Li (2002). To the best of our knowledge, Tian and Yu (2015) is the only previous study using this approach in a case study for China.

In addition to dealing with non-linearity, we also address the problems of measurement error and simultaneity bias. In fact, meta-analyses by Ogundari and Abdulai (2013) and Santeramo and Shabnam (2015) have pointed to differences in dealing with these two biases as key factors behind the heterogeneity in estimated nutrient-income elasticities. We use the approach developed by Lewbel (2012), recently extended to panel data by Meijer et al. (2017), to address the measurement error problem, and standard panel IV estimation to tackle the problem of simultaneity bias.

The remainder of the paper is organized as follows: In section 2, we discuss the data used in the empirical analysis and provide some descriptive statistics. Section 3 introduces the estimation methodology, while Section 4 discusses the results. Section 5 summarizes the main findings and concludes.

2. Data and descriptive statistics

The data we use was collected for the evaluation of the Kenya cash transfer program for orphans and vulnerable children (Kenya CT-OVC in short). The Kenya CT-OVC is a safety net program run by the government with the main objective of offering social protection to ultra-poor households with orphans and vulnerable children. It takes the form of a regular monthly income transfer of initially Ksh 1500 (about $20) per household, which has been increased over time to capture price changes. An OVC household is defined as a household that satisfies one of the following three criteria: there exists at least one single or double orphan; the primary caregiver is chronically ill; or the head is chronically ill (Handa et al., 2016). OVC status plus other variables such as education level, asset ownership, access to clean water and sanitation were the main indicators used in the program targeting (see Handa et al., 2012, for details).

The program was piloted by the government of Kenya and UNICEF in 2004 and since then has reached over 350,000 beneficiaries. Followings its success in the piloting phase, it was adopted into the national budget in 2007 as a government flagship social protection program and consequently there was an agreement for it to be expanded to other regions of the country. Prior to this expansion an evaluation of the program was commissioned by UNICEF and contracted to a private consulting firm, the (OPM) Oxford Policy Management (The Kenya CT-OVC Evaluation Team, 2012). Three rounds of surveys were carried out between 2007 and 2011: a baseline survey in 2007 and two follow-up surveys in 2009 and 2011. The number of eligible households surveyed in the baseline was 2,294 and there was, respectively, a 17% and 5% attrition between baseline and first follow-up in 2009 and between first follow-up and second follow-up in 2011 (Handa et al., 2016).

Since one of the objectives of the program was to improve food security in OVC households, information on household food consumption was collected alongside other household indicators on health, education, and child welfare. In this study, we use data on household characteristics, village-level characteristics, and food consumption to study the calorie or nutrient income elasticity for the households. Specifically, we use age of the household head, gender of the household head, education status of the household head (measured as the number of adults in a household with at most 8 years of education), household size, household total expenditure (measured as the sum of food and nonfood expenditure), and different household demographic ratios (such as proportion of children under 5, under 14, under 16, proportion of adults, and proportion of old people) as household characteristics. Distance to the nearest market and access to a road network serve as village level characteristics.

In collecting the food consumption data, households were requested to make a recall of 29 food items that they consumed in the last seven days prior to the day of their interview. For each food item, information on total outlays, the main source as well as the quantity and unit consumed were collected. In addition, local market prices of 19 food items were collected using a questionnaire that was administered at the community level. We use this information to construct food availability in a household, from which per-capita caloric and nutrient intake from 13 food items (maize flour, Irish potato, beans, bananas, kale, beef with bones, dried fish, eggs, milk, cooking oil, sugar, salt, and tea leaves) consumed by the household is computed. In doing so, we first divide household monthly total expenditure on each food item by the minimum locality price per unit (kilogram or liters) of that food item, where minimum prices are computed from the report of local prices at the community level. This allows us to obtain quantities of the different food items consumed by each household.[3] We consider food from all sources: purchased, own production, and received as gifts. Food from own production and received as gifts are valued at the prevailing local market price. Since the information available is not enough to compute actual consumption of food by each household as wastage and leftovers are not accounted for, we only capture food availability, which could result in an overestimation of our calorie- or nutrient-wealth elasticities. Second, we convert the unit of measurement of each food item from kilogram and liters to grams and milliliters. Third, to derive total calorie or nutrient intakes from a food item, we multiply the total number of grams or milliliters of household consumption of that food item by the number of kilocalories or nutrient available in a 100 grams edible portion. We use the food composition table of Tanzania by Lukmanji et al. (2008) for the conversions. The total caloric or nutrient intake from a food item can then be summarized as

$$C = \sum_i q_i c_i \tag{1}$$

where q_i is the total number of grams or milliliters of food i(i=1,2,3,...,13) consumed by a household and c_i is the calorie or nutrient equivalent unit of food item i . Finally, we compute per-capita calorie or nutrient intake by dividing total caloric or nutrient intake (given as C_j) of each household by the adult equivalent unit of that household using the adult equivalent scales from Anzagi and Bernard (1977).

For the analysis below, we categorize the food items into food groups. The total number of calories from a food group is given as the sum of per-capita calories from all the items in that food group, i.e. for food group g with j food items we obtain

$$C_g = \sum_j C_j \tag{2}$$

We define nine food groups: cereal, tuber, pulse, fruit, vegetable, meat, milk, oil, and sugar. To minimize the effect of outliers, we winsorize the values of all the measures computed at their 2*nd* and 98*th* percentile; that is all values below and above the 2*nd* and 98*th* percentile are set to these percentiles. Given that the design of the evaluation was longitudinal, the data collected for the three periods can be used to construct a panel dataset on households' food consumption. To ensure that we are dealing with a balance panel, we focus on

[3] Even though some of the questionnaires collected information on the exact quantities that were consumed and the units involved, this information was not collected in all survey rounds. Thus, for the sake of consistency, we rely on the above procedure to derive quantities consumed.

households that were surveyed in all three rounds; households surveyed in the baseline but not surveyed in the follow-ups are dropped from the analysis. Pooled over the three rounds, this gives a total of 5346 observations. Note that the evaluation of the program targeted only poor rural households. Our sample is therefore not representative of all households in Kenya, but represents only those with characteristics similar to the targeted group.

Table 1 reports summary statistics for the household and village-level characteristics. Each statistic is shown for OVC households in the bottom and top quartile of the wealth distribution (given in terms of per-capita household total expenditure), as well as for all the OVC households in the sample and for each period. It can be seen that the average age of OVC household heads is about 61 years in the baseline (2007) and a bit lower in the subsequent years. As expected, poorer OVC households are much bigger in size than their richer counterparts. However, in terms of the composition of the different age groups in the households, the two samples do not differ significantly. Household per-capita expenditures increase over time for all households. In the overall sample, there was a rise of about 75% and 25%, respectively, in 2009 and 2011, which may at least partly be due to the positive income shock created by the cash transfer. Increases are more pronounced for richer than poorer OVC households; in fact, between 2009 and 2011, the poor OVC households didn't experience any rise in per-capita expenditure. Education levels also improve over time for both poor and rich OVC households as well as in the overall sample. The price of calories, given as the ratio of total expenditure to total calories of all food items, is used to measure food quality in the two groups. In general, food quality improves over time for all households, but OVC households in the top quartile invest more in quality food than those in the bottom quartile. The village-level characteristics (distance to market and access to road) do not differ significantly across the two groups of households.

