Institute for Fiscal Studies

Food and Cash Transfers: Evidence from Colombia

IFS Working Paper W09/15

Orazio Attanasio
Erich Battistin
Alice Mesnard

E·S·R·C
ECONOMIC & SOCIAL RESEARCH COUNCIL

Food and Cash Transfers: Evidence from Colombia[♠]

Orazio Attanasio[#], Erich Battistin[*], Alice Mesnard[♣]

June 2009

Abstract

We study food Engel curves among the poor population targeted by a conditional cash transfer programme in Colombia. After controlling for the endogeneity of total expenditure and for the (unobserved) variability of prices across villages, the best fit is provided by a log-linear specification. Our estimates imply that an increase in total expenditure by 10% would lead to a decrease of 1% in the share of food. However, quasi-experimental estimates of the impact of the programme on total and food consumption show that the share of food increases, suggesting that the programme has more complex impacts than increasing household income. In particular, our results are not inconsistent with the hypothesis that the programme, targeted to women, could increase their bargaining power and induce a more than proportional increase in food consumption.

Key words : Demand patterns, food Engel curves, evaluation of welfare programme.

JEL classification : C52, D12, I38.

[♠] This research was financed with ESRC grant RES-000-22-0583 on "Household and Individual Consumption Decisions under Extreme Poverty". We received useful comments from Richard Blundell and Arthur Lewbel.

[#] IFS, UCL, NBER and CEPR.

[*] Università di Padova and IRVAPP.

[♣] IFS, CEPR and Toulouse School of Economics.

1. Introduction

The description of demand patterns is one of the oldest endeavours in applied economics. And yet, many unresolved problems still make the estimation of a demand system a difficult exercise. When considering, for instance, how expenditure shares vary with total expenditure and prices, there is no consensus on the specific functional form to be used for the relationships to be estimated, how to address the endogeneity of total expenditure, how to model the effect of prices when they are not observed, what is the estimation approach that is more effective. All these issues are key for a correct estimation of demand systems and the relationship between expenditure shares and total expenditure. These relationships are not only of academic interest but have important implications for the design of policies.

In this paper, we study expenditure patterns among poor households in rural Colombia. This study has three main goals. First and foremost, we want to characterize the demand for food for our population. This is an interesting exercise *per se* because of the very nature of the population. The extreme poverty of the households in our sample makes some facts taken for granted among other populations, in particular that the food income elasticity is less than one, questionable. Such elasticity is relevant for the design of policies aimed at improving the nutritional status of children and other poor and vulnerable individuals.

Second, as the data were collected for the evaluation of a welfare programme, one can assess the extent to which the structural equation defined by the demand for food can predict the changes in expenditure patterns implied by the quasi-experimental variation in our sample. Since the latter can be estimated with some confidence given the way in which the evaluation was designed, we can use these results and our demand estimates to validate the latter. The identification of specific inadequacies of the demand system we estimate in predicting how the structure of consumption changes with the policy intervention might be suggestive of the channels through which the policy operates and of richer behavioural models that should be fitted to the data.

Finally, by addressing the various methodological problems we will be dealing with and by exploring alternative modelling choices, the paper gives a methodological contribution to the study of demand patterns. In particular, we will be addressing the issue of the functional form for the demand system, the appropriate instrumenting of total expenditure and how to control for price differences when prices are not observed.

The main results of the paper can be summarised as follows. After investigating different econometric techniques we conclude that estimates of the structural parameters of the Engel curve obtained using a control function approach seem to be the most reliable. It is clear that taking into account the endogeneity of total expenditure is important and affects in an important fashion the shape of the curve. OLS estimates seem to indicate that food is a luxury at very low levels of total expenditure and become a necessity only at sufficiently high levels. This evidence, however, disappears when we instrument for total expenditure. We find that food consumption is indeed a necessity for almost every household in our sample. This inference is important given that our sample is made of very poor individuals. We also show that it is important to take into account the variability of relative prices across villages (which we do not observe perfectly). However, by far the most important aspect turns out to be to control for the endogeneity of total expenditure.

An issue that we discuss and about which there is no consensus in the literature is what type of instrument one should use for total expenditure. The data set we use, which was collected to evaluate a conditional cash transfer in Colombia, is particularly useful in this respect as it contains an interesting variable (that we will refer to as "expected income" in what follows) that seems particularly appropriate to instrument total expenditure in the context of Engel curve estimation. Because of the way the survey questions from which we derive expected income are formulated, the instrument is likely to be valid even in the case of non separability between consumption and leisure choices. To the best of our knowledge, such a variable has not been used before in other studies.

Having obtained a preferred specification for the Engel curve, we use it, together with quasi-experimental estimates of the increase in total expenditure induced by the programme, to predict changes in food shares induced by the programme. The comparison of structural and quasi-experimental estimates reveals that the Engel curve does a very poor job at predicting changes in food shares. We argue that a possible explanation is the fact that the grant is targeted to women and therefore is likely to change the balance of power within the household and, in general, change the way choices are made. Implicit in this argument is a mis-specification of the Engel curves. We discuss possible alternatives in the conclusions.

The remainder of the paper is organized as follows. In Section 2 we introduce the conceptual framework within which we will be discussing the various estimation problems. In Section 3 we present the data we will be used in the analysis. In this section we also briefly describe the welfare programme for whose evaluation the data were collected and present estimates of its effects on household consumption. In Section 4, we present the results of our empirical analysis of demand patterns in Colombia using different approaches. Having established in Section 5 which of the alternative approaches considered yields our preferred specification, in Section 6 we relate the impacts estimates to the estimates of the Engel curves, as discussed above. Section 7 concludes.

2. Estimating Engel curves

In this paper we study Engel curves for food, that is, the relationship between the share of total expenditure devoted to food and total consumption. Such a relationship can be derived within a standard demand system. If one assumes that individual households (conceived as a single decision unit) maximize utility subject to a budget constraint, one can obtain demand curves where expenditure (shares) on individual commodities depend on total expenditure, prices and preference shifters that might include demographic and other variables. The tension in an exercise of this type is between equations that are flexible enough to fit the data and yet are consistent with the restrictions implied by the theory.

Deaton and Muellbauer's (1980) Almost Ideal Demand System (AIDS) has been widely used and, for a given level of prices, implies a linear relationship between expenditure shares and the log of total expenditure. Banks, Blundell and Lewbel (1997) (BBL from now on) have proposed a quadratic generalization of such a system (the Q-AIDS). It could be argued that the AIDS and its quadratic generalization constitute one of the most flexible theory consistent functional forms available in the literature. Therefore, in our discussion, we use the BBL specification as a starting point.

2.1. Functional forms and price effects

As detailed in BBL, a Q-AIDS demand system can be derived from the following indirect utility function V:

$$\text{(1)} \qquad \ln V = \left\{ \left[\frac{\ln m - \ln(a(\mathbf{p}))}{b(\mathbf{p})} \right]^{-1} + \lambda(\mathbf{p}) \right\}^{-1},$$

where m is total expenditure, $\mathbf{p}$ the vector of prices and the functions $a(\mathbf{p})$, $b(\mathbf{p})$ and $\lambda(\mathbf{p})$ are defined as follows:

$$\ln a(\mathbf{p}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \ln p_i + 0.5 \sum_{j=1}^{n} \sum_{i=1}^{n} \gamma_{ij} \ln p_i \ln p_j,$$

$$b(\mathbf{p}) = \prod_{i=1}^{n} p_i^{\beta_i},$$

$$\lambda(\mathbf{p}) = \sum_{i=1}^{n} \lambda_i \ln p_i; \qquad \sum_{i=1}^{n} \lambda_i = 0.$$

By applying Shephard's lemma to this indirect utility one can get the following share equations:

$$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln p_j + \beta_i (\ln m - \ln a(\mathbf{p})) + \frac{\lambda_i}{b(\mathbf{p})} \left[\ln \frac{m}{a(\mathbf{p})} \right]^2 .$$

To these equations one can add demographics (either as affecting the intercept α_i or the price coefficients or even the coefficients on total expenditure. We re-write this equation so to include a residual term $\mathrm{u_i}$ to reflect unobserved taste shocks and measurement error:

(2) $$w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln p_j + \beta_i (\ln m - \ln a(\mathbf{p})) + \frac{\lambda_i}{b(\mathbf{p})} \left[\ln \frac{m}{a(\mathbf{p})} \right]^2 + \mathrm{u_i}$$

As discussed in BBL, the demand system in (2) combines functional form flexibility and consistency with theory, in that it is integrable. The last term in (2) makes the demand system of rank 3, the highest admissible rank for a theory-consistent demand system that is exactly aggregable, in that it is linear in function of total expenditure.

BBL discuss extensively the importance of a quadratic term in the demand system, as such a term allows some commodities to be necessities at certain levels of total expenditure and luxuries at others. This aspect is potentially very important in our context. We will be particularly interested in the expenditure elasticity of food. We will therefore want to avoid imposing ex-ante the linearity implied by a standard AIDS system and allow for the additional flexibility afforded by the quadratic term in (2).

BBL show that any theory-consistent system is of rank 3 (and therefore allows some commodities to have quadratic terms and some not) only if the coefficient on the quadratic term is a function of prices, as is the case in equation (2). This issue is of particular relevance for us because, although the data we use to estimate versions of equation (2) are from a single cross section, they come from more than 100 small villages that exhibit a substantial amount of variation in relative prices. Moreover, as in our data price information is limited to some food items and we do not have any price information on other commodities, we cannot compute the relative price of 'food'. Therefore, we will have to work under the assumption that prices are unobservable.

One possibility, of course, is to assume the problem away. If one uses data from a single cross section and is willing to assume that prices faced by the consumers in that cross section are uniform (within and across towns), one does not need to worry about the issue of unobservable prices. It should be remembered, however, that in such a situation, the size of the coefficient on the quadratic term cannot be extrapolated to different contexts, as it would depend on the level of prices prevalent in the cross section used for estimation and would vary in different situations. Moreover, this assumption seems very strong.[1]

If we expand the square in equation (2), we see that prices enter in three places, that is as an intercept shifts and as shifters of the coefficients of both the linear and quadratic terms:

$$(3) \qquad w_i = \alpha_i + \sum_{j=1}^{n} \gamma_{ij} \ln p_j + \beta_i (\ln m - \ln a(\mathbf{p})) + \frac{\lambda_i}{b(\mathbf{p})} \left[\ln m^2 + \ln a(\mathbf{p})^2 - 2 \ln m \ln a(\mathbf{p}) \right] + \mathrm{u_i}$$

Notice, however, that if one is willing to assume that the coefficients on the quadratic term λ_i are zero, then prices enter as a simple intercept shift.

In the absence of detailed information on (relative) prices, we consider two alternative strategies other than the simple strategy that assumes no variation in prices in the cross section. A first and flexible approach is to control for prices by village dummies. While this approach is robust, testing for the presence of quadratic terms in (2) in this context becomes problematic as the coefficient on the quadratic term becomes village specific and varies with prices. Notice that if the quadratic term has no effect on expenditure shares, the estimation procedure is greatly simplified as village dummies enter only as intercept shifts.[2]

Alternatively, one can try to capture differences in relative prices across villages by means of village level variables reflecting the economic environment that are relevant for the determination of relative prices. These variables might include the size and population density of the villages, the number of shops, the altitude and the level of some representative prices on which information is available. Of course such an approach implies that all the systematic variability in relative prices across villages is captured by these variables.

In what follows we will be looking at these different approaches. Obviously, if one does not reject the hypothesis that the quadratic terms are absent, the analysis, even in the presence of unobservable prices, is greatly simplified.

