Food Quality, Calories, and Household Income

Emmanuel Skoufias, World Bank, USA
Vincenzo Di Maro, University College London, UK
Teresa González-Cossío, INSP, MX
Sonia Rodríguez Ramirez, INSP, MX

Dec, 2009

Abstract

We investigate the relationship between calories, food quality and household per capita expenditure using regression and semiparametric methods on a sample of poor households from rural Mexico, where *Programa de Apoyo Alimentario* (PAL), a targeted nutritional program, is operating. The semiparametric method yields an estimate of the elasticity between calorie and expenditure of 0.39 below the median and 0.28 above. The corresponding estimates of the elasticity of the calorie price are 0.48 below the median and 0.45 above. We also examine the extent to which the expenditure elasticity of the calorie price is explained by substitution between and within major food groups. We find that there is a very high incidence of substitution within cereals (especially for poor households) and that between-group substitution explains at most 59 % of the income elasticity for food quality. These estimates suggest that the potential of a cash transfer program to have a positive impact on the food diversity and the nutritional status of households is quite limited.

JEL classification: *012; C14*
Keywords: Calorie-income elasticity, price of calories, food consumption, food quality, Mexico.

Corresponding author: Emmanuel Skoufias, The World Bank, 1818 H Street N.W., Washington D.C. 20433-USA. tel: (202)458-7539. email: eskoufias@worldbank.org.

Acknowledgements: The findings, interpretations, and conclusions in this paper are entirely those of the authors. They do not necessarily reflect the view of the World Bank

1. Introduction

The extent to which nutrition responds to income has been a much debated issue in the recent literature. In particular, a lot of attention has been devoted to estimating the size of the income elasticity of calories. A number of studies argue that this elasticity is close to zero (e.g Behrman and Deolalikar, 1987, Bouis and Haddad, 1992, Bouis, 1994) while other studies find an elasticity significantly different from zero (e.g. Subramanian and Deaton, 1996 ; Hoddinott and Skoufias, 2004; Abdulai and Aubert, 2004; Gibson and Rozelle, 2002). Accordingly, the evidence on whether increases in income will result in substantial improvements in the nutritional status of poor populations is inconclusive and thus open: on the one extreme there are those who believe that policies that raises income will also eliminate hunger as well as improve nutritional status and, on the other extreme, those who support interventions targeted specifically to nutrition since increases in income may have no substantial effect on nutritional status.

This paper investigates the relationship between nutrition and household income by focusing on the inter-relationship between calories, the quality of food and household income. The size of the calorie income elasticity depends on the extent to which the quality food also increases with increases in income (Behrman and Deolalikar, 1987). As long as increases in household income induce shifts towards food items of higher quality that typically contain fewer calories (e.g. consuming more fruits and vegetables instead of more rice and cassava) then the income elasticity for calories is likely vary with the level of household income and typically lower at higher levels of income. The price of calories, defined as the ratio of the value of food consumed and the total calories available in the household, can provide one indicator of food quality that is relatively easy to construct and analyze.

An additional motivation for the focus of this paper on food quality is to understand the determinants of household behavior towards the quality of food consumed. While useful towards explaining a low calorie income elasticity, the relationship between the cost of calories and income does not reveal much about substitution between foods. For example, a rising calorie price with income may be explained either by households continuing to consume rice of higher quality (within-group substitution) or by households switching towards new food groups, such as fruits and vegetables, that enhance the variety and the nutritional content of their diet (between-group substitution).

These two types of behavior have very different policy implications. For example, in the event that within-group substitution is prevalent in poor households, interventions that are aimed at increasing the income of poor households, such as conditional or unconditional cash transfer programs, could be more successful in increasing the caloric intake but less so in fostering better diets (Behrman, 1995). In contrast, in the case that between-group substitution prevails, cash transfer programs may be quite effective in increasing the consumption of foods with essential micronutrients. In general, better prior knowledge of the patterns of substitution within and between-food groups associated with increases in income can substantially improve the design and potential impact of interventions aimed at improving the nutritional status of poor households.

We explore these issues with the baseline data from a program targeting nutrition, *Programa de Apoyo Alimentario* (PAL). PAL is being implemented in poor rural Mexico in those areas not covered by other programs such as *Oportunidades* and *Abasto Social de Leche*. Consistently with the argument raised above (calorie-income elasticity not being very informative), and following Subramanian and Deaton (1996), we study the relationship between calories, the price of calories, used as a measure of food quality, and income. In addition to a standard regression analysis with locality fixed effects, we complement the results in this paper with a semiparametric approach. A number of studies in the literature – Strauss and Thomas (1995), Subramanian and Deaton (1996), and Hoddinott and Skoufias (2004) – estimated the relationship between calories, the price of calories and income/expenditure using a nonparametric approach. One major drawback of a completely non parametric approach is that it is not possible to control for many other independent variables such as household composition[1]. On the contrary, the semiparametric estimation approach (partially linear model) adopted here can, in principle, accommodate very large sets of independent variables and, in particular, household composition.

The rest of the paper is organized as follows. Section two describes the data, the econometric specification and estimation approaches used and presents and discusses descriptive evidence on the intake of calories and the price of calories in the sample. The relationship between calories, the cost of calories and household expenditure is studied using regression and semiparametric methods in section 3. In this section we also examine the extent to which the expenditure elasticity of the calorie price is explained by substitution between and within food groups. Section four concludes by summarizing the policy implications of our findings.

2. Data and Descriptive Statistics

The data set used a sample of 7553 households in 240 poor rural localities from eight Mexican states (Campeche, Chiapas, Guerrero, Oaxaca, Quintana Roo, Tabasco, Veracruz and Yucatan), surveyed between October 2003 and April 2004 . This sample has been collected for

[1] Subramanian and Deaton (1996) show the importance of controlling for household composition in the context of the non parametric estimation of the elasticity between calorie and expenditure.

the purposes of evaluating the *Programa de Apoyo Alimentario* (PAL). [2] This program has as its major objective the improvement of the nutritional status of poor households living in rural localities of Mexico and it is targeted to localities that are not covered by other food programs, or programs with a substantial nutrition component, such as *Oportunidades* and *Abasto Social de Leche*. [3]

The analysis in this paper is based on the baseline survey round that took place before the start of the program. The survey collects extensive socioeconomic information, as well as information about food and non food expenditures. Specifically, the consumption module collects information on the quantity consumed (including that out of own production) in the last seven days for sixty one food items.[4] We use per-capita expenditure (PCE), and not current income as a measure of household welfare and income because, in general, current expenditure and consumption tend to be a more reliable estimate of a household's permanent income than current income. PCE is derived by dividing total food and nonfood expenditures by household size. Total household expenditure (per month) is defined as the sum of value of food consumption, value of meals consumed away from home and total expenditure for goods other than food (excluding expenditures on health services). Deaton and Zaidi (2002) stress that in cases where the amount of food consumed can be distinguished from food purchased (as is the case with our data), it is the value of food consumed that should go into the consumption aggregate. The value of food consumed at home is constructed, following the guidelines above,

[2] Since one of the purposes of the evaluation of PAL is studying is impact on the nutrition of children of age less than 5, it was decided from the beginning that 40% of households interviewed in each locality had to have children less than 5.

