Modelling Zero Expenditures on Italian Household Consumption

Economic Notes by Banca Monte dei Paschi di Siena SpA, vol. 27, no. 1-1998, pp. 55-96

Modelling Zero Expenditures on Italian Household Consumption

ROSSELLA BARDAZZI* - MARCO BARNABANI*

This paper discusses the problem of dealing with zero expenditures in consumption analysis. Various causes of zero expenditures are analyzed and alternative econometric models are applied to highly disaggregated household consumption data. The source of zero expenditures matters in the choice of the appropriate model. Tobin's model, generally suggested, is not appropriate when the percentage of zeros in the sample is very high. Our findings lead us to agree on the necessity of using different models in the explanation of zeros for different goods in order to pinpoint the presumed major reason of zero purchases for each item.

(JEL Classification: C24, C25, D12)

Introduction

In this paper we present a cross-sectional study of expenditure behaviour at the household level. Much work has been done in this area, however the purpose of this study is to focus on the problem of dealing with zero expenditures in the estimation of an integrated cross-section and time-series demand system. Deaton (1986), among others, argues that "the problem of dealing appropriately with zero expenditures is currently one of the most pressing in applied demand analysis." (p. 1809).

Applied cross-section analysis relies upon household budget surveys. The most casual inspection of the observed data shows that, during the survey period, many households report zero purchases on various commodities. These zeros may or may not indicate no consumption of these goods. This major difficulty is very well known since the pioneering work of Tobin (1958). However the problem of dealing with

* University of Florence. The paper is the result of joint research by the authors, however Sections 1-5 are by R. Bardazzi, Sections 6-8 by M. Barnabani. The authors are grateful to C. Almon for encouragement, advice, and generous help.

Financial support from MURST (Italian Ministry of the University and Scientific Research) in gratefully acknowledged.

zero expenditures has often been neglected in estimating demand systems although a wide variety of econometric methodologies has been proposed.

The main problem is that *consumption is not observable*. What we observe instead is household expenditure: the models we analyze are based on different hypotheses on how to link these two variables.

The first issue has been to distinguish various reasons why zero expenditures are observed. The second issue has been to specify alternative statistical models appropriate for treating zero expenditures due to the various causes.

In this study, we concentrate on Italian disaggregated consumption as a first step in a model involving both cross-section and time-series data. The result is to be used to forecast personal consumption expenditures in a long-term interindustry model. Work reported here is the first step towards equations with stronger micro foundations using both cross-section and time-series data. Cross-section data provide a perfect environment in which to measure the effects of household composition on consumption patterns. However, they are not overly useful when one wants to estimate the price effects. Time-series data, on the other hand, can be used for this purpose. Because the cross-section work must contribute to the time-series analysis, some understanding of the entire framework is necessary to understand the characteristics of the cross-section work presented here.

The main results of this paper are the following. First of all, we find that Tobin's model, which is generally suggested as a solution to the problem of zeros, is not appropriate when the percentage of zeros in the sample is very high. This result is consistent with what has been observed by Maddala (1990). Moreover, our findings lead us to agree with Pudney (1990) on the necessity of using different models in the explanations of zeros for different goods: the source of zeros matters in the choice of the appropriate model.

The paper is organized as follows. In Section 1, a cross-section consumption function is explained which uses both income and demographic variables. The linkage with the time-series analysis is shown in Sections 2 and 9. The problem of zero expenditures is tackled in Section 3 where the data set is illustrated. An analysis of some explanations of zero expenditures is developed and some models are proposed to interpret different sources of this phenomenon (Sections 4 and 5). The models used in this paper are fully analyzed along with their estimation schemes in Sections 6 to 8. In Section 10 the estimation results of the models are shown and discussed.

1. *General Form of the Cross-Section Consumption Function*

Our approach to the study of consumer behaviour is to combine information from different sources in establishing the system of demand functions using both cross-section and time-series data. In particular, in this work we concentrate on the cross-section analysis. We follow the procedure of establishing some properties of the Engel functions (the relationship between income and consumption) and quantifying the effect of demographic characteristics on the basis of cross-section information [1]. With the functional form thus specified as our starting point, we investigate various explanations of zero expenditures.

The basic idea of the consumption function used in our work is that a household's demand for a particular good is the product of two components: expenditure per household member and the "size" of the household specific to that good. The first component is determined by per capita household income and demographic characteristics of the household. The second component, the product-specific size of the household, depends on the age structure of the household, more particularly, it is a weighted sum of the household members grouped by age. For each consumption item, the relation may be summarized by the following function:

$$(1) \qquad c_i = (x_i\beta + d_i\delta)\, n_i w \qquad i = 1, ..., N$$

where

c_i : consumption of household i;

x_i : 1xk vector. Per capita income within household i divided in $k = 7$ brackets.

d_i : 1xm vector. Zero/one dummy variables used to show inclusion of household i in $m = 15$ demographic groups.

n_i : 1xg vector. Number of household members for $g = 8$ age groups.

β, δ, w : vectors of parameters to be estimated for each commodity.

N : number of households in our sample (34273).

[1] The system of consumption expenditures equations adopted in this study is based on work done at INFORUM (Interindustry Forecasting Project University of Maryland). The time-series demand system was originally designed by Almon (1979), then Devine (1983) and Chao (1991) expanded the model including the cross-section analysis and some empirical applications to the U.S. economy. Recently Almon (1996) revised and extended the time-series system of demands.

The term in parentheses on the right side of (1) is the per capita consumption within the family: it is a function of per capita income and non-age demographic variables. The weighted size of the household, the last term on the right of (1), is introduced to obtain the household consumption of each good.

1.1. *Engel Curves*

The first step in constructing the component of per capita household consumption is to establish a relationship between income and expenditure on a particular commodity.

The specification of the appropriate functional form for the Engel curve is a traditional topic in the literature of demand analysis and a wide selection of functional forms has been explored.

A classic study is that of Prais and Houthakker (1955) who listed and compared several functional forms to be suitable for different types of goods. This methodology has been followed by other scholars (for instance Brown and Deaton, 1972) and more recently defined "unashamedly pragmatic" because the functional forms were chosen on the grounds of fit[2]. With respect to the functional form used here we "unashamedly" place our approach in the stream of a pragmatic orientation: the goal of this work is to develop a cross-section analysis to be linked to a time-series demand system endogenously solved in a disaggregated long-term forecasting model.

The choice of the functional form has been based on some properties which the Engel curve should have.

An Engel curve should be flexible in the sense of being able a) to represent different types of goods: luxuries, necessities and inferior goods; b) to express different propensities to consume for different income levels. By using different forms for different commodities one can describe the behaviour of all types of goods. However it would be very convenient to find a simple form common to all goods which could be transformed in different shapes and so would approximate different functional forms.

A linear spline function should have the necessary flexibility. Splines have been traditionally used in approximation theory because they combine ease of handling with great flexibility[3].

A spline is a piecewise function in which the pieces are joined together

[2] See Deaton (1986), p. 1799.

[3] Applications of splines to estimate consumer demand can be found in Diewert and Wales (1992, 1993). Another use of splines in economics is concerned with the analysis of structural change, see Poirier (1976).

in a smooth fashion. To apply this function to design the Engel curve, income is divided in brackets. The relationship between income and expenditure can be thought as linear in each of these brackets, but if we run separate regressions for each bracket, the lines would not necessarily connect at the "knots", the points of transition from one linear segment to the next. Forcing the curve to be continuous at the knots is an application of the spline idea. The profile of this curve has the appearance of a sequence of linear segments which are continuous at the knot points. Such a curve is described as a **Piecewise Linear Engel Curve** (PLEC). To specify this idea more concretely, we define an arbitrary number of income brackets whose boundaries (B_L, $L = 1, \ldots, k-1$) are designed so that each bracket contains the same percentage of total households in the sample. (Nothing much depends on this choice of the boundaries, but it gives a good dispersion of the boundaries over the observed incomes.) These boundaries are the knots where the segments of our function must be joined.

The consumption of the *i*th household with a per capita income R_i falling in the *j*th bracket is predicted by[4]:

$$c_{ij} = b_{01} + \sum_{L=1}^{J-1} \beta_{1L}(B_L - B_{L-1}) + \beta_{1j}(R_i - B_{j-1}) \tag{2}$$

Equation (2) may be formulated as a conventional regression model whose deterministic term is

$$c_i = b_{01} + \beta_{11}x_{i1} + \ldots + \beta_{1j}x_{ij} + \ldots + \beta_{1k}x_{ik}$$

where per capita household income, R_i, is transformed into a vector in which each component represents the amount of the household's income in a particular bracket. That is,

$$x_{ij} \begin{cases} = B_j - B_{j-1} & \text{if } R_i \geq B_j \\ = R_i - B_{j-1} & \text{if } B_j > R_i \geq B_{j-1} \\ = 0 & \text{if } B_{j-1} \geq R_i \end{cases}$$

for $j = 1, \ldots, k$.

The coefficients β_{1j} represent the slope of the spline over each income bracket: the marginal propensity to consume is not only product-specific but is also different over different income levels. For necessities, the

[4] See the appendix in Bardazzi and Barnabani (1996).

marginal propensity to consume will be relatively low at the upper end of the curve and large at the lower end; while for luxuries, the slope will be higher for the upper income levels[5].

1.2. *Modelling Demographic Effects*

The non-age demographic characteristics of the household are included in the estimation of the per capita household consumption with zero/one dummy variables to indicate inclusion of the household in different demographic groups. The characteristics of the household considered here are the region of residence, the family size, the age of householder and his/her education and occupation, and the number of workers within the family besides the householder[6].

The effect of these variables is to shift the Engel curve up or down changing the intercept of the PLEC. In other words, we assume no interactions between the demographic variables so that their total effect on household consumption is additive. Therefore the effect of being a one-earner family on the consumption of alcohol is the same for households living either in the North or in the South of the country. This assumption is necessary for the transition from cross-section to the time series. To make this transition possible in case of full interaction, we would need historical data for all the demographic groups – for instance a time series of households which reside in the south, with a householder aged more than 55, retired, etc. – and these data are not available from our official statistics.

To avoid the dummy variables trap we drop one of the categories for each demographic variable. The **reference household** thus specified is a two-earner family composed of three or four members, residing in Central Italy, with a non-college educated householder aged between 35 and 55, and working in a non- professional white-collar job.

[5] Even though throughout the explanation of PLEC we have referred to household income, in this study per capita household total expenditures are used to measure R. Since the purpose of this cross-section study is to obtain Engel curves that will be used in the time-series analysis to measure the long run consumption behaviour, total expenditures are probably a better proxy of permanent income than current income, which may include all sorts of transitory components. In the light of the considerations on zero expenditures, an additional issue arises concerning the use of total expenditure as a measure of income. In fact "once it is recognized that expenditure on any particular commodity is liable to be contaminated by infrequent purchase, the total of such expenditures becomes questionable as a conditioning variable in individual equations" (Keen, 1986, p. 285).