2.2. Endogeneity of expenditure

There are several reasons why the terms in log total expenditure might be correlated with the residuals of the demand system. The usual interpretation of a static system is as the second step of a two stage budgeting, where the first step determines the allocation of total expenditure across time periods, and the second determines the allocation within the period. If heterogeneity

[1] Attanasio and Frayne (2005) show that, in the same data we are using, there is a substantial amount of variability of unit values for individual food items both within and across villages.

[2] Price interacts with the linear term in log expenditure only through the cross product in equation (3). If $\lambda_i = 0$ there are no interactions between prices and log expenditure.

in intertemporal preferences are correlated with (unobserved) taste shifters in the demand system, one would obtain that the residuals of the latter are correlated, across individuals, with the allocation of resources over time and therefore with log expenditure. It is possible, for instance, that individuals that have a relatively stronger preference for food are also relatively impatient and therefore have a higher level of current consumption as well as a high share of food expenditure.

Another reason for the possible correlation between residuals and log expenditure is the presence of measurement error. A useful source of exogenous variation in this context, therefore, may come from a variable that explains the cross sectional variability of log expenditure but is unlikely to be correlated with taste variables and/or with measurement error. In the literature, income is often used for such a purpose. However, if labour supply enters the utility function in a non-separable manner, income might be correlated with taste shifters in the same way as total expenditure is. Moreover, in the presence of large transitory shocks, current income can constitute an inefficient instrument for total expenditure even if it is uncorrelated with taste shifters. A possible alternative to the use of income is the use of wages, which may be considered as a price that the individual household takes as given An even more conservative stance would be to use village level wages as instruments for total expenditure of an individual household and is unlikely to be correlated with measurement error or taste shifters. Such an approach has been tried, for instance, by Attanasio and Lechene (2002).

Wages, however, are an invalid instrument if leisure and consumption are not separable. Moreover, as our analysis allows consumption patterns to vary depending on village prices, we cannot use village wages as an instrument for total expenditure. Instead, the instrument we choose is a proxy for household expected income, using information relating to the variation in households' future income stream. This variable is defined by using two questions on the lowest and highest income a household is expecting to receive in the next month.[3] Given the way these questions are asked, these bounds have the advantage of being exogenous to labour supply choices. Moreover, as the variables we try to construct capture *expected* income, it should be uncorrelated with transitory shocks. In Section 3, we will show some descriptive statistics on this somewhat unusual variable and show how it co-varies with income and, most importantly, with total consumption.

---

[3] The two questions are: (1) Suppose that next month the members of your family who want to work, get a good job (alternatively: imagine the harvest is good). How much money do you think would be earned/would come into the household in that month? (2) Now imagine the opposite: that they have very little work next month (alternatively: imagine the harvest is bad) and that they have just this to live on, as well as what people give them (which is very little). How much money do you think the household would receive in that month?

The presence of quadratic terms in equation (2) introduces additional problems to the instrumenting approach. Once one has established which the instruments one wants to use are, one can use powers in this instruments to take into account the presence of non linear terms. Such an approach, however, often yields very imprecise estimates. To overcome this problem we adopt an alternative strategy based on a control function (CF) approach as proposed in this context, for instance, by Blundell, Duncan and Pendakur (1998). According to such an approach, one uses the residuals of the first step regression for total expenditure to control for the endogeneity of this variable in equation (2) by introducing a polynomial in the residuals as additional regressors.

2.3. Unitary or non unitary models?

The structure described above is derived under the assumption that decisions are taken by a single decision unity that maximizes a well defined utility function. In what follows, we will suggest that such a representation might not be accurate as the expenditure patterns of poor Colombian families might be the result of the interactions of more than one decision maker. A model that has been proposed to deal with these issues is the so-called collective model of Chiappori (1988) which imposes the restriction that decisions are made in an efficient fashion. Browning and Chiappori (1998) have studied some of the features of household demand systems that emerge from such a framework and used a QAIDS model very similar to the one we use to exemplify their results.

In the collective model, efficiency implies that the household maximizes a weighted average of the utility functions of the household members with weights reflecting the relative power the different members have within the family. We will argue below that a conditional cash transfer targeted to women might shift the weights in favour of women and therefore change the nature of the demand system. Browning and Chiappori (1998) show that under certain circumstances such a shift can take the form of a change in the intercept and possibly the slope of the Engel curves.[4] If one does not allow for the effects of what Browning and Chiappori (1998) call 'distribution factors', of which a conditional cash transfer can be one, these would be reflected in changes in the unobserved component u.

3. Consumption in rural Colombia.

The survey we use for the estimation was collected to evaluate a welfare programme sponsored by the World Bank and the Inter American Development Bank. The programme, modelled after the Mexican PROGRESA, consists of conditional cash transfers targeted to poor

[4] The coefficient on prices, which we do not estimate explicitly here, also changes.

households living in small towns with certain features. To evaluate the effects of the programme, two waves of a large data set were collected over a period of one year. In the first part of this section we describe the programme and discuss its potential effects on consumption. We then move to describe in detail the dataset that we use in the estimation of Engel curves. We finally present the effects of the programme on household consumption.

3.1. The programme and its potential effects on consumption

In 2001 the World Bank and the Inter American Development Bank decided to sponsor a large welfare intervention in Colombia, inspired by the Mexican conditional cash transfer programme PROGRESA. As in PROGRESA, the *Familias en Acción* (*FeA* from now on) programme consists of three components: health, nutrition and education. The 'nutrition' component is a cash transfer (40,000 pesos per month or 15 US$) eligible households receive if they have children under the age of 6 and participate into the health component. The latter consists of a number of growth and development check ups for young children, a vaccination programme and some courses for mothers on various health issues. The education component consists of grants for school age children that are received by each child who attends regularly school. The grant is 14,000 monthly pesos (5.5 US$) for primary school children and 28,000 (11 US$) for secondary school children. As in PROGRESA, the money is received by mothers.

The programme was first targeted geographically. Of the 900 odd municipalities in Colombia, 627 were chosen as targets. The targeted municipalities had to have less than 100,000 inhabitants, they could not be department capitals, had to have enough education and health infrastructure, had to have up-to-date lists of welfare recipients and had to have a bank in town. Within each municipality eligibility was established using the so-called SISBEN indicator. SISBEN is an indicator of economic well-being that is used throughout Colombia for targeting welfare programmes as well as for the pricing of utilities. In theory, each Colombian household is classified periodically in one of six levels, on the basis of an indicator determined by the value of several variables periodically measured. In the case of *FeA*, only households in the first level of SISBEN as of December 1999 were eligible. Eligible households, which in what follows will be referred to as SISBEN 1, constitute approximately the bottom twenty percent of Colombian households living in rural areas (see Vélez, Castaño and Deutsch, 1998).

The programme started, with a few exceptions, in the second half of 2002 and the take up among eligible households was over 90%. By 2003 about 340,000 households throughout Colombia were covered by the programme. *FeA* was subsequently expanded to larger towns and as of the end of 2008, more than 1.5 million households were involved in it.

The program is now very visible and probably constitutes the largest social intervention in Colombia. For the households in our sample, the grant received constitutes typically about 20% of household monthly consumption for participant households, and is thus likely to have an important impact on their consumption.

3.2. The data set

As the *FeA* programme was started, the Colombian government decided to launch a large scale evaluation of its effects. The evaluation work started with the collection of a large scale data base in 2002. The evaluation is based on the comparison of SISBEN 1 households in municipalities targeted by the programme (hereafter 'treatment areas') to SISBEN 1 households living in 'control' municipalities. As the random allocation of the programme was not feasible, the evaluation survey was constructed by first choosing a stratified random sample of targeted communities. The stratification was done on the basis of geographic areas and the level of health and education infrastructure, for a total of 25 strata. Within each of these strata, the evaluation team chose 'control' municipalities that were as similar as possible to the municipalities included in the 'treatment' sample in terms of size, population, an index of quality of life as well as health and education infrastructure. In each of the municipalities in the sample, 10 geographic clusters were randomly drawn, with weights proportional to the population, of which three clusters are in the urban center (*cabecera municipal*) and seven are more rural. Finally, in each of the clusters, about 20 households were drawn from the SISBEN 1 lists. Given non response rates and mobility about 10 households per cluster entered the final evaluation sample, which was, in the end, made of about 11,500 households living in 122 municipalities, of which 57 were 'treated' and 65 used as 'controls'.

The data set, in addition to standard demographic variable, contains very detailed information on consumption, which allows us to use these data for the estimation of Engle curves. The data set contains information on both quantities and, if bought, value of many commodities. This structure allows us to measure accurately food consumption by imputing the value of consumption in kind, which, as we shall see, is very important for this population.

For the estimation of the Engel curves, we use data from a *baseline* survey -collected *before* the programme started- in order to investigate household consumption patterns that are not affected by the programme. Political pressure resulted in the programme starting in 26 out of 57 treatment municipalities before the fieldwork commenced. Because of this we dropped from our sample the households living in such 'early-treatment' areas. We will make use of data from the baseline and follow-up surveys (collected *before* and *after* the implementation of the programme on the *same* households) to compare the estimated effect of *FeA* on consumption to the implications of Engel curves on the same outcome variable.[5]

We kept in our working sample only households for which we do not have missing information for expected income, which, as we mentioned above, we use as an instrument for total expenditure. Our chosen instrument is missing for about 20% of the sample, thus reducing the size of our final sample to 5,218 households.[6] The results obtained by estimating Engel Curves via OLS before and after this selection step turned out very similar. Moreover, by comparing changes in consumption from before to after the implementation of the programme across treatment and control areas using a difference in differences approach (see Section 3.3), the results we find are similar to those in Attanasio and Mesnard (2006). This we took as informal evidence to rule out selection issues in the data due to the missing instrument.

Tables 1a and 1b summarize area and household level variables for our final sample. From now on we will consider geographical clusters defined by *villages* instead of municipalities, as some neighbouring municipalities that are very small and adjacent were grouped together to form the same cluster for the analysis. This geographical definition of areas led us to 75 villages in our working sample, 25 of which were treated. In Table 1a we notice that about 55% of our sample lives in the '*cabecera municipal*' (the urban centre of the municipality). We will be defining these as 'urban' households, although the villages in our sample are relatively rural and small, the average village having less than 30,000 inhabitants. The location of the villages in our sample is spread all over Colombia, with a relatively smaller proportion of villages (12%) in the Pacific Region. The average altitude of our villages is about 650 meters above sea level. However, there is a large dispersion around this mean (750 meters) reflecting the large geographic diversity of Colombia. A large proportion of households in the villages in our sample are not connected to the sewage system (44% on average) and 13% of them do not have piped water. These numbers are indicative of the relatively high poverty levels in these villages. It is worth noting that, perhaps not surprising given the way control municipalities were chosen for the evaluation, the distribution of the variables reported in Table 1a turned out the same in treatment and control villages: a binary regression of the indicator for treatment villages on the full set of village characteristics led to a p-value of the F statistic for their joint significance of about 32%.

[5] The attrition rate between the baseline and follow-up surveys is particularly low (less than 6%) due to the very low mobility rate in our sample of very poor households living in rural areas and to the special efforts that have been put in place to track households who do not live in the same place at follow-up (see Attanasio and Mesnard, 2006).

[6] As a consequence of this selection, tests showed that variables such as the number of children aged 7-11, number of adults above 60, head having less than a primary level, total consumption and total food are significantly different between the sample we use and the full sample.

Mean and standard deviation of household characteristics are reported in Table 1b. On average, the households in our sample have 6 members. The large majority of them do not have a good health insurance, as only 5% of them benefit from an unsubsidised health insurance, which is typically associated to having a good job in the formal sector. Moreover, only 4% of their heads or spouses have more than a secondary education level and just above 20% of them have less than a primary education. Only few household characteristics were found to significantly explain the probability of living in treatment villages, the latter being characterized by a larger proportion of households with spouses having less than a primary education level and children under 7 years old.