[3] For instance, the localities that do not fulfill the requirements in terms of education and health infrastructures in order to be included in *Oportunidades* can be included in PAL.

[4] We did not use the information collected on purchases of food, as this would provide information on nutrient availability instead of consumption. The PAL questionnaire also contains a module based on the alternative approach of measuring food consumption through a 24 hour recall survey, whereby respondents are asked to recall all the foods consumed by each household member during the previous day.

using the quantity of food consumed at home and expressing it in monthly value using as prices the median unit value for each food at the locality level[5]. We use a food composition database compiled by the National Institute of Public Health of Mexico (INSP) that contains information on the macronutrient content per 100 grams of all the major food items in Mexico to convert the quantity consumed of each of the sixty one food items by each household into its equivalent content of calories. The quantity of calories consumed is then aggregated at the household level.[6]

The remainder of this section provides a descriptive analysis of the sources of calories and of the price of calories. To allow for heterogeneity in the source of calories and price of calorie across households at different points of the per capita expenditure distribution we study the daily per capita calorie intake and the price of calorie for three different samples: the full sample of households, the households at the bottom 25% of the PCE distribution (sample A) and the households at the top 25% of the PCE distribution (sample B).

The total daily per capita calorie is disaggregated here by foods and food groups: we calculate the calories coming from a particular food converting the quantity consumed of that food into calories. A preliminary answer to the question of where the calories are coming from can be gained from table 1, which reports the total daily per capita calories together with the calories coming from the food major food groups: fruit and vegetables, cereals and grains, meat and dairy, and industrialized food products (e.g. sugar, oil, coffee, refreshment drinks, etc.). The major source of calories is cereals especially for the poorest 25% population. There are some

[5] We also have the information of the market price for the food items at the locality level. However, we do not have the market price for all the food items that are included in the list of foods consumed (either some items are not included in the market price list or the definition of the food item is different).

[6] Thus, total calories are constructed as $K = \sum_i q_i k_i$, where q_i denotes the quantity of item i and k_i is the weight-to-calorie conversion factor for the same item. The calories for a specific food group G, are derived for the food items $i \in G$.

patterns worth noting: some food aggregates differ remarkably between the bottom 25% and the top 25% of the PCE distribution in the sample; for example, cereals on average represent 52.1% of the total daily per capita caloric intake but this percentage rises from 43.9% for the top 25% to 63.8% for the bottom 25% population.

On the contrary, the share of other food group aggregates does not vary as much (see the percentage out of the total for vegetables, fruit, meat and dairy, fish and coffee in table 1). Passing from the bottom 25% to the top 25% of the PCE in the sample, the decrease in the percentage of total daily per capita calorie intake provided by cereals and grains (from 73.8% to 51%) is almost completely absorbed by an increase in the percentage of meat and dairy (from 4% to 14.7%) and of industrialized foods (from 20.4% to 29.1%). Fruits and vegetables only contribute for a modest share (3.7% for all households) and it only goes up from 1.9% to 5.2%.

Table 2 presents summarizes the cost of calories in Pesos per 1000 calories from the different food groups. The cost per calorie is calculated by dividing the value of household consumption for each food by the total quantity consumed. Overall, the price of calorie reflects the quality of food consumed; for example, a rising price of calorie means that consumers are moving from foods with many calories that are cheap (low quality) to foods with less calories that are expensive (high quality). In the full sample of households, the calorie price from all types of food is 4.2 pesos per 1000 calories. This mean, however, masks remarkable differences in the quality of food consumed by poorer and better off households. The calorie price for household in the bottom 25% of the PCE distribution is 2.6 Pesos whereas it is 5.9 Pesos for households at the top 25% of the PCE distribution, suggesting that increases in income are associated with increases in the overall quality of food consumed by households.

Given the extremely high incidence of cereals as source of caloric intake table 2 includes the relative-to-cereals price of 1000 calories for each food. The results for the "all households" sample show that cereals are the cheapest major source of calorie (only oil and sugar provide cheaper calories) with vegetables and fruits being the most expensive with respect to cereals (fruit and vegetables as a group are 13.2 times more expensive than cereals; considering fruit and vegetables separately we notice that vegetables are the most expensive source of calories, 18.8 times more expensive than cereals). It is interesting to note that as income increases the price of calories from cereals does not increase substantially while the price of calories from fruit and vegetable is decreased.

3. Calories, the Price of Calories, and Expenditure

A major economic and nutritional research issue is to understand which mechanism is the most efficient for improving nutrient intake among poor households. In an effort to study this issue, a relevant part of the literature has discussed the magnitude of the elasticity of nutrient intakes with respect to income, in particular of calories. The estimates of the magnitude of the elasticity between caloric intake and income provided in the literature span a wide range; the policy implications of the elasticity can substantially change according to the magnitude estimated. Some studies argue that the elasticity is close to zero (see Behrman and Deolalikar, 1987, Bouis and Haddad, 1992, Bouis (1994) while other studies find an elasticity significantly different from zero (see Subramanian and Deaton, 1996, Hoddinott and Skoufias, 2004 and Gibson and Rozelle, 2002). Accordingly, there is no conclusive evidence on whether increases in income will result in substantial improvements in nutrient intake and, therefore, the debate remains open between those who believe that policies that raises income will also eliminate hunger and those, on the other extreme, that support interventions targeted specifically to nutrition since increases in income may not affect at all the nutritional status.

We argue in this paper that the relationship between calories and expenditure provides only an incomplete view of the more general relationship between nutrition and income. Following this stand we will analyze in the next sections the relationship between price of calorie and expenditure and the degree and the behavior of within and between food group substitution as expenditure increases.