[6] The full list of income and demographic variables used in this study is presented in Figure 1.

1.3. *Family Size*

The specification of the effect of per capita income and demographic characteristics leads to the **per capita consumption** on each good within the family. In order to obtain the **household consumption** of each commodity, the family size has to be provided.

The approach adopted here is based on the **adult equivalency weights** originally designed by Prais and Houthakker (1955)[7]. The basic idea is that individual members of the household contribute differently to the household consumption of specific commodities according to their age[8]. If g age groups are distinguished, then a set of weights w_j can be estimated to express the importance of the household members in different age groups in contributing to the consumption of a specific item. Then the weighted size of the household i for each item is given by

$$size_i = \sum_{j=1}^{g} n_{ij} w_j = n_i w$$

where n_{ij} is the number of persons in age group j in household i. The product of per capita consumption and weighted family size will give the household consumption for each product. To perform this last step a reference age group has been chosen to avoid the problem of under-identification of the parameters arising in the final equation. The weight for adults aged 31-40 has been set equal to one for each item. Therefore, the remaining weights indicate how much an individual in each age group counts relative to an adult in the reference group[9].

In Figure 1 a complete list of explanatory variables for the Italian model is shown.

[7] In fact this procedure derives directly from Engel's work. Sydenstricker and King (1921) introduced the model that was later rediscovered and popularized by Prais and Houthakker (1955).

[8] The Prais-Houthakker methodology is applied to investigate the age structure of the household but it could be applied to analyze the differential effect on consumption patterns with respect to any significative characteristics of the family members: occupation, sex, etc.

[9] This weighing scheme could be extended to create a more accurate measure of *per unit* rather than *per capita* household income (Prais and Houthakker, 1955; Singh and Nagar, 1973). This approach has not be adopted in this study, therefore per capita household income has been obtained dividing household total expenditures by the unweighted sum of family members.

Income Brackets for the Piecewise Linear Spline Function
(millions of lire)

k_1 = 0 - 6.069
k_2 = 6.069 - 7.868
k_3 = 7.868 - 9.729
k_4 = 9.729 - 12.035
k_5 = 12.035 - 15.246
k_6 = 15.246 - 21.156
k_7 = 21.156 - INFINITY

Non-age Demographic Variables

Region *(Central = base)*:
Northwest
Northeast
South and Islands
Family size *(3-4 = base)*:
single
2 components
5 or more components
Age of Householder *(35-55 = base)*:
less than 35
more than 55
Education *(high school = base)*:
lower than high school
bachelor degree or higher
Occupation *(Other non-production workers = base)*:
unemployed, retired, unoccupied
professionals
production workers
Workers other than the householder *(1 = base)*:
no workers
more than one

Age Groups

0 to 5 years old
6 to 15 years old
16 to 20 years old
21 to 30 years old
31 to 40 years old (reference age group)
41 to 50 years old
51 to 65 years old
> 65 years old

Figure 1: Explanatory Variables

2. *From the Cross-section to the Time-series Model*

Because the cross-section work must contribute to the time-series analysis, an explanation about the linkage between these two steps is needed. Moreover, the way we intend to introduce the micro-results into the time-series consumption model has generated an aggregation problem that we have solved with a procedure explained in Section 8. At this point, an introduction to the linkage is presented[10].

A "prediction" of consumption over all households for each good in year t, c_t^*, may be constructed by using the estimated $\hat{\beta}$ and $\hat{\delta}$ parameters of equation (1) applied to historical data on income and demographic trends: $c_t^* = X_t \hat{\beta} + D_t \hat{\delta}$, where X_t is a (1xk) vector of average per capita total expenditure and D_t is a (1xm) vector of population proportions of each demographic group. The variable c_t^* captures the effects of demographic and income variables over time. The Piecewise Linear Engel Curve allows us to consider the effects of a change in the distribution of income[11], therefore a forecast of income distribution is needed. Hence to forecast c_t^* for each good, the distribution of total expenditure has to be constructed through an income distribution model; and the demographic composition of the population may be forecasted by a demographic model or exogenously specified[12].

3. *The Data: the Problem of Zero Expenditures Arises*

The data used for the cross-sectional analysis is obtained from the "Household Expenditure Survey", 1993, conducted by the Italian Institute of Statistics (ISTAT). The registered expenditures have different periodicity: a ten day period for food, gasoline and sundries; a quarter for durables, house maintenance and energy expenses; a month for services and other goods (clothing, shoes, furniture, etc). All these expenditures are converted to a per-month rate.

The survey consists of two separate components. One is the diary

[10] Almon (1996) has designed the time-series system of demands which makes use of this cross- section work.

[11] As mentioned above, when we discuss income we mean total expenditure.

[12] A commodity specific weighted population is constructed through the adult equivalency weights, w, so that a more relevant population size for each good is provided. A weighted population time series for a commodity, WP_t, is defined by: $WP_t = P_t w$ where P_t is a (1xg) vector representing the individuals divided in age groups in year t. The availability of a demographic model will provide forecasts of changes in the age structure of the population. The WP_t variable will be used to construct the time-series per capita consumption.

survey in which each household in the sample is requested to register current expenditures for ten days. The second is the interview survey when the household is asked about the monthly expenditures and the purchase of durables over the last quarter.

The sample consists of 34273 households. A list of the expenditure categories is shown in Table 1[13].

[13] The original detailed data are aggregated by ISTAT into 76 categories, 12 of which are subtotals that include not only the total of the group of sectors to which they refer, but also some items of the group not elsewhere specified. We have substituted these subtotals with a balance sector (B. S. in Table 1) that include only the items not elsewhere specified in the group. It is a major weakness of our data that some important goods are mixed with

Table 1

Cross-Section Consumption Items: percentage of zero observations

Consumption categories	*Zeros %*
01 BREAD	1.5
02 PASTA	12.5
03 B.S. (NOT LISTED IN 01-02): CEREALS	10.7
04 BEEF	11.5
05 POULTRY	29.9
06 OTHER MEAT (CHARCUTERIE EXCLUDED)	49.9
07 B.S.(NOT LISTED IN 04-06): PRESERVED MEATS	16.3
08 FISH	21.8
09 OLIVE OIL AND OTHER COOKING OILS	27.2
10 B.S. (NOT LISTED IN 09): OTHER FATS	45.3
11 MILK AND DAIRY PRODUCTS	6.7
12 CHEESE	8.6
13 EGGS	23.1
14 FRESH AND DRIED FRUIT	3.9
15 VEGETABLES	4.7
16 SUGAR	3.2
17 COFFEE AND TEA	2.5
18 SWEETS AND SPICES	25.6
19 MINERAL WATER	39.2
20 WINE	48.8
21 B.S. (NOT LISTED IN 14-20): OTHER FOOD	52.1
22 TOBACCO PRODUCTS	60.0
23 CLOTHING	53.1
24 SHOES	55.0
25 B.S. (NOT LISTED IN 23-24): DRESS ACCESSORIES	30.8
26 RENT	75.8
27 IMPUTED RENT FOR OWNER-OCCUPIED HOUSE	24.3
28 HOUSE MAINTENANCE AND REPAIRS	92.3
29 B.S. (NOT LISTED IN 26-28): OTHER REPAIRS	34.9
30 ELECTRICITY	1.4

(continued)

(continued)

Cross-Section Consumption Items: percentage of zero observations

Consumption categories	*Zeros %*
31 NATURAL GAS	25.5
32 OTHER FUELS	96.5
33 HEATING	89.2
34 B.S. (NOT LISTED IN 30-33): ENERGY	96.2
35 FURNITURE	97.7
36 HOUSEHOLD LINEN	77.6
37 CHINA, GLASSWARE AND TABLEWARE	81.2
38 STOVES, HEATERS	99.0
39 FRIDGE	99.2
40 WASHING-MACHINES	99.1
41 DISHWASHERS	99.7
42 WASHING POWDERS	22.3
43 DOMESTIC SERVICE	97.7
44 LAUNDRIES	66.7
45 B.S.(NOT LISTED IN 35-44): DURABLES	42.9
46 PHYSICIANS	86.2
47 PRESCRIPTION DRUG AND SUNDRIES	47.8
48 B.S. (NOT LISTED IN 46-47): HEALTH CARE	84.2
49 SCOOTERS AND MOTORBIKES	99.5
50 BICYCLES	98.9
51 CAR INSURANCE	87.8
52 GASOLINE	25.1
53 PUBLIC TRANSPORTATION	73.4
54 TELEPHONE	10.6
55 PUBLIC TELEPHONE	79.2
56 B.S. (NOT LISTED IN 49-55): TRANSPORTS	53.2
57 NEWSPAPERS AND MAGAZINES	28.0
58 BOOKS	84.6
59 EDUCATION	92.7
60 RADIO, TV, RECORDERS, HIFI	82.7
61 PHOTOGRAPHIC EQUIPMENT	85.4
62 SPORT EQUIPMENT	97.6
63 FLOWERS AND PLANTS	71.2
64 TOYS	84.9
65 RECREATIONAL SUPPLIES, EQUIPMENT AND SERVICES	73.6
66 RADIO AND TV RENTAL	93.7
67 B.S. (NOT LISTED IN 57-66): OTHER RECREATIONAL EXPENSES	76.2
68 CASH CONTRIBUTIONS	66.2
69 PERSONAL CARE	26.3
70 HAIRDRESSING	54.5
71 LEATHER ARTICLES	95.9
72 SILVERWARE AND JEWELRY	95.0
73 STATIONERY	80.1
74 HOTELS AND MOTELS	34.0
75 B.S. (NOT LISTED IN 68-74) OTHER GOODS AND SERVICES	60.9
76 FOOD ON PREMISE	34.4

Note: B. S. is for Balance Sector.

Inspection of the survey data has immediately revealed the presence of many zero purchases for most of the consumption items considered in the classification. As is shown by the percentages presented in the second column of Table 1, the problem of zero expenditures is severe in more than half of the expenditure categories. In fact 42 out of 76 consumption categories have zero expenditures for over half of the households.

This characteristic is not peculiar to the Italian data: in general all household surveys show large fractions of households reporting zero purchases for some commodities [14]. Therefore, this problem is well known to researchers in the field, and a large literature has been devoted to the formulation of so-called limited dependent variable models. However, these econometric techniques developed by statisticians are able to give a convincing theoretical explanation only for some commodities and are therefore usually employed for studying the consumption of only a few specific goods. The aim of this paper is to interpret why zero expenditures arise, to examine and estimate some econometric models.

