A Long-run Disaggregated Cross-Section and Time-Series Demand System: An Application to Italy

Economic Systems Research, Vol. 13, No. 4, 2001

A Long-run Disaggregated Cross-section and Time-series Demand System: an Application to Italy

ROSSELLA BARDAZZI & MARCO BARNABANI

(Received September 2000; revised May 2001)

ABSTRACT *In this paper, we study a long-run disaggregated model of consumption following an approach based on an integrated cross-section and time-series demand system. The study consists of three steps. First, a cross-section analysis is performed on data from household budget surveys. At this stage, the problem of 'zero expenditures' is solved. The cross-section results are transformed into variables for use in the time-series system of demand. Then, this demand system is built and estimated. Some results for Italy concerning both the cross-section and the time-series analyses are presented.*

KEYWORDS: *Demand systems, cross-section and time series models, long-run disaggregated models*

1. Introduction

Consumer demand analysis has always held a central position in economic analysis on both the theoretical and the empirical sides. In this work, we study a long-run model of personal consumption to be used to forecast personal consumption expenditures in a modern multisectoral model for Italy. Therefore, the ultimate goal of this study is to develop a demand system that will forecast the (time-series of) household consumption in the national accounts framework. This imposes some constraints on the analysis and on the data sets to be used but allows the interaction of this fundamental GDP component with all the other relevant variables of the economy.

Although the estimation of demand functions on time-series data has certainly been the dominant concern in demand analysis in the 1980s, a much older literature is concerned with the analysis of family budgets using data on cross-sections of households. In the last few years, as more and more such microeconomic

Rossella Bardazzi, Dipartimento di Studi sullo Stato, University of Florence, Italy, E-mail: rbardaz@studistato.unifi.it; Marco Barnabani, Dipartimento di Statistica 'Giuseppe Parenti', University of Florence. The paper is the result of joint research by the authors. However, Sections 1, 2, 3, 6 and 7 are due to R. Bardazzi; Sections 4 and 5 are due to M. Barnabani. We wish to thank Clopper Almon and two anonymous referees for helpful comments on this paper. We are also grateful to Terry Barker, the chairman of the session on 'Econometrics in Input–Output Modeling' at the Twelfth International Conference on Input–Output Techniques (New York, May 1998) for useful suggestions on an earlier draft of this paper. All remaining errors are exclusively ours.

ISSN 0953-5314 print; 1469-5758 online/01/040365-25 DOI: 10.1080/09535310120089761

data sets became available and as computing power increased and econometric techniques were developed to deal with them, interest in this topic has become intense. It has now become clear that, to understand aggregate consumption, it is necessary to conduct a detailed analysis of individual (household) behaviour.[1] This does not mean, however, that aggregate time-series analysis is not useful. The specification of a time-series model for consumption is a useful instrument for summarizing the main features of the data and for forecasting. Moreover, modelling personal consumption expenditures in a macroeconomic model of the economy built upon national accounts requires an accounting consistency that is guaranteed by the time-series data on consumption, not by the microeconomic sample data.

The time-series system of consumer demand equations used in this study is the model developed by Almon (1996) and known as PADS (Perhaps Adequate Demand System).[2] This demand system has been designed to be suitable in a long-term forecasting model. When choosing a system, one must take into account the purpose for which it will be used. Since we will use the demand system to forecast consumption by goods in the long-term, our primary means to evaluate whether a system is appropriate are the long-run properties of the model and its suitability for forecasting. The Almost Ideal Demand System (AIDS) is perhaps the most popular of the recent systems of demand. In fact, this system has supplanted nearly all other systems in applied work. The PAD system is radically different from AIDS in that it is not derived by utility maximization and therefore rejects both the concept of a representative agent and that of a specific class of preferences, such as PIGLOG.[3] Almon (1979, 1996) shows that such derivation does not automatically imply reasonable properties and that the link to a utility function is neither a necessary nor sufficient condition for deriving a good demand system. The original AIDS, by Deaton & Muellbauer (1980), while possessing flexibility and ease of estimation, shows undesirable long-run properties. One problem is known as the irregularity of the class of PIGLOG cost functions. In fact, given the restriction on the real income parameters—their sum must equal zero—either the coefficients must all equal zero or at least one of them is less than zero. In this latter case, spending on the good is driven toward zero as income increases and may even become negative. This property implied by AIDS may be particularly undesirable depending on the time-horizon of the model and on the level of disaggregation at which the consumption functions are estimated. AIDS does not allow the response of an expenditure share to growth in real expenditure to be modified for higher income. To preserve regularity in a wider region of expenditure-price space, some authors have suggested a modification to the class of preferences—termed MPIGLOG—which allows an amelioration of the effect of growth in real expenditures on shares, avoiding violation of the (0, 1) interval (Cooper & McLaren, 1992). A property implied by AIDS, as well as by other models derived from utility maximization, is Slutsky symmetry in the market demand functions. On the other hand, a common criticism of the Almon system is that it lacks absolute Slutsky symmetry. Although it is well known that there is, in fact, no reason why aggregated data should respect the same rules as the individual data, the most recent demand systems impose the theoretical constraints derived from the theory of utility maximization: additivity, homogeneity and Slutsky symmetry. It has been shown that homogeneity and symmetry constraints are valid only if the assumption of a representative agent is true (Muellbauer, 1975, 1976) or, more recently, consumers' preferences are of a specific form so as to generate cost functions of the PIGLOG family (Deaton & Muellbauer, 1980). Almon (1996) has shown that while Slutsky symmetry is not a necessary property of market demand curves, 'it probably does not do great violence to reality to impose symmetry to reduce the number of parameters to be estimated' (Almon, 1996, p. 4). Therefore, the Almon system does possess base point symmetry, an assumption to reduce the number of estimated parameters in the system.[4] The system is homogeneous in prices and income, and it possesses additivity at a fixed base set of prices. As prices move from that fixed point, a 'spreader' allocates the difference in the sum of the individual equations from total expenditures in accordance with the marginal propensities to consume at current prices. Therefore, the Almon system can be faulted for lacking automatic additivity but, in defence of the system, experience has shown that the magnitude of the *ex-post* scaling factor is very low; therefore, the spreading does not materially affect the behaviour of the system.[5] The system allows complementarity/substitutability between goods and this may vary for a second good. That is, a commodity may be a complement for some goods and a substitute for other goods. Furthermore, special attention is paid to the effects of prices on the marginal propensity to consume each good as income increases.