Summary statistics on nutrient intakes are reported in Table 2. Average per-capita energy intake by OVC households in the bottom quartile is generally much lower than energy intakes by OVC households in the top quartile of the per-capita consumption distribution. A similar pattern is observed for both the intake of macro nutrients (protein, fat, and fiber) and micro nutrients (vitamins, minerals, and amino acids). For example, in the case of a mineral like iron, a deficiency of which is a cause for anemia that remains a burden for many developing countries, there was a difference in favor of the top quartile of about 133% in 2007, 141% in 2009 and 215% in 2011. This evidence points to non-linearities in per-capita nutrients intake among OVC households, which we explore below using a partial linear panel data model.

Table 3 reports descriptive statistics on the share of calories attributable to the food groups, which is an indicator of food energy sources of the household. Both poor and rich OVC households receive a large share of their calories from cheap sources like maize flour, indicating their ultra-poor status, but the proportion of calories from cereal is higher for the former group. Healthy components of diets, such as fruits and vegetables, do not seem to be important sources of calories for these households; still, they are more important for top quartile than for bottom quartile OVC households. Oil and sugar, which are associated with dietary diseases like Diabetes and heart-related diseases, are more important sources of calories for both poor and rich OVC households than fruits and vegetables.

Given the results from above, it is perhaps not surprising that a significant chunk of the OVC household's food budget - e.g. about 34% in 2011 in the overall sample - is on average spent on cereals (see Table 4). After cereals, oil, sugar, and vegetables are the food groups that account for the highest portions of food budgets. However, they all decreased in importance over time, in particular for poorer OVC households.

The overall picture emerging from the descriptive statistics is that nutrient intakes, calorie shares, and food budget shares partly vary over time and across households with different wealth status. We will explore these observations in more detail in the econometric analysis below.

3. Methodology

As highlighted above, three methodological issues need to be addressed when estimating the calorie or nutrient-expenditure elasticity: Measurement error, simultaneity bias, and non-linearity in the relationship between the two variables. We employ the heteroskedasticity-restriction IV estimation proposed by Lewbel (2012) and recently extended to a panel data framework by Meijer et al. (2017) to address the endogeneity of total expenditures due to measurement error; standard panel IV methods to deal with simultaneity bias; and the partial linear panel data approach of Baltagi and Li (2002) to address non-linearity in the nutrient-wealth relationship.

3.1. Measurement Error and Endogeneity

We use total per-capita household expenditure on food and non-food items as a measure of wealth in OVC households. Total expenditure is a better proxy of wealth than total income as it comes closer to measuring permanent income. Its main drawback is that it may suffer from measurement error, e.g. because total expenditure does contain items that are not regularly purchased (Meghir and Robin, 1992). When total expenditure is combined with caloric intake or availability in a regression model this may lead to the common measurement error problem, which arises if total expenditure and caloric intake are derived from the same data source and as a result measurement error in one could correlate with measurement error in the other. This phenomenon is likely to occur in survey data where tendencies for error affect series of self-reported variables (De Nadai and Lewbel, 2016). Bouis and Haddad (1992) were the first to indicate that the common measurement error leads to a positive bias in estimated calorie-income elasticities. Thus, it is different from the classical error-in-variables problem that biases estimated coefficients towards zero, commonly known as the attenuation bias. Griliches and Hausman (1986) have shown that in the presence of measurement error the use of panel data techniques such as the within or between estimator will likely exacerbate such bias.

3.1.1. Conceptual Framework

The model we estimate can be described as:

$$Y_{nt} = X_{nt}\beta + S'_{nt}\delta + \alpha_n + \varepsilon_{nt} \tag{3}$$

where Y_{nt} denotes the logarithm of per-capita calorie or nutrient intake by household *n* at time *t*; X_{nt} is the logarithm of per-capita expenditure by household *n* at time *t*; S_{nt} is a vector of exogenous covariates (both household characteristics such as household size, age, and demographic ratios, and village-level indicators including access to roads and markets) for household *n* at time *t*, which are assumed to be measured free of error; α_n denotes household fixed effects; and ε_{nt} is the error term, assumed to have a zero mean. X_{nt} cannot be observed as it is measured with error. Instead what we observe is X^*_{nt} with error v_{nt}. The relationship between X_{nt} and X^*_{nt} is given as:

$$X^*_{nt} = X_{nt} + v_{nt} \tag{4}$$

Assuming that the measurement error v_{nt} is independent of X_{nt}, we are dealing with a case of classical measurement error. The reduced-form model from above is then given as:

$$Y_{nt} = X^*_{nt}\beta + S'_{nt}\delta + \alpha_n + \mu_{nt} \tag{5}$$

where $\mu_{nt} = X^*_{nt} + v'_{nt}\beta$. It is obvious that X^*_{nt} is endogenous (correlated with v_{nt}) even if X_{nt} itself is not correlated with v_{nt}. A standard way to address this endogeneity problem is to use instrumental variable methods where the endogenous regressor is instrumented with a variable such as income that is assumed to be exogenous (like in Gibson and Rozelle, 2002). In our case, such instruments are not available. But even if standard external instruments are hard to come by, one can use internally generated instruments to identify the model. This was first highlighted by Griliches and Hausman (1986), who showed that series of error-in-variables models can be identified in a panel data framework without the need for external instruments. Meijer et al. (2017) suggest three approaches that rely on GMM estimated moment-based conditions with different assumptions: restrictions on the intertemporal matrix of covariance errors of the model, third-moment restrictions, and heteroskedasticity restrictions. In this paper, we apply the last method, following Lewbel (2012). Meijer et al. (2017) have provided a general account of how such an approach could be implemented in a panel-data model. In particular, they have shown that information from exogenous variables in different time periods as well as their relationship with the endogenous variable can be used to identify mismeasured regressors. Next we briefly discuss this approach. For details regarding the method, see Meijer et al. (2017).

Suppose S_n contains exogenous – i.e. not mismeasured – variables that can be excluded from the equation for Y_n but not from the equation for X_n; then we can use the relationship between these variables and X_n to identify the structural model. Meijer et al. (2017) have shown that this can be done when the relationship between S_n and X_n is heteroskedastic, an observation first made by Lewbel (2012) for a cross-sectional data context. Suppose a linear relationship between X_n and S_n is as follows:

$$X_n = S_n\kappa + w_n \tag{6}$$

If q_n equals $v_n + w_n$, this implies from equation 4 that $X^*_n = S_n\kappa + q_n$. Given is a restriction of the form $E\left(w_n w'_n \middle| S_n\right) \neq 0$ that can be made stronger by assuming that $E\left(S_n \otimes w_n \otimes w_n\right) \neq 0$. If we assume further that all the variables are centered based on time, the latter condition is an indication that X_{nt} and S_{nt} form a heteroskedastic relationship. To identify β and δ in equation 3, we can use this heteroskedasticity condition together with the assumption that v_{nt} and ε_{nt} are independent from X_{nt} and S_{nt} as well as the exogeneity condition for S_{nt}, which is given as $E\left(S_{ns}\mu_{nt}\right) = 0$ ∀ *s* and *t*. Now suppose that $Z'_n\mu_n$ is a moment condition for identification involving q_n; then, if q_n is observable, we obtain

$Z_n = S_n \otimes (1, q'_n) \otimes I_T$. Yet, q_n is unobservable but can be replaced by $\hat{q}_{nt}$ that is estimated from the regression of X^*_n on S_n. $\hat{q}_{nt}$ is a valid instrument to identify β and δ. Hence, in our case of mismeasured *m*, Z_n is a generated instrument and can be used to consistently identify the model.

