Sobre o comportamento do consumidor na zona urbana colombiana: o caso de Bogotá

Jorge Barrientos Marín[*]

Neste artigo eu estou interessado em estimar as curvas Engel semi-paramétrico para alimentação, roupa e transporte público, usando a ECV-DANE (1997). A maior parte dos artigos sobre o tema não inclui o transporte, o que é uma motivação, pois em países de baixo rendimento como a Colômbia, o transporte público é o meio que usa a maioria das pessoas para se mobilizar, em especial para ir ao trabalho. É provável que enfrentemos problemas de processos endógenos ao estar conjuntamente determinados ao gasto total e ao gasto nos diferentes bens. Além disso, realizamos contrastes não-paramétricos. Nossas estimativas produzem em alguns casos curvas de Engel quase quadráticas para alimentos e transporte, indicando que os alimentos podem ser para alguns um luxo e o transporte um bem inferior.

Classificação JEL: C14, D12, D1.

Palavras chave: curva de Engel, gasto em bens, modelos semi-paramétricos, endogeneidade, variáveis instrumentais, mostra.

Sobre el comportamiento del consumidor en la zona urbana colombiana: el caso de Bogotá

Jorge Barrientos Marín[*]

Este artículo estima las curvas de Engel semiparamétricamente para alimentación, ropa y transporte público, usando la ECV-DANE de 1997. La mayoría de los artículos sobre el tema excluyen el transporte, lo que constituye una de nuestras motivaciones, pues en países con bajos ingresos como Colombia la mayoría de las personas se movilizan utilizando el transporte público, en especial para ir a sus trabajos. Es probable que enfrentemos problemas de endogeneidad al estar conjuntamente determinados el gasto total y el gasto en los diferentes bienes. Adicionalmente, llevamos a cabo contrastes no paramétricos. Nuestras estimaciones produjeron en algunos casos curvas de Engel casi cuadráticas para alimentos y transporte, lo cual es un indicativo de que los alimentos pueden ser un lujo para algunos y el transporte, un bien inferior.

*Este artículo fue presentado en la sesión de Ingreso y Consumo de la reunión anual de la LACEA, Bogotá, Colombia 2007.

Soy único responsable por la interpretación y posibles errores. Departamento de Economía, Universidad de Antioquia.

Correo electrónico: jbarr@economicas.udea.edu.co

Documento recibido: 1 de octubre de 2009; versión final aceptada: 31 de enero de 2009.

On the Consumer Behavior in Urban Colombia: The Case of Bogotá

Jorge Barrientos Marín[*]

In this paper we are interested in estimating semi-parametric Engel curves for different categories of goods using data drawn from the 1997 Family Expenditure Survey (ECV1997). Most of the papers about consumer behavior include food and clothing shares in the analysis. Since people in developing countries spend a great deal of time going to work, through their use of public transportation, then it is interesting to take this into account in this study. Due to joint determination of expenditure shares and total expenditure, we adjust for any possible endogeneity, which is a familiar problem in demand analysis. So, not only estimation but also adaptive testing is provided by non-parametric procedures. Our estimations produce quadratic food and transportation Engel curves, which is an interesting result.

JEL classification: C14, D12, D1.

Keywords: Engel curve, expenditure share, semi-parametric model, endogeneity, instrumental variable, bootstrap.

I. INTRODUCTION

Empirical economics analysis of Engel curves has been an important topic for almost 50 years, since the early studies of Working (1943), Leser (1963), Deaton and Muellbauer (1980a, 1980b), in which they have developed parametric structures for investigating demand and Engel functions. From semi- and non-parametric points of view, Banks, Blundell and Lewel (1997), Blundell, Duncan and Pendakur (1998), Blundell, Browning and Crawford (2003) and Lyssouto, Pashardes and Stengos. (2002) have developed non-parametric and semi-parametric approaches to investigate consumer behavior in the United Kingdom.

In this study we are interested in estimating the relationship between total expenditure and budget shares using data drawn from the 1997 Family Expenditure Survey[1]. Additionally, we estimate the conditional mean of share expenditures by controlling the regression model by family characteristics such as gender, age, schooling (of household head) and the social stratum. It is worth remarking that the Engel curve analysis allows us to carry out welfare analysis. Moreover, Engel curve play an important role in models of income distribution and poverty.

Most of the papers concerned with consumer behavior have in common three broad categories of goods: food, drink/tobacco and clothing shares. In this paper we have included public transportation share instead of drink/tobacco share. However, most papers do not make special reference to public transportation. We introduce transporta-

---

1 Family Expenditure Survey, or ECV for its Spanish name, contains information of households collected by Departamento Nacional de Estadística (DANE).

tion because people, especially in developing countries, are transported by local public service. We mean, each day there are many people (especially workers) who travel throughout the city of Bogotá to work, usually from the south to the north of the city.