Finally, in the last three rows of Table 1b we also report the mean, at baseline, of total consumption, food consumption and the share of food. Food is a very important component of consumption of the very poor households in our sample, with an average share of about 70%. Note that these figures do not include only expenditure, as we imputed the value of commodities consumed but not purchased (i.e. produced or received as a gift) using information on food prices in the municipality. This measure of 'food-in-kind' represents around 18% of food consumption for the 75% of households in the sample who report consumption-in-kind. All these household characteristics point to the fact that the households in our sample are amongst the poorest households in rural Colombia.

As we mentioned above, to instrument total household consumption we use a measure of 'expected future income' constructed from answers to questions about subjective expectations about next month income. As already argued, because of the way the questions are phrased, our measure of expected income is exogenous to household labour supply choices and reflects household expectations of labour market conditions and wages that should be taken as given by the individual household. Because of this consideration and because it should not reflect temporary shocks, such a variable should represent a good candidate to instrument total expenditure while modelling Engel curves. However, as the variable is somewhat unusual and these types of questions are not often found in standard household surveys, it is important to check how this variable co-varies with standard variables, such as income and consumption.

In Table 2 we report the estimates of the coefficients of a regression of consumption and income on our measure of expected income. These regressions are run both at the individual and at the village level. In either case, we can see that our measure of expected income covaries significantly with both income and consumption. For identification, of course, what is important is the significance of the instrument in a regression of total consumption after controlling for additional exogenous variables. When we run such a first stage regression with the additional controls we will use in the estimation of the Engel curve, we find that the instrument is strongly significant, as shown in the last row of Table 2. This first stage regression, whose complete results are not reported for the sake of brevity, has an R-squared of 21% and the t-statistic for the instrument is 21.

3.3. Effects of the Familias en Acción programme on household consumption

A first step in our analysis is to quantify the effect of the programme on total consumption and food consumption. As the programme was not allocated randomly between treatment and control municipalities, we need to control for differences among them. We do so using a conditional difference in difference approach that exploits follow up information for the households in our working sample as in Attanasio and Mesnard (2006), who look explicitly at the effect of the programme on consumption.

We estimate the impact of the programme on eligible households by using the following parametric specification:

$$\Delta y = \alpha_0 + \alpha_1 D + \alpha_2 X + \varepsilon \ , \tag{4}$$

where Δy is the change in the outcome variable between baseline and follow up, X is the set of control variables at both individual and village level described in Table 1 (measured at baseline) and D is a dummy for households living in treatment areas. The programme impact is given by the coefficient α_1, which estimates the average treatment effect for households living in treatment villages. Estimation of equation (4) is an application of a difference in difference approach and provides consistent estimates of the programme impact under the assumption that, conditional on the control variables X, there would have been no differential trends in consumption between control and treatment villages in the absence of the programme. The validity of this assumption in the context of *FeA* is discussed at length in Attanasio and Mesnard (2006).

To check the sensitivity of results, we estimate equation (4) using different techniques. First, we consider a simple OLS regression and implicitly assume homogeneous impacts of the programme. Second, we allow for heterogeneous programme effects by adding interactions of D with the X's, and calculate the effect of interest by taking the average of the X's across treatment observations. Finally, we check the robustness of our results to possible support problems in the distribution of the X's in treatment and control areas. To this end, we compare the average of Δy for households in treatment villages to the weighted average of Δy for "similar" households in control villages, the similarity being defined with respect to the *propensity score* P[D=1|X]. Weights are defined using a Gaussian kernel truncated at a 1 percent distance, resulting in higher weights for households that are most similar with respect to the propensity score. As a result of this matching procedure only about 1% of the households in treatment villages were discarded from the analysis.

Table 3 reports the impact of the programme on logged total and food consumption, as well as on the share of budget devoted to food consumption, using OLS, fully interacted OLS and matching. The impact of the programme on total consumption is estimated at 13.3% while that on food consumption is estimated at 15.9%. Both effects are statistically different from zero. These results are reasonably similar to those obtained with the other two methods.

13.3% of total consumption is about 60,000 pesos, which compares with an average monthly grant (conditional on being paid) of about 100,000 pesos. This implies either that a part of the grant is saved or that there is a reduction in other sources of income. These point estimates are slightly below those obtained for other conditional cash transfer programmes, such as PROGRESA, where the effect on consumption corresponded to roughly 80% of the grant.

The effect of the programme on the share of food is estimated at around 1% but is not statistically different from zero. The questions as to whether this would be consistent with a linear or quadratic Engel curve and what can be inferred from such results are examined in the remainder of the paper. Here we only notice that, although we used a selected sample relative to the one used in Attanasio and Mesnard (2006) (because for some households the 'expected income variable was missing), the results we obtain are largely consistent with those reported in that study.

4. Estimating Engel curves for food

In this section, we present the estimation results for the food Engel curve. The purpose of this part of the analysis is mainly methodological. We present the results obtained with some approaches commonly used in the literature and that are valid only under very stringent assumptions. We then argue that some of these assumptions are violated and show how the use of appropriate methods makes a substantial difference in practice.

4.1. Basic specification: no price effects and exogenous expenditure

As we have mentioned above, following BBL, we use a Q-AIDS model for the share of food. One of the main purposes of the analysis will be to determine the shape for the food Engel curve in our population. In particular, we will be interested in whether for a sizeable fraction of the population food is actually a luxury, so that its share increases with total spending. A finding of this nature could potentially explain why, on average, the share of food stays constant or increases slightly.

We start by re-writing equation (3) under a very strong assumption, namely that consumers in our cross section face the same relative prices, which we normalize to one, regardless of the town in which they live. This implies the following specification for the food share:

(5) $$w_f = \alpha_f + \beta_f \ln m + \lambda_f \ln m^2 + \theta_f ' z + u_f \, ,$$

where the *f* index stands for food and we omit the individual index for simplicity. The vector z includes controls, such as demographic variables, that enter the demand system as determinant of the intercept of equation (3).[7]

If we are willing to assume that total consumption is uncorrelated with the residual term u_f, we can estimate equation (5) by OLS. We report the results of such an exercise in Table 4. In Figure 1, we plot how the share of food varies with (logged) total expenditure according to the estimates reported in Table 4. Here and in what follows, in plotting the Engel curve, we set the intercept at an arbitrary point, so that the only relevant information is the shape of the curve. A remarkable feature of the estimated Engel curve is that both coefficients on the linear and quadratic are strongly significant. Moreover, the estimates imply that the share of food increases at low levels of total expenditure and starts decreasing at levels that are close to the 10th percentile of expenditure in our sample (between 5 and 5.5 in Figure 1). Effectively, these estimates imply that food is a luxury for 10% of our sample and a necessity for the remaining households.

This result appears to be a feature already discussed in other studies that use data from developing countries. For example, Kedir and Girma (2008) using data from the Ethiopian Urban Household Survey find that for a non-negligible proportion of households the share of food increases at low levels of the total expenditure distribution. There are several possible explanations for this finding. First, it could be that for these very poor households the basic necessity is constituted by housing expenditures (rent, utilities) and that the rest of what they consume goes into food. Increases in total expenditure, therefore, are translated into increases in the share of food, as these households start switching from diets based almost entirely on basic staples (rice, potatoes and so on) and start introducing more frequently foods richer in proteins (chicken, beef and so on).

Another possible explanation could be data quality for the households reporting very low levels of total expenditure. However, Kedir and Girma (2008) find that the curvature of the Engel

[7] We also experimented with the possibility that the demographic variables enter the 'constant' of the price index a(p) of the same equation (3). In this case, as with the price effects, the demographics were interacted with the log of total expenditure and its square. The results we obtained with this richer specification, which we do not report for the sake of brevity, were very similar to those presented here.

curve is robust to the presence of measurement error in the data. Errors in survey reports of food expenditures induce a complicated form of non-classical measurement error in the estimation of Engel curves, as both the right hand side and left hand side variables are error ridden. To account for this, the study by Kedir and Girma (2008) exploits the approach suggested by Lewbel (1996). By adapting the same approach to our particular analyses with and without price heterogeneity we found no significant effect of measurement error on the results presented in this and the next section, thus concluding that the non-linearity of food Engel curves estimated by OLS is robust to the presence of such non classical type of measurement errors (results available on request).

A third explanation may be that the shape is induced by quality and food composition effects, so that households at low level of total consumption increase the consumption of better and more expensive food items when total consumption increases.[8] Or, finally, it might be induced by the fact that we are ignoring endogeneity of total consumption and possible price effects. It is to these issues that we now turn.

4.2. Heterogeneous prices across villages

We start with prices and, as we mentioned above, we try two different approaches. First, we proxy prices with village level dummies. As it is clear from equation (3), in the case in which the coefficient on the quadratic log expenditure term is non-zero, prices interact both with the linear and quadratic terms. We therefore estimate the following version of equation (3):

$$(6) \qquad w_f = \alpha_f^v + \beta_f^v \ln m + \lambda_f^v \ln m^2 + \theta_f{}' z + u_f$$

where the v superscript stands for village. The village specific coefficients in equation (6) are estimated adding to the regression village dummies and their interactions with the linear and quadratic expenditure terms. We use the same vector of controls as in the previous specification and, as before, we ignore the possible endogeneity of $\ln m$ and estimate equation (6) by OLS. We parametrized the village dummies so to interpret the coefficients on the linear and quadratic terms as the average coefficients in the sample. We report them in Table 5 and plot the profile of the Engel curve in Figure 2.

Remarkably, the average coefficients in Table 5 are not too dissimilar from those in Table 4. As a consequence, the shape of the implied Engel curve implied by the average coefficients is not very different from that in Figure 1. However, there is substantial variation in the estimated coefficients across villages. To document this heterogeneity, in Figure 3 we plot the deviation of

[8] For example, Bhalotra and Attfield (1998) present evidence of quadratic Engel curves for milk consumption in rural Pakistan.

the village coefficients (that is, the coefficients of village dummies and their interaction with the linear and quadratic terms in log expenditure) from their average.

While Figure 3 indicates a substantial amount of heterogeneity, it is not very informative about the variation in the shape of the village level Engel curves, which is ultimately what we are interested in. Rather than plotting the 75 Engel curves implied by these coefficients, we predicted for each observation in the data the value of the first derivative of the Engel curve. This will vary because total expenditure varies across households and because the coefficients of the Engel curve vary across villages. In our sample, both the mean and median first derivative of the Engel curve across households are negative (at -0.030 and -0.032 respectively). The standard deviation of first derivatives is substantial (0.076) and around 50% of the cross sectional variance is explained by variability across villages. A substantial fraction of the households (27%) has a positively sloped Engel curve for food.

To investigate the issue further, in Table 6 we report the first derivative of the Engel curve estimated from (6), that is $\beta_f^v + 2\lambda_f^v \ln m$, calculated at specific percentiles of the distribution of $\ln m$ in our sample. As the value of this derivative is allowed to vary through the village specific coefficients β_f^v and λ_f^v, we report its mean and median across villages weighted by the size of the village population. In the last row of the table we also report the percentage of negative first derivatives that at each percentile of the total expenditure distribution.

The evidence emerging from this table indicates that for a substantial part of our sample food is a luxury. This is particularly true at low levels of total expenditure: in this sense, this evidence is consistent with the one in Figure 1. What is remarkable is that even at relatively high levels of total expenditure, the heterogeneity of coefficients across villages induces the derivative of the Engel curve to be positive for more than 20% of the observations. To sum up the evidence so far, there is evidence of heterogeneity across villages, induced by differences in relative prices, although price differences does not affect the shape of the Engel curve on average.