We follow Lewbel (2012) and include the number of assets owned by the household, mean community nonfood expenditure, age of the household head, education level of the household head, gender of the household head, and household size as variables in $S_n(s_{in}; i = 1,2,...,5)$ that can be used to generate the internal instruments. For these variables to be valid for identification, their relationship with X_{nt} should be heteroskedastic.

In line with Meijer et al. (2017), we proceed as follows to generate the instruments: First, for each period $t = 1,2,3$, we regress x_{nt} on s_{int} (i.e. $s_{in1}; s_{in2}; s_{in3}$) and the other exogenous variables in s_n. In this step, we also check for heteroskedasticity. Second, for each of the three regressions in step one, we generate the residual $\hat{q}_{nt}$. Third, we generate the internal instruments Z_{it} (i.e. $z_{i1}; z_{i2}; z_{i3}$) from $\hat{s}_{int} q_{nt}$. This leaves us with a total of fifteen generated instruments. Finally, we perform a panel IV estimation with the generated instruments. Note that when the data is transformed into a panel structure, the dimension of the instruments reduces from fifteen to five. Our panel IV model is thus over-identified. We apply the Sargan-Hansen test to check for over-identification as an indication of the validity of the instruments.

Apart from measurement error, another problem that might affect the identification of the elasticity parameter β is endogeneity due to simultaneity bias. Specifically, it might be the case that the direction of causation between caloric or nutrient intake and household wealth occurs in both directions, i.e. higher intake of calories or nutrients leads to more productivity due to better health and this in turn can cause higher wealth (Skoufias et al., 2011). To address this problem, we follow Skoufias et al. (2011) and instrument total expenditure with mean locality household nonfood expenditure via a standard panel data instrumental variable technique.

3.2. Non-Linearity

As discussed above, previous research has shown that nonlinearities in the relationship between calories or nutrients and income might also affect the elasticity parameters of interest. To address this issue, we use a semi-parametric panel data model. Among the different classes of semi-parametric models suggested in the literature (see Yatchew, 1998), we choose Baltagi and Li's (2002) variant of the partial-linear model. This model allows us to capture nonlinearity in the relationship between nutrient or calories and income while at the same time making it possible to parametrically include other variables than income that might affect caloric or nutrient intake.

3.2.1. Conceptual framework

The model we estimate can be specified as:

$$y_{nt} = x_{nt}'\gamma + g\left(m_{nt}\right) + u_{nt} \tag{7}$$

where x_{nt} denotes the energy, macro-nutrient or micro-nutrient index variable for household n in year t. x_{nt} represents a vector of other exogenous time-variant characteristics of household n in period t. $g\left(m_{nt}\right)$ is a non-parametric function of per-capita household wealth for household n in period t; it captures possible non-linearities in the relationship between Y_{nt} and m_{nt}. u_{nt} is a one-way error component disturbance given as $u_{nt} = \mu_n + v_{nt}$. μ_n denotes household time-invariant fixed effects and v_{nt} is a random component that is assumed to follow a normal distribution.

To estimate the model, we first perform first-differencing to eliminate the household fixed effects μ_n. We then have

$$Y_{nt} = X_{nt}\gamma + G\left(m_{nt}, m_{nt-1}\right) + U_{nt} \tag{8}$$

where $Y_{nt} = y_{nt} - y_{nt-1}$ $X_{nt} = x_{nt} - x_{nt-1}$, $G\left(m_{nt}, m_{nt-1}\right) = g\left(m_{nt} - m_{nt-1}\right)$, and $U_{nt} = u_{nt} - u_{nt-1}$. In estimating γ, we follow Baltagi and Li (2002) and use a series-based method. This method avoids two shortcomings that are associated with the alternative kernel-based method suggested by Li and Stengos (1996): (1) the fact that first-differencing increases the dimension of the problem (curse-of-dimensionality problem); and (2) the fact that it cannot be used to obtain the original non-parametric function of interest $g\left(m_{nt}\right)$. We proceed by using a series of the form $p^{k(m)}$ with dimension *K*x1 to approximate $g(m)$. If $p^{k(m)}$ can approximate any function in $g \in \mathcal{G}$ as K grows, then $p^{k(m)}$ can approximate $g(m)$ and $p^k\left(m_{nt} - m_{nt-1}\right) = p^k\left(m_{nt}\right) - p^k\left(m_{nt-1}\right)$ can approximate $G(m_{nt}, m_{nt-1})$, where $p^{k(m)}$ are the first k terms of a sequence of functions $(p_{1(m)}, p_{2(m)}, \ldots\ldots)$. Hence, equation 5 can be re-written as

$$y_{nt} - y_{nt-1} = \left(x_{nt} - x_{nt-1}\right)\gamma + \left(p^k\left(m_{nt}\right) - p^k\left(m_{nt-1}\right)\right)\Theta + u_{nt} - u_{nt-1} \quad (9)$$

OLS estimates of γ and Θ from equation 9 are consistent. Having estimated Θ and γ, we follow Libois and Verardi (2013) and use equation 7 to fit the fixed effects term of the model, get an estimate of the disturbance term residual, $\hat{u}_{nt}$, and then estimate the non-parametric term using standard nonparametric regression techniques. An important consideration in estimating the model is to select the right series for p^k. p^k can be considered a spline. We use a particular kind of splines called B-splines that avoids the problem of high correlation between successive terms associated with linear splines. For details regarding this procedure, see Newson (2000). As a robustness check, we also employed a kernel-weighted local polynomial technique to estimate $g\left(m_{nt}\right)$.

It is important to note that, following our discussion in section 3.1, m_{nt} might also be endogenous. If this is due to the common measurement error problem, it will make the elasticity parameter upward-biased, whereas if it is due to classical measurement error this will lead to attenuation bias. In the above specification, we only address the bias in the elasticity parameter that is due to possible nonlinearity in Y_{nt} and m_{nt}. This caveat should be kept in mind when interpreting the results.

4. Results

Our findings from the parametric estimations are reported in Table 5. As a baseline, column 1 shows the results from a fixed effects model that is estimated using the within estimator, i.e. we assume that individual specific unobserved effects are not independent of observed covariates.[4] Extensions of the baseline model that account for simultaneity bias and measurement error as discussed in Section 3.1 are reported, respectively, in columns 2 and 3 of Table 5. While the IV model for addressing simultaneity bias is exactly identified, the IV model for addressing measurement error is over-identified. Hence, in the latter case, we checked for over-identifying restrictions - that is the validity of instruments - using the Hansen-Sargan test; the p-values from the test are also reported in the table. Each of the nutrient-elasticity estimates shown in Table 5 were obtained controlling for other observable

[4]We checked for the validity of the assumption by performing a Hausman test that compares this estimator with the FGLS estimator of an error component model. The results, not reported, indicate that non-randomness of the unobserved effect cannot be rejected; hence, the fixed-effect model is the preferred specification.

characteristics at the household and village level as discussed in Section 2. The results for the control variables are not reported here but are available from the authors upon request.