4. *Some Interpretations of Zero Expenditures*

Why then do zero expenditures appear?

First of all, we must stress that what we observe with budget surveys is not **consumption** but **expenditure**. Although consumption behaviour is the object of applied cross-section demand analysis, what is currently observed is total purchases over the survey period. In theory, the variable consumption "is to be interpreted as the average rate of consumption that we would arrive at by observing the household over a very long period during which all external conditions (prices, income, family composition, etc.) remain unchanged." (Pudney, 1990, p. 268). The problem in cross-section analysis is that we cannot observe households for long periods under stable conditions, so **consumption, the variable we are trying to explain, is not observable. What we observe instead is total expenditure on the good over some short observation period.**

Following this line of reasoning, we might argue that zero expenditures may arise when the underlying consumption is either zero or positive. Moreover, we distinguish two different kinds of unobservable consumption: affordable and desired. *Affordable consumption* can be viewed as the potential level of consumption that a household can afford given its income. Beside that, a household may desire or not to be a consumer of an item although this is affordable: we refer to the result of this choice as *desired consumption*. Therefore, for each item, the observed expenditure of the household is denoted $\mathbf{y}_i$, the unobserved affordable consumption is $\mathbf{c}_i$ while the desired consumption is $\mathbf{d}_i$. In the following, when we refer to the consumption variable without any additional specification we assume that $\mathbf{c}_i = \mathbf{d}_i$.

The relationship between expenditure and consumption is one of the keys to explaining zero expenditures. In fact, the models we analyze are based on different hypotheses on how to link these two variables in order to estimate consumption. We will realize that the distinction between expenditure and consumption matters when we come to analyze the most common sources of this problem: infrequency of purchase, economic reasons, conscientious abstention and misreporting.

Case 1: Infrequency of Purchase

For some goods consumption must be positive: everyone wears clothes and eats food. Therefore we know that **consumption underlying zero expenditures is positive**, and the expenditure is merely a poor indicator of consumption. In fact, survey duration is the reason for this problem: almost every good is durable relative to an interview period of ten days or a month, and a quarter is not the most appropriate period to register the purchase of a car.

Consumption is not directly observed: "a positive expenditure will represent *a purchase of stock whose services will be consumed over future periods* typically longer than the period of observation" (Blundell and Meghir, 1987, p. 182). A family reporting a zero expenditure of cars or shoes does not mean that the household is not a consumer of these commodities but simply that it is consuming the services of a car and some pairs of shoes bought outside the survey period.

The consumption underlying expenditure can be studied by a three-step procedure. Firstly, over all households in the sample, we determine the probability of purchase as a function of one set of variables. Secondly, over the households which purchased the item, we estimate the amount spent as a function of a second, perhaps different, set of variables. Finally, for all households, we calculate the expected consumption of the household as the product of the expected expenditure if any is purchased multiplied by the probability of purchase. The probabilities to buy may be

others in these balance sectors. For example, purchases of used and new cars are included in balance sector 56 along with train and plane tickets, vans, boats, etc. This classification has been adopted by ISTAT to prevent households in the sample from being identified by the data, as provided by the Italian law (D.L. 322, September 1989). It is hard to conceive how the purchase of a car could identify a household in the sample.

[14] A recent study on Italian data focused on the treatment of zero expenditures is Grassini and Viviani (1995).

viewed as a set of weights allowing us to go from the "stock accumulation" (expenditures) to the "consumption of services" (consumption) of the good. For those families with a positive observation, the underlying consumption will be lower than the expenditure in the survey period because part of the good bought will be consumed in the future; on the other hand, households with a zero expenditure will show a positive consumption of the commodity since we assume they bought outside the survey period.

Case 2: Economic Decision

Some goods are not consumed by every household given current income and prices. This phenomenon is quite familiar and was first analyzed by Tobin (1958). In his seminal article, he analyzed the case of those categories of goods, luxuries, for which zero expenditures at low income levels are often recorded. He then applied his model to the case of durable goods where most of the problem falls in Case 1.

In this approach, expenditure is assumed to be a correct indicator of underlying consumption: a zero observation corresponds to a zero consumption. Non-consumption is always the result of an economic decision: under changing conditions – lower prices, higher income – a non-consumer could become a consumer of luxuries. Here consumption is assumed to be the result of a single choice: the decision to buy and the amount spent are determined by the same variables.

Case 3: Conscientious Abstention

Zero expenditures may also arise from variation in preferences across the sample: some households may simply not consume some commodities at any prices or income level. The population can be divided into different groups: abstainers and non-abstainers. The former is a group of non-consumers as a result of a conscientious decision. Obvious examples are tobacco, alcohol, and meat.

As in the case of infrequent purchase, consumption may be figured as the result of a two-stage process. The decision to be a consumer rather than an abstainer is the first hurdle to overcome to get a positive expenditure: this is to say that the unobservable desired consumption, d_i, must be positive. Then non-abstainers should decide how much to spend for the commodity: this decision concerns the level of affordable consumption, c_i. When this second hurdle is passed, consumption of the good is observed.

Case 4: Misreporting

This is the case when a positive expenditure occurred during the survey period has not being reported by the consumer. This problem is a serious concern for budget surveys of developing countries. Zeros of this kind are not directly analyzed in this paper.

The classification of the first three cases of zero expenditures is summarized in Figure 2. Here the explanations described above are correlated to the relationship between zero expenditure and consumption. Case 1 is the only one where the underlying consumption is supposed to be positive while the other interpretations of zero expenditures allow the consumption to be zero. The shadowed boxes in the figure indicate meaningless cases: for instance the assumption of a positive consumption where the zero observation is due to a conscientious abstention from purchasing.

Zero Expenditures

Reasons	*No Consumption*	*Positive Consumption*
Economic Decision	**Case 2**	
Conscientious Abstention	**Case 3**	
Infrequency of Purchase		**Case 1**

Figure 2: Some Interpretations of Zero Expenditures

5. *From Interpretations to Some Models*

The scheme shown above is a simplified representation of a more complicated problem. There are goods where zeros may arise for any of these reasons or as a combination of two different sources. A zero purchase of a durable good could be due either to an economic decision or to an infrequent purchase. Alcohol is a storable good that could be bought very seldom. Moreover, there could be a mixture of conscientious abstention and economic non-consumption. Therefore, a unique model appropriate for all commodities and representing all zeros interpretations is hard to find.

For these two reasons – the complexity of models proposed and the necessity to treat different goods with different models – the analysis developed in this paper is for each good separately. We do not attempt a simultaneous estimation for all goods.

The models proposed are independent from the analytical form of cross-section function specified in the previous sections: consumption could be measured in quantity or in budget shares, the underlying Engel curve could be different as the specification of demographic variables. The problem of zero expenditures is a general one, and the consumption model adopted is not particularly important in how it is handled.

Among the models for limited dependent variables[15] proposed in the literature, we have decided to adopt three specifications which represent the interpretations of zero expenditures described in the previous section.

The **Tobit Model** is supposed to interpret consumption behaviour when zero expenditures are mainly ascribed to economic reasons. This model was designed by Tobin in his article of 1958 and widely used since then to model cross-section consumption even of commodities where infrequent purchase is expected to be a very likely explanation of zero observations, such as durable goods. In fact, as has been stressed by Maddala (1990): "... the Tobit model is inappropriate for almost all the applications in which it has been used (including that by Tobin)" (p. 54). And furthermore, "... it is tempting to use the simple Tobit model every time that one has a bunch of zero (or other limit) observations on y. However this is inappropriate" (p. 55).

A wide class of sequential choice models has been designed following the **Double-Hurdle model** originally developed by Cragg (1971). The two-step statistical structure of this model can be adopted to interpret the case when households could be consumers or abstainers[16]. For this purpose the distinction between affordable, c_i, and desired consumption, d_i, is fundamental to explain the different behaviour of households.

A **non-linear probability model** is used to treat zero expenditures due to infrequent purchasing. This is also a sequential choice model but it is based on the assumption that $c_i = d_i$ and the unobservable consumption is always positive. However, these two models could be used to explain both

15 Variables having a non-negligible probability exactly equaling zero.

16 Cragg proposed his model to interpret infrequent purchasing but later other authors applied his model to interpret the case of conscientious abstention (see Pudney, 1990).

sources of zeros, conscientious abstention and infrequent purchase, if appropriate explanatory variables are used to perform the first step, that is the decision to buy.

Although a precise and unique connection between models and interpretations of zeros cannot be made, we here propose a scheme that is meant to guide us in the empirical applications. In Figure 3, models are superimposed on the interpretations of zero expenditures suggested in Figure 2: the associations we propose are not strict but could be useful in understanding the purpose of each model in the formal analysis that will be carried out in the following sections.

Zero Expenditures

Reasons	*No Consumption*	*Positive Consumption*
Economic Decision	**Tobit Model**	
Conscientious Abstention	**Cragg Double-Hurdle Model**	
Infrequency of Purchase		**Non-Linear Probability Model**

Figure 3: Some Models of Zero Expenditures

The models have been applied to the consumption categories listed in Table 1 and a complete set of results is available. However, for the purpose of this study, we have selected the results of the consumption categories listed in Figure 4. This subset of consumption items should be representative of all the zero expenditure interpretations described in Section 4. For instance, zero observations for clothing and shoes are presumably due to infrequent purchases while consumption of tobacco is a matter of individual preferences. Some of these categories are an aggregation of several consumption items such as durables because the huge percentage of zero observations does not allow an analysis on every single item. Food products have been aggregated to show that with such a broad category the percentage of zero expenditures is negligible and therefore the models should have similar results.

The linkage between models and interpretations of zero expenditures suggested in Figure 3 can be used to explain the association between specific items and models proposed in Figure 4. However, to investigate the performance of the models, each of them has been applied to all the consumption items listed in Figure 4.

Goods	Model
Pasta *Clothing* *Shoes* *Food* (*items*: 1-19)	*Non Linear Probability Model*
Durable House Furnishing (*items*: 38-41) *Durables for Transportation* (*items*: 49-51 *and* 56) *Recreational Durables* (*items*: 60-62)	*Non Linear Probability Model*
Silverware and Jewelry *Recreational Durables* (*items*: 60-62)	*Tobit Model*
Beef *Wine* *Tobacco*	*Cragg Model*

Figure 4: A Proposal for a Choice of Models for Specific Goods

6. *The Models*

In the previous sections we pointed out that cross-section consumption is a variable that is not observable. We explain the underlying economic process by the following non-linear model:

$$\begin{aligned} c_i &= (x_i\,\beta + d_i\,\delta)\,n_i\,w + u_i \\ &= g\,(t_i, \gamma) + u_i \end{aligned}$$

where $\gamma' = [\beta \mid \delta \mid w]$ is the $1x\,(k+m+g)$ vector of parameters;

$t_i = [x_i \mid d_i \mid n_i]$ is the set of independent variables of the cross-section function as explained in Section 1 (see formula (1)); $u_i \sim N(0, \sigma^2)$ and i indexes the household[17]. Actually, what we observe is the household expenditure, y_i, and we face the problem of modelling the underlying relationship between expenditure and consumption.