It is well known that food share is very important in consumer behavior analysis because access to food (nutritional component, calories etc) is considered an important measure of social welfare; so it is important to know what the pattern of food consumption is in the urban zone in a low-income country. Bogotá is a city in which there are many differences (among households); additionally Bogota attracts people from different regions of the country with different preferences and tastes. Clothing share is also included in this study in order to compare the food Engel curve with the clothing Engel curve. We are motivated to compare both curves because we think that food is a *basic good* while clothing is a necessity but not a *basic good.* Some co-variates, which usually have a linear effect such as age, gender, size and stratum (households are classified, by the local government, according to some characteristics especially those referred to as the quality of house building) are introduced parametrically. We make note that our estimations produce quadratic food and transportation Engel curves, which is a key feature because the current empirical evidence displays linearity of food Engel curve for poorest people in richer countries.

It should be remarked that a problem we may well have to consider is the endogeneity of regressors. Note that in this context total expenditure may well be jointly determined with expenditure shares. The approach used to solve this problem is instrumental variable estimation. We remark on two recently developed procedures in the context of semi- and non-parametric regression to tackle the problem of endogenous regressors: the so-called non-parametric two step least square (NP2SLS) according to Newey and Powell (2003), and the non-parametric two steps with generated regressors and constructed variables (NP2SCV) according to Sperlich (2005). Newey's approach is a cumbersome procedure involving the choice of a basis expansion at the first step. However, Sperlich's approach only requires a non-, semi- or even parametric construction of regressors of interest in the first step. An empirical paper where this approach is applied can be found in Barrientos (2006).

We don't examine the share-total expenditure relationship when the household survey data contains a fraction of reported zero expenditure shares. There are many explanations for observed zero expenditure. Firstly, it can result from false reporting by the respondent or the enumerator. Secondly, it is possible to have so low frequency of purchase that at the moment of the survey no purchase is recorded for some house-

holds. Finally, it could be possible that some households do not actually purchase the good during the period required by the survey. In general, the source of the zero expenditure shares is unobservable. We need other robust methods (different to non-parametric ones) to get robust Engel curves. Consequently, we get rid of zero expenditure observations.

The contribution of this study can be summarized as follows. Firstly, we investigate consumer behavior (through Engel curves) by estimating a system of partial linear models (PLM). We include in our regressions some demographic characteristics concerning the household head. Secondly, to adjust our estimations for endogeneity we use Sperlich's approach by estimating a non-parametric model; family income and other exogenous covariates are our main instrumental variables set. In the context of Engel curve it is a common practice to correct endogeneity by instrumenting total expenditure with family income. Finally, we consider testing both linearity and quadratic effects of the estimated non-parametric function (in PLM) by performing adaptive testing hypothesis.

Papers related to this one are Banks *et al.* (1997), Blundell *et al.* (1998), Blundell *et al.* (2003), Lyssiotou *et al.* (2002), Barrientos (2006) and references therein. Concerning the Colombian case, it is worth noting that there are not many papers about consumer behavior. Nevertheless, Lasso (2003), among others, tries to get non-parametric estimation of Engel curve. However, we found some problems in their estimations. Firstly, they do not adjust estimated models for endogeneity. Secondly, any non-parametric estimation should be accompanied with non-parametric testing of non-linearity or quadratic specification.

The layout of this paper is as follows. Section II makes a short introduction to the literature about estimation of Engel curves. In Section III we introduce the methodology for estimating the underlying (system of) conditional mean. In Section IV we present the data and results about estimating Engel curves. And finally, we conclude.

II. THE SHAPE OF ENGEL CURVES

An usual structure in consumer behavior analysis is the so-called Working-Leser specification. In this model each expenditure share is defined over the logarithm of total expenditure. Thus the model has a simple structure given by:

$$w_i = g\left(\ln X_i\right) + \varepsilon_i \quad i = 1,...,N \tag{1}$$

where w_i is the budget share, $\ln X_i$ is the log total expenditure and ε_i is an error term satisfying $E\left(\varepsilon_i \mid \ln X_i\right) = 0$. Empirical analysis using parametric specification as in (1) can be found in the literature on consumer behavior, see for instance Deaton and Muellbauer (1980a, 1980b) and references therein. Note that this model has a one-dimension estructure. In this paper, we make an extension of model (1) given by

$$w_i = \psi + \varphi' Z_i + m\left(\ln X_i\right) + \varepsilon_i \tag{2}$$

where w_i denotes expenditure share, $m(\ln X_i)$ is an unknown function of the total expenditure, Z_i is a matrix of household backgrounds (age, schooling, gender), including some demographic variables like household size and household stratum. The term ε_i is supposed to satisfy the standard assumption $E\left(\varepsilon \mid \ln(X), Z\right) = 0$. Most of the papers about consumer behavior using fully-parametric approaches, usually include $\left(\ln\left(X_i\right)\right)^2$ to capture possible non-linearities in the Engel curve. Note that in this paper, such hypothesis is not needed, because, by definition $m(.)$ could have any functional form; moreover, leaving $m(.)$ unrestricted permits a commodity that is initially a luxury to become a necessity at higher income levels.