The results from the fixed-effects estimation show that the calorie total expenditure elasticity in the OVC households is about 0.49, and highly statistically significant. This implies that for every percentage increase in wealth - expressed here in terms of total per capita expenditure - these households increase their intake of calories by about 49 percentage points. If we control for a possible simultaneity bias, we obtain a lower calorie-expenditure elasticity of about 0.39. It decreases further to about 0.19 if measurement error is addressed using the IV/GMM estimator.[5] The fixed-effects baseline estimates thus exhibit a considerable upward-bias, but even after accounting for simultaneity and measurement error the relationship between calorie intake and per-capita expenditure remains positive and statistically significant.

A similar pattern holds for the majority of nutrients, with some notable exceptions: First, the intake of two macro nutrients – fat and fiber – is found to be unresponsive to changing per-capita expenditures in the IV/GMM regression; and second, elasticities for vitamin A and D as well as calcium rise compared to the fixed-effects model when accounting for both simultaneity and measurement error. Overall, elasticities of nutrients tend to be higher than the elasticity of calories, which points to a tendency among households to switch to food richer in nutrients as their wealth rises. With an expenditure elasticity of above unity, calcium and vitamin A stand out in this regard.

Tests for weak instruments in the IV/GMM model are reported in Table 6. F-statistics for all the nutrient elasticity models are above the recommended threshold of 10, pointing to sufficiently strong instruments (Staiger and Stock, 1997). An alternative test for weak instruments is based on the Cragg-Donald (CD) statistic (Stock and Yoko, 2005). Since we are using five external instruments to identify a single endogenous regressor, allowing for 5 percent maximal IV bias means that the relevant CD critical value is 18.7. For all the nutrient elasticity models we obtain a CD statistic that is markedly higher than the critical value, which again indicates that our IV/GMM estimators do not appear to suffer from a weak instruments problem. The Anderson-Rubin p-values shown in the last row of Table 6 corroborates this finding.

Now we turn to the results from the semi-parametric model. In line with the discussions above, we estimate the non-parametric component, $g(m_{rt})$, employing a kernel-weighted local

[5] With a p-value of 0.435, the Hansen-Sargan test indicates that the model is over-identified.

polynomial regression. For our discussions here, we focus on the fitted curves from the estimations reported in Figure 1, which shows how the various nutrients relate to log per-capita expenditure. For details regarding the relationship between nutrients intake and the other household characteristics, see Appendix Table 1.

We find that per-capita calorie intake increases monotonically with log per-capita expenditure, but that the growth rate slows down at higher levels of expenditure, which implies that with the same growth in expenditure, the increase in calorie-intake is lower at higher expenditure levels. This result is consistent with previous studies (e.g. Behrman and Deolalikar, 1987; Tian and Yu, 2015). Food quality as measured by the price of calories also increases continuously with per-capita expenditures of the OVC households, with one notable exception: at very low levels of expenditure, an increase in expenditure is associated with a statistically significant decline in of food quality of up to 20% before it starts to increase. This is in contrast to what Skoufias et al. (2011) found for poor households in Mexico. The rising part of the curve is steepest at very high expenditure levels, i.e. at the upper tail the elasticity of food quality with respect to per-capita expenditures is highest.[6]

The general pattern that emerges for per-capita calories also holds for most nutrients. The intake of all macro-nutrients considered here (fiber, protein, and fat) is rising with per-capita expenditures. As shown by the dashed lines in the graphs, the macronutrients’ responsiveness is stronger for per-capita expenditures in the bottom quartile than for expenditures in the top quartile. Likewise, with the exception of vitamin D, for which the curve first declines and then flattens, micronutrient intake increases as per-capita expenditure rises, the slope of the curves being somewhat flatter for households in the top quartile than for households in the bottom quartile.

Overall, all graphs in Figure 1 point to a non-linear relationship between caloric and nutrition indices and log per-capita expenditure that our semi-parametric model is able to capture.

5. Summary and Conclusion

Despite a large existing literature, the issue of whether and to what extent additional income can help poor households meet their dietary needs is still not fully resolved. In this paper, we provide new evidence on calorie-expenditure and nutrient-expenditure elasticities in the so far largely overlooked setting of ultra-poor households in Sub-Saharan Africa. Employing panel

[6] Note that the confidence bands tend to be wider at the tails. This implies less precise estimates, which is likely to be due to the presence of fewer households.

data estimation techniques that control for simultaneity and measurement error bias, we find for a sample of close to 1800 Kenyan OVC households that higher per-capita expenditures are generally associated with higher calorie and nutrient intake. Richer OVC households tend to consume food richer in macro and micro nutrients than poorer OVC households. Food quality also tends to be better in richer OVC households than in poorer ones. A policy implication of these findings is that social protection schemes such as social assistance in the form of income transfers have the potential to help poor households get access to better diets.

From a methodological point of view, our results suggest that biases resulting from simultaneity and measurement error are considerable, rendering it important to control for them. We also find that even in extremely poor households there are non-linearities in calorie and nutrient intake, which we accommodate by using a semi-parametric model. A shortcoming of our paper is that we deal with the three methodological issues one at a time and not simultaneously. This means that each estimation challenge is addressed while isolating the other two challenges. We therefore cannot identify what happens to the elasticity estimates when the two sources of bias and non-linearity are all addressed in the same model. This is something we leave for future research.

References

Abdulai, A. and Aubert, D. (2004). Nonparametric and parametric analysis of calorie consumption in Tanzania. *Food Policy*, 29(2):113–129.

Anzagi, S. and Bernard, F. E. (1977). Population pressure in Kenya: A preliminary report. *Central Bureau of Statistics, Nairobi*.

Aromolaran, A. B. (2004). Household income, women's income share and food calorie intake in South Western Nigeria. *Food Policy*, 29(5):507–530.

Baltagi, B. H. and Li, D. (2002). Series estimation of partially linear panel data models with fixed effects. *Annals of economics and finance*, 3(1):103–116.

Banerjee, A. and Duflo, E. (2011). *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. Public Affairs.

Behrman, J. R. and Deolalikar, A. B. (1987). Will developing country nutrition improve with income? a case study for rural South India. *Journal of Political Economy*, 95(3):492–507.

Bouis, H. E. and Haddad, L. J. (1992). Are estimates of calorie-income elasticities too high? A recalibration of the plausible range. *Journal of Development Economics*, 39(2):333–364.

De Nadai, M. and Lewbel, A. (2016). Nonparametric errors in variables models with measurement errors on both sides of the equation. *Journal of Econometrics*, 191(1):19–32.

DiNardo, J. and Tobias, J. L. (2001). Nonparametric density and regression estimation. *The Journal of Economic Perspectives*, 15(4):11–28.

Gibson, J. and Rozelle, S. (2002). How elastic is calorie demand? parametric, nonparametric, and semiparametric results for urban Papua New Guinea. *Journal of Development Studies*, 38(6):23–46.