As already stated in previous sections of this paper, we take into consideration the following models:

— The Tobit model
— The Cragg or Double-Hurdle model
— The Non-Linear Probability model

The **Tobit model** (also known as censored regression model), may be seen as a way to analyze the relationship between household expenditure and affordable consumption that may be expressed in the following way:

$$y_i = \max\,\{c_i, 0\}$$

In this formulation, zeros arise only if the household genuinely does not consume, that is when $c_i < 0$ or $c_i = 0$. In both cases, we observe an expenditure equal to zero that presumably may be ascribed to an economic decision.

We recall that associated with this model we could consider three conditional mean functions. For the household affordable consumption (our latent variable), $E(c_i)$ is $g(t_i, \gamma)$; for the expenditure computed only on positive values, $E(y_i \mid y_i > 0)$ is obtained as the expected value of a normal density function truncated at zero from below. Finally, the expected expenditure on all observations, $E(y_i)$, is:

$$E(y_i) = P(y_i > 0)\;E(y_i \mid y_i > 0)$$

given by the product of two quantities: the probability of purchasing computed with reference to the normal distribution function, and the amount spent if the household buys. In this way the Tobit model may be seen as a one-step procedure where the decision to buy and the amount to spend is simultaneous. Because of this simultaneity, the variables involved in the computation of the household expenditure are the same.

"It is an unresolved question which of these functions should be used for computing predicted values from this model" (Greene, 1993, p. 694).

[17] We stress a characteristic of the consumption function (1). It is linear in "per adult equivalent" term that is when the household consumption is divided by $n_i w$, while it is intrinsically non linear in the household consumption.

However, our objective is to obtain an estimate of the affordable consumption so the expected value we are interested in is $E(c_i)$.

Unlike the Tobit, the **Double-Hurdle model** may be seen as a two-stage decision approach where the household first decides whether or not to buy on the basis of some variables here called x_i^*, and then, if the decision was to buy, decides how much to spend[18].

Let I_i be a binary variable which takes on value of one or zero according to whether the household *i desires* to buy or not to buy. We suppose that a set of factors gathered in a vector x_i^* explains this unknown desirable consumption, d_i, so that

$$p_i = P(d_i > 0) = P(I_i = 1) = \Phi(x_i^*) = \Phi_i$$

where Φ_i is a cumulative distribution function.

Under the assumption that the process generating I_i is independent of y_i conditional on x_i, the Tobit specification is now modified to:

$$y_i = I_i \max \{c_i, 0\}$$

From this rule we may observe that a zero expenditure occurs either if the household desires to consume ($d_i > 0$, that is, $I_i = 1$) but its affordable consumption is zero ($c_i = 0$) because of economic reasons; or if the household is a conscientious abstainer ($d_i \leq 0$, $I_i = 0$) although it could afford the consumption, ($c_i > 0$).

Finally, the **Non-Linear Probability model** may be seen as a three- step approach where in the first step we compute the probability to buy, then we fit a non-linear regression model for the subsample for which the consumption is positive, and finally we compute the expected purchases, or consumption, for all households[19].

In this model, we assume that the affordable consumption c_i is always positive, that is $P(c_i > 0) = 1$, and that

$$c_i = y_i\, P(y_i > 0) = y_i\, p_i + u_i \tag{3}$$

[18] We do not fully exploit the potential capacity of this two-stage decision approach because we assume that the variables explaining the decision to buy and those determining the amount spent are the same. This assumption is to be relaxed in a further development of the work.

[19] This approach may be traced back to the beginning of the sixties. Due to the linearity hypothesis assumed both for the probability and the regression function, it was called by Goldberger (1964) "twin linear probability function".

where u_i is the usual random component. This model implies that the observed expenditure, when positive, will exceed the level of services consumed: households are stocking up during the survey period. The parameter p_i may be seen as a specific depreciation rate and should be modeled on the basis of specific variables[20]. When the observed expenditure is zero, the household is not making any purchase but it is *consuming* from stocks.

Because of our hypothesis on consumption, we have

$$E(y_i) = P(c_i > 0)\, E(c_i \mid c_i > 0) = E(c_i \mid c_i > 0)$$

This relationship suggests how to estimate consistently the expected expenditures when a zero is observed. The estimate of the parameters can be obtained using only positive observations. Under this hypothesis, from (3) the expected value of consumption is given by:

$$E(c_i) = E(p_i\, y_i) = p_i\, E(c_i \mid c_i > 0)$$

for all households.

The models for zero expenditures usually presented in the literature, such as the Tobit and Cragg models, are strictly dependent on the normality assumption of the random component. It has been shown that if this hypothesis is not verified the usual estimator is inconsistent. Research is ongoing on the use of alternative distributions. The setting of the Non-Linear Probability Model may be seen as a way to overcome this problem as shown in the following paragraph.

Besides, if the assumption of a simultaneous choice for the decision to buy and the amount spent underlying the Tobit model has been criticized, the alternative consumption behaviour of the sequential models may not be accepted for some specific items. One can imagine that the decision on whether or not to purchase a good is not independent of the decision to spend on the good, having decided to buy it. These considerations support our argument that, in a highly disaggregated analysis, there is not a unique model to explain zero expenditures for all items.

7. *The Estimation Schemes*

To estimate the parameters of the Tobit model we applied the maximum likelihood procedure. With regard to this approach we must

[20] As a first step of our work, we assumed the same set of variables for explaining both p_i and $E(y_i)$.

observe that the likelihood of the sample has a component for the observations that are positive and one for those that are zero.

For the observation $y_i = 0$ all we know is that $g(t_i, \gamma) + u_i \leq 0$ so

$$P(y_i = 0) = P\left(\frac{u_i}{\sigma} \leq -\frac{g(t_i, \gamma)}{\sigma}\right)$$
$$= 1 - F_i$$

where F_i is to be interpreted as the cumulative distribution function of the standard normal random variable. If $y_i > 0$ the usual component of the likelihood function is present, consequently, as known, the log-likelihood function is

$$\log L(\gamma) = \Sigma_0 \log(1 - F_i) + \Sigma_1 \log f_i \tag{4}$$

where f_i is the probability density function of a standard normal random variable; Σ_0 denotes the sum over the zero observations and Σ_1 the sum over positive observations.

The maximization of (4) causes many problems:

a) this function is "doubly non-linear" both in the specification and in the argument $- g(t_i, \gamma)$ is a non-linear function with respect to the parameters – so we have no more guarantee that whatever algorithm we use it will converge to the maximum of (4) as happens in the linear case;

b) the parameter w_5 must be equal to one, therefore we need to maximize equation (4) under this assumption;

c) because of the numerosity of our observations, it is not possible to use standard procedures of maximization available in the usual statistical packages which use numerical algorithms to compute the first and second derivatives[21].

To compute the maximum of the likelihood function we decided to work with the Levenberg-Marquardt algorithm because of the possibility, acting on the so called lambda parameter, to drive the algorithm towards an "optimum". The problem of the constraint has been solved "by substitution" imposing the parameter w_5 of the function $g(t_i, \gamma)$ to be equal to one. Finally, the problem of the derivatives was tackled by directly giving the analytic form of the gradient of the likelihood function (4) while the Hessian matrix was avoided, instead computing its approximation matrix suggested by Berndt, Hall, Hall and Hausmann (1974) (from now on BHHH matrix) that requires, as known, only the computation of the first derivatives of the log likelihood for a single observation.

21 In our experience with N=34273 observations and 31 parameters, after about three hours we were not able to obtain an estimate of the γ vector using the numerical procedure for the computation of the first and second derivatives available in the Gauss package.

As to the estimation of the Double-Hurdle model, we recall that the log-likelihood function to be maximized is given by Cragg (1971)

$$\log L(\gamma) = \Sigma_0 \log(1 - p_i F_i) + \Sigma_1 \log p_i + \Sigma_1 \log f_i$$

Also in this case we have had the same problems as those illustrated for the Tobit model because of the similarity of the two functions. Therefore, we used the Levemberg-Marquardt algorithm with the BHHH matrix approximating the Hessian matrix; we constrained the parameter w_5 to be equal to one and we gave to the program the analytic function of the first-order derivatives.

The estimation of the Non-Linear probability model is based on the following steps. First, we compute the probability of purchasing; second, we applied the Levemberg-Marquardt algorithm to non-linear least squares using only the non-zero observations; third, the expected value of consumption is computed. The estimation of the household probability of purchasing need some further explanations.

Suppose we let the binary choice of whether to consume to be represented by a dichotomous variable I_i which takes the value 1 when the family consumes and 0 when it does not consume. The methods we could use to model this binary choice are:

1) the linear probability model in which the binary choice is explained using the linear regression model:

$$I_i = (x_i \beta + d_i \delta) s_i + u_i \tag{5}$$

where $s_i = \Sigma_j n_{ij}$ is the scalar size of household and u_i is the usual random component. The conditional expectation of (5) is interpreted as the probability that the family purchases;

2) the probit analysis model that may be seen as a two-step procedure where firstly we assume the following regression relationship

$$c_i = (x_i \beta + d_i \delta) s_i + u_i$$

with c_i unobservable variable, then an observable dummy variable I_i is defined by

$$I_i = \begin{cases} 1 & if\ c_i > 0 \\ 0 & if\ c_i \leq 0 \end{cases}$$

In this formulation

$$P(I_i = 1) = P(u_i > -x_i s_i \beta - d_i s_i \delta) = \Phi_i (x_i s_i \beta + d_i s_i \delta) \quad (6)$$

where Φ_i is a cumulative distribution function of u_i.

Faced with these two possibilities that is the linear probability model and the probit model, we first tried to apply the linear probability model because of its simplicity. Unfortunately, in several items of consumption we found a value of the probability to purchase outside the limits (0,1)[22] and, although in large samples this problem could be overcome, we left this approach in favour of the widely used probit analysis.