4.3. Allowing for endogenous total expenditure without price heterogeneity.

So far we have assumed that the log of total expenditure is uncorrelated with the residuals of the share equation. We now allow for the presence of correlation, which can be caused by any of the reasons discussed in Section 2.2. To obtain consistent estimates of the coefficients of interest we used a Control Function approach. We experiment with two instruments: average town wages, as in Attanasio and Lechene (2002), and expected future income, as we discussed above.[9] The results were very similar, whatever instrument we chose. Therefore, in the remainder of the paper, we only report the results obtained using expected future income, which allows us to estimate village level intercepts.

To implement the Control Function approach we first run a regression of log total expenditure on the same controls included in the share equation and the instrument. We then add to equation (5) a third order polynomial in the estimated residuals of the first stage regression.[10] We report the results we obtain in Table 7. In Figure 4, which compares with Figure 1, we plot the estimated Engel curve.

The striking feature of the results in Table 7 is that, unlike in Tables 4 and 5, both the sign of the coefficient on the linear term and the sign of the one on the quadratic term are negative, implying a very different profile in Figure 4, and one that is now close to being linear. No households are on an increasing portion of the Engel curve. However, it should also be noticed that, while jointly significant, neither of the two coefficients is statistically different from zero. This is particularly true for the coefficient on the linear term. This evidence seems to indicate, therefore, that a linear specification would probably achieve the same fit as the quadratic specification. We will discuss the relative fit of a quadratic and linear Engel curve in Section 5. Here we notice that the hump that characterized the food Engle curve has now disappeared once we control for the possibility that total expenditure is correlated with the residuals of the share equation.

4.4. *Allowing for price heterogeneity and endogenous expenditures*

In Sections 4.2 we allowed prices to be different across villages and the resulting estimates indicated a substantial amount of heterogeneity in the shape of the Engel curves. The next step in our analysis is to allow simultaneously variable prices and endogeneity of $\ln m$. To do so, we now estimate equation (6) by a control function approach. In particular, to the variables in equation (6) we added a third order polynomial in the residuals of the first stage regression (obtained as described in the previous section and controlling for village dummies) and its interactions with village dummies. A summary of the results obtained in this estimation is reported in Table 8 and in Figures 5 and 6, which can be compared to Table 5 and Figures 2 and 3, respectively.

The estimates reported in Table 8 are different from those in Table 7. However, the shape of the implied Engel curve does not change substantially, as can be seen comparing Figure 5 and

[9] We have already discussed in Section 3 the correlation between 'expected income' and total expenditure (see Table 2).

[10] Our results proved robust with respect to the choice of the order of the polynomial.

Figure 4. As with Figure 4, over the range of expenditures observed in our sample the Engel curve declines monotonically. This implies that, if there were no heterogeneity in coefficients, food would be a necessity for the whole sample. Moreover, the fit obtained from the quadratic model is not too different from that from a linear model. Figure 6, however, indicates the presence of a substantial amount of variability in the coefficients on $\ln m$ and its square. It is therefore important to consider the distribution of the first derivative of the Engel curves across income levels and across villages. We did so by replicating the same analysis presented in Section 4.2.

While both the mean and median first derivative of the Engel curve across households are negative (-0.10 and -0.11, respectively), we found that it is positive for 14% of our sample. The standard deviation of first derivative is even larger than in the OLS case at 0.129. Of the total variance of these derivatives, 43% is accounted for by variability across villages. In Table 9, which replicates the analysis in Table 6, the mean and median derivative (across villages) is negative at all percentiles considered. However, not the entire distribution is negative. A remarkable number of households are on an increasing portion of the Engel curve. Surprisingly, the proportion of negative values of the first derivative of the Engel curve does not increase monotonically with total expenditure. Instead, after the 75th percentile, this percentage decreases and is only equal to 76.9.% at the 99th percentile. Taken at face value these results indicate that, while not as common as indicated by the simple OLS regressions, there is a non negligible number of households for which food is a luxury in our sample, even after allowing for variable prices and endogenous total expenditure.

The remaining issue, of course, is to establish the extent to which this result is driven by imprecisely estimated coefficients. The first derivative of the Engel curve can only change sign if the (village specific) coefficients on the quadratic term are different from zero and with a different sign from the coefficient on the linear term. From a statistical point of view, the quadratic terms are jointly significant. From an economic point of view, the interesting issue is not just whether the quadratic terms are significantly different from zero, but whether the coefficients imply, for a sizeable fraction of households, a non-monotonic Engel curve. It could, for instance, be the case that the observations with a positively sloped Engel curve correspond to insignificant quadratic terms and extreme values for expenditure.

4.5. Parametrizing price effects

In this section we put some structure on the village dummies that control for the variability of (relative) prices across villages. In particular, we assume that the vector of relative prices faced by villages can be completely spanned by a linear combination of a vector of nominal prices of certain important commodities and some village level variables. In particular, we assume that the variability across villages of the price indexes in the demand system can be controlled for by the log price of potatoes, rice, coffee as well as the average level of men wages and some other village level variables. We chose potatoes, rice and coffee as we have good quality data on their prices, they are widely consumed by most households and their prices are not too correlated. Village level variables include population size, altitude and its square, an index of quality of life in 1993. We therefore estimate the following equation:

$$(7) \qquad w_f = \alpha_f(\xi_v) + \beta_f(\xi_v)\ln m + \lambda_f(\xi_v)\ln m^2 + \theta_f{}'z + u_f ,$$

where ξ_v is a vector of village level variables, including the representative prices mentioned above. As in equation (6), we allow the demographics to enter only the intercept of the demand for food and not the price indexes, although allowing for these effects does not change the results substantially. We express the village level variables ξ_v in terms of deviation from the mean, so that we can interpret the coefficients on log total expenditure and its square as the average coefficient across villages. In Table 10 we report estimates of (7) using a control function approach and results can thus be compared to those presented in Table 8. The profile of the Engel curve implied by these coefficients is in Figure 7.

The average profile is monotonically decreasing for the whole sample, very similarly to what we observed in Figure 5. The estimates reported in Table 10 are very similar to those in Table 8. The quadratic terms are mostly not statistically different from zero, while being jointly significant. This suggests once again that a linear specification would probably achieve the same fit as the quadratic specification. Table 11, which is the equivalent to Table 9 for the case of equation (7), implies that 99.5% of the first derivatives of the Engel curve at the household level (that is $\beta_f(\xi_v) + 2\lambda_f(\xi_v)\ln m$) are negative: for almost all observations food is a necessity. This is particularly true when excluding the extreme observations in the lowest and highest percentiles in our sample. This result contrasts with what we obtained when using village dummies to control for price effects, in which case the first derivative of the Engel curve was positive for about 14% of the households of the whole sample. This difference may suggest that the result obtained with village dummies is probably driven by the imprecision with which some coefficients are estimated and by the fact that some extreme values of expenditure can yield peculiar slopes for the Engel curve.

In general, it is interesting to establish the extent to which the shape of the Engel curve will change (and how much the fit of the overall equation worsens) when we impose a linear specification. In the next section we tackle this issue directly after assessing the robustness of our results.

5. Preferred specification

Given the evidence presented so far, it is worth exploring the possibility that the Engel curve for food is linear in log total expenditure. An advantage of such a specification is that, even in the presence of heterogeneous prices across villages, we do not need to interact their proxies with the total expenditure terms. Intercept shifts will be sufficient.

In this section, we thus run a formal test to compare linear vis-à-vis quadratic specifications when prices are controlled for by means of village dummies (as in Section 4.4) or using the set of variables considered in Table 10 (as in Section 4.5). To this end, we re-estimated specifications (6) and (7) using a control function approach and imposing that the coefficients associated to the quadratic terms in logged expenditure are all zero. Results for the specification with village dummies and specification with representative proxies for prices are reported in the first and second column of Table 12, respectively. The estimated linear coefficients are remarkably similar.

To test for differences between the linear and quadratic models, we compared the slope of the estimated Engel curves implied by the quadratic specification with village dummies (6), $\beta_f^v + 2\lambda_f^v \ln m,$ to the slope implied by the linear specification reported in the first column of Table 12. Similarly, we compared the slope of the quadratic specification with village level variables (7), $\beta_f(\xi_v) + 2\lambda_f(\xi_v) \ln m,$ to the slope reported in the second column of Table 12. In Table 13 we report the 95% confidence intervals for the difference in slopes at different percentiles of total expenditure, obtained via bootstrap and taking into account possible dependences within villages. As it is clear from the table, the linear specification does not impose strong restrictions on the data and its fit is similar to the quadratic specification. There is some weak evidence that linearity may matter at low percentiles of the total expenditure distribution in our sample, but only for the specification with village level variables, as it is shown in the second column of Table 13. The conclusion of this section is that a linear specification provides, in our data, an adequate specification. To estimate adequately the food Engel curve with this specification, however, we need to take into account price heterogeneity and the endogeneity of total expenditure.

6. Using Engel Curves to predict changes in the structure of consumption

Having estimated the Engel curve for food, we can now use it to predict changes in the structure of consumption induced by the programme. The aim of this section is twofold. First, we will consider a general setting that allows us to compare the quasi-experimental results discussed in Section 3 with the "structural" implications of our preferred specification for the Engel curve. Second, we will show how the comparison can be implemented for the case at hand, and how discrepancies between *ex-ante* (structural) and *ex-post* (quasi-experimental) results may be informative on the behavioural response of individuals to the programme.

6.1. A potential outcomes framework

Let the following *potential* Engel curves (PECs) be defined:

(8) $$E\left[w_j(t) \mid m_j(t), u_j(t), d\right] = \alpha_j + \beta_j \ln m_j(t) + u_j(t),$$

where d is a dummy for treatment villages, t=1,2 (baseline and follow up, respectively) and j=0,1 ("no programme" and "programme" regimes, respectively). For notational simplicity we omit from equation (8) variables, such as demographics, that are actually considered in our Engel curve (see, for instance, equation (5) or equation (6)). The quantities $w_j(t), m_j(t), u_j(t)$ are potential outcomes that are observed depending on whether or not the programme is in operation (see Rubin, 1974). For example, $m_0(t)$ denotes the value of total expenditure that would be observed at time *t* in the absence of the programme, and $m_1(t)$ the value of total expenditure that would be observed at time *t* the programme being in operation. A similar interpretation applies to the remaining quantities. The variables $u_j(t)$ can be interpreted as representing unobserved preference heterogeneity. The possibility that $u_j(t)$ is correlated with total expenditure motivated the necessity to instrument such a variable, discussed at length in Section 4.

Equation (8) implies that, if all quantities involved were observed, a standard regression would identify the structural parameters of the Engel curves relevant in the different time periods and in the two groups of villages. Implicit is the assumption that the shape of the Engel curve is constant over time, though it is allowed to change as a result of the programme. In the empirical application we allow for village specific intercepts, which could be justified, for instance, by differences in relative prices. The Engel curves are allowed, therefore, to be somewhat different between treatment and control villages, as the difference in difference results reported in Table 3 allow for baseline differences. However, in the setup considered the *shape* of the curve does not depend on the village status with respect to the programme, and the linear specification is maintained as discussed in Section 5.