Griliches, Z. and Hausman, J. A. (1986). Errors in variables in panel data. *Journal of Econometrics*, 31(1):93–118.

Handa, S., Huang, C., Hypher, N., Teixeira, C., Soares, F. V., and Davis, B. (2012). Targeting effectiveness of social cash transfer programmes in three African countries. *Journal of Development Effectiveness*, 4(1):78–108.

Handa, S., Martorano, B., Halpern, C. T., Pettifor, A., and Thirumurthy, H. (2016). Time discounting and credit market access in a large-scale cash transfer programme. *Journal of African Economies*, 25(3):367–387.

Lee, V. C., Muriithi, P., Giberl-Nandra, U., Kim, A. A., Schmitz, M. E., Odek, J., Mokaya, R., Gallbraith, J., and KSG (2014). Orphans and vulnerable children in kenya: Results from a nationally representative population-based survey. *Journal of Acquired Immune Deficiency Syndromes*, 66(1):S89–S97.

Leibenstein, H. (1957). *Economic Backwardness and Economic Growth*. Wiley.

Lewbel, A. (2012). Using heteroscedasticity to identify and estimate mismeasured and endogenous regressor models. *Journal of Business & Economic Statistics*, 30(1):67–80.

Li, Q. and Stengos, T. (1996). Semiparametric estimation of partially linear panel data models. *Journal of Econometrics*, 71(1):389–397.

Libois, F. and Verardi, V. (2013). Semiparametric fixed-effects estimator. *Stata journal*, 13(2):329–336.

Lukmanji, Z., Hertzmark, E., Mlingi, N., Assey, V., Ndossi, G., and Fawzi, W. (2008). Tanzania food composition tables. *Dar es Salaam, Tanzania*.

Meghir, C. and Robin, J.-M. (1992). Frequency of purchase and the estimation of demand systems. *Journal of Econometrics*, 53(1-3):53–85.

Meijer, E., Spierdijk, L., and Wansbeek, T. (2017). Consistent estimation of linear panel data models with measurement error. *Journal of Econometrics*, 200(2):169 – 180.

Newson, R. (2000). bspline: Stata modules to compute b-splines parameterized by their values at reference points. statistical software components s411701, Department of Economics, Boston College.

Ogundari, K. and Abdulai, A. (2013). Examining the heterogeneity in calorie–income elasticities: A meta-analysis. *Food Policy*, 40(C):119–128.

Rodriguez-Poo, J. M. and Soberon, A. (2017). Nonparametric and semiparametric panel data models: Recent developments. *Journal of Economic Surveys*, 31(4):923–960.

Santeramo, F. G. and Shabnam, N. (2015). The income-elasticity of calories, macro-and micro-nutrients: What is the literature telling us? *Food Research International*, 76:932–937.

Skoufias, E. (2002). The sensitivity of calorie-income demand elasticity to price changes: evidence from Indonesia. Technical report, FCND Discussion Paper.

Skoufias, E., Di Maro, V., González-Cossío, T., and Ramirez, S. R. (2009). Nutrient consumption and household income in rural Mexico. *Agricultural Economics*, 40(6):657–675.

Skoufias, E., Di Maro, V., González-Cossío, T., and Ramirez, S. R. (2011). Food quality, calories and household income. *Applied Economics*, 43(28):4331–4342.

Stiglitz, J. E. (1976). The efficiency wage hypothesis, surplus labour, and the distribution of income in LDCs. *Oxford Economic Papers*, 28(2):185–207.

Strauss, J. and Thomas, D. (1990). The shape of the calorie-expenditure curve. Technical report, Center Discussion Paper.

Subramanian, S. and Deaton, A. (1996). The demand for food and calories. *Journal of Political Economy*, 104(1):133–162.

The Kenya CT-OVC Evaluation Team (2012). The impact of Kenya's cash transfer for orphans and vulnerable children on human capital. *Journal of Development Effectiveness*, 4(1):38–49.

Tian, X. and Yu, X. (2015). Using semiparametric models to study nutrition improvement and dietary change with different indices: The case of China. *Food Policy*, 53:67–81.

Yatchew, A. (1998). Nonparametric regression techniques in economics. *Journal of Economic Literature*, 36(2):669–721.

Table 1: Household and Village Indicators

	Bottom Quartile			Top Quartile			Pooled		
	2007	**2009**	**2011**	**2007**	**2009**	**2011**	**2007**	**2009**	**2011**
Age (head)	61.46	56.99	59.43	57.28	58.50	60.06	61.34	57.58	58.99
	(19.18)	(16.87)	(14.23)	(20.25)	(14.93)	(14.16)	(19.31)	(15.06)	(14.79)
Gender (head)	0.622	0.585	0.558	0.670	0.697	0.652	0.635	0.648	0.628
	(0.485)	(0.494)	(0.498)	(0.473)	(0.460)	(0.477)	(0.482)	(0.478)	(0.483)
Education	1.22	1.45	1.74	0.647	0.696	1.09	1.11	1.12	1.28
	(1.22)	(1.296)	(1.517)	(0.817)	(0.769)	(1.00)	(1.140)	(1.062)	(1.199)
Child under 6	0.130	0.126	0.090	0.139	0.101	0.071	0.129	0.113	0.080
	(0.145)	(0.1320)	(0.1070)	(0.1670)	(0.152)	(0.121)	(0.150)	(0.143)	(0.117)
Child under 14	0.275	0.227	0.276	0.316	0.294	0.270	0.282	0.256	0.270
	(0.180)	(0.161)	(0.155)	(0.198)	(0.205)	(0.205)	(0.188)	(0.187)	(0.190)
Teenager	0.202	0.034	0.204	0.119	0.042	0.180	0.187	0.0405	0.193
	(0.171)	(0.0322)	(0.141)	(0.161)	(0.046)	(0.183)	(0.174)	(0.040)	(0.169)
Adult	0.216	0.308	0.279	0.176	0.201	0.199	0.156	0.154	0.166
	(0.180)	(0.172)	(0.166)	(0.200)	(0.188)	(0.179)	(0.161)	(0.161)	(0.165)
Old	0.144	0.108	0.124	0.176	0.201	0.199	0.156	0.154	0.166
	(0.151)	(0.117)	(0.119)	(0.200)	(0.188)	(0.179)	(0.161)	(0.161)	(0.165)
Road	0.726	0.952	0.755	0.795	0.911	0.758	0.723	0.919	0.756
	(0.446)	(0.214)	(0.431)	(0.406)	(0.285)	(0.429)	(0.447)	(0.273)	(0.430)
Market (minutes)	41.95	34.41	38.30	39.75	30.75	42.57	43.41	32.12	39.34
	(37.60)	(43.61)	(39.99)	(29.97)	(34.17)	(37.95)	(38.15)	(35.44)	(36.86)
Per capita expenditure (Ksh)	2731.7	3168.1	3170.6	11448.6	13167.9	14667.1	4502.6	7891.8	9916.9
	(801.5)	(654.1)	(625.9)	(1268.3)	(3204.6)	(4913.6)	(2468.7)	(3966.1)	(5437.8)
Household size	6.17	7.43	7.76	3.99	4.34	4.66	5.62	5.70	5.73
	(2.740)	(2.963)	(2.932)	(1.643)	(1.725)	(2.218)	(2.632)	(2.532)	(2.727)
price of calorie (Ksh, 00)	1.86	2.23	4.24	2.10	2.86	4.78	1.90	2.59	4.55
	(1.437)	(0.864)	(3.114)	(0.579)	(0.819)	(2.104)	(1.126)	(0.800)	(3.300)

Reported in the table are the means of the variables in each sub-sample by year; in parentheses are standard deviations. Note that all the continuous variables were winsorized at 2%; this was done to get rid of outliers. Ksh=Kenyan Shillings; km=Kilometer. Demographic variables are in proportions.