The use of the probit analysis implies the choice of the cumulative distribution function of the random component. As a first step, we have tried only two hypotheses, the normal and the logistic function and the choice was made on the basis of what we call "ex-post test-procedure":

1) under the hypothesis of normal (or logistic) distribution we have estimated the parameter of equation (6). Let $\hat{\beta}$ and $\hat{\delta}$ be the estimates respectively of β and δ;
2) we have ranked the households by their $x_i s_i \hat{\beta} + d_i s_i \hat{\delta}$ score and we have grouped them into ventiles. Subsequently we have calculated an average of this probability, $\hat{p}_m$, in each group, m = 1, ..., 20;
3) we have computed the average (in this case simply a percentage) of households who bought ($I_i = 1$) in each group, $\bar{p}_m$, that may be seen as an empirical probability;
4) we compared the two averages obtained (both with normal and logistic distributions) using different indicators such as the χ^2, the mean of the absolute (square) difference. The rationale for using this procedure is to detect if the hypothesis formulated on the cumulative distribution of the random component is supported by the data acting in a simple descriptive way. We would expect to have a low value of the indicators if the approximation chosen is good. Unfortunately, in the consumption items analyzed we observed quite a low value in all the indicators used, both in the case of the normal hypothesis and in the logistic one, indicating a lack of an objective criterion of choice with the two assumptions. This fact persuaded us to select (at least as a first step of our work) the normal function (from which Normit model) to compute the probability of purchasing. However, our plan for the future is to use an empirical distribution function set up *ad hoc* for each item.

22 We found this situation in those items where there are many zeros, for example in Jewelry (about 10% of observations shows a negative probability) or where there are few zeros, for example in Food (about 25% of probability greater than one). We verified that this situation arises when we move close to the tails of the distribution function.

Due to the properties of the normal function, equation (6) is equal to:

$$P(I_i = 1) = F_i \left(x_i s_i \frac{\beta}{\sigma} + d_i s_i \frac{\delta}{\sigma} \right) \quad (7)$$

From (7) we can observe that in the normit model we are able to estimate only the ratios β/σ and δ/σ, then there is a problem of identification of the parameters, but in our case this indeterminacy is not a problem because our goal is to obtain an estimate of the probability to purchase, not the parameter values.

The normit approach is based on the maximization of the following log likelihood function:

$$\log L(\gamma) = \sum_{i=1}^{N} I_i \log P(I_i = 1) + \sum_{i=1}^{N} (1 - I_i) \log [1 - P(I_i = 1)] \quad (8)$$

where $P(I_i = 1)$ is given by (7). The method we used to maximize (8) is the well known Newton-Raphson procedure.

8. *From the Cross-Section to the Time-Series: the Analytical Linkage*

One way to take advantage of the information coming from a cross-section analysis in a time-series consumption function may be to use the expected value of per capita consumption from the cross-section as an independent variable in the time-series model. This procedure allows us to include income and demographic effects on consumption indirectly whereas other models that do not rely upon cross-section analysis must include these variables explicitly, thus facing the constraint of scarce degrees of freedom of the time-series.

The linkage between the disaggregated and the aggregate analysis is performed by summing the estimate of per capita expected value across all households. If we refer to the non-linear regression model (see Section 7), because of its linearity in "per adult equivalent" terms and the equality between expenditure and consumption, we have:

$$c^* = \sum_{i=1}^{N} \frac{c_i}{n_i \hat{w}} = \sum_{i=1}^{N} \left(\sum_{j=1}^{k} \hat{\beta}_j x_{ij} + \sum_{j=1}^{m} \hat{\delta}_j d_{ij} \right) = \sum_{j=1}^{k} \hat{\beta}_j X_j + \sum_{j=1}^{m} \hat{\delta}_j D_j$$

where X_j and D_j are respectively aggregates of income and demographic variables, $\hat{\beta}_j$ and $\hat{\delta}_j$ are the non-linear estimates of the parameters and the superscript star is for "per adult equivalent" c* may also be viewed as the consumption per person component of the cross-section equation.

The inclusion of c* as an explanatory variable in a time-series model requires the availability of this variable over time. Usually, however, it is available from the cross-section only in isolated periods, so the only thing we can do is to build an estimate of it. If we have a time series of the aggregates X_{jt} and D_{jt} (and usually this is the case) we can estimate c^*_t.

This procedure may be followed for the models that are linear in per capita terms, such as the Tobit and the Cragg model. In fact for these models, the expected value of per capita consumption is: $E(c_i / n_i w) = x_i\beta + d_i\delta$. How should we proceed if we use the non-linear probability model?

In this case the problem is that the estimate of the expected value of consumption is strictly non linear, that is

$$c^*_i = p_i\left(\sum_{j=1}^{k}\hat{\beta}_j x_{ij} + \sum_{j=1}^{m}\hat{\delta}_j d_{ij}\right)$$

The aggregation by summation over all the households is

$$c^*_i = \sum_{j=1}^{k}\hat{\beta}_j\left(\sum_{i=1}^{N} p_j x_{ij}\right) + \sum_{j=1}^{m}\hat{\delta}_j\left(\sum_{i=1}^{N} p_j d_{ij}\right) \qquad (9)$$

and the quantities between brackets are not available in time series. Therefore, the problem we face is how to do an aggregation in such a way that the resulting equation may be written in terms of X_j and D_j, the only variables usually available in time series.

We could still use the variables X_j and D_j instead of the two quantities within brackets in formula (9) but in this case the result is no longer consistent with the estimate obtained in the cross-section analysis.

A way to solve this problem could be to modify the estimate of the parameters so as to reach this compatibility. Following a hint given by C. Almon, this may be done following a two-step estimation procedure. In the first, we estimate the model; in the second, the estimate of the expected value obtained in the first step is used as A dependent variable of the non-linear regression model and then the parameters are re-estimated.

The final results shown in the tables at the end of this paper, as far as the non-linear probability model is concerned, are estimated following this procedure.

9. *The Results*

The cross-section consumption function (1) can be summarized as follows:

$$\begin{matrix}\textit{Family Consumption} \\ \textit{for each product}\end{matrix} = \left(\begin{matrix}\textit{Function of Income per capita} \\ \textit{within the ith Family}\end{matrix} + \begin{matrix}\textit{Demographic} \\ \textit{Effects}\end{matrix}\right) * \begin{matrix}\textit{Family} \\ \textit{Size}\end{matrix}$$

Consumption is the result of three components: per capita household income, demographic effects and the specific size of the household. Therefore our comments can be organized following this distinction.

The estimation results of the models are shown in Tables 2-5 and in Figures 5-26.

The tables report, for each consumption category, the coefficients of demographic variables[23]. The figures in the parentheses are the t-statistics.

Instead of showing the coefficients of income in the tables, the relationship between consumption and income is highlighted using the plots of the Piecewise Linear Engel Curves, PLEC (Figures 5-15). As to the demographic effects, because of the high number of results we will try to summarize them by underlying similarities and distinctive features between goods and models. Finally, the specific size of the household is analyzed by the bar graphs of adult equivalency weights (Figures 16-26).

[23] We refer to the consumption categories specified in Figure 4. Note that the explanatory variables are those listed in Figure 1.