We estimate here the calorie-expenditure relationship with both a semiparametric and a standard OLS approach. In order not to impose *a priori* a linear form to this relationship few studies – Strauss and Thomas (1995), Subramanian and Deaton (1996), and Hoddinott and Skoufias (2004) – estimated this relationship in a completely non parametric way[7]. A completely non parametric estimation has the major drawback of not being especially suitable to explore multivariate relationships; even with big samples of households the number of independent variables that can be included in the estimation may be limited (especially if one wants to take into account locality effects). In addition to this, the sparseness of data when the number of independent variables in the model is large causes the variances of the estimates to be unacceptably large; the problem of rapidly increasing variance for increasing dimensionality is sometimes referred to as the curse of dimensionality. On the contrary, our semiparametric estimation approach is not affected by this drawback and can, in principle, accommodate very large sets of independent variables (see section 3 for details).

In what follows we estimate the objects of interest using a semiparametric as well as a completely parametric approach. Each specification includes a set of variable at the individual and at household level as controls. The regression approach includes one dummy variable for each locality (village fixed effect) in order to control for possible locality effects, especially

[7] It is interesting to notice that all these studies find an elasticity different from zero, at least for the sample of those households that are below the median of the per capita expenditure. In poor rural Mexico, Di Maro (2004), as in Hoddinott and Skoufias (2004), also finds a positive elasticity below the median of per capita expenditure.

spatial price variability. Controls included are age-sex family composition dummies, head and spouse's age, schooling, speaking indigenous language, working status, average grade in Spanish and Math class, dummy for household owning radio, owning television, dummies for month of data collection.

Semiparametric estimation: partially linear model

A semi-parametric approach is useful for identifying the functional form that best describes the relationship between calories and income. The model we estimate below is a partially linear model:

$$y_i = z_i \beta + m(x_i) + \varepsilon_i \qquad (1)$$

where y_i denotes the *ln* of the quantity of calories consumed, z_i is a vector of the variables that we would like to control for in a linear function, β is a vector of parameters and *m(x)* is a nonlinear function of x in this case of the lnPCE.

This model has been traditionally estimated with the Robinson (1988) estimator, which is especially suitable for the estimation of the vector β in (1). Since we are primarily interested in the estimation of *m(x)* we implement an estimator based on a differencing approach (first suggested by Yatchew, 1997, and discussed by DiNardo and Tobias, 2001). The procedure for estimating (2) consists of the following steps: first, the data are sorted by ascending values of the x variable (in our case *ln*PCE) and the *m*-th order[8] differences are calculated on the sorted data. The idea here is that if x_i and x_{i-1} are close enough in the sorted data, then so will $m(x_i)$ and

[8] As noted in Yatchew (1997) the differencing order is important as far as the efficiency of the estimator is concerned. In order to maximize the efficiency of the estimator, we use the optimal differencing weights, as tabulated in Yatchew (1997), to compute differences of the sorted data. We set the differencing order to 3 to compute differences in the semi-parametric estimation. We also tried other differencing orders and the results did not change substantially.

$m(x_{i-1})$. Accordingly, the differenced version of the model (1) on the sorted data will remove the nonparametric component $m(x_i)$. Then the vector β can be estimated with a regression of the differenced y's on the differenced z's. With the estimated vector $\hat{\beta}$ in hand it is then possible to derive a new "adjusted" dependent variable net of the linear effect of the z variables, i.e.,

$$y_{adjusted} = y_i - z_i\hat{\beta} \qquad (2)$$

The final step is to perform a local linear regression using the variable defined in (2) as dependent variable. The vector *z* in eq. (1) includes the number of males and females age 0 to 4, 5 to 9, 10 to 14, 15 to 54 and more than 55. In particular, we use a smooth local regression technique similar to that used by Subramanian and Deaton (1996).

The semiparametric estimate of the regression function that links daily per capita calories and per capita expenditure is presented in figure 1 together with confidence intervals that show two standard errors on either side calculated with a clustered bootstrap. Even though the confidence bands are quite wide around the regression line (and so the line is not tightly estimated) the bands do not widen substantially at the extreme of the distribution of the per capita expenditure. From the figure it emerges that the functional form that best describes this relationship is a linear one; this means that also for households that are at the top of the distribution of the per capita expenditure per capita calorie intake continue to rise with income, with this being an indication that households in this rural setting are extremely poor (calories never become unimportant).

Figure 2 displays the estimate of the elasticity of calories to PCE; the elasticity is on average below the median 0.39 and it remains different from zero (on average 0.28) also above the median; it is interesting to note that in the previous studies that estimated the relationship in a non parametric fashion (see above) a common result was an elasticity close to zero for the sample of households above the median, with a notable exception being Subramanian and Deaton (1996).

As a complement to the semiparametric estimates we also provide estimates of the calorie expenditure elasticity from a linear regression model, i.e.

$$\ln Y_{i,v} = \alpha_0 + \alpha_1 FE_v + \beta Z_{i,v} + \gamma \ln PCE_{i,v} + \varepsilon_{i,v} \qquad (3)$$

where $Y_{i,v}$ denotes caloric consumption per capita in household i in locality v, FE is a vector of binary variables summarizing village-specific fixed effects, Z is a vector of household characteristics and ε is an error term.

The inclusion of locality-specific fixed effects, FE, is intended to control for price differences across villages and other village-specific characteristics that may have also a direct impact on nutrient consumption. The vector Z includes binary variables summarizing the age and gender composition of the household, binary variables for the age and educational level of the household head and of his/her spouse, the number of individuals within the household with access to health services, binary variables for ownership of assets (radio and television) and binary variables indicating whether the household head and/or his/her spouse speak an indigenous language, the average test score in Spanish and Math in the last school grade attended, and binary variables for the month of data collection.

These estimates of the parameter γ from equation (3) are presented in table 3. In general, the elasticity estimates obtained using fixed effects are higher than the semiparametric estimates[9]. Table 3 also reports how the expenditure elasticity of calories varies depending on the presence of young children in the household. Specifically, the estimates in the even-

[9] It is important to note that the linear regression model in equation (4) uses a larger set of controls than the semiparametric estimation that uses as controls only age-sex composition.

numbered columns reveal how the expenditure elasticity for caloric consumption varies with the presence of children of three different age categories: ages 0 to 4, 5 to 9 and 10 to 14.[10] Overall, it seems that the households with young children exhibit a significantly smaller expenditure elasticity for calories. However, for the sample of households below the median of PCE the interaction terms are all insignificant (see estimates in column 4 of table 3).