As we remarked in the introduction of this work, family income is used to adjust our Engel curves for endogeneity, which is caused by the joint determination of the total expenditures and the expenditure share. Additionally, we use as instruments some exogenous variables contained in the matrix Z. For this purpose we used the non-parametric method developed by Sperlich (2005), which gets rid any assumption about the range of matrix of instruments. Moreover, Sperlich's approach could be performed by using a single instrument.

III.   ESTIMATING AND TESTING

A.   PARTIALLY LINEAR MODEL AND INSTRUMENTAL VARIABLES

In order to get an appropriate approach to Engel curves, we estimate the PLM given in (2), which is motivated because they allow us to include other regressors that typically show linear effects (dummies or any categorical covariate). At the same time we consider reducing the *curse* of dimensionality, which may be the main weakness of non-parametric technique, by estimating this semi-parametric PLM. To get a $\sqrt{n}$-consistent estimator of ϕ we use Robinson's approach (1998). So let $\hat{\phi}$ be an estimator of ϕ, then (2) can be rewritten as:

$$\omega_i = m\left(\ln X_i\right) + \xi_i \tag{3}$$

where $\omega_i = w_i - Z'_k \hat{\phi}_k$ and $\xi_i = \varepsilon_i + Z'_i \left(\hat{\phi}_k - \phi_k \right)$ is the new composite error term for all i and k. The intercept ψ term can be $\sqrt{n}$-consistently estimated by $\hat{\psi} = \overline{Y} - \overline{Z}^T \hat{\phi}$ where $\overline{Y}$ and $\overline{X}$ are the sample mean. Then (3) can be estimated consistently using standard kernel regression estimator developed independently from each other by Nadaraya (1964) and Watson (1964).

Now we turn to the problem mentioned in the Introduction about constructing regressors to overcome the endogeneity problem. A detailed explanation can be found in Sperlich (2005). Let x_i be an unobservable or endogenous variable and let $\hat{X}_i$ be a generated regressor[2], it is then possible to write $\hat{x}_i = x + b(x) + \sigma(x)$, where $b(.)$ is the bias term such that $b(.) \to 0$ as $n \to \infty$ and $\sigma(.)$ is the variance term. In order to obtain consistent estimates of the conditional mean and thus construct $\hat{x}_i$, with the help of instruments or even with help from different data sets it is possible to estimate the reduced regression form, semi-, non- or even parametrically (first step) and then use it in the structural regression (second step), instead of the original regressor. The procedure can be described as follows: let $\{W_i\}$ be the set of exogenous variables. We carry out the estimation of Engel curves by constructing total expenditure using all exogenous variables as instruments in the non-parametric (multidimensional) regression of x_i on W_i, so at the first step we get an estimator of x given by $\hat{x} = \hat{g}(W)$ and after, in a second step, we use $\hat{x}$ to estimate our PLM.

As we mentioned above we use family income as instrument, and additionally we include as instruments age and schooling of household head. Note that we are worried about endogeneity due to joint determination of total expenditure, $\ln X_1$, and expenditure share (*endogeneity due to simultaneity*). Suppose that $\ln X_1$ is endogenous such that:

$$\ln(X_1) = g(W) + U \tag{4}$$

where $E(\xi \mid W) = E(U \mid W) = 0$, but $E(\xi \mid \ln X) \neq 0$. Putting (3) and (4) together we get

$$\omega = \psi + m \left(g(W) + U \right) + \xi \tag{6}$$

---

2 Both bias and variance have orders of $O(g^2)$ and $O\left(\dfrac{1}{ng^\delta}\right)$ respectively, to fulfil the assumptions of Theorem 2 in Sperlich (2005).

applying the modeling $m\big(g(W)+U\big)=m_g\big(g(W)\big)+\lambda(U)$, which somehow means that we assume additivity in the exogenous impact of the explanatory variables, which is possible as an assumption. So, we get:

$$\omega = \psi + m_g\big(g(W)\big) + \lambda(U) + \xi$$
$$\omega = \psi + m_g\big(g(W)\big) + \tilde{\xi}$$

(7)

where $E\big[\lambda(U)\big]=0$ and $E\big[\xi \mid g(W)\big]=0$. Then, the expression (2) is the model that we estimate only with the additional burden of a consistent estimation of such function $g(.)$.

B.   TESTING PARAMETRIC V.S. NON-PARAMETRIC EFFECTS

We are interested in testing hypothesis of linearity or quadratic effect, which depends on the form of estimated regression function in model (2). So we establish the null hypothesis as follows: $H_0 : m(x) = m_\gamma(x)$ for some γ, with $m_\gamma(x) = \gamma x^l$, where $l=1$ and $l=2$ denotes linear and quadratic relationship between total expenditure and budget share.