PIECEWISE LINEAR ENGEL CURVES

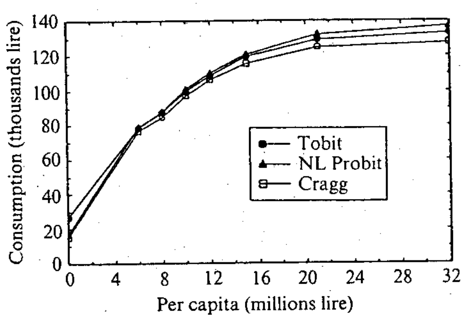


Figure 5: Pasta

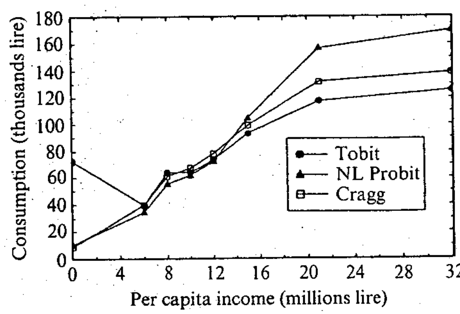


Figure 6: Wine

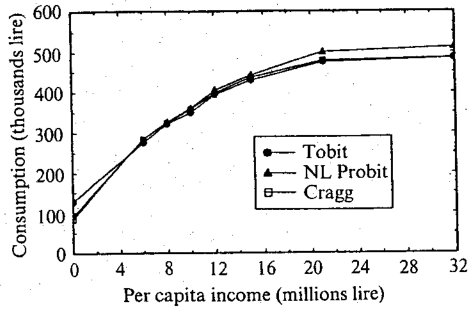


Figure 7: Beef

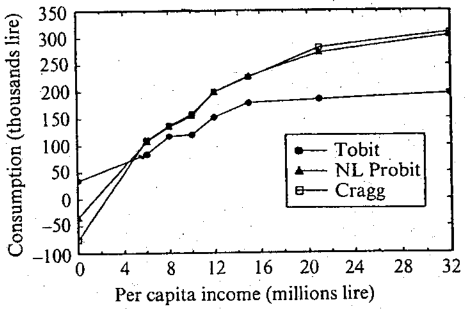


Figure 8: Tobacco

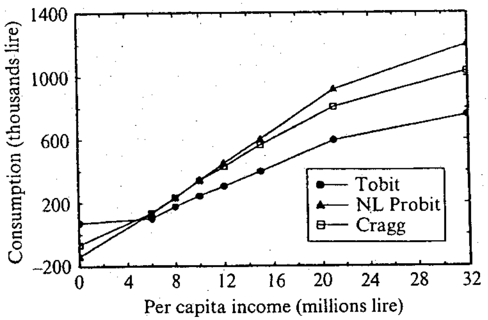


Figure 9: Clothing

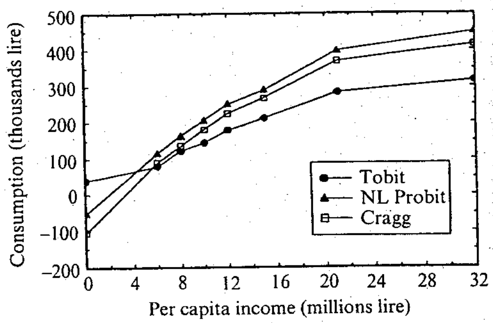


Figure 10: Shoes

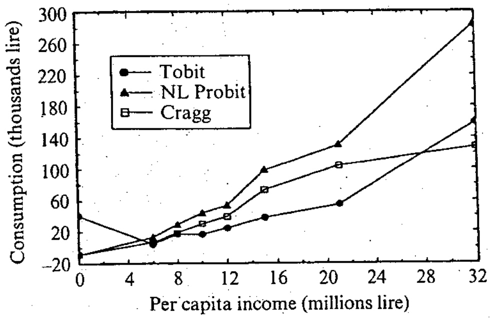


Figure 11: Silverware and Jewelry

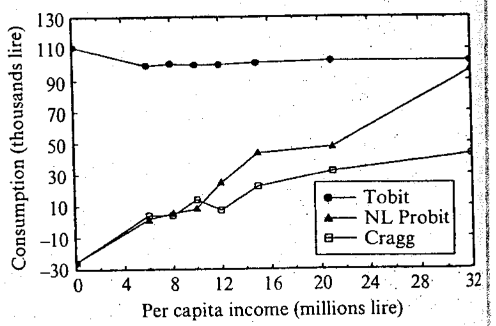


Figure 12: Durable House

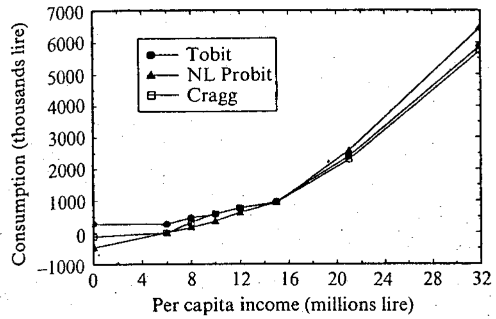


Figure 13: Durables for Transportation

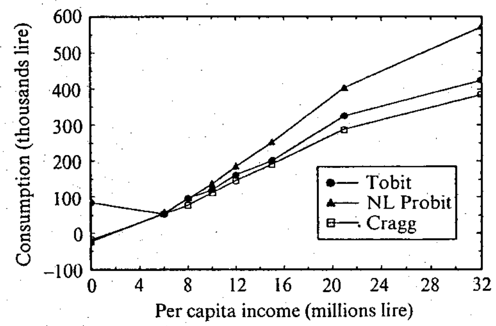


Figure 14: Recreational Durables

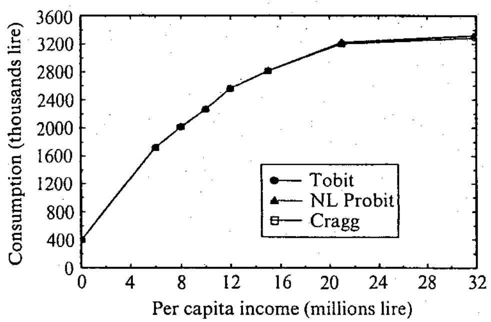


Figure 15: Food Products

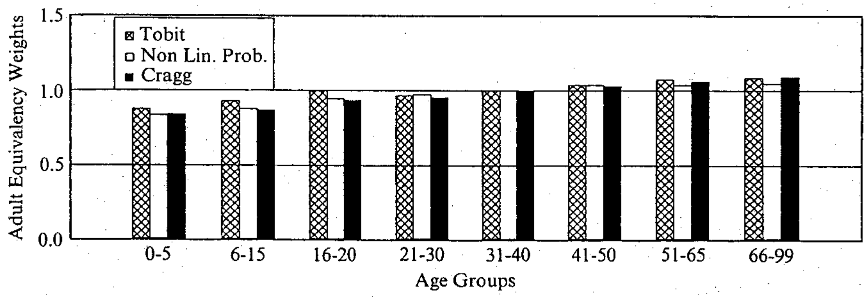


Figure 16: Pasta

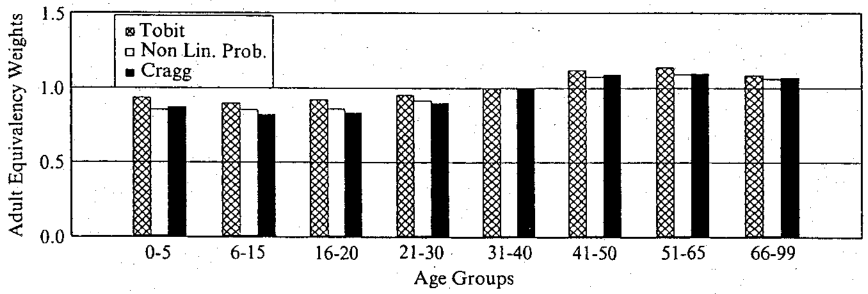


Figure 17: Beef

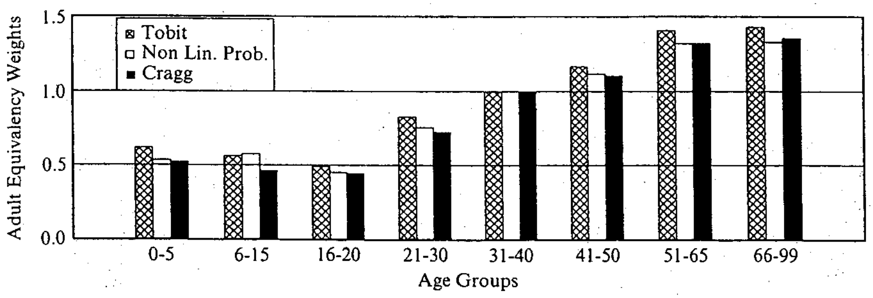


Figure 18: Wine

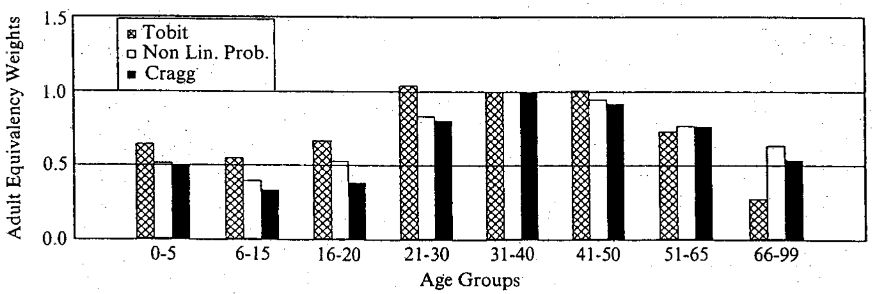


Figure 19: Tobacco

ADULT EQUIVALENCY WEIGHTS

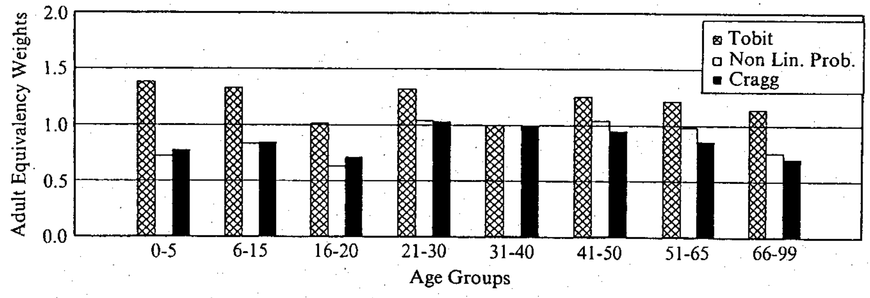


Figure 20: Clothing

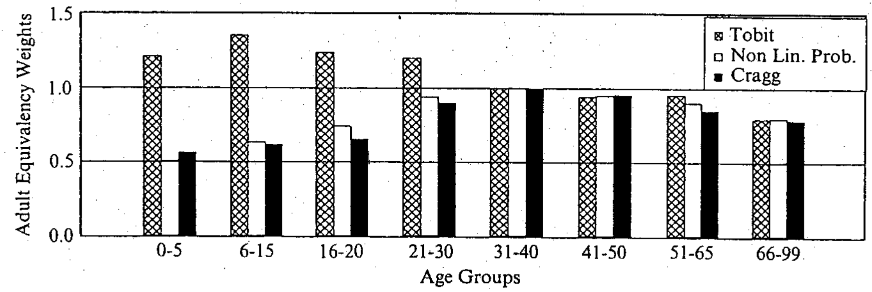


Figure 21: Shoes

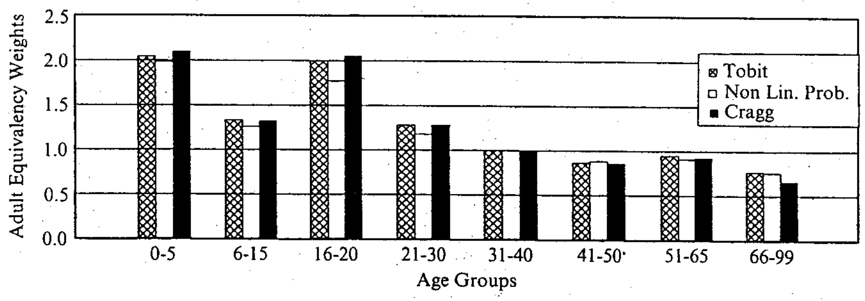
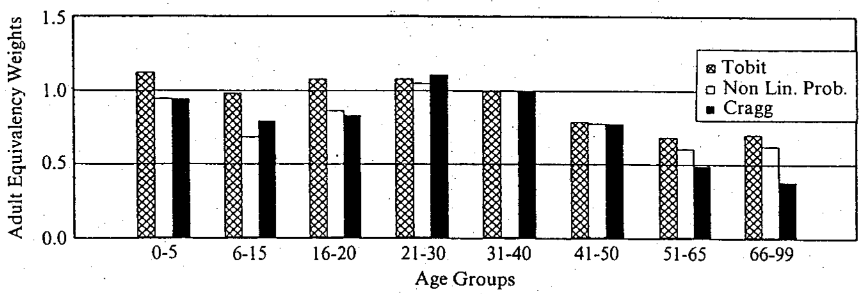