Table 2: Macro and Micro Nutrient Intake

	Bottom Quartile			Top Quartile			Pooled		
	2007	2009	2011	2007	2009	2011	2007	2009	2011
Total calorie (kcal)	3399.7	3484.6	2279.1	9014.7	9181.6	6663.7	4687.0	6441.5	5146.7
	(1698.5)	(1550.1)	(1087.2)	(2576.8)	(3427.5)	(2803.8)	(2540.3)	(3228.3)	(2723.0)
Fiber (kcal)	211.6	245.4	139.9	482.5	589.9	405.1	281.9	428.4	316.8
	(130.3)	(124.2)	(91.42)	(150.0)	(227.2)	(216.1)	(165.4)	(219.6)	(202.4)
Fat (kcal)	637.7	528.4	580.7	2092.4	1681.7	1710.9	947.1	1103.1	1291.9
	(370.7)	(306.9)	(659.4)	(792.2)	(897.7)	(1070.9)	(621.8)	(736.6)	(1022.9)
Protein (kcal)	293.7	305.8	222.4	856.1	1012.4	667.3	432.7	648.1	513.1
	(176.9)	(149.6)	(99.23)	(203.7)	(463.2)	(266.9)	(266.9)	(407.8)	(272.6)
Iron (mg)	26.59	31.55	17.40	62.14	76.30	54.88	35.58	55.33	41.71
	(16.93)	(17.48)	(10.92)	(18.83)	(30.48)	(28.36)	(21.43)	(29.60)	(26.66)
Zinc (mg)	14.15	15.22	9.440	38.68	41.98	29.40	20.05	28.83	22.54
	(8.43)	(7.42)	(5.58)	(8.79)	(14.29)	(14.65)	(11.93)	(14.87)	(13.91)
Calcium (mg)	425.1	307.0	363.3	1663.2	1265.4	1172.4	725.4	785.2	862.5
	(578.3)	(363.5)	(351.7)	(963.0)	(865.3)	(773.6)	(823.3)	(754.6)	(719.8)
Vitamin A (µg)	44.26	170.2	51.45	162.4	489.4	485.2	69.47	332.8	292.3
	(95.74)	(448.3)	(122.2)	(135.6)	(647.0)	(645.2)	(111.4)	(553.0)	(515.4)
Vitamin D (µg)	0.279	0.149	0.471	1.248	0.718	1.661	0.512	0.416	1.210
	(0.604)	(0.371)	(0.451)	(1.161)	(0.893)	(1.130)	(0.849)	(0.703)	(1.028)
Folate (µg)	312.3	369.3	253.0	1064.7	1142.6	852.1	488.4	763.3	619.6
	(254.5)	(217.1)	(201.0)	(436.3)	(434.5)	(476.3)	(392.6)	(447.4)	(437.1)
Arginine (mg)	3302.9	3438.6	2326.5	9019.9	9660.4	7780.4	4734.0	6599.3	5834.4
	(2270.5)	(1899.9)	(1470.2)	(3243.5)	(4144.2)	(3943.5)	(3227.7)	(3926.9)	(3771.4)
Histidine (mg)	2298.9	2442.6	1669.8	7021.8	7356.4	5783.0	3445.8	4914.5	4291.6
	(1462.4)	(1219.9)	(930.8)	(1977.9)	(2645.2)	(2848.8)	(2236.8)	(2689.1)	(2714.1)
Lysine (mg)	2993.2	2936.8	2618.1	10697.2	10731.2	9042.6	4880.3	6749.1	6592.0
	(2303.6)	(1507.2)	(1471.9)	(3469.6)	(4085.2)	(4551.7)	(3725.3)	(4175.7)	(4341.1)

Reported in the table are the means of the variables in each sub-sample by year; in parentheses are standard deviations. Note that all the variables were winsorized at 2%; this was done to get rid of outliers. µg =microgram; mg=milligram; kcal=kilo calorie.

Table 3: Share of Calories

	Bottom Quartile			Top Quartile			Pooled		
	2007	2009	2011	2007	2009	2011	2007	2009	2011
Cereal	0.615	0.705	0.549	0.483	0.568	0.501	0.585	0.6340	0.537
	(0.2780)	(0.2051)	(0.3320)	(0.2091)	(0.1992)	(0.2631)	(0.2520)	(0.2050)	(0.2780)
Tuber	0.007	0.002	0.006	0.034	0.009	0.009	0.011	0.006	0.008
	(0.0313)	(0.0104)	(0.0288)	(0.0693)	(0.0323)	(0.0305)	(0.0372)	(0.0241)	(0.0298)
Pulse	0.036	0.046	0.068	0.065	0.076	0.066	0.046	0.061	0.063
	(0.0667)	(0.0643)	(0.1160)	(0.0630)	(0.0804)	(0.0759)	(0.0706)	(0.0730)	(0.0840)
Fruit	0.006	0.005	0.004	0.024	0.007	0.012	0.010	0.007	0.010
	(0.0277)	(0.0204)	(0.0243)	(0.0501)	(0.0189)	(0.0257)	(0.0348)	(0.0265)	(0.0234)
Vegetable	0.006	0.008	0.014	0.003	0.007	0.013	0.006	0.008	0.012
	(0.0337)	(0.0283)	(0.0365)	(0.0180)	(0.0161)	(0.0333)	(0.0321)	(0.0220)	(0.0324)
Meat	0.024	0.014	0.040	0.078	0.059	0.069	0.037	0.035	0.057
	(0.0557)	(0.0348)	(0.0828)	(0.0995)	(0.0690)	(0.0750)	(0.0653)	(0.0561)	(0.0724)
Milk	0.018	0.008	0.023	0.023	0.026	0.041	0.021	0.020	0.034
	(0.0656)	(0.0173)	(0.0105)	(0.0499)	(0.0333)	(0.0533)	(0.0604)	(0.0313)	(0.0513)
Oil	0.137	0.079	0.099	0.140	0.078	0.090	0.133	0.080	0.095
	(0.1600)	(0.0886)	(0.1300)	(0.1270)	(0.0804)	(0.0971)	(0.1390)	(0.0839)	(0.1080)
Sugar	0.148	0.105	0.074	0.144	0.143	0.084	0.150	0.127	0.084
	(0.1920)	(0.0933)	(0.1020)	(0.0970)	(0.1080)	(0.0237)	(0.1660)	(0.1050)	(0.0925)

Reported in the table are the means of the variables in each sub-sample by year; in parentheses are standard deviations. Note that all the continuous variables were winsorized at 2%; this was done to get rid of outliers.