Testing hypothesis can be performed by modifying the Hardle, Huet, Mammen and Sperlich (2004)'s procedure. To this end we use three test statistics. The first one is defined as the square of the differences between the partially linear model and the fully-parametric estimator, extending the concept introduced by Hardle and Mammen (1993). In order to test the validity of our hypothesis, we also consider the test statistics introduced by Gozalo and Linton (2001) and Rodriguez-Poo, Sperlich and Vieu (2005) given by:

$$\hat{T}_1 = \frac{1}{n}\sum_{i=1}^{n}\big[\hat{m}(X_i)-\hat{m}_\gamma(X_i)\big]^2 w(X_i)$$

(8)

$$\hat{T}_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\varepsilon}_i\big[\hat{m}(X_i)-\hat{m}_\gamma(X_i)\big]w(X_i)$$

(9)

$$\hat{T}_3 = \sum_{i=1}^{n}\left[\frac{1}{nk^d}\sum_{j=1}^{n}K_h\big(X_i-X_j\big)\big(Y_j-\hat{m}_\gamma(X_j)\big)\right]^2 w(X_i)$$

(10)

where $\hat{\varepsilon}_i = Y_i - \hat{m}_\gamma(X_i)$ are the residuals of the model under the null hypothesis. Now we discuss the performed procedure to compute the critical values. Note that our idea is based on a combination of adaptive test statistics with bootstrap scheme (Horowitz and Spokoiny (2001)). Having estimated semi-parametric and non-parametric models, $\hat{m}_\gamma(\cdot)$ and $\hat{m}(\cdot)$ respectively, we construct the original test statistics denoted by $\hat{T}_{jk}$. As the distribution of $\hat{T}_{jk}$ varies with k we define the standard test statistic denoted by

$$\hat{\tau}_{jk} = \frac{\hat{T}_{jk} - \hat{\mu}_j}{\hat{v}_j} \tag{11}$$

where $\hat{\mu}_j$ and $\hat{v}_j^2$ are the estimated mean and variance of the test $\hat{T}_{jk}$ for $j = 1,2,3,$. Then we compute the test statistics based on the resampling data (boootstrap), denoted by:

$$\hat{\tau}_{jk}^* = \frac{\hat{T}_{jk}^* - \hat{\mu}_j^*}{\hat{v}_j^*} \tag{12}$$

This procedure creates a family of test statistics $\{\tau_k, k \in K_n\}$ where the choice of k makes the difference between the null and global alternative hypotheses. In order to maximize power we take the maximum of $\hat{\tau}_{jk}^*$ over a finite set of bandwidth values K_n with cardinality L. Then we define the final test statistics by:

$$\hat{\tau}_{jk}^{**} = \max_{k \in K_n} \hat{\tau}_{jk}^* \tag{13}$$

where $K_n = \left\{ k = a_{(l)} n^{-1/5} \quad l = 1,...,L \right\}$, $a(l) = \left[l + \left(c_X (l-1)^{-1} \right) \right] n^{-1/5}$, and $c_X = C \left(\max(X_i) - \min(X_i) \right)$ with $C \in (0,1)$. The testing procedure rejects H$_0$ if at least one of the $k \in K_n$ causes original test statistic to be significantly larger than the bootstrap analogues. In Horowitz and Spokoiny (2001) the estimators for variance and bias are asked to be consistent under alternative hypotheses. Note that this is only necessary for efficiency; for consistency of the test, it is sufficient for the difference between real variance and estimate to be bounded.

In summary, to obtain bootstrap critical values we consider the following steps. 1) Estimate $\hat{m}_\gamma(X) = \hat{\psi} + \sum_{q \in Q} \beta_q X_q$. 2) Use the bootstrap scheme to get ε_i^* for each $i = 1, ..., n$. 3) For each $i = 1, ..., n$ generate $Y_i^* = (\hat{m}_\gamma) X_i + \varepsilon_i^*$, where ε_i^* is sampled randomly and we use the data $\{Y_i^*, X_i\}_{i=1}^n$ to estimate $\hat{m}_\gamma(x)$ under H$_0$. 4) Repeat the

process 1-3 B times to obtain $\{\tau^*_{jk}\}$ and use these B values to construct the empirical bootstrap distribution. The bootstrap errors ε^*_i are generated by multiplying the original estimated residuals from the model under null hypothesis, $\hat{\varepsilon}_i = Y_i - \hat{m}_\gamma(X_i)$, by a random variable with standard distribution. This procedure provides exactly the same first and second moments for $\hat{\varepsilon}_i$ and for ε^*_i.