Figure 22: Durables for Transportation



ADULT EQUIVALENCY WEIGHTS

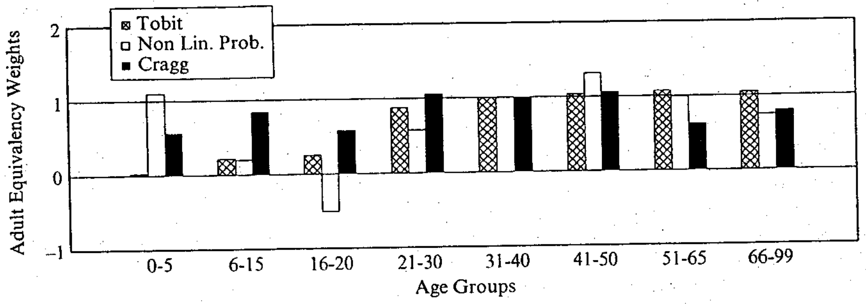


Figure 24: Silverware and Jewelry

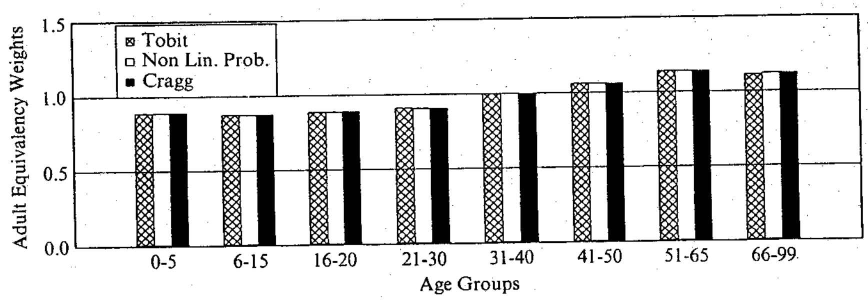
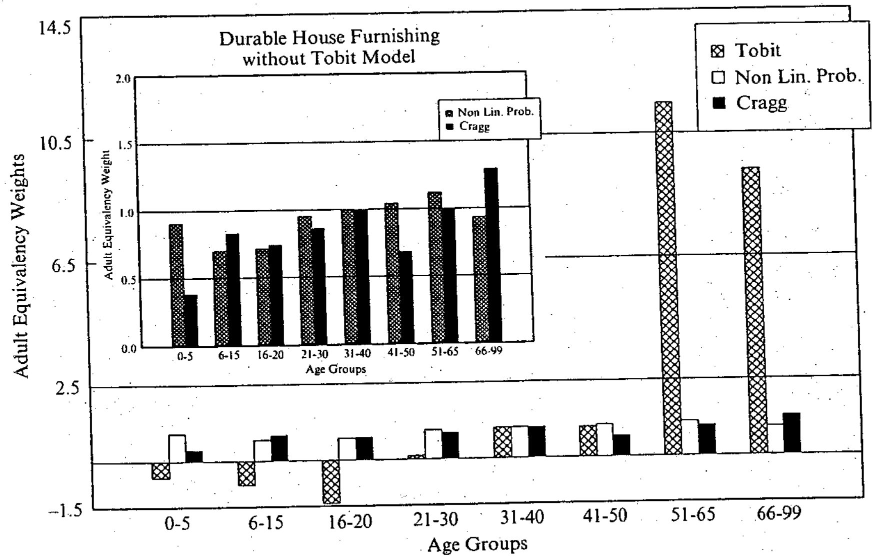


Figure 25: Food Products



9.1. *Engel Curves*

Plots of estimated Engel Curves are drawn for the reference household which is a two earner family composed of 3 or 4 members and residing in central Italy with a non-college educated householder aged between 35 and 55 working as an employee. The consumption expenditure per adult equivalent is plotted against per capita household income.

The coefficents of income are the slope of the PLEC within each bracket. The changing slope along the curve represents the variation of the marginal propensity to consume for different income levels allowing the underlining of specific patterns for certain groups of consumption goods. For instance, Pasta, Food Products, Beef and Tobacco show a definite necessity pattern of consumption. Shoes and Footwear, Clothing and Recreational Durables show a relatively constant slope of the Engel curve throughout all the income brackets. Durables for Transportation are luxuries showing a steeper slope in the upper levels of income. In the case of Silverware and Jewelry and Durable House Furnishings we would have expected a pattern typical of luxuries. Instead, the plots do not show this profile and, moreover, the results are different according to the model used. A possible explanation is the huge share of zero observations for these two consumption categories, in both cases higher than 95 percent.

9.2. *Demographic Effects*

The models do not show a uniform behavior on the sign, the magnitude of the impact on the consumption category and the significance of the t-statistics. This prevents us from a general comment about the demographic effects on consumption, therefore we can only try to trace some similarities between the models.

The results of demographic variables shown in the tables have to be considered as deviations with respect to the base category omitted from the regression to avoid the dummy variable trap. The null hypothesis is that, for each demographic variable – for instance Region – there is no difference between the coefficient of one of the specific categories – North East – and the coefficient of the base – Central Region[24].

A negative sign of the dummy variable coefficient – for example in Table 2, for Pasta, the parameter of Northeast (–23.738) – means that the consumption of this item for a household living in this part of the country

[24] The level of significance has been assumed 0.05.

Table 2

NON LINEAR PROBABILITY MODEL

DEMOGRAPHIC VARIABLES	PASTA (12.5%)*	CLOTHING (53%)*	SHOES (55%)*	DUR. HOUSE (97.5%)*
REGION (*Central = base*):				
Northwest	-6.1459 (-1.7447)	-42.2514 (-2.4447)	-24.8157 (-3.1794)	6.8146 (4.6929)
Northeast	-23.7388 (-6.2012)	-3.8069 (-0.2151)	-9.5682 (-1.1969)	12.7227 (8.0805)
South and Islands	8.8043 (2.6924)	48.5641 (3.0528)	35.6119 (4.8676)	6.0016 (4.5064)
FAMILY SIZE (*3-4 = base*):				
single	19.0369 (2.3142)	-166.3720 (-3.9688)	23.8564 (1.3199)	-23.5907 (-6.9725)
2 components	8.6925 (2.0013)	-54.5560 (-2.6624)	-15.1036 (-1.6596)	-12.0849 (-7.0122)
5 or more components	-1.8649 (-0.6252)	33.3959 (2.2816)	16.5907 (2.4426)	2.4641 (2.0211)
AGE OF HOUSEHOLDER (*35-55 = base*):				
less than 35	-0.5418 (-0.1182)	32.0097 (1.6209)	7.8750 (0.8136)	-0.4411 (-0.2771)
more than 55	1.1990 (0.2962)	-36.1165 (-2.1115)	-13.7742 (-1.6914)	2.0821 (1.5164)
EDUCATION (*high school = base*):				
lower than high school	4.8918 (1.5142)	-13.2071 (-0.8519)	-6.5867 (-0.9313)	4.7299 (3.5690)
bachelor degree or higher	-5.8046 (-1.1213)	84.0759 (3.3148)	1.6569 (0.1449)	3.2218 (1.5249)
OCCUPATION (*Employee = base*):				
unemploy, retired, unoccupied	3.3751 (0.8679)	-4.1975 (-0.2258)	-14.6810 (-1.7309)	3.8792 (2.4992)
professional class	2.7240 (0.8181)	44.2227 (2.7266)	13.2855 (1.8033)	-3.8930 (-2.8456)
working class	2.8957 (0.8515)	19.8345 (1.2047)	3.4438 (0.4587)	-0.6135 (-0.4415)
WORKERS OTHER THAN HOUSEHOLDER (*1 = base*):				
no workers	0.0691 (0.0263)	-28.7071 (-2.2801)	-13.9795 (-2.4386)	-1.2899 (-1.2240)
more than one	-2.6375 (-0.6925)	42.4229 (2.3249)	-3.0471 (-0.3710)	-3.2943 (-2.1696)

(*) Percentage of Zero Observations.

is lower compared with that of a family living in the Center. Viceversa if the coefficient is positive.

The Region is significant for at least one region category for all models and for all goods except for Durables for Transportation. Also the Family Size has a significant impact on the household consumption although there are some differences among the models.

In the case of the other demographic variables – Age of Householder, Education and Occupation of the Householder, number of Workers in the family – these differences are more evident. Following the association between goods and models as assumed in Figure 4, we limit our attention to the results of the Non-Linear Probability model for Clothing, Durable House Furnishings, Pasta and Shoes (Table 2). We observe that for the latter none of these demographic variables has a significant impact on the household consumption, while the level of expenditure for Clothing is higher if the householder has a college education and belongs to the professional class. Moreover we observe a direct relationship between the number of workers in the family and the amount spent for Clothing.

As to the Durable House Furnishings, the model shows a non significance for the Age of the Householder, while for the other variables there is at least one significant category although the coefficients present a sign not consistent with our *a priori* expectations. In fact according to our results, the consumption of House Furnishings is higher when the householder has no education and is retired or unemployed. We stress that an analysis with 97.5 per cent of zero observations is not very reliable.

For Transportation Durables, in general all the models show that these demographic variables are not relevant in explaining consumption (Table 3).

Following our choice in linking some models of zero expenditures to specific goods, we have proposed to explain the consumption of Recreational Durables either by the Non-Linear Probability Model or by the Tobit model (Table 4). This was motivated by the possibility of a double interpretation of zeros: these could be ascribed both to infrequent purchase and to economic reasons. The Education of Householder is not discriminant in the household consumption of Recreational Durables, and, as far as the other variables are concerned, the sign of coefficients compared with our expectations could indicate a better representative performance of the infrequent purchase model.

The demographic variables are not very meaningful in determining the household consumption of Silverware and Jewelry. Here, the Tobit model shows that only the Family size and the Age of householder are fully significant in the explanation of expenditure.

Table 3

Durables for Transportation (48%)*

DEMOGRAPHIC VARIABLES	NL PROB. MODEL	TOBIT	CRAGG
REGION (*Central = base*):			
Northwest	-120.3292 (-1.6902)	-65.6325 (-0.9520)	-29.3754 (-0.4557)
Northeast	22.3686 (0.3089)	9.0253 (0.1284)	16.7770 (0.2551)
South and Islands	40.8310 (0.6425)	-8.6754 (-0.1408)	56.7969 (0.9848)
FAMILY SIZE (*3-4 = base*):			
single	-1642.8506 (-7.6857)	947.7652 (4.7708)	-806.9660 (-4.1422)
2 components	-480.8429 (-5.0441)	384.8416 (4.2389)	-112.0783 (1.2994)
5 or more components	149.6413 (2.7257)	-147.0101 (-2.7981)	4.5698 (0.0937)
AGE OF HOUSEHOLDER (*35-55 = base*):			
less than 35	-7.8986 (-0.1230)	-45.6633 -0.7139)	-28.9935 (-0.4953)
more than 55	22.6009 (0.3392)	128.5640 (1.9672)	44.3233 (0.7342)
EDUCATION (*high school = base*):			
lower than high school	109.6923 (1.7912)	76.0272 (1.2798)	60.8414 (1.1010)
bachelor degree or higher	-490.2513 (-4.8338)	-268.3670 (-2.7570)	-221.3334 (-2.4457)
OCCUPATION (*Employee = base*):			
unemploy, retired, unoccupied	45.1958 (0.5938)	116.1259 (0.5716)	33.8372 (0.4922)
professional class	-61.6214 (-0.9758)	-46.5136 (-0.7595)	-53.4816 (-0.9388)
working class	95.1280 (1.4854)	28.8282 (0.4653)	49.7320 (0.8631)
WORKERS OTHER THAN HOUSEHOLDER (*1=base*):			
no workers	-9.4265 (-0.1901)	15.7103 (0.3252)	-5.7354 (-0.1278)
more than one	134.2055 (1.7784)	-22.3577 (-0.3088)	50.1222 (0.7407)