In our attempt to offer a more complete view of the food demand and nutrition patterns than that provided by the elasticity calorie-expenditure we focus next on the relationship between the cost of calories and per capita expenditure. As mentioned above, the cost of calories, defined here as the ratio of the value of food consumed and the total calories consumed by the household, may be considered as an indicator of the quality of food purchased by the household. A higher cost of calories at higher levels of per capita expenditure suggests that households are switching towards foods that are more expensive and contain less calories. An indication of the extent to which households substitute towards better quality foods as PCE increases is given by the expenditure elasticity of the price of calories.

Figures 3 and 4 (that present the estimated regression function and elasticity together with two bootstrapped standard error bands that allow for cluster design) allow us to shed some light on the issues discussed above. The price of calories is steadily rising as per capita expenditure increases (see figure 3) and the relationship seems appropriately described by a linear functional form (see figure 5). The elasticity is gently declining as PCE increases, even though there are no big differences on average below and above the median of PCE (elasticity is on average 0.48 below and 0.45 above the median). This result suggests that households that are above the median of PCE will still spend additional income to buy higher quality food; however, figure 4 suggests that for households at the top of the distribution of PCE the elasticity

[10] The reference category is having no children at all, or having children 15 years of age or older.

is declining faster.

Table 4 contains the estimates of the elasticity of the price of calories with respect to PCE based on the linear regression model of equation (4), (with $Y_{i,v}$ denoting the price of calories in household i in locality v). The estimates of how the expenditure elasticity of the price of calories varies with the presence of children of three different age categories reveal a very interesting pattern. The elasticity for food quality is higher when households have very young children (age 0-4 and 5-9) while there is no significant effect for children of age 10-14.

In combination, the estimates of tables 3 and 4 reveal that the expenditure of quality (cost of calories) is higher for families with younger children while the expenditure of calories in families with younger children is not much different from the elasticity of calories for families with older or without children at all. Thus, among poor families with children between 0 and 4 years of age,, increases in income are associated with an improvement in the quality of food that is significantly higher than the improvement in the quality of food observed for other families.

Between and within group substitution

The preceding results are suggest that increases in income may indeed induce poor families (and especially those with young children) to consume more diverse foods with higher micronutrient content, such as fruits and vegetables. In order to establish this in a more convincing manner it is important to examine what happens to the diversity of food consumption as PCE increases. Changes in the price of calories can be driven either by households starting to consume more expensive or higher quality foods within the same food group (within-group substitution) or by switching to another more expensive food group, e.g. from cereals to fruits and vegetables, (between group substitution). To investigate this matter further, we implement an estimation strategy that allows us to breakdown the total elasticity of the price of calories (or the quality of food) into two components: the between- and the within-food group elasticity.

To accomplish this decomposition, we follow Subramanian and Deaton (1996) who propose one simple way of how to calculate the price of calories elasticity excluding the total within-group elasticity. We slightly improve on that approach since we estimate also the contribution of the single food groups to the total within- and between-group elasticity.

Begin with the definition of the inverse of the price per calories, (i.e., calories per peso),

$$(P)^{-1} = \frac{K}{x_f} = \sum_G w_G \pi_G = \sum_G w_G \sum_{i \in G} \frac{k_i q_i}{x_G} \tag{4}$$

where x_f is the total expenditure on food, K is the total calorie consumption, w_G is the share of the food budget devoted to food group G, and π_G which is defined by the second equality, is calories per peso devoted to food group G (or the inverse of the price of calories of food group G).[11] As we move from poor to rich, the ratio of K to x_f and thus the (inverse of the) price per calorie, is influenced both by changes in the budget share w_G, as households substitute between food groups, and by changes in the food group-specific calorie price inverses π_G as quality and nutrient substitution takes place within groups.

One can neutralize the within-group effect by calculating the values of the π 's at the sample mean and using these means in place of the household-specific π 's to recalculate (4). This yields an adjusted log calorie price, $\ln P_{BG}$, that excludes within-group substitution, i.e.,

$$\ln P_{BG} = -\ln\left(\sum_G w_G \overline{\pi}_G\right). \tag{5}$$

[11] Note that even though the calorie price is household-specific, a subscript identifying households is omitted in order to keep the notation simple.

Figure 5 displays the expenditure elasticity of the calorie price that excludes within group substitution using the semiparametric approach described earlier. This calorie price, estimated based on equation (6) simply reflects between-group (BG) differences in the shares of the four food groups. The figure also contains two bootstrapped standard error bands that allow for the clustered design of the sample. As it is apparent from figure 5, the between group substitution is higher among poorer households, a finding that is not surprising considering that increases in income are likely to induce households to complement their diet of staple foods (cereals, grains and root crops) with different foods.

To provide a better perspective, figure 6 presents in the same graph the elasticity for food quality, along with the elasticity of the calorie price capturing substitution between groups. The difference between these two elasticities provides an estimate of the contribution of the within-group substitution to the elasticity of the total calorie price. Figure 6 reveals that substitution between groups explains about fifty percent of the elasticity of food quality. Moreover, the contribution of within -group substitution is considerably higher for households at the top half of the distribution of PCE. This pattern is in contrast to the one observed by Subramanian and Deaton (1996) in rural Maharashtra in India. Using identical methods, they find that the contribution of within-group substitution falls down to zero for households in the top half of the PCE distribution. Thus, in rural Mexico, increases in income are likely to be associated with considerable upgrading of the quality of foods consumed within a given food group and not so much with shifts between food groups. Given that cereals dominate the diet of poor households in rural Mexico, (73.8% of the total calories of the bottom 25% of the PCE distribution) these findings suggest that there is little scope for cash transfer programs and economic growth to increase the diversity of the diet of rural populations in Mexico.

To complement our analysis so far based on semiparametric methods we also report some estimates based on regression methods. Using $\ln P_{BG}$ as the dependent variable in regression (3), table 5 reports the estimates of the parameter γ without and with interaction terms between lnPCE and binary variables identifying the presence of young children in the household.

Two interesting patterns emerge regarding the income elasticity of the calorie price excluding within-group substitution. First, between-group substitution accounts for about 50% of the elasticity of food quality with respect to PCE in the full sample of households. Second, between-group substitution seems to be slightly higher in the sample below the median of PCE (i.e. 58-59% of the elasticity of food quality). These results support the earlier semiparametric findings.