There are *four* groups of observations defined by the evaluation design considered in Section 3, resulting from the combination of control and treatment villages in the baseline or follow-up periods. For each group equation (8) defines *two* potential Engel curves resulting from participation and non-participation. The following variables are observed:

$$w(t) = w_0(t) + d \times 1(t=2)[w_1(2) - w_0(2)],$$
$$m(t) = m_0(t) + d \times 1(t=2)[m_1(2) - m_0(2)],$$

where $1(t=2)$ is an indicator for observations in the post-programme period. In words, $w_0(1), m_0(1)$ are observed at baseline for households in all villages, while $w_0(2), m_0(2)$ and $w_1(2), m_1(2)$ are observed at follow up for households in control and treatment villages, respectively. Using this notation, the causal effect of the programme on $\ln m$ (log expenditure) and w (budget share for food) was identified in Section 3 by making the following two assumptions:

$$E[\ln m_0(2) - \ln m_0(1) \mid d=0] = E[\ln m_0(2) - \ln m_0(1) \mid d=1],$$
$$E[w_0(2) - w_0(1) \mid d=0] = E[w_0(2) - w_0(1) \mid d=1].$$

Note that equation (8) and the definition of the PCEs imply that the latter condition is equivalent to:

$$E[u_0(2) - u_0(1) \mid d=0] = E[u_0(2) - u_0(1) \mid d=1].$$

The assumptions above embody the idea that any change in total expenditure and preferences over time would have been the same in the two groups of villages had the programme not been implemented. The impact estimates presented in Section 3, effectively, make use of a version of this restriction, conditional on a large set of observed pre-programme household and village characteristics, which are not considered explicitly here. Under the conditions stated the difference in differences estimates identify the following parameters (or impacts):

$$\Delta_m = E[\ln m_1(2) - \ln m_0(2) \mid d=1],$$
$$\Delta_w = E[w_1(2) - w_0(2) \mid d=1],$$

which we interpret as the causal effect of the programme on (logged) total expenditure and on the budget share for food, respectively. These were the results presented in Table 3. Using the definition of the PCEs it is easy to show the following relationship between the Engel curve parameters and the estimated impacts:

(9) $$\Delta_w = (\alpha_1 - \alpha_0) + \beta_1 \Delta_m + (\beta_1 - \beta_0) E[\ln m_0(2) \mid d=1] + \Delta_u,$$

where:

(10) $$\Delta_u = E[u_1(2) - u_0(2) | d = 1].$$

Equation (9) establishes the relationship between the change in the budget share Δ_w that we should expect through the estimated Engel curve and the change in total expenditure Δ_m induced by the programme. Notice that, in general, the left-hand side of equation (10) is not observable as $u_0(2)$ is not observable. Δ_u cannot be obtained performing a difference in difference of the estimated residuals without additional assumptions.

If the programme does not induce a change in the parameters of the Engel curve, so that $\alpha_0 = \alpha_1$ and $\beta_0 = \beta_1$, then equation (9) simplifies to:

(11) $$\Delta_w = \beta_0 \Delta_m + \Delta_u,$$

If one is willing to assume that the program does not affect the two parameters of the Engel curve, then one can use equation (11) to estimate Δ_u. And, if one thinks that the programme does not induce a change in preference shifters *u,* one could use equation (11) and the impact estimates to get an estimate of the slope of the Engel curve, which, given the evidence in Table 3, would be zero or positive.[11]

To check for possible changes in the shape of the Engel curves induced by the programme, we estimate the Engel curve in each of the four cells (treatment vs control, baseline vs follow-up) separately and test for the equality of the coefficients across cells.[12] The results presented in Table 14 suggest that the coefficients are not statistically different across cells at the conventional level (see also Figure 8 where we report the four estimated curves). We can therefore proceed under the tentative assumption that the intercept and the slope of the estimated curves are not different across the four cells. We then also discuss the alternative of taking the point estimates of the parameters at face value.

6.2. Structural vis-à-vis non-experimental estimates of programme effects

Table 3 indicates that log consumption increased between 13% and 15% depending on the estimation strategy adopted. We can then input this figure for Δ_m into (11) to predict the effect

[11] Notice that even if we pool all the data together (baseline and follow up, treatment and control) we will not be using any of the *variability induced by the programme* in total expenditure and food shares in estimating the slope of the Engel curve. This is because we allow for village and time specific intercepts to capture the effect of prices. The Engel curve is identified by individual variation in food shares, total expenditure (and the instrument) within treatment/ control and baseline/ follow up cells.

[12] Because of the strategy adopted, we need to make the assumption that these parameters are heterogeneous only across villages, but constant within each village. This assumption rules out any discussion about the LATE (Imbens and Angrist, 1994) interpretation of our point estimates.

that the programme should have had on the share of food. As we have discussed at the end of Section 5, our favourite specification for the Engel curve is piglog linear. In particular, we have argued that the results reported in Table 12 represent our favourite specification, implying a value for β_0 around -11%.

The results on the estimation of the Engel curve reported in Table 12 imply that, food being a necessity, the share of food would *decrease* as a consequence of a positive shift in Δ_m induced by the programme. In contrast, the quasi-experimental estimate of Δ_w in Table 3 indicates a slight *increase* in the share of food, as its value is set at about 1% (though not statistically different from zero).

How does one interpret and reconcile this evidence?[13] Using the relationship in (11), the difference in the two predictions can be explained by a positive value of Δ_u, that is, by a positive effect of the programme on the U's that offsets the negative value of $\beta_0\Delta_m$. The U terms can be thought as "preferences", which for example might have changed because the programme is likely to change the balance of power within the household and, in general, change the way choices are made. One possible reason for this discrepancy is thus a failure of the unitary model behind the derivation of the Engel curve. It is possible, instead, that household decisions are reached taking into consideration the utility function of more than one agent. The result would then be evident in the case of the Familias en Acción because the programme does not only increase household resources but also targets them to women. Attanasio and Lechene (2002, 2008) propose this explanation in the case of the Mexican PROGRESA.

Our results, which are also consistent with the evidence reported from all Conditional Cash Transfers programs about the impacts of these programmes on consumption and its composition (see for instance, Attanasio and Lechene (2002) for rural Mexico, Schady and Rosero (2006) for Ecuador; Angelucci and Attanasio (2009) for urban Mexico, Macours et al. (2008) for Nicaragua and Attanasio and Mesnard (2006), for Colombia), might suggest that such programmes change, somehow, the decision process. One possibility, mentioned in several of these papers, is that the programme changes the relative weight given to women in the decision process. Another possibility is that information changes. Without additional evidence and a more structural analysis it is difficult to establish how.

[13] The results in Table 14 rule out the possibility that the difference between the predictions from Table 12 and those obtained by diff in diff in Table 3 are due to the fact that the former refer to treatment *and* control towns while the latter refer only to control towns.

A closer look at Table 14 indicates that the point estimates of the α's and the β's are remarkably similar for three of the four cells. The one cell that stands out, although the relatively low precision of the estimators imply that the estimates are not statistically different from the others, is the 'treatment' cell at follow up: the only one where the programme was operating. For this particular cell, the intercept goes from 2.0 to 1.4 (a decrease of more than 25%) and the slope almost halves (in absolute value) from -0.10 to -0.06. These changes can explain, potentially, the finding, even without a change in the *U*'s. To see this, we can use equation (9). If we take Δ_w and Δ_m from Table 3 and the α's and β's with and without the programme from the second and fourth row of Table 14, we find that the small positive (or zero) left-hand side of equation (9) could be generated by the combination of the first two terms of the right hand side ($\alpha_1 - \alpha_0$ and $\beta_1 \Delta_m$) being negative and the third term ($(\beta_1 - \beta_0) E[\ln m_0 | d = 1]$), being positive.

To conclude, we speculate that the misspecification of the Engel curve that we detect because of their inability to predict the change in food shares induced by the program can stem from a misspecification of the model that fails to take into account distributional factors that shift the power within the family. As we suggested in Section 2.3, conditional cash transfers targeted to women could be equivalent to shifts in the unobserved component of the Engel curve captured in equations (2) and (3). The small increase (and insignificantly different from zero) in the share of food estimated in Table 3 combined with a sizeable increase in total consumption is suggestive of shift in preferences. This might be linked to a shift in the balance of power within the household, as speculated in Attanasio and Lechene (2002).

7. Conclusions

In this paper, we have studied the shape of food Engel curves in a data set collected to evaluate the impact of a large welfare programme in Colombia. *Familias en Acción* is a conditional cash transfer (CCT) that has become one of the main social programmes in Colombia. We use the evaluation data set to study consumption patterns among the poor population targeted by the programme. In particular, we estimate food Engel curves for this population. Such an exercise, and the availability of quasi-experimental estimates of the impact of the programme on total expenditure can be used to predict the effect of the programme on the share of food. This prediction and the quasi-experimental estimates can then be used to validate the specification of the Engel curves.

The first aim of the paper was, therefore, methodological. We wanted to establish the best specification for food Engel curves in our population and the best technique to estimate their parameters. In this respect we established that it is crucial to control for the endogeneity of total expenditure and to control for the (unobserved) variability of prices across towns (at the same point in time).

We conclude our analysis of food Engel curve by saying that, in our data set, the best fit seems to be provided by a log-linear specification (estimated by a control function method) with a coefficient of -0.1 on (log) total expenditure. This implies that, *coeteris paribus*, an increase in total expenditure by 10% would lead to a decrease of 1% in the share of food.

The introduction of the conditional cash transfer programme *Familias en Acción* is a useful testing ground for our specification of Engel curves. The introduction of the programme led to an increase in total consumption expenditure of about 13.3%. Our estimates would therefore imply a decline in the share of food in total expenditure by about 0.013. Instead, our quasi-experimental estimates imply an increase by 0.010 from a rather large level of 0.72. This evidence, on the effect of CCTs on the share of food is consistent with that of other CCTs in different countries, such as Mexico, Nicaragua and Ecuador. We speculate that the mis-specification of the food Engel curve might be explained by the fact that these are targeted to women. Attanasio and Lechene (2002, 2008) suggest in the context of the Mexican programme PROGRESA, that a failure of the unitary model could explain this type of observations. A shift in power towards the women would lead an increase in total expenditure to induce a more than proportional increase in food consumption because in addition to the income effect, the CCT would imply a modification of weights towards mothers' preferences. The evidence we present here is not inconsistent with that hypothesis.

Further work is surely needed. In particular, it would be interesting to repeat our exercise for subcomponents of food consumption. For these, prices are observable and, under the assumption of separability between food and non-food, one could estimate a demand system that controls for prices explicitly. One could then compare the predictions of the Engel curves derived from a unitary model to the quasi-experimental evidence.

The allocation of resources across commodities (and within the household) is important not only from an academic point of view but also from a policy perspective. CCT have explicitly targeted women because of the perceived need to improve the standing of women within households. Moreover, these programmes have a strong emphasis on nutrition and provide, in addition to cash, advice on best health and nutrition practices. It is therefore important to check whether these programmes are having the desired effects. Understanding the mechanisms at play behind the effects is important to the design and re-design of policy interventions. This paper is a first attempt towards an understanding of these mechanisms.

References

Angelucci, M. and O. Attanasio (2009): "The effects of *Conditional Cash transfers* on consumption and its components: evidence from urban areas and implications for household behaviour" Mimeo, University of Arizona.

Attanasio O.P. and C. Frayne (2005): "Do the poor pay more?" Mimeo, Institute for Fiscal Study, London.

Attanasio, O.P. and V. Lechene (2002): "Tests of Income Pooling in Household Decisions", Review of Economic Dynamics, Vol. 5, No. 4, pp. 720-748.

Attanasio, O.P. and V. Lechene (2008): "Efficient Responses to Targeted Cash Transfers", in progress.

Attanasio, O.P. and A. Mesnard (2006): "The Impact of a Conditional Cash Transfer Programme on Consumption in Colombia", Fiscal Studies 27 (4), pp. 421–442.

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

Banks, J., Blundell, R. and A. Lewbel (1997): "Quadratic Engel curves, welfare measurement and consumer demand", Review of Economics and Statistics, Vol. 79, No. 4, pp. 527-539.