IV.   DATA USED IN THIS APPLICATION AND RESULTS

In this study we consider three broad categories of goods: food, clothing and public transportation shares. We use data from Bogota D.C. Total income, total expenditure and expenditure shares are measured in pesos (yearly) at constant 1997 prices. Our ECV1997 sample contains households in which the household head was working at the moment of the survey. For clothing and transport shares there are expenditures reported at zero, the fraction of which is 12% and 10%, respectively. If we get rid of zeros, these restrictions leave us with a sample of 610 observations. Each Graph presents the estimated curve together with 95% bootstrap pointwise confidence bands (dashed lines).

In the estimation of our regression functions, normal bootstrap errors are used to generate bootstrap sample for computing confidence bands, as well as for computing critical values to test linear and quadratic null hypotheses. Bandwiths for bootstrapping are obtained by adaptive methods. This parameter is different to that to get the estimation of non-parametric curve which is obtained by the leave-one-out cross-validation method to automatic bandwidth choice. In all cases we present kernel regression for the quartic kernel $\frac{15}{16}\left(1-u^2\right)^2 I\left(|u|\leq 1\right)$ where $I(.)$ is the indicator function. Table 1 gives us some descriptive statistics for expenditure data. According to this one, we have that in this sample, on average, a typical home consists of a male (household head) in his forties (45.3) who has 9.52 years of education and 4 members per home, all of them residing mainly in strata 3 and 4. Graphs 1-3 and Table 2 comprise results about Partial Linear Model.

Graph 1 displays the semi-parametric estimations of the *food* Engel curve. The vertical axis refers to the budget share adjusted by family background. Partially consistent with Engels' law, the curve slopes downwards for higher expenditure levels. Food is a broad aggregate of 21 commodities which, on average, consumes 24.3% of the household budget in Bogotá. Table 3 presents parametric OLS estimates of the Engel curve which show that the quadratic in log expenditure is highly significant. According to Gorman (1981) a quadratic Engel curve implies a demand

system of rank three, so Almost Ideal Demand System of Deaton and Muellbauer (1980) and any Linear Expenditure System are inappropriate for demand analysis in Bogota, and in general in regions with strong inequalities among the population.

Table 1
Descriptive Statistics for Expenditure Data

| Variables	Mean	Std. dev	Min	Max
Food share	0.243	0.137	---	---
Clothing share	0.042	0.038	---	---
Transportation share	0.078	0.068	---	---
Log total Expenditure	14.1	1.08	---	---
Log total income	13.5	1.42	---	---
Size	4.19	1.80	1	13
HHAge	45.3	13.4	17	87
HH Schooling	9.52	4.87	0	20
HHSEX	0.76	---	0	1
Stratum_1	0.001	---	0	1
Stratum_2	0.028	---	0	1
Stratum_3	0.28	---	0	1
Stratum_4	0.47	---	0	1
Stratum_5	0.10	---	0	1
Stratum_6	0.067	---	0	1
Observations		610		

Source: author's calculation.

Graph 1
Nonparametric Estimated Effect of Total Expenditure on Foodshare.
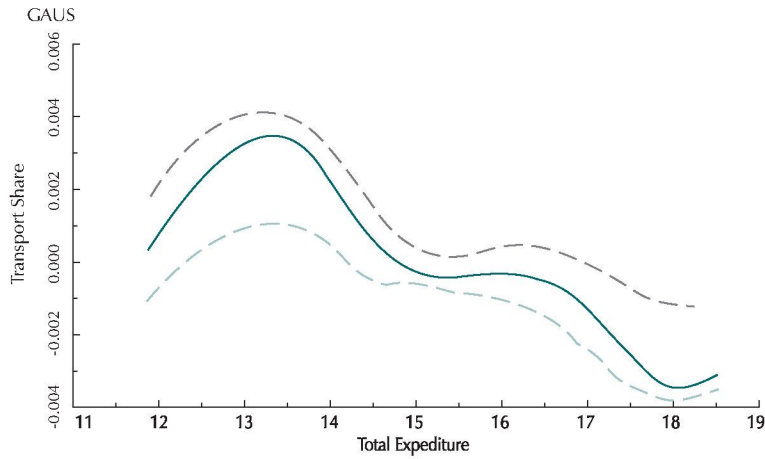Bandwitdth $k = 1.9$
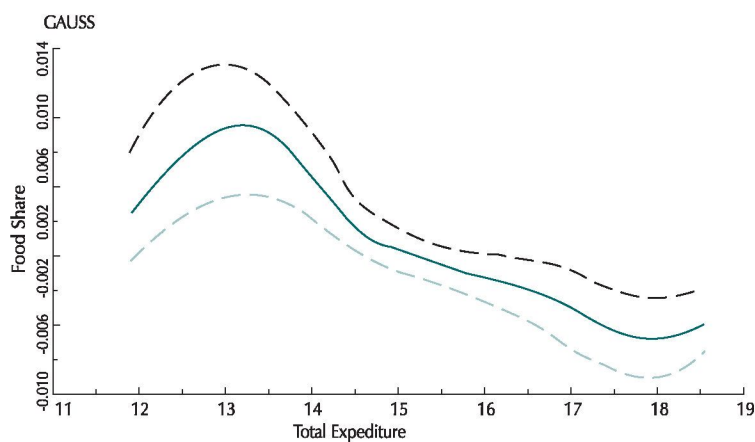
Source: author´s calculations.