The interaction terms suggest that the degree of between-group substitution is higher when households have younger children (ages 0-4 and 5-9). In particular, for the sample below the median of PCE results show that between-group substitution is increased only by the presence of very young children, age 0-4 (coefficients for interaction with age 5-9 are not significant). Thus, even though between-group substitution plays only a mild role in explaining increases in food quality with changes in income, it appears that increases in income are likely to have a stronger effect in improving the diet of families with infant children.

4. Conclusions and policy implications

The findings of this paper shed light on a number of issues that are policy relevant. In the majority of the targeted population groups the primary source of calories are cereals and grains (rice, bread, tortillas etc.) and roots crops. Prior knowledge of the within and between group substitution responses to increases in income, such as those resulting from an income transfer program, provides valuable information about the potential of a transfer program to have a positive impact on the food diversity and the nutritional status of households. In addition, a better understanding of the consumption patterns of poor households can assist in the design of an appropriate food basket aimed at increasing the food diet of the target population.

The semiparametric estimation delivers an estimate of the elasticity between calorie and expenditure of 0.39 below the median and 0.28 above. Regarding the relationship between the quality of food (measured by price of calories) and expenditure, our results indicate that the quality of food rises steadily with PCE and that the best functional form describing this relationship seems to be a linear one. The estimates of the elasticity are 0.48 below the median and 0.45 above. Coupling the results from the calorie-PCE and price of calorie-PCE relationships our evidence implies that both the intake of calories and the quality food rise steadily with expenditure; the elasticity calorie-expenditure is substantially different from zero and, in particular, it remains high also above the median of PCE; finally the incentive to buy higher quality is only slight declining as PCE increases.

A finer analysis of the price of calorie allows to study the issue of the behavior of the food diversity. Changes in price of calorie can be driven either by households starting to consume different foods within the same food group (within-group substitution) or by households switching to another food group (between-group substitution). The policy implications of this issue are remarkable. For example, an intervention that only targets income of poor households could be extremely successful in increasing the caloric intake but not in fostering the consumption of foods with essential micronutrients (see also Behrman, 1995). In addition to this, interventions that have as stated objective of improving the diet of poor households can be better designed (and/or evaluated) with knowledge of the baseline relationship between food substitution and expenditure as well as the knowledge of this relationship can substantially improve the task of selecting the food basket composition in case the program provide for it.

Our estimates suggest that between-group substitution explains at most 59 % of the income elasticity for food quality among poor households in rural Mexico. Moreover, our analysis reveals that there is a very high incidence of substitution within cereals (especially for poor households). Given that poor households take their calories mostly from cereals (as confirmed by our descriptive evidence) these estimates suggest that the potential of a cash transfer program to have a positive impact on the food diversity and the nutritional status of households is quite limited. The income obtained from a transfer program is likely to result in households consuming more calories obtained through the consumption of higher quality and thus more expensive cereals rather than in inducing them to buy foods that are richer in micronutrient content such as fruits, vegetables, meat and dairy products.

Our findings also imply that if the objective of policy is to promote a more diverse diet in poor rural areas of Mexico then direct nutritional interventions are more appropriate. Considering that in Mexico micronutrient deficiencies remain widespread while obesity and chronic diseases, such as diabetes and heart disease, are becoming leading causes of morbidity and mortality (Rivera et al. 2002; Eckhardt, 2006), the provision of a carefully designed food basket composed of micronutrient-rich food items and informational strategies focused on improving knowledge about nutrition can be effective in improving nutrition and resulting in healthier outcomes.

References

Abdulai., A., and D. Aubert., 2004, "Nonparametric and Parametric Analysis of Calorie Consumption in Tanzania", *Food Policy,* Vol. 29, No. 2, April 2004, pp. 113-129

Behrman, J., 1995. "Micronutrients and Household Behavior: What we Know and What we Don't Know." Working Papers Series on Agricultural Strategies for Micronutrients, Number 2, International Food Policy Research Institute, Washington DC.

Behrman, J., and A. Deolalikar., 1987 "Will Developing Country Nutrition Improve With Income? A Case Study for Rural South India." *Journal of Political Economy* vol. 95, pp. 492-507.

Behrman, Jere R., and Barbara L. Wolfe., 1987 "How Does Mother's Schooling Affect the Family's Health, Nutrition, Medical Care Usage, and Household Sanitation?" *Journal of Econometrics* 36 (1987): 185-204.

Blundell, R., and J. Powell, 2004, "Censored Regression Quantiles with Endogenous Regressors", mimeo

Bouis, H., 1991 "The Determinants of Household-level Demand for Micronutrients: An Analysis for Philippine Farm Households." International Food Policy Research Institute. Washington, D.C. Mimeo.

Bouis, H., and L. Haddad., 1992 "Are Estimates of Calorie-Income Elasticities too high? A Recalibration of the Plausible Range." *Journal of Development Economics,* vol. 39, no. 2, pp.333-364.

Deaton, A., 1997, "*The Analysis of Household Surveys. A Microeconometric Approach to Development Policy*", John Hopkins University Press, Baltimore MD USA

Deaton A., and S. Zaidi, 2002, Guidelines for Constructing Consumption Aggregates for Welfare Analysis, *no. 135 in Living Standards Measurement Study Working Paper*, The World Bank, Washington, DC

Di Maro, V., 2004 "Evaluation of the Impact of Progresa on Nutrition: Theory, Econometric Methods and an approach to deriving Individual Welfare Findings from Household Data", UCL, mimeo, Sep 2004

Di Nardo, J., and J. Tobias, 2001, "Nonparametric Density and Regression Estimation" *Journal of Economic Perspectives*, vol. 15, no. 4, (Fall), pp. 11-28.

Eckhardt, C., 2006, "Micronutrient Malnutrition, Obesity, and Chronic Disease in Countries Undergoing the Nutrition Transition: Potential Links and Program/Policy Implications," FCND Discussion Paper 213 (November), International Food Policy Research Institute, Washington DC.

Gibson, J. and S. Rozelle, 2002, "How Elastic is Calorie Demand? Parametric, Nonparametric and Semiparametric Results for Urban Papua New Guinea", *The Journal of Development Studies*, Vol. 38, No.6, August 2002, pp. 23-46

Hoddinott, J. and E. Skoufias, 2004. "The Impact of PROGRESA on Food Consumption.", *Economic Development and Cultural Change*, Vol. 53, No. 1, October, pp 37-61.