Blundell, R.,Duncan, A. and K. Pendakur (1998):"Semiparametric Estimation and Consumer Demand", Journal of Applied Econometrics, Vol. 13, No 5, pp.435-461.

Browning , M and P.A. Chiappori (1998): "Efficient Intra-Household Allocations: A General Characterization and Empirical Tests," Econometrica, Vol. 66(6), pp. 1241-1278.

Chiappori, P.A. (1988): "Rational Household Labor Supply" Econometrica, Vol. 56(1), pp. 63-90.

Deaton, A. and J. Muellbauer (1980): "Economics and Consumer behaviour", Cambridge University Press.

Kedir A. and S. Girma (2008): "Quadratic Engel Curves with Measurement Error: Evidence from a Budget Survey", Oxford Bulletin of Economics and Statistics, Vol. 69, No. 1, pp. 123-138.

Imbens, G.W. and J.D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects", Econometrica, Vol. 62, No. 2, pp.467-475.

Lewbel A. (1996): "Demand Estimation with Expenditure Measurement Errors on the Left and Right Hand Side", The Review of Economics and Statistics, Vol. 78, No. 4, pp. 718-725.

Macours, K., Schady, N. and R. Vakis (2008): "Can Conditional Cash Transfer Programs Compensate for Delays in Early Childhood Development?" Mimeo, Johns Hopkins University.

Rubin, D.B. (1974): "Estimating Causal Effects of Treatments in Randomised and Non-randomised Studies", Journal of Educational Psychology, 66, pp. 688-701.

Schady, N. and J. Rosero (2007): "Are cash transfers made to women spent like other sources of income?," Policy Research Working Paper Series 4282, The World Bank.

Velez, C. E., Castano, E. and R. Deutsch (1998): "An Economic Interpretation of Colombia's SISBEN: A Composite Welfare Index Derived from the Optimal Scaling Algorithm" Unpublished manuscript, Inter American Development Bank.

Table 1a

Village characteristics

	Mean	Std. Dev.
Dummy for households in town centre	0.55	0.28
Dummy for village in Atlantic Region	0.36	0.48
Dummy for village in Oriental Region	0.25	0.44
Dummy for village in Central Region	0.28	0.45
Dummy for village in Pacific Region	0.12	0.33
Altitude of the village (in meters)	646.18	753.90
Total population of the village	28,066.16	23,472.64
Dummy for households with piped water	0.87	0.14
Dummy for households connected to sewage system	0.56	0.36
Number of villages	75	

Table 1b **Household characteristics**		
	Mean	**Std. Dev.**
Number of household members	6.09	2.42
Number of children under 7	1.18	1.16
Number of children aged 7–11	1.08	0.97
Number of children aged 12–17	1.03	1.03
Number of adults above 60	0.29	0.59
Number of female adults in the household	1.37	0.73
Single headed household	0.19	0.39
Affiliated to a good social security	0.05	0.22
Age of head	44.63	13.06
Head has less than a primary education	0.24	0.43
Head has more than a secondary education	0.04	0.2
Spouse has less than a primary education	0.28	0.45
Spouse has more than a secondary education	0.04	0.2
Total consumption	420,778	249,825
Total food consumed	296,134	174,289
Share of consumption devoted to food	0.72	0.14
Log of total income	12.80	0.57
Log of expected income	12.47	0.68
Number of households	5,218	

Note:

Sample selection criteria: we excluded all households leaving in "early treatment" areas or for which we were not able to compute the "expected income" variable due to missing values in the data. Only baseline data are considered. The exchange rate between the US dollar and the Colombian peso was about 2,600 at the date of the survey. The values of consumption reported by the households have been converted in monthly amounts.

Table 2

Correlation of expected income with consumption and income

	Log of Total Consumption	Log of Household Total Income
Household level	0.303 (0.023)***	0.498 (0.034)***
Village level	0.227 (0.060)***	0.495 (0.081)***
First stage regression	0.222 (0.011)***	

Note:

Sample selection criteria: see Table 1. Regressions computed from observations for 5,218 households in 75 villages. The first row reports results from regressions of logged consumption and logged total income on logged expected income using household level data, clustering standard errors by village (in parentheses). The second row reports results from the same regressions calculated at the village level. The last row reports results from the regression of logged consumption on logged expected income and the additional controls we use in the estimation of the Engel curve (see Table 1 and Table 4). The latter regression will define the first stage equation in all instrumental variables like regressions reported in what follows.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 3

Difference in differences estimates of the effect on total consumption and food consumption

Estimate (standard error)	Log Total Consumption	Log Food Consumption	Share of Consumption Devoted to Food
OLS	0.133 (0.043)***	0.159 (0.045)***	0.010 (0.010)
Fully interacted OLS	0.148 (0.048)***	0.170 (0.050)***	0.009 (0.011)
Matching	0.148*** (0.053)	0.176 (0.055)***	0.009 (0.013)

Note:
Estimation results obtained from (4) using baseline and follow-up information and the sample selection criteria described in Table 1. Standard errors clustered at the village level are given in parentheses and obtained from 1000 bootstrap replications. Only treatment and control households on the "common support" are considered (5,163 out of 5,218), the latter being defined from the regression of a dummy for living in treatment areas on the controls described in Table 1.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 4

OLS estimates

without controlling for price effects and with exogenous expenditure

Parameter	β_f	λ_f
Estimate (standard error)	0.2392 (0.0589)***	-0.0228 (0.0051)***

Note:
OLS estimation results obtained from (5) using baseline data and the sample selection criteria described in Table 1. Standard errors are clustered at the village level and given in parentheses. Covariates included: number of people in the household, number of elderly adults, number of children less than 6, number of children between 7 and 11, number of children between 12 and 17, number of female adults, education dummies for head and spouse, age of the household head and its square, dummy for single household head, dummies for affiliation to social security.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 5

OLS estimates

controlling for village dummies and with exogenous expenditure

Parameter	β_f^v	λ_f^v
Estimate (standard error)	0.2130 (0.0091)**	-0.0206 (0.0008)***

Note:
OLS estimation results obtained from (6) using baseline data and the sample selection criteria described in Table 1. Standard errors are clustered at the village level and given in parentheses. Covariates included: in addition to those described in Table 4, village dummies and their interactions with logged total expenditure and its squared. Numbers reported are the average coefficients across villages.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 6

OLS estimates with village dummies

Distribution of 1st derivatives of Engel Curve

	1st	5th	25th	50th	75th	95th	99th
Mean	0.032	0.009	-0.016	-0.030	-0.045	-0.066	-0.081
Median	-0.003	-0.004-	-0.027-	-0.032	-0.043	-0.060	-0.069
% of negative	47.9	52.8	70.4	77.8	80.2	77.8	73.4

Note:
Mean, median and percentage of negative values across villages of the first derivative of the Engel curve estimated from (6), calculated at the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentile of the distribution of total expenditure. Weighted statistics by the size of the village population. Results based on Table 5.

Table 7 Control Function estimates without controlling for price effects, allowing for endogenous expenditure			
Parameter	β_f	λ_f	F- statistic for residual polynomial (p-value)
Estimate (standard error)	-0. 0373 (0.1116)	-0. 0105 (0. 0095)	17.00 (0.000)

Note:
CF estimation results obtained from (5) using baseline data and the sample selection criteria described in Table 1. Standard errors are clustered at the village level and given in parentheses. Covariates included: in addition to those described in Table 4, a third order polynomial in the residuals of the first stage regression. Instrument: expected future income. The F statistic reported refers to the joint significance of the coefficients of the polynomial in the residuals of the first stage equation.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 8 Control Function estimates controlling for village dummies and allowing for endogenous expenditure			
Parameter	β_f^v	λ_f^v	F- statistic for residual polynomial (p-value)
Estimate (standard error)	0.0320 (0.0198)	-0.0104 (0.0014)***	438.27 (0.000)

Note:
CF estimation results obtained from (6) using baseline data and the sample selection criteria described in Table 1b. Standard errors are clustered at the village level and given in parentheses. Covariates included: in addition to those described in Table 5, a third order polynomial in the residuals of the first stage regression and their interaction with village dummies. Instrument: expected future income. Numbers reported are the average coefficients across villages. The F statistic reported refers to the joint significance of the coefficients of the polynomial in the residuals of the first stage equation.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 9 **Control Function estimates with village dummies** **Distribution of 1st derivatives of Engel Curve**							
	1st	5th	25th	50th	75th	95th	99th
Mean	-0.060	-0.071	-0.084	-0.091	-0.099	-0.109	-0.117
Median	-0.062	-0.023	-0.092	-0.098	-0.115	-0.149	-0.168
% of negative	52.1	69.5	82.2	85.9	89.1	83.3	76.9

Note:
Mean, median and percentage of negative values across villages of the first derivative of the Engel curve estimated from (6), calculated at the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentile of the distribution of total expenditure. Weighted statistics by the size of the village population. Results based on Table 8.

Table 10 **Control Function estimates** **Modelling price effects by representative village prices interactions and allowing for endogenous expenditure**			
Parameter	$\beta_f(\xi_v)$	$\lambda_f(\xi_v)$	F- statistic for residual polynomial (p-value)
Estimate (standard error)	0.0236 (0.0956)	-0.0127 (0.0080)	5.13 (0.000)

Note:
CF estimation results obtained from (7) using baseline data and the sample selection criteria described in Table 1b. Standard errors are clustered at the village level and given in parentheses. Covariates included: in addition to those described in Table 4, log prices of coffee, potatoes, rice, sugar, male wages, altitude and its square, index of quality of life and their interaction with a third order polynomial in the residuals of the first stage regression. Instrument: expected future income. Numbers reported are the average coefficients across villages. The F statistic reported refers to the joint significance of the coefficients of the polynomial in the residuals of the first stage equation.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 11 **Control Function estimates with representative village prices** **Distribution of 1st derivatives of Engel Curve**							
	1st	5th	25th	50th	75th	95th	99th
Mean	-0.088	-0.101	-0.117	-0.126	-0.135	-0.149	-0.158
Median	-0.078	-0.097	-0.114	-0.122	-0.135	-0.154	-0.165
% of negative	92.1	98.2	100	100	100	100	98.5

Note:
Mean, median and percentage of negative values across villages of the first derivative of the Engel curve estimated from (7), calculated at the 1st, 5th, 25th, 50th, 75th, 95th and 99th percentile of the distribution of total expenditure. Weighted statistics by the size of the village population. Results based on Table 10.

Table 12 **Control Function estimates** **Linear specification using the full sample**		
Parameter	β_f^v Village dummies	$\beta_f(\xi_v)$ village level variables
Coefficient	-0.1063 (0.0169)***	-0.1208 (0.0169)***

Note:
The first column replicates the analysis in Table 8 imposing a linear specification for the Engel curve. The second column replicates the analysis in Table 10 imposing a linear specification for the Engel curve.

****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

Table 13

Testing linear versus quadratic specifications

Expenditure percentile	(1) : village dummies	(2) : village level variables
1st	(-0.1347,0.1803)	(0.0067,0.1003)
5th	(-0.0813,0.1205)	(0.0034,0.0654)
10th	(-0.0543,0.0931)	(0.0026,0.0487)
25th	(-0.0186,0.0559)	(-0.0002,0.0255)
50th	(-0.0122,0.0388)	(-0.0087,0.0095)
75th	(-0.0425,0.0624)	(-0.0270,0.0042)
90th	(-0.0811,0.0930)	(-0.0462,0.0028)
95th	(-0.1000,0.1147)	(-0.0600,0.0017)
99th	(-0.1379,0.1478)	(0.0840,0.0005)

Note:
The first column compares estimation results in the Table 8 to estimation results in the first column of Table 12. Similarly, the second column compares estimation results in Table 10 to the second column of Table 12. Numbers reported are 95% confidence intervals for the difference in the slope of the Engel curve calculated at the 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 99th percentile of the distribution of total expenditure using a linear vis-à-vis a quadratic specification. These were derived using a bootstrap procedure based on 1000 replications for the difference of the slope under the two specifications on each pseudo sample. The confidence intervals reported are based on the 2.5th and 97.5th percentiles of the bootstrap distribution from the 1000 replications.