Graph 2
Nonparametric Estimated Effect of Total Expenditure on Clothing Share.
Bandwidth $k = 2.2$
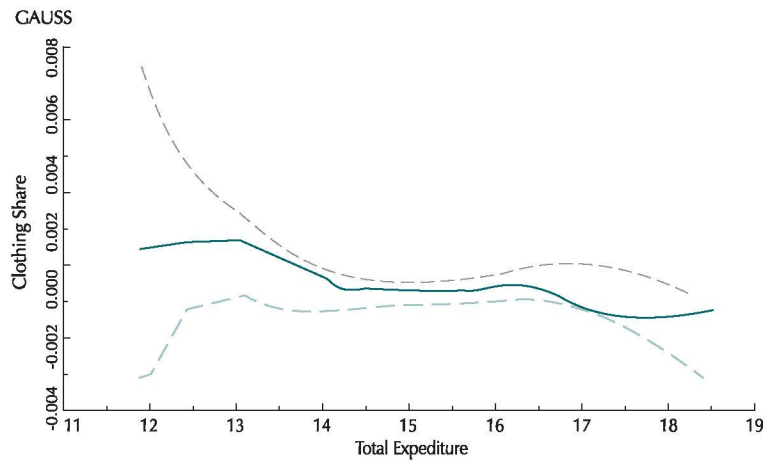
Source: author´s calculations.

Graph 3
Nonparametric Estimated Effect of Total Expenditure on Public Transportation
Share. Bandwitdth $k = 2.0$

Source: author´s calculations.

A quadratic food Engel curve is an interesting result. The food Engel curve is usually (almost) linear in the richer populations of Spain (see Delgado and Miles, 1996), the U.K. (Banks *et al.*, 1997) and the USA (Lewbel, 1991), suggesting that the quadratic is a feature of the income levels in our sample ECV1997 and the stage of economic development. In our data, as expenditure rises, food share increases to attain a maximum, but for higher expenditures levels the food share decreases. The shape of this Engel curve suggests that food is a normal good for low-income families in the sample and an inferior one for the others. According to this feature we could think that food basket of the poorest people should be deficient in either quantity or quality.

Table 2
Parametric Estimation in PLM and OLS Models

| Variable | Estimated Parametric Component in PLM | | | | | |
| | Food Share | | Clothing Share | | Trans. Share | |
	Parameter	Z*	Parameter	Z*	Parameter	Z*
Size	0.015	3.68	0.0003	0.27	0.006	2.60
Schooling	-0.003	2.58	0.0001	0.36	0.001	2.22
Age	-0.0005	-1.43	-0.0004	3.66	0.0002	1.19
Gender	0.02	1.73	0.003	0.87	0.009	1.32
R2	0.13		0.05		0.041	
Variable	Estimated Parametric Model OLS					
	Food Share		Clothing Share		Trans. Share	
	Parameter	Z*	Parameter	Z*	Parameter	Z*
Size	0.14	3.83	0.0005	0.51	0.006	2.96
Schooling	-0.0035	2.55	0.0002	0.57	0.001	2.51
Age	-0.0003	0.8	-0.0004	3.5	0.0002	1.15
Gender	0.23	1.91	0.003	0.94	0.009	1.8
Ln(X)	0.214	2.21	-0.041	1.41	0.034	0.88
$(Ln(X))^2$	-0.008	2.82	0.001	1.29	-0.001	1.23
R2	0.25		0.06		0.048	

(*) *Z:* Standard Normal Distribution.
Source: author's calculation.

Graph 2 displays the semi-parametric estimate of the *clothing* Engel curve. This commodity comprises clothing and footwear (for men, women and children) and, on average, consumes 4.2% of the household budget. Note the (almost) linear effect of total expenditure on clothing share. This behavior is explained because clothing, unlike food or transport, is a necessary good but not a basic one. Though clothing is important and necessary, people have to feed and take the bus or a taxi to get their workplace. The stable behavior of clothing share across households means that people prefer meeting their basic needs.

Graph 3 displays the semi-parametric estimation of the *transport* Engel curve. This commodity comprises of tickets for buses (urban and rural) and money spent on taxi journeys and, on average, consumes 7.8% of the household budget. It is worth noting that as long as total expenditure is increasing the amount devoted to transport share is decreasing for high level of expenditure. This behavior is typical of households in developing regions where the form of the Engel curve usually implies that public transportation is an inferior good for many people. It is possible that a typical household head, who improves his income, considers buying a car for private transportation.