Rivera, J. A., S. Barquera, F. Campirano, I. Campos, M. Safdie, and V. Tovar, 2002, "Epidemiological and Nutritional Transition in Mexico: Rapid Increase of Non-Communicable Chronic Diseases and Obesity," *Public Health Nutrition* 5 (1A), pp. 113-122.

Robinson, P., 1988, "Root-N-consistent semiparametric regression", *Econometrica* 56(4), pp. 931-954.

Strauss, J., 1984, "Joint Determination of Food Consumption and Product in rural Sierra Leone: Estimates of a Household-Firm model," *Journal of Development Economics*, vol. 14 (1-2), pp. 77-104

Strauss, J., and D. Thomas., 1995 "Human Resources: Empirical Modeling of Household and Family Decisions." Chapter 34 in *Handbook of development economics* vol. 3, eds. J. Behrman and T.N. Srinivasan. Amsterdam: North-Holland.

Subramanian, S. and A. Deaton., 1996 "The Demand for Food and Calories." *Journal of Political Economy* vol. 104, pp. 133-162.

Yatchew, A., 1997, "An Elementary Estimator of the Partial Linear Model," (1997), *Economics Letters*, 57, 135-43. See also Vol. 59, 1998 403-5.

Figure 1 – Caloric Consumption and PCE

Short-dashed lines are two bootstrapped standard error bands that allow for cluster design

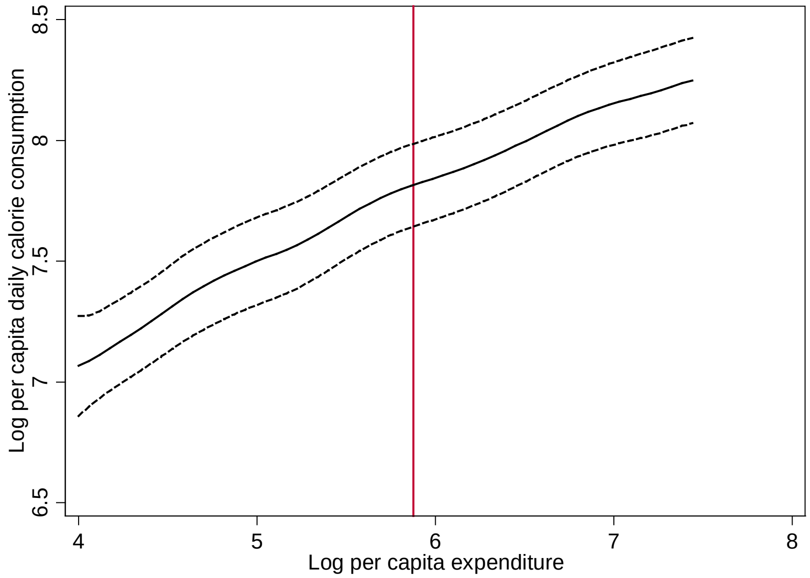

Figure 2 –Elasticity of caloric consumption to PCE

Short-dashed lines are two bootstrapped standard error bands that allow for cluster design

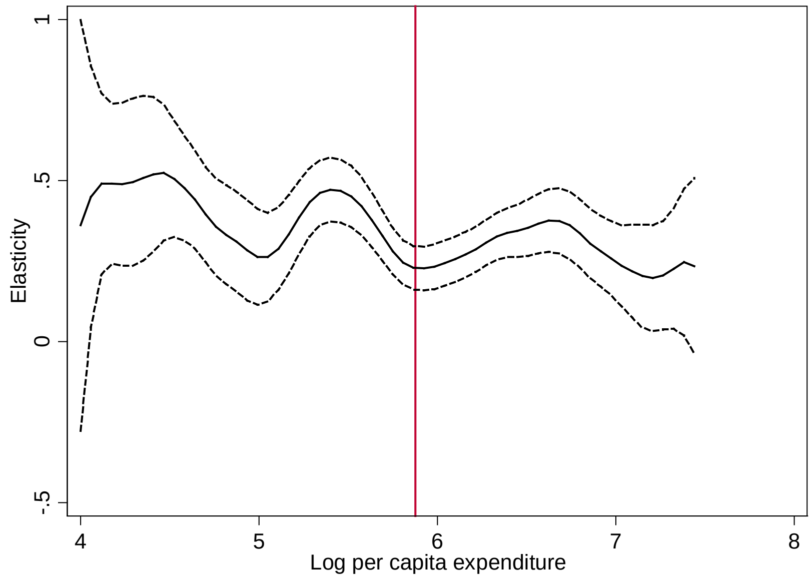

Figure 3 – Food quality (price of calories) and PCE

Short-dashed lines are two bootstrapped standard error bands that allow for cluster design

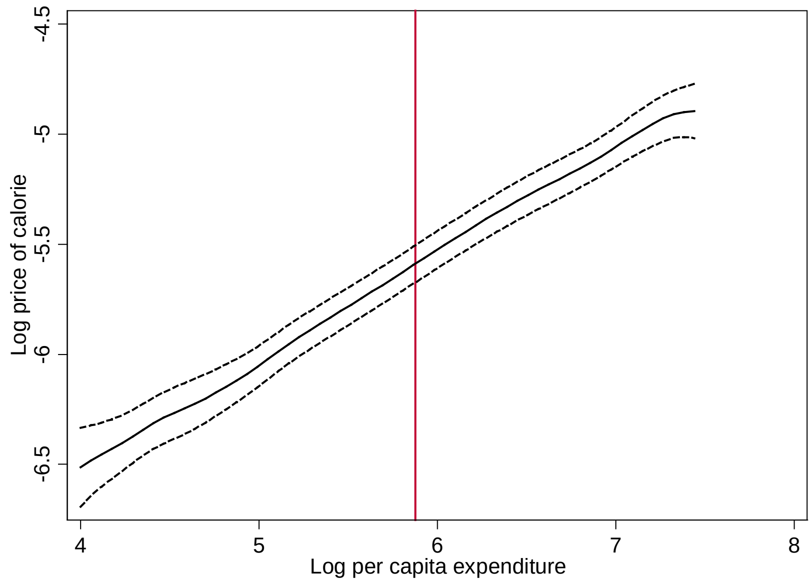

Figure 4 – Elasticity of food quality (price of calories) with respect to PCE
Short-dashed lines are two bootstrapped standard error bands that allow for cluster design