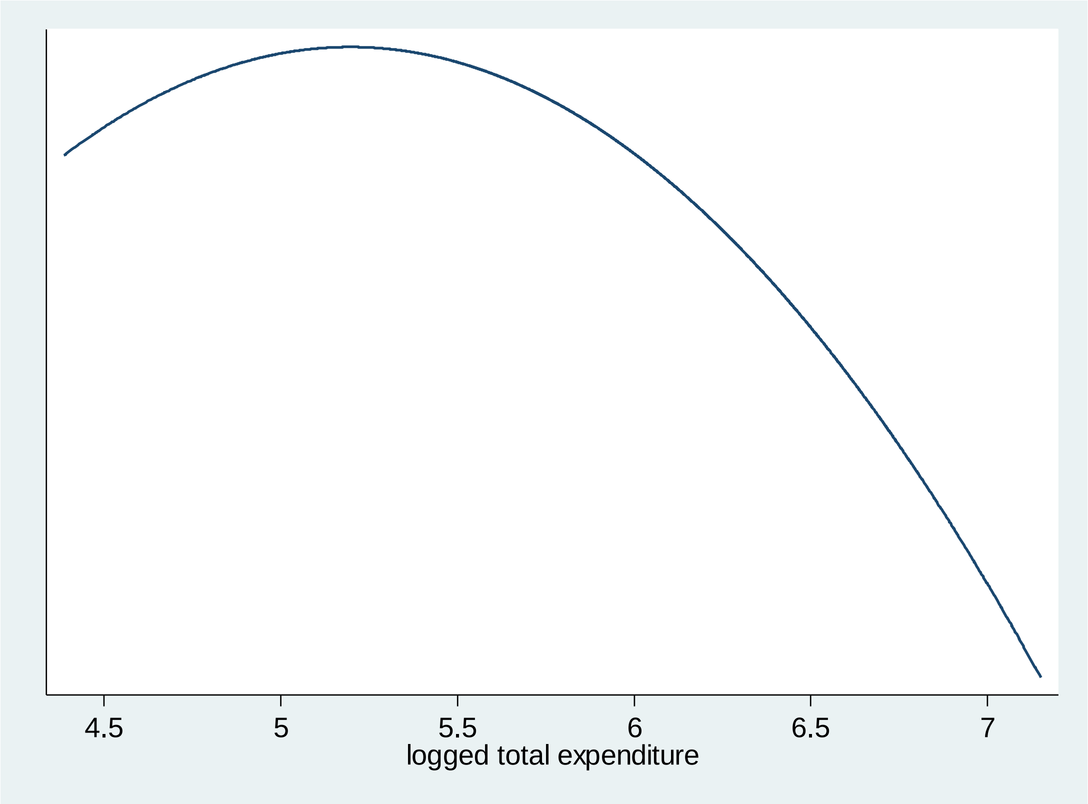
Table 14 **Control Function estimates** **Linear specification with village dummies** **estimated separately in each of the four cells**		
Parameters	α_f^v	β_f^v
Control areas at baseline (standard errors) N = 3,071	2.1161 (0.2692)***	-0.1093 (0.0210)***
Treatment areas at baseline (standard errors) N = 2,232	2.0012 (0.3627)***	-0.0991 (0.0283)***
Control areas at follow up (standard errors) N = 3,071	1.9332 (0.3015)***	-0.1034 (0.0234)***
Treatment areas at follow up (standard errors) N = 2,232	1.4201 (0.3120)***	-0.0597 (0.0240)**
Difference between coefficients at baseline and follow up in control areas (95% conf. interval)	0.1828 (-0.4966,0.9272)	-0.0059 (-0.0635,0.0468)
Difference between coefficients at baseline in control areas and at follow up in treatment areas (95% conf. interval)	0.6960 (-0.1765,1.4817)	-0.0496 (-0.1104,0.0178)
Difference between coefficients in control and treatment areas at baseline (95% conf. interval)	0.1148 (-0.6045,1.1971)	-0.0102 (-0.0950,0.0462)

Note:
The estimates result from the specification with village dummies. Bootstrapped confidence intervals using 1000 replications.
****denotes statistical significance at the 1 percent level or less.*
***denotes statistical significance at the 5 percent level or less.*

`

FIGURE 1

Estimated profile of the Engel curve without controlling for price effects and assuming that total expenditure is exogenous

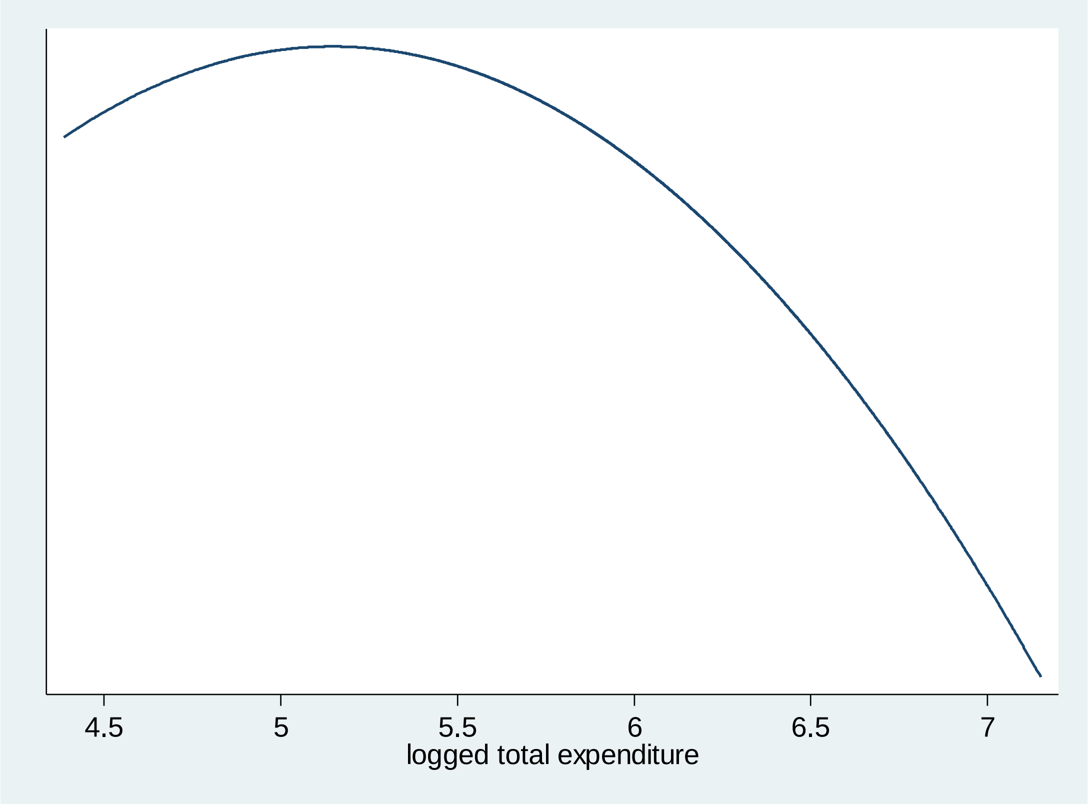


Note:

Share of food profile with respect to logged total expenditure obtained from (5) using the estimation results in Table 4.

FIGURE 2

Estimated profile of the Engel curve controlling for village dummies and assuming that total expenditure is exogenous

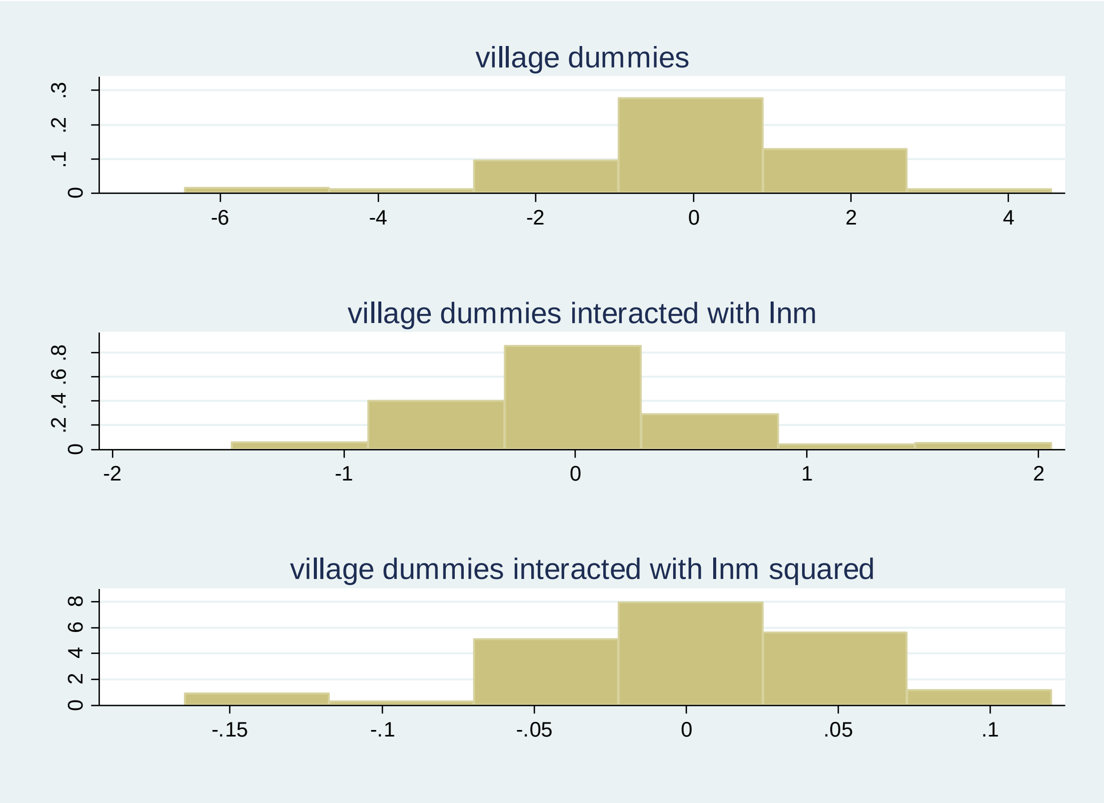


Note:

Share of food profile with respect to logged total expenditure obtained from (6) using the estimation results in Table 5.