Age has an interesting effect on consumer behavior. Each estimation produces a significant and negative effect; Table 1 shows us that 76% of household heads in our ECV1997 sample are men: from a sociological point of view men do not care as much about fashion as women; Barrientos (2006b) estimates the non-parametric effect of age on consumer behavior coming to the same conclusion obtained in this investigation: spending in clothing decreases with age, in particular for household heads of 40 and over. Note that non-parametric estimation gives us the possibility of observing the effect of whole age range on expenditure commodities whereas parametric procedures produce an estimated parameter; thus a single effect (linear) can be observed.

Table 2 shows us the semi-parametric effect of total expenditure on budget shares. Note that size has a positive effect in all cases but is significant for food and transport shares. This behavior is clear: more people imply more expenditure share. In general, it seems that from a parametric point of view no variable is important for clothing share. Schooling has different but significant effects. On the one hand, schooling has a positive effect on transportation share; but on the other hand, it has a negative effect on food share. Note that food share does not include meals out of home, so people (household head and maybe other members) are likely to feed out of home if they have increasing income, so schooling affects food share negatively via family income. This result is similar for fully-parametric estimation reported in Table 2. The Engel curves include stratum effects which are small and only marginally significant.

If the model is chosen correctly, these results quantify the extent to which each variable affects consumer behavior. We remark some results about the linear and quadratic effects of total expenditure on budget shares. To perform adaptive testing procedure we define a set of K_n (with cardinality $L = 10$) of bandwidths k's to range from 0.24 to 0.87. As we mentioned above, optimal bandwidth to estimate under-

lying regression function is obtained by cross-validation (these are the same as those used in the estimation). Moreover, it is worth mentioning that bootstrap for testing hypotheses are generated in a different way from the one for generating confidence bands. In practice (we mean in applications) we have to over-smooth for bootstrap tests and under-smooth for confidence bands (Barrientos and Sperlich 2006).

We used 1000 normal bootstrap samples and performed testing procedure according to aforementioned algorithm. Because over-smothing parameter to generate bootstrap sample is crucial to compute critical values, we carried out test hypothesis with different bandwidths, but we did not find significant differences in results. Therefore, we report critical values for a single over-smoothing parameter for each underlying regression $g_{fs} = 3.0$ for food share, $g_{cs} = 2.3$ for clothing share and $g_{ts} = 2.1$ for transport share. According to performed testing procedure, for clothing share we are able to reject linearity of log total expenditure at $\alpha = 0.1\%$ with τ_1 and τ_2; with τ_3 we reject linear effect of total expenditure on clothing share at $\alpha = 6.5\%$.

On the other hand, we tested quadratic effect for the food and transportation Engel curve. According to the evidence, both transportation and food Engel curve displays imperfect quadratic forms. Of course, linearity is strongly rejected; we are unable to reject quadratic effect, however. Note that neither τ_2 nor τ_3 rejects the null hypothesis of quadratic effect of total expenditure on food expenditure ($\alpha = 20\%$ and $\alpha = 95\%$ respectively). According to the testing procedure, there is no evidence against quadratic effect of total expenditure on public transportation ($\alpha = 48\%$ and $\alpha = 96\%$ with τ_2 and τ_3). These results are coherent with the shape of the estimated curves in Graphs 1-3. This fact gives us an idea about the robustness and reliability of our methods.

V. CONCLUSIONS AND FUTURE RESEARCH

General results obtained from the estimation of Engel curves show that modelling the effects of total expenditure on the different budget shares deserves a better treatment than usually found in one-dimensional semi-parametric analysis. Bhalotra and Cliff (1998) ask if the non-linearity of the Engel curves could reflect specification error. It is a point to be included in the future.

Note that in this paper we only take into account a partial household composition (we only control for size, gender, schooling and strata). Therefore, a reasonable

extension of empirical analysis with PLM could be carried out by introducing more demographic variation to obtain variety in behavior (regions, labor market, temporal dummy to capture price effects, etc.). Moreover, we could be interested in allowing Z_{ij} vary in any way with j and testing Stlusky symmetry hypothesis, for that it would be necessary to impose a function to get general equivalence scale in order to fulfill conditions of proposition 5 in Blundell *et al.* (2003).

Another interesting point to investigate is whether changes in consumer preferences take place over time and make an extension to dynamic models. One can take data from the 2003 Family Expenditure Survey, for example, to compare consumer behavior. Additionally, more categories of goods (health, furniture, house, rent, etc) should be included to investigate, for instance, whether health expenditure is a luxury good.

Finally, every paper about non-parametric applications to consumer behavior should report non-parametric elasticities. It is not easy, because this requires estimations of the first derivatives. In the parametric world elasticities often are constant between families, whereas in nonparametric universe, elasticity is a function (involving the first derivative of the Engel curve), so by definition the elasticity would be different among individuals. So future research should include non-parametric elasticities.