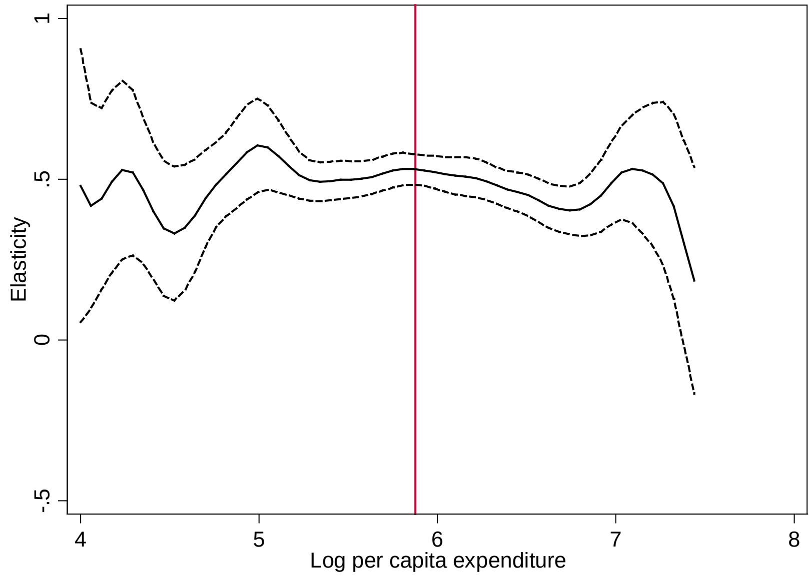

Figure 5 – Elasticity of the calorie price excluding within-group substitution

Short-dashed lines are two bootstrapped standard error bands that allow for cluster design

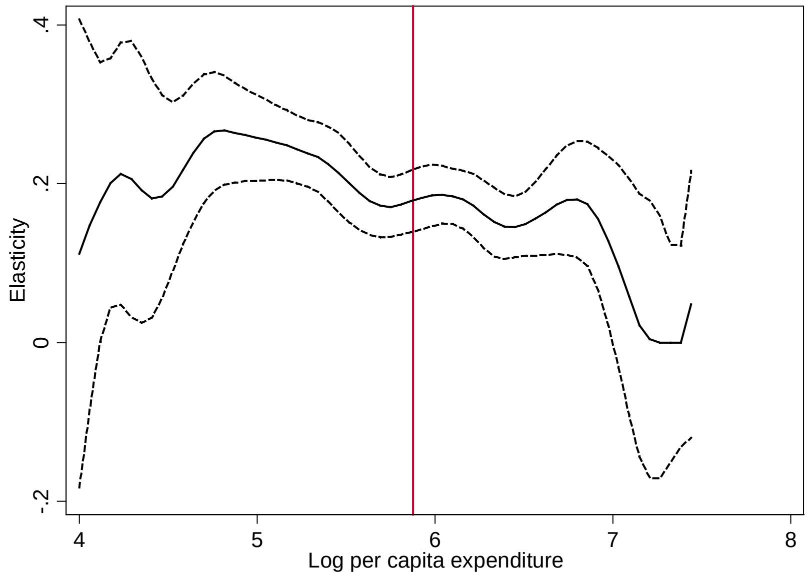

Figure 6– Total Calorie-Price elasticity and elasticity excluding within-group substitution