FIGURE 3

Distribution of estimated coefficients across villages

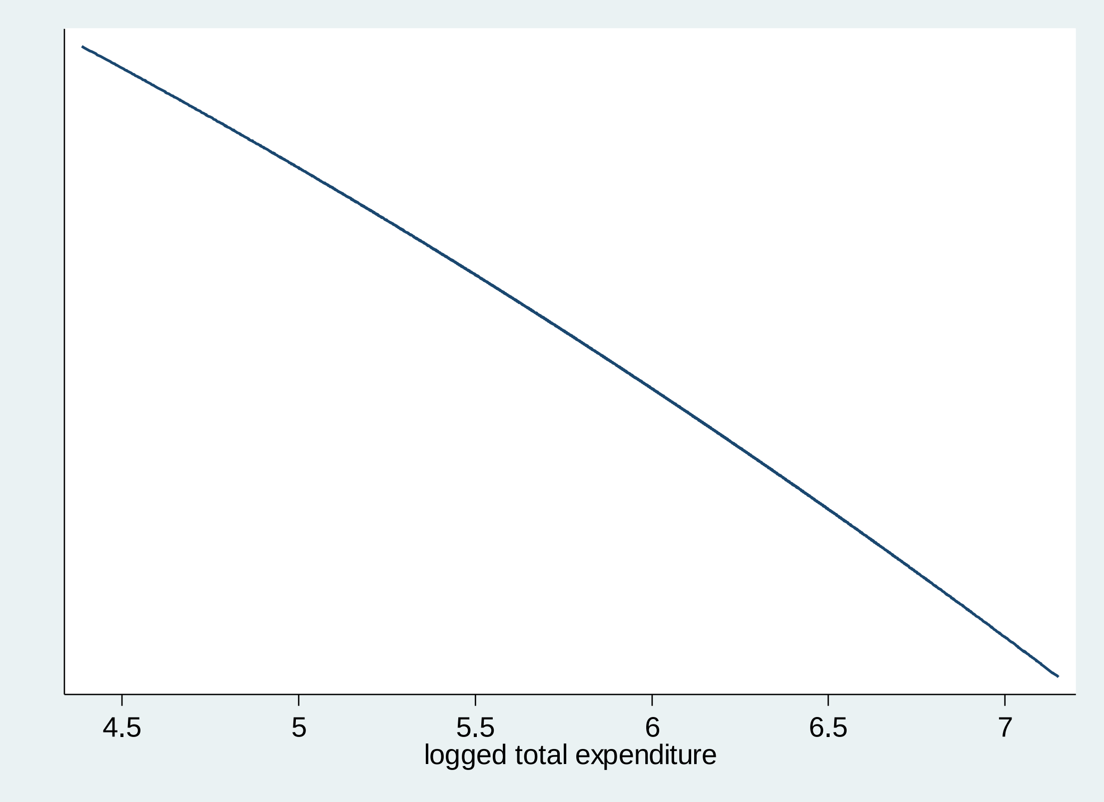


Note:

Distribution of the estimated coefficients from (6) for village dummies (top panel), their interaction with logged total expenditure (central panel) and with the square of logged total expenditure (bottom panel). Coefficients are expressed as deviations from the average.

FIGURE 4

Estimated profile of the Engel curve without controlling for price effects and allowing for endogenous total expenditure

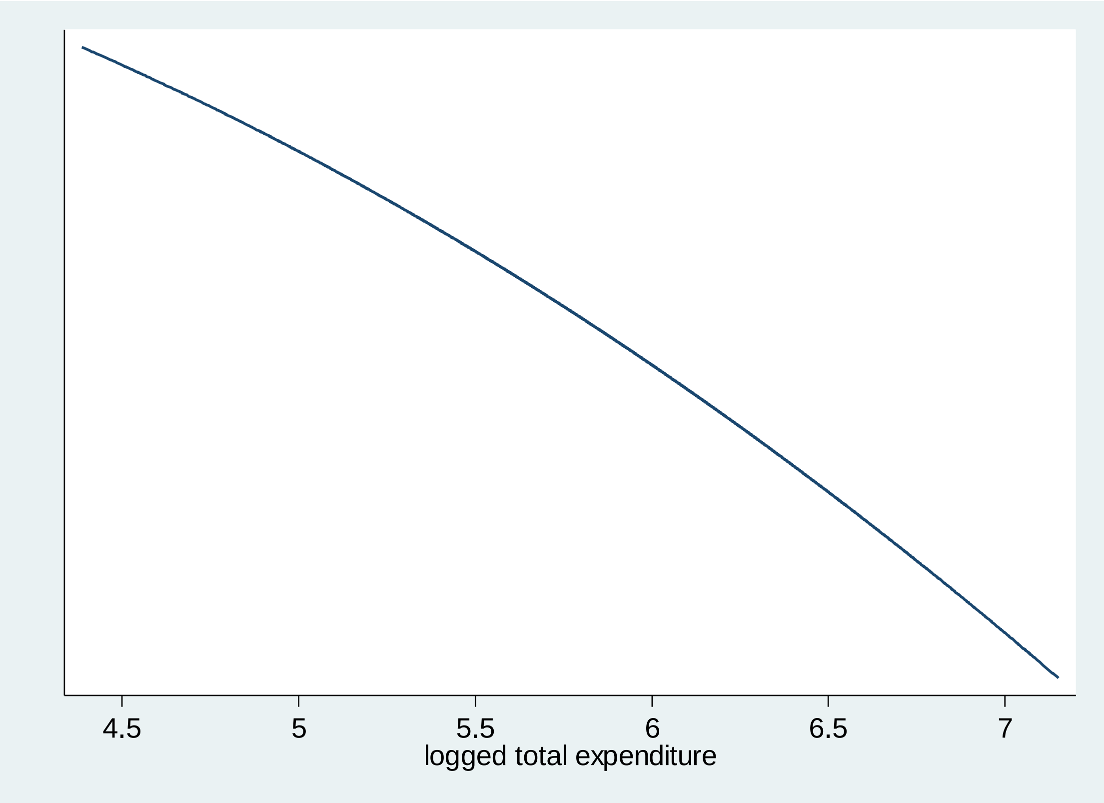


Note:

Share of food profile with respect to logged total expenditure obtained from (5) estimated with a control function approach, using the estimation results in Table 7.

FIGURE 5

Estimated profile of the Engel curve controlling for village dummies and allowing for endogenous total expenditure



Note:

Share of food profile with respect to logged total expenditure obtained from (6) estimated with a control function approach, using the estimation results in Table 8.

FIGURE 6

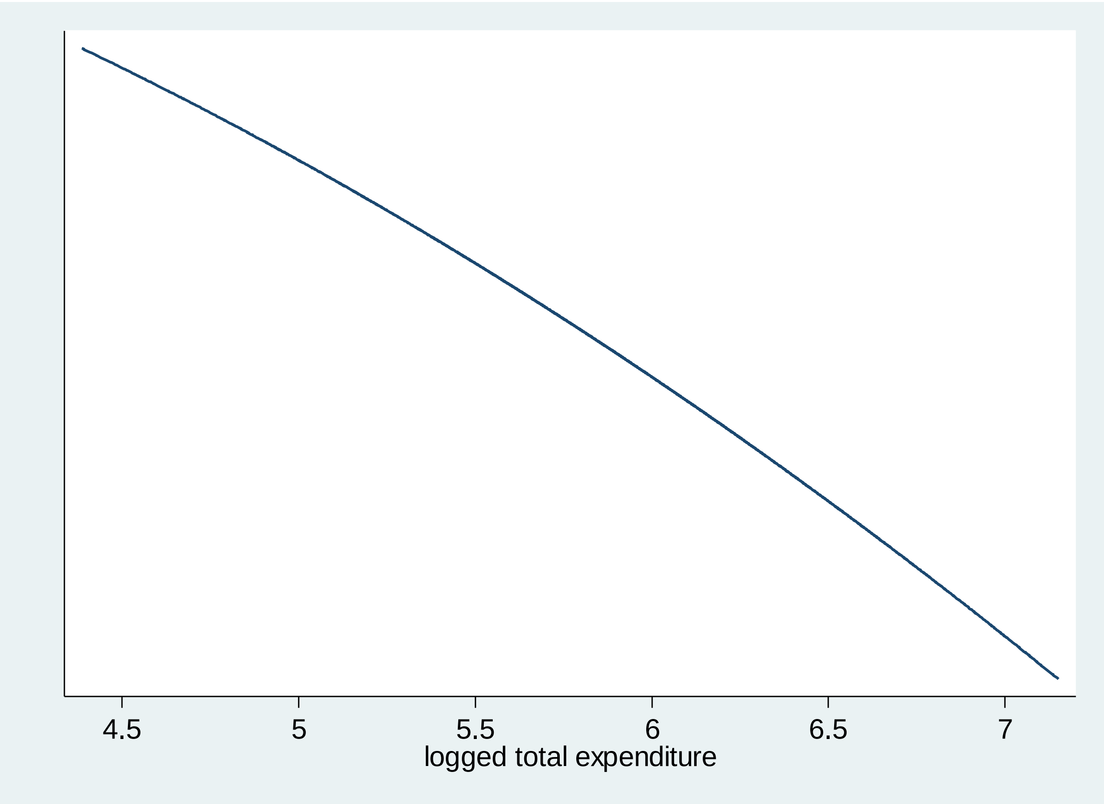
Distribution of estimated coefficients across villages, allowing for endogenous total expenditure

Note:

Distribution of the estimated coefficients from (6) estimated with a control function approach for village dummies (top panel), their interaction with logged total expenditure (central panel) and with the square of logged total expenditure (bottom panel). Coefficients are expressed as deviations from the average.

FIGURE 7

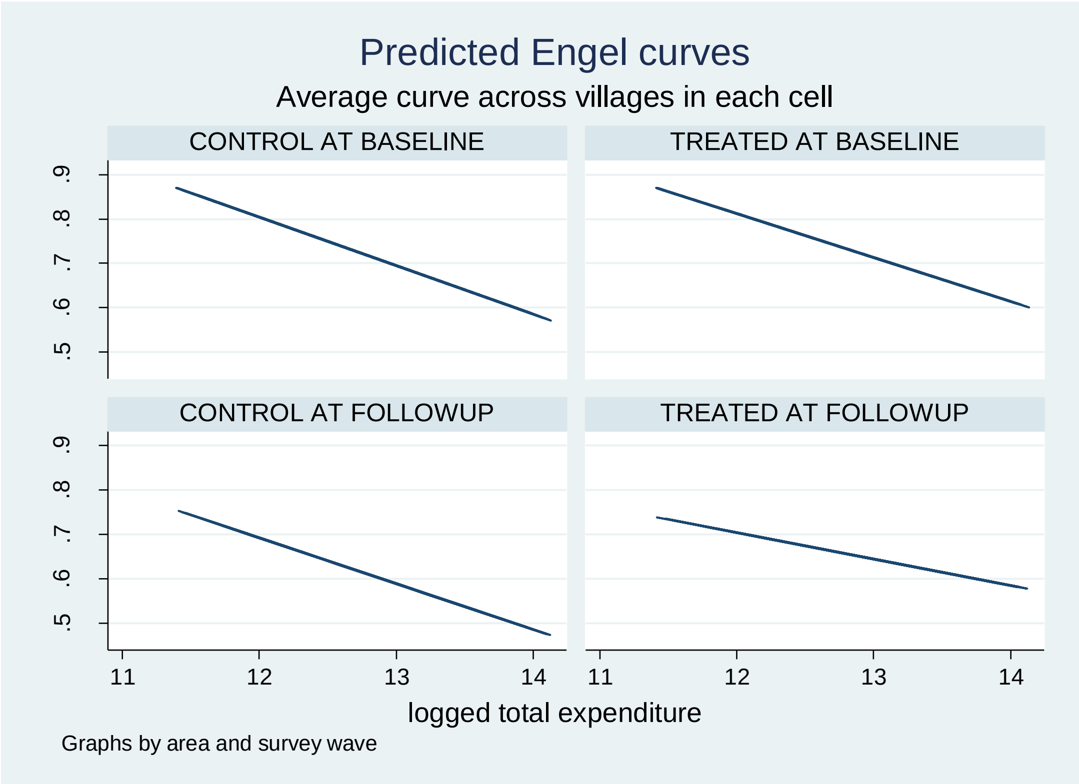
Estimated profile of the Engel curve using representative village prices interactions and allowing for endogenous total expenditure



Note:
Share of food profile with respect to logged total expenditure obtained from (7) estimated with a control function approach, using the estimation results in Table 10.

FIGURE 8

Engel curves by baseline/follow-up and treatment/control cells



Note:

Engel curves estimated as explained in Section 6.1.

List of Footnotes