REFERENCES

1. Bhalotra, S.; Cliff, A. "Intrahousehold Resources Allocation in Rural Pakistan: A Semi-parametric Analysis", *Journal of Applied Econometrics*, vol. 13, num. 5, pp. 463-480, 1998.

2. Banks, J.; Blundell, R.; Lewbel, A. "Quadratic Engel Curves and Consumer Demand", *The Review of Economics and Statistics*, vol. 79, num. 4, pp. 527-539, 1997.

3. Barrientos Marín, J.; Sperlich, S. *The Size Problem of Kernel Based Bootstrap Test When the Null Is Non-parametric*, University of Alicante, Social Science Research Network-SSRN: *http:// ssrn.com/abstract=1010939 Submitted*, 2006.

4. Barrientos Marín, J. "Estimation and Testing An Additive Partially Linear Model in a System of Engel Curves", University of Alicante, IVIE *Working Paper*, WP-AD 2006-23.

5. Blundell, R.; Duncan, A.; Pendakur, K. "Semiparametric Estimation and Consumer Demand", *Journal of Applied Econometrics*, vol. 13, num. 5, pp. 435-461, 1998.

6. Blundell, R.; Browning, M.; Crawford, I. "Non-parametric Engel Curve and Revelead Preference", *Econometrica*, vol. 71, num. 1, pp. 205-240, 2003.

7. Deaton, A.; Muellbauer, J. "An Almost Ideal Demand System", *American Economic Review*, vol. 70, pp. 321-326, 1980a.

8. Deaton, A.; Muellbauer, J. *Economic and Consumer Behavior*, Cambridge, Cambridge University Press, 1980b.

9. Delgado, M. A.; Miles, D. Households Characteristics and Consumption Behavior: A Nonparametric Approach. *Empirical Economics*, 1996.

10. Gorman, W. M. "Some Engel Curves". A. Deaton (ed), *Essays in the Theory and Measurement of Consumer Behavior in Honor of Sir Richard Stone*, Cambridge, Cambridge University Press, pp. 7-29, 1981.

11. Gozalo, P. L.; Linton, O. B. "Testing Additivity in Generalized Non-parametric regression models with Estimated Parameters", *Journal of Econometrics*, num. 104, pp. 1-48, 2001.

12. Hardle, W.; Mammen, E. "Comparing Nonparametric Versus Parametric Regression Fits", *Annals of Statistics*, vol. 21, num. 4, 1926-1947, 1993.

13. Hardle, W.; Huet, S.; Mammen, E.; Sperlich, S. "Semi-parametric Additive Indices for Binary Response and Generalized Additive Models", *Econometric Theory*, num. 20, pp. 265-300, 2004.

14. Härdle, W.; Müller, M.; Sperlich, S.; Werwatz, A. *Non-parametric and Semi-parametric Models*, Springer Series in Statistics, Springer-Verlag, 2004.

15. Horowitz. J, L.;Spokoiny, V. "An Adaptive, Rate-optimal Test of Parametric Mean-Regression Model Against A Non-parametric Alternative", *Econometrica*, vol. 69, num. 3, pp. 599-631, 2001.

16. Lasso, F. "Economías de escala en los hogares y pobreza", *Revista de Economía del Rosario*, vol. 6, núm. 1, pp. 23-52, 2003.

17. Lewbel, A. "The Rank of Demand Systems: Theory and Non-parametric Estimations", *Econometrica*, num. 59, pp. 711-730, 1991.

18. Leser, C. E. V. "Form of Engel Functions", *Econometrica*, vol. 31, num. 4, pp. 694-703, 1963.

19. Nadaraya, E. A. "On Estimating Regression", *Theory Probability Applied*, num. 10.1, 1964.

20. Newey, W. K.; Powell, J. "Instrumental Variables Estimation of Non-parametric Models", *Econometrica*, num. 71, pp. 1565-1578, 2003.

21. Lyssiotou, P.; Pashardes, P.; Stengos, T. "Age Effects on Consumer Demand: An Additive Partially Linear Regression Model", *The Canadian Journal of Economics*, vol. 35, num. 1, pp. 153-165, 2002.

22. Robinson, P. "Root N-Consistent Semiparametric Regression", *Econometrica*, vol. 56, pp. 931-54, 1988.

23. Rodriguez-Póo, J. M.; Sperlich, S.;Vieu, P. "An Adaptive Specification Test For Semi-parametric Models", *Working Paper*, University of Carlos III de Madrid, 2005.

24. Sperlich, S. *A Note on Non-parametric Estimation with Constructed Variables and Generated Regressors*, Social Science Research Network-SSRN: http://ssrn.com/abstract=1010923.2005

25. Watson, G. S. "Smooth Regression Analysis", *Sankhy ā Ser.* A 26, 1964.

26. Working, H. "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, vol. 38, pp. 4-56, 1943.