Table 4: Share of Food Expenditure

	Bottom Quartile			Top Quartile			Pooled		
	2007	2009	2011	2007	2009	2011	2007	2009	2011
Cereal	0.309	0.494	0.365	0.223	0.352	0.311	0.279	0.406	0.341
	(0.131)	(0.178)	(0.167)	(0.108)	(0.118)	(0.114)	(0.124)	(0.148)	(0.134)
Tuber	0.050	0.030	0.034	0.049	0.043	0.038	0.048	0.035	0.036
	(0.064)	(0.051)	(0.062)	(0.055)	(0.051)	(0.040)	(0.057)	(0.050)	(0.046)
Pulse	0.032	0.0563	0.065	0.056	0.078	0.059	0.040	0.068	0.056
	(0.046)	(0.063)	(0.083)	(0.058)	(0.056)	(0.053)	(0.038)	(0.064)	(0.059)
Meat	0.077	0.074	0.071	0.247	0.202	0.196	0.124	0.147	0.152
	(0.110)	(0.095)	(0.090)	(0.165)	(0.127)	(0.116)	(0.134)	(0.126)	(0.119)
Vegetable	0.137	0.117	0.085	0.073	0.083	0.069	0.122	0.097	0.073
	(0.103)	(0.095)	(0.073)	(0.045)	(0.055)	(0.042)	(0.093)	(0.070)	(0.051)
Milk	0.089	0.078	0.085	0.056	0.078	0.061	0.079	0.084	0.066
	(0.068)	(0.050)	(0.058)	(0.036)	(0.059)	(0.055)	(0.062)	(0.058)	(0.053)
Fruits	0.016	0.023	0.018	0.030	0.023	0.027	0.020	0.024	0.024
	(0.036)	(0.035)	(0.035)	(0.037)	(0.030)	(0.044)	(0.036)	(0.036)	(0.040)
Oil	0.077	0.071	0.076	0.055	0.056	0.058	0.070	0.062	0.063
	(0.050)	(0.045)	(0.05)	(0.0554)	(0.0404)	(0.038)	(0.046)	(0.042)	(0.042)
Sugar	0.085	0.082	0.069	0.068	0.072	0.063	0.083	0.078	0.067
	(0.090)	(0.063)	(0.057)	(0.047)	(0.046)	(0.043)	(0.079)	(0.056)	(0.054)

Reported in the table are the means of the variables in each sub-sample by year; in parentheses are standard deviations. Note that all the continuous variables were winsorized at 2%; this was done to get rid of outliers.

Table 5: Nutrient-Income Elasticities from Panel Regressions

Nutrient	FE	IV	IV/GMM
Calorie	0.485 (0.0145) ***	0.387 (0.0478) ***	0.188 (0.0698) *** [0.435]
Price (calorie)	0.516 (0.0133) ***	0.647 (0.0437) ***	0.648 (0.0651) *** [0.020]
Protein	0.683 (0.0149) ***	0.617 (0.0488) ***	0.357 (0.0752)*** [0.540]
Fat	0.682 (0.0199) ***	0.577 (.0691) ***	-0.154 (0.1612) [.087]
Fiber	0.573 (0.0184) ***	0.502 (0.0596) ***	0.095 (0.1002) [0.143]
Iron	0.600 (0.0186) ***	0.532 (0.0605) ***	0.347 (0.1014)*** [0.259]
Zinc	0.622 (0.0162) ***	0.536 (0.0537) ***	0.342 (0.0860)*** [0.359]
Calcium	0.965 (0 .0286) ***	0.754 (0.0932) ***	1.155 (0.1422)*** [0.547]
Vitamin A	0.976 (0.0583) ***	1.675 (0.2122) ***	1.124 (0.1985) *** [0.431]
Vitamin D	0.497 (0.0481)***	0.095 (0.2107)	0.599 (0.1593)*** [0.400]
Folate	0.810 (0.0202) ***	0.756 (0.0666) ***	0.437 (0.1053)*** [.094]
Arginine	0.715 (0.0202) ***	0.650 (0.0688) ***	0.540 (0.1088)*** [0.494]
Histidine	0.710 (0.0156) ***	0.605 (0.0530) ***	0.432 (0.592)*** [0.0827]
Lysine	0.847 (0.0170) ***	0.733 (0.0563) ***	0.530 (.0877)*** [0.529]

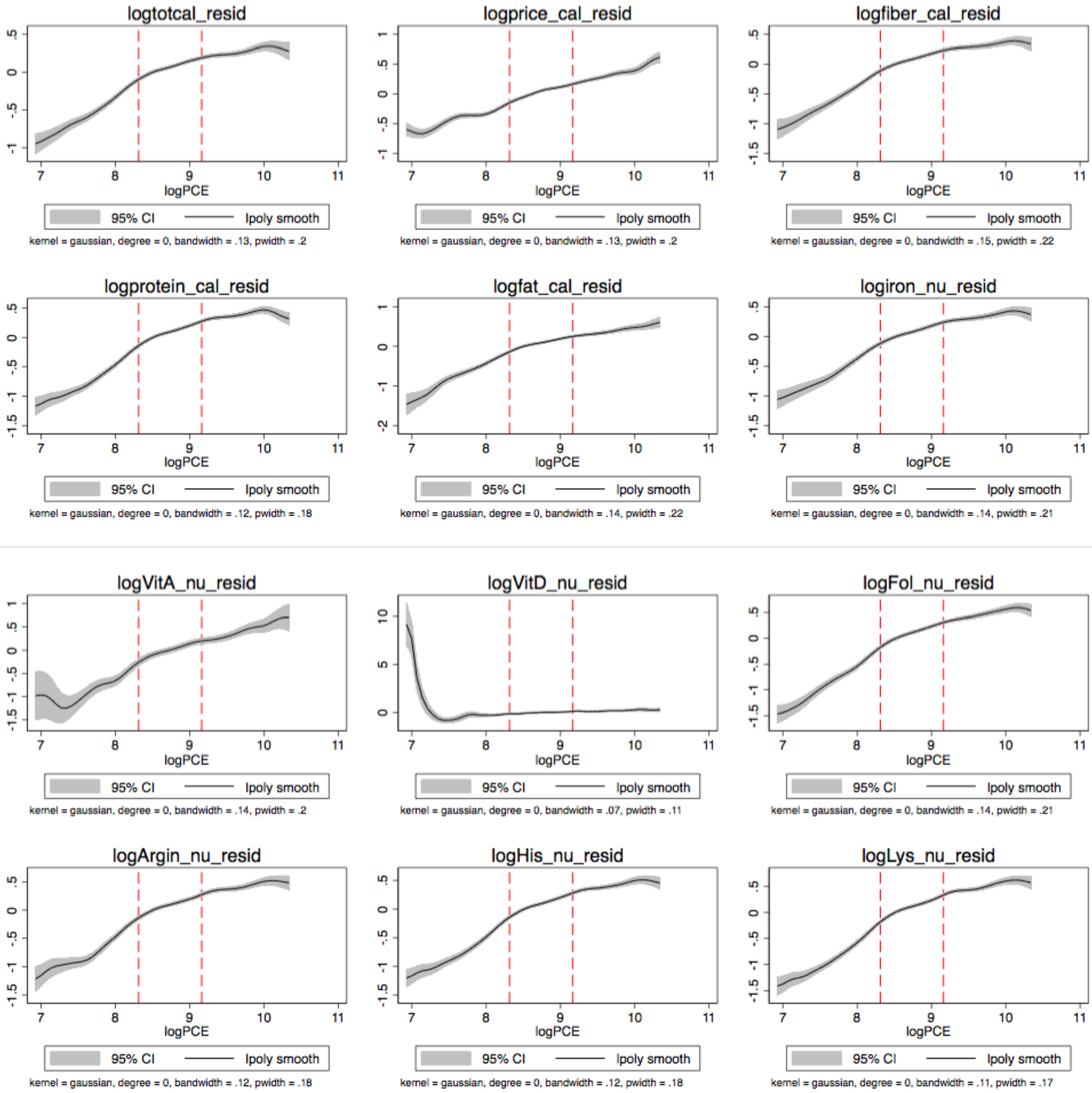
Note. Reported in the table are the elasticity estimates from the panel regressions. The robust standard errors are reported in the parentheses. For the last column, also reported in brackets are the p-values from the Hansen-Sargant over-identification test. FE = fixed effect panel data model using the within estimator; IV=panel instrumental variable; GMM= Generalized method of moments estimation.
*** 99% significance; ** 95% significance; *90% significance.