(*) Percentage of Zero Observations

Table 4

Recreational Durables, Silverware and Jewelry

DEMOGRAPHIC VARIABLES	RECREATIONAL DURABLES (88.6%)*		SILVERWARE (95%)*
	NL PROB. MODEL	TOBIT	TOBIT
REGION (*Central = base*):			
Northwest	20.4599 (2.7264)	5.0017 (0.7194)	-3.6776 (-1.5187)
Northeast	30.0468 (3.8850)	8.7822 (1.2341)	-3.2372 (-1.2842)
South and Islands	12.9140 (1.9026)	-1.5777 (-0.2520)	2.1721 (0.9583)
FAMILY SIZE (*3-4 = base*):			
single	-69.3681 (-3.6012)	347.0722 (15.8855)	113.7311 (19.0487)
2 components	-21.3677 (-2.3026)	113.0758 (11.9563)	16.0049 (6.4228)
5 or more components	24.1846 (3.8872)	-27.8633 (-5.0780)	-7.4145 (-3.3128)
AGE OF HOUSEHOLDER (*35-55 = base*):			
less than 35	7.4904 (1.0114)	-8.0875 (-1.1577)	31.6334 (9.0112)
more than 55	-18.0300 (-2.4158)	18.7054 (2.6009)	-12.5177 (-5.5539)
EDUCATION (*high school = base*):			
lower than high school	-4.4190 (-0.6768)	5.3002 (0.8758)	-1.8168 (-0.7698)
bachelor degree or higher	-6.8597 (-0.6458)	5.3512 (0.5447)	-3.9227 (-1.0247)
OCCUPATION (*Employee = base*):			
unemploy, retired, unoccupied	-28.0778 (-3.4321)	0.4421 (0.0584)	-5.2118 (-1.9575)
professional class	-16.0677 (-2.3549)	-8.3507 (-1.3308)	-0.0710 (-0.0285)
working class	-12.7811 (-1.8467)	-12.4552 (-1.9527)	0.8972 (0.3486)
WORKERS OTHER THAN HOUSEHOLDER (*1=base*):			
no workers	-20.3694 (-3.8021)	-4.7985 (-0.9682)	3.0626 (1.6007)
more than one	4.9607 (0.6261)	-14.8595 (-1.9961)	-0.7452 (-0.2968)

(*) Percentage of Zero Observations

Table 5

CRAGG DOUBLE-HURDLE MODEL

DEMOGRAPHIC VARIABLES	BEEF (11.5%)*	WINE (48.7%)*	TOBACCO (60%)*
REGION (*Central = base*):			
Northwest	22.5858 (-1.9486)	-4.6231 (-1.5097)	-19.0078 (-2.9046)
Northeast	-123.5433 (-9.0822)	-19.1044 (-5.7401)	-56.2353 (-7.9735)
South and Islands	-18.8365 (-1.7682)	-4.7857 (-1.6684)	25.8109 (4.1970)
FAMILY SIZE (*3-4 = base*):			
single	35.7807 (1.3572)	-24.1256 (-4.0339)	-42.4204 (-2.7693)
2 components	6.3963 (0.4559)	2.7772 (0.8271)	-7.9531 (-1.0456)
5 or more components	-7.8779 (-0.7963)	3.2367 (1.1076)	12.5816 (2.1244)
AGE OF HOUSEHOLDER (*35-55 = base*):			
less than 35	-4.0322 (-0.2649)	-4.1222 (-0.8907)	6.2440 (0.7853)
more than 55	2.8509 (0.2154)	2.1440 (0.6200)	-24.2723 (-3.4054)
EDUCATION (*high school = base*):			
lower than high school	14.1836 (1.3289)	14.5685 (4.6829)	18.0501 (3.0128)
bachelor degree or higher	-17.4680 (-1.0240)	-12.0897 (-2.5036)	-42.1254 (-4.3918)
OCCUPATION (*Employee = base*):			
unemploy, retired, unoccupied	6.7810 (0.5288)	6.8418 (1.9672)	-9.4365 (-1.3197)
professional class	-8.3355 (-0.7565)	10.0504 (3.1522)	18.5930 (3.0061)
working class	-12.5491 (-1.1117)	12.6377 (3.8167)	35.3498 (5.4877)
WORKERS OTHER THAN HOUSEHOLDER (*1 = base*):			
no workers	21.3143 (2.3897)	6.3117 (2.5130)	-4.2100 (-0.8664)
more than one	-3.8809 (-0.3086)	0.9826 (0.2910)	39.8618 (5.6246)

(*) Percentage of Zero Observations

Beef, Wine and Tobacco are the consumption items that we suggest be analyzed by a double-hurdle model because the most plausible explanation of zero observations could be conscientious abstention (Table 5). While we observe that the consumption of Beef does not depend on demographic variables except the Region, for Wine and Tobacco all the variables used have at least one category significantly different from the base. For example a family with a householder aged below 35 has a lower consumption of Wine which increases with the age of the householder. The reverse is true in the case of Tobacco. The level of education affects the consumption of both goods: a college educated householder spend's less in Wine and Tobacco than the families whose heads are not college-educated.

9.3. *Adult Equivalency Weights*

The adult equivalency weights represent the third and last part of equation (1).

The bar charts (Figures 16-26) can be helpful in showing the results of all the models.

We can observe that Pasta, Beef and Food products show a similar pattern in all the models. For these items the age structure of the family does not make very much difference in the household consumption: the coefficients vary approximately between 0.85 and 1.10.

In the other cases the patterns vary across goods and, to some extent, also across the models.

The age groups from 5 to 20 years old do not contribute very much to the household consumption of Tobacco and Wine while their weight in the expenditure for Transportation Durables is two times higher than the reference age group. Continuing with Wine and Tobacco, we observe a constant weight for the age 20-50 then the trend diverges for the two items: the consumption of Wine increases because of Italian habits more deeply rooted in the old generations, while, for the age group over 50, most of the smokers are perhaps not in the sample anymore.

In general, as age increases the consumption of Durables is lower. In the case of Clothing and Shoes, the profile of weights is constant for the age groups between 20 and 65 years and we observe a significant decrease for the oldest class.

A final comment on Silverware and Jewelry and Durable House Furnishings. These two consumption categories show a percentage of zeros higher than 95 per cent. This fact could explain the indefinite pattern of equivalency weights and the wide differences among models.

Concluding Remarks

The results presented in this paper cannot be considered conclusive. The major outcome of our work is the inevitability of dealing properly with the problem of zero expenditures in a cross-section/time-series analysis of consumption. Our work underlines the necessity of using different models in the explanation of zeros for different goods in order to pinpoint the presumed reason for zero purchases for each item. The major problem is how to choose the right model. The simulation results may be useful as a guide in the selection of explanatory variables but unfortunately they are not very helpful in the choice of the most appropriate model. The procedure followed here is based upon an *a priori* linkage between consumption items, plausible interpretation of zeros and models. Our hypothesis has been supported by the results in most cases, although some difficulties remain to be solved as we have mentioned in this study.

REFERENCES

C. ALMON (1979), "A System of Consumption Functions and Its Estimation for Belgium", *Southern Economic Journal*, 46 (1) pp.85-106.

C. ALMON (1996), "A Perhaps Adequate Demand System", *INFORUM Working Paper*, Series 96 (7).

R. BARDAZZI - M. BARNABANI (1966), "A Cross-Sectional Analysis of Italian Household Consumption", *Inforum Working Paper*, Series 96 (6).

E. R. BERNDT - H. HALL - R. HALL - J. A. HAUSMAN (1974), "Estimation and Inference in Non-Linear Structural Models.", *Annals of Economic and Social Measurement*, 3, pp. 653-665.

R. W. BLUNDELL - C. MEGHIR (1987), "Bivariate Alternatives to the Tobit Model", *Journal of Econometrics*, 34, pp. 179-200.

J. A. BROWN - A. DEATON (1972), "Models of Consumer Behaviour: A Survey", *Economic Journal*, 82, pp. 1145-1236.

C. CHAO (1991), *A Cross-Sectional and Time-Series Analysis of Household Consumption and a Forecast of Personal Consumption Expenditures*, Unpublished Ph. D. Dissertation, University of Maryland, College Park.

J. CRAGG (1971), "Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods", *Econometrica*, 39, pp. 829-844.

A. DEATON (1986), "Demand Analysis", in *Handbook of Econometrics*, vol. III, edited by Griliches Z. and Intriligator M. D., North Holland, Amsterdam, pp. 1767-1839.

P. DEVINE (1983), *Forecasting Personal Consumption Expenditures with Cross-Section and Time-Series Data*, Unpublished Ph. D. Dissertation, University of Maryland, College Park.

W. E. DIEWERT - T. J. WALES (1992), "Quadratic Spline Models for Producer's Supply and Demand Functions", *International Economic Review*, 33 (3) pp. 705-722.

— (1993), "Linear and Quadratic Spline Models for Consumer Demand Functions", *Canadian Journal of Economics*, 26 (1) pp. 77-106.

A. S. GOLDBERGER (1964), *Econometric Theory*, Wiley, New York.

L. GRASSINI - A. VIVIANI (1995), "Analisi del comportamento di spesa da dati individuali in presenza di errori di registrazione", *Working Papers* (55), Department of Statistics, University of Florence.

W. H. GREENE (1993), *Econometric Analysis*, MacMillan, New York.

M. KEEN (1986), "Zero Expenditures and the Estimation of Engel Curves", *Journal of Applied Econometrics*, 1, pp. 277-286.

G. S. MADDALA (1983), *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, London.

G. S. MADDALA (1990), "Censored data models", in Eatwell, J., Milgate, M. and Newman P. (eds.) *The New Palgrave: Econometrics*, W. W. Norton & Company, New York, pp. 54-57.

S. E. PUDNEY (1990), "The Estimation of Engel Curves", Myles G.D. (ed) *Measurement and Modelling in Economics*, North Holland, pp. 267-305.

D. J. POIRIER (1976), *The Econometrics of Structural Change*, North Holland, Amsterdam.

S. J. PRAIS - H. S. HOUTHAKKER (1955), *The Analysis of Family Budgets*, Cambridge University Press, London.

G. SINGH - A. L. NAGAR (1973), "Determination of Consumer Unit Scales", *Econometrica*, 41 pp. 347-355.

E. SYDENSTRICKER - W. I. KING (1921), "The Measurement of the Relative Economic Status of Families", *Quarterly Publication of the American Statistical Association*, 17, pp. 842-857.

J. TOBIN (1958), "Estimation of Relationships for Limited Dependent Variables", *Econometrica*, 26, pp. 24-36.