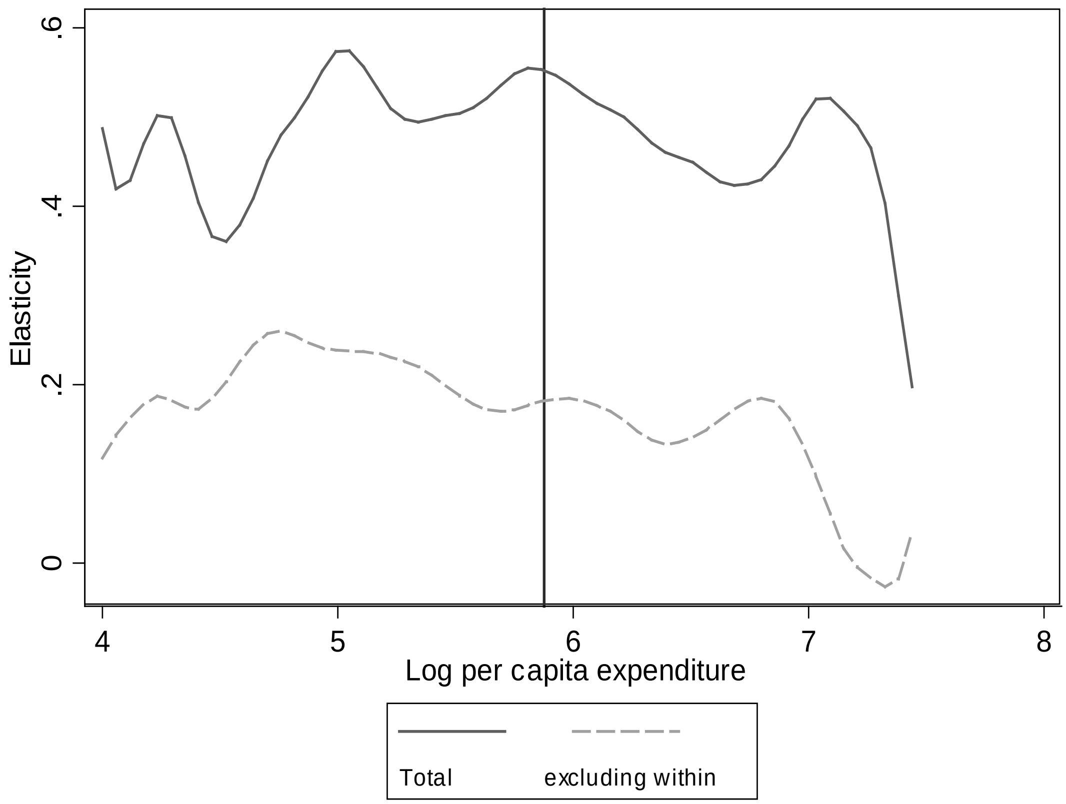

Table 1 – Daily Calories per capita

	All		*Bottom 25%*		*Top 25%*	
Calories from	***Mean***	***% of total***	***Mean***	***% of total***	***Mean***	***% of total***
Vegetables	31.5	*1.4*	13	*0.8*	52.1	*1.9*
Fruit	51.1	*2.3*	18	*1.1*	90.3	*3.3*
Cereals	1155.1	*52.1*	1057.7	*63.8*	1196.3	*43.9*
Pulses	178.3	*8.0*	164.7	*9.9*	195	*7.2*
Meat	73.4	*3.3*	21.5	*1.3*	135.1	*5.0*
Fish	10.4	*0.5*	2.9	*0.2*	20.5	*0.8*
Dairy	74	*3.3*	14.1	*0.9*	147.1	*5.4*
Oil	287.9	*13.0*	155.6	*9.4*	399.8	*14.7*
Sugar	212.8	*9.6*	132.9	*8.0*	270.6	*9.9*
Coffee	25.4	*1.1*	22.1	*1.3*	28.8	*1.1*
Other foods	115.5	*5.2*	54.5	*3.3*	191.8	*7.0*
Total calories	**2215**	***100.0***	**1657**	***100.0***	**2727**	***100.0***
Food Groups: Fruit and Vegetables	82.6	*3.7*	31.1	*1.9*	142.5	*5.2*
Cereals	1333.4	*60.2*	1222.4	*73.8*	1391.3	*51.0*
Meat and Dairy	224	*10.1*	66	*4.0*	401.4	*14.7*
Industrialized products	575.7	*26.0*	337.8	*20.4*	792.5	*29.1*

Table 2 – Price per Calorie (Pesos per 1000 calories)

Calories from:	*All* Mean	*All* Relative to cereals	*Bottom 25%* Mean	*Bottom 25%* Relative to cereals	*Top 25%* Mean	*Top 25%* Relative to cereals
Vegetables	37.6	*18.8*	39.8	*30.6*	36.4	*13.0*
Fruit	17.1	*8.6*	17.4	*13.4*	16.8	*6.0*
Cereals	2	*1.0*	1.3	*1.0*	2.8	*1.0*
Pulses	3.1	*1.6*	2.8	*2.2*	3.3	*1.2*
Meat	16.8	*8.4*	16.6	*12.8*	16.8	*6.0*
Fish	30.4	*15.2*	24.7	*19.0*	33.3	*11.9*
Dairy	12.4	*6.2*	10.5	*8.1*	13.3	*4.8*
Oil	1.2	*0.6*	1.2	*0.9*	1.2	*0.4*
Sugar	1.8	*0.9*	1.8	*1.4*	1.8	*0.6*
Coffee	21.5	*10.8*	15.8	*12.2*	25.7	*9.2*
Other foods	11.5	*5.8*	10.4	*8.0*	12.6	*4.5*
All foods:	4.2	*2.1*	2.6	*2.0*	5.9	*2.1*
Food Groups: Fruit and Vegetables	26.3	*13.2*	29.6	*22.8*	24.4	*8.7*
Cereals	2.1	*1.1*	1.5	*1.2*	2.8	*1.0*
Meat and Dairy	13	*6.5*	12.1	*9.3*	14.2	*5.1*
Industrialized products	3.9	*2.0*	3.5	*2.7*	4.6	*1.6*

Table 3 – The Elasticity of Calorie Consumption with respect to PCE (Per Capita Expenditure)

	Full sample		**Sample A**		**Sample B**	
Log Calories per capita	(1)	(2)	(3)	(4)	(5)	(6)
log PCE	0.470	0.465	0.521	0.515	0.445	0.438
	(0.012)***	(0.012)***	(0.020)***	(0.021)***	(0.023)***	(0.023)***
Interactions with log PCE						
Children age 0-4		-0.005		-0.003		-0.008
		(0.002)***		(0.003)		(0.002)***
Children age 5-9		-0.002		-0.003		-0.002
		(0.002)		(0.003)		(0.002)
Children age 10-14		-0.003		-0.002		-0.005
		(0.002)**		(0.003)		(0.002)**
Observations	6029		2988		3041	
R-squared	0.44	0.44	0.42	0.42	0.35	0.35

* significant at 10%; ** significant at 5%; *** significant at 1%
Sample A: Households below the median of PCE
Sample B: Households above the median of PCE
See text for more details on the regression model and method used

Table 4 – The elasticity of food quality (price of calories) with respect to PCE (Per Capita Expenditure)

	Full sample		Sample A		Sample B	
Log Price of Calories	(1)	(2)	(3)	(4)	(5)	(6)
log PCE	0.383	0.373	0.361	0.353	0.400	0.389
	(0.012)***	(0.012)***	(0.020)***	(0.020)***	(0.022)***	(0.023)***
Interactions with log PCE						
Children age 0-4		0.009		0.008		0.013
		(0.002)***		(0.003)***		(0.002)***
Children age 5-9		0.008		0.005		0.009
		(0.001)***		(0.003)*		(0.002)***
Children age 10-14		0.001		0.000		0.002
		(0.001)		(0.003)		(0.002)
Observations	6029		2988		3041	
R-squared	0.60	0.60	0.50	0.50	0.42	0.42

Notes: *significant at 10%; ** significant at 5%; *** significant at 1%
Sample A: Households below the median of PCE
Sample B: Households above the median of PCE
See text for more details on the regression model and method used

Table 5 – Between-group elasticity estimates

	Full sample		Sample A		Sample B	
Log Price of Calories	(1)	(2)	(3)	(4)	(5)	(6)
log PCE	0.195	0.192	0.212	0.210	0.183	0.179
	(0.008)***	(0.008)***	(0.013)***	(0.013)***	(0.016)***	(0.016)***
% explained of total calorie price elasticity wrt PCE	**50.9**	**50.1**	**58.7**	**58.2**	**45.8**	**44.8**
Interactions with log PCE						
Children age 0-4		0.006		0.006		0.007
		(0.001)***		(0.002)***		(0.002)***
Children age 5-9		0.003		0.002		0.003
		(0.001)***		(0.002)		(0.001)**
Children age 10-14		-0.001		-0.003		0.000
		(0.001)		(0.002)*		(0.002)
Observations	6029		2988		3041	
R-squared	0.38	0.38	0.37	0.37	0.29	0.28

Notes: * significant at 10%; ** significant at 5%; *** significant at 1%
Sample A: Households below the median of PCE
Sample B: Households above the median of PCE
See text for more details on the regression model and method used