Table 6: IV/GMM Weak Instrument Test Statistics

	Calorie	Price calorie	Protein	Fat	Fiber	Iron	Zinc	Calcium	Vit. A	Vit. D	Folate	Arginine	Histine	Lysine
F-stat	19.4	15.88	15.10	15.43	15.43	15.43	15.32	15.43	16.79	12.48	15.34	15.32	15.32	15.32
CD-stat+	34.05	28.56	27.09	27.58	27.58	27.58	27.41	27.58	24.72	26.02	27.42	27.41	27.41	27.41
AR	0.090	0.000	0.160	0.008	0.078	0.017	0.011	0.000	0.000	0.005	0.002	0.003	0.002	0.000

Note. The table reports the various test statistics for weak instruments for our over-identified Panel IV model reported in the last column of table 5. F-stat= first stage F-statistics; CD-stat= Cragg-Donald Statistic; AR=Anderson Rubin test p-value.
+ Since we are instrumenting one endogenous variable with five external instruments, the applicable Stock-Yogo critical value at 5% maximal IV relative bias is 18.37. Therefore, the Cragg-Donald statistics should be compared with this critical value.

Figure 1: Nonparametric Estimation Results



Shown in the figure are estimates of the nutrient expenditure elasticity from our semiparametric model. The shaded areas on the graphs are the 95% confidence bands. The two red dashed vertical lines represent the bottom and top quartile of per capita expenditure.

Appendix Table 1: Other Covariates in the Semi-parametric Model

	Calorie	Price calorie	Protein	Fat	Fiber	Iron	Zinc	Calcium	Vit. A	Vit. D	Folate	Arginine	Histine	Lysine
Age	-0.003***	-0.000	-0.003	-0.002	0.001	-0.004	-0.002	0.000	-0.014	-0.002	-0.004	-0.002	-0.001	-0.001
	(0.0010)	(0.0012)	(0.0019)	(0.0011)	(0.0019)	(0.0024)	(0.0011)	(0.0020)	(0.0140)	(0.0060)	(0.0021)	(0.0017)	(0.0011)	(0.0010)
Gender	0.032	0.016	0.046	0.037	0.047	0.039	0.037	-0.092	0.209	-0.013	0.036	0.070	0.040	0.027
	(0.0503)	(0.0272)	(0.0544)	(0.0452)	(0.0655)	(0.0625)	(0.0625)	(0.0883)	(0.1551)	(0.1711)	(0.0322)	(0.0722)	(0.0519)	(0.0387)
Education	0.019	-0.054	0.066	0.011	-0.039	0.069	0.022	-0.081*	0.209	-0.053	0.019	0.026	0.007	-0.024
	(0.0244)	(0.0231)	(0.0434)	(0.0271)	(0.0478)	(0.0610)	(0.0339)	(0.0348)	(0.2563)	(0.1142)	(0.0578)	(0.0354)	(0.0259)	(0.0316)
HH Size	-0.320***	0.091	-0.227***	-0.241***	-0.459***	-0.229***	-0.238***	-0.215***	0.039	-0.542***	-0.216*	-0.179***	-0.216***	-0.213***
	(0.0491)	(0.0785)	(0.0471)	(0.0434)	(0.0701)	(0.0631)	(0.0430)	(0.0537)	(0.2600)	(0.1241)	(0.0914)	(0.0443)	(0.0236)	(0.0509)
Child U6	0.150	-0.4122***	-0.066	-0.013	0.145	-0.071	0.017	0.120	-0.553	0.714	-0.131	-0.054	0.046	0.001
	(0.0876)	(0.0482)	(0.1509)	(0.1264)	(0.1550)	(0.1426)	(0.1346)	(0.1146)	(0.5406)	(0.4070)	(0.2359)	(0.1235)	(0.0971)	(0.1113)
Child U14	0.026	-0.097	-0.096	0.060	0.367***	-0.094	-0.005	0.337***	-0.377	0.713***	-0.072	0.041	0.042	0.107
	(0.0924)	(0.0953)	(0.1421)	(0.0938)	(0.1356)	(0.1470)	(0.1080)	(0.0820)	(0.5318)	(0.1553)	(0.1924)	(00928)	(0.0708)	(0.0705)
Teenager	-0.669***	0.484***	-0.9141***	-0.457***	0.373	-0.884***	-0.646***	0.564*	-0.882	1.753***	-0.790***	-0.461*	-0.407***	-0.111
	(0.1697)	(0.1867)	(0.2109)	(0.1479)	(0.2890)	(0.2163)	(0.1820)	(0.2368)	(0.8840)	(0.2691)	(0.2807)	(0.2018)	(0.1220)	(0.1059)
Road	0.071	0.051	0.103*	0.092*	0.029	0.141***	0.095*	-0.042	0.851***	-0.087	0.125	0.069	0.067*	0.034
	(0.0434)	(0.0724)	(0.0430)	(0.0458)	(0.0476)	(0.0472)	(0.0420)	(0.1419)	(0.3088)	(0.0527)	(0.0651)	(0.0459)	(0.0324)	(0.0383)
Market	0.000	0.000	0.000	0.000	0.001	0.000	0.000	-0.000	0.003	0.001	-0.000	0.000	0.000	0.000
	(0.0005)	(0.0007)	(0.0005)	(0.0005)	(0.0013)	(0.0005)	(0.0005)	(0.0006)	(0.0017)	(0.0014)	(0.0010)	(0.0006)	(0.0004)	(0.0006)
R-sq	0.33	0.13	0.29	0.35	0.21	0.29	0.31	0.17	0.13	0.10	0.31	0.25	0.34	0.32
N	2982	3080	3067	3084	3084	3084	3083	3084	1695	944	3081	3083	3083	3083

Note. This table reports the results from the semi-parametric model. Reported in parantheses are the robust standard errors. HH= Household; U6= Under 6; U14= Under 14; R-sq=Adjusted within R-squared; N=number of observations used in the estimation
*** 99\% significance; ** 95\% significance; *90\% significance.