The result of this analysis is to be used to forecast personal consumption expenditures in a modern multisectoral model for Italy. Prior to this work, consumption equations in this model have been based on time-series alone (Grassini, 1983).[6] Our work differs from the cited paper in two ways. First, the time-series function used here is a development by Almon (1996) of the original function presented by the same author in 1979. Second, we have estimated an integrated cross-section/time-series system while microeconomic data were not present in the previous version of the demand system.[7] In that framework, demographic characteristics and income at the household level could not be considered and the equations measured only the effects of average income and prices on personal consumption expenditures. The work reported here is based on equations with stronger microfoundations in that they incorporate additional information relative to the distribution of income, the age structure of the population and the demographic composition. Our demand system accounts for these effects by estimating a set of cross-section expenditure functions. It is likely that demographic variables and income play a key role in determining household consumption. In the aggregate, however, demographic characteristics move quite slowly so that their effect cannot be estimated precisely from those data. On the contrary, differences across the sample of cross-section households can profitably be exploited to identify these effects. Since these cross-section functions are estimated for a single year, there are no differences in relative prices over the sample. The effect of relative price movements is determined when the time-series analysis is undertaken because the price response is generally observed over time. Therefore, our demand analysis requires that both time-series and cross-section data be utilized. It should be stressed that the aggregate data are not obtained by summing over the data for individual households. In fact, the aggregated individual data and the National Accounts aggregates differ for several reasons. First of all, they are based on different methods of calculation. That is, the individual sample data are collected through a direct survey while the National Accounts assess household expenditures indirectly using different sources.[8] Moreover, the two data-sets differ for the reference unit, the definitions and the classification criteria.[9]

This paper is organized as follows. In Section 2, the integrated cross-section/time-series model is described. The linkage between the two models is presented in Section 3, while in Section 4 the data set is illustrated. The estimation procedure is explained in Section 5 and a selection of the estimation results of the cross-section work is presented in Section 6. Section 7 outlines some comments on the results obtained from the time-series demand system. Finally, conclusions are presented in Section 8.

2. The Cross-section/Time-series Model

The model is an integrated cross-section/time-series demand system and will be described following three steps: the cross-section equation, the time-series consumption system and the linkage between the two.

For each consumption item, the cross-section function is, in essence, a simple equation that relates per person consumption to per person income and demographic characteristics, such as:

$$\frac{c_{ih}}{N_h}=f(x_h, d_h) \quad \text{or} \quad c_{ih}=f(x_h, d_h)N_h$$

where c_{ih} is consumption of good i by household h; x_h is per-capita income within household h; d_h is a set of demographic characteristics of household h and N_h is the size of household h. To this simple form we generalize on the size of the family to make it a function of the number of members in separate age categories, say $N_h(n_{h1},\ldots,n_{hg})$. That is, the equation above can be represented as:

$$c_{ih}=f(x_h, d_h)N_h(n_{h1},\ldots,n_{hg})$$

By assuming a linear relationship between consumption, income and demographic characteristics, the cross-section equation for each commodity i may be expanded in the following form:[10]

$$c_{ht}=\left(\sum_{j=1}^{k} x_{hjt}\beta_{jt}+\sum_{j=1}^{m} d_{hjt}\delta_{jt}\right)\sum_{j=1}^{g} n_{hjt}w_{jt} \qquad h=1,\ldots,H \tag{1}$$

where

c_{ht} = consumption of household h at time t;
x_{hjt} = per-capita income within household h divided in $k=10$ brackets; j is the bracket index and t is the time index;
d_{hjt} = dummy variable j used to show inclusion of household h in $m=15$ demographic groups at time t;
n_{hjt} = number of members of household h for $g=8$ age groups at time t;
β_{jt}, δ_{jt}, w_{jt} = parameters to be estimated for each commodity at time t;
H = number of households in the sample.

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 household income and non-age demographic variables. The relationship between consumption and income is modelled through a linear spline.[11] The choice of this function is due to its flexibility in the sense of being capable (a) of representing different types of goods, such as luxuries, necessities and inferior goods; (b) of expressing different propensities to consume for different income levels. A linear spline is a piecewise function in which the linear pieces are joined together in a smooth fashion. To apply this function to

design the Engel curve, income must be divided in brackets. In our study, the boundaries of the income brackets are determined such that each of the brackets contains one tenth of the total households in the sample. Such a curve will be called Piecewise Linear Engel Curve (PLEC). 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 the household head, his/her education and occupation, and the number of workers within the family besides the household head. The effect of these variables in the equation is to shift the Engel curve up or down changing the intercept of the PLEC. No interaction among the demographic variables is assumed.[12]

The last term on the right of (1) is the product-specific size of the household and depends on the age structure of the household, more particularly, it is a weighted sum of the household members grouped by age. It is introduced in the equation by a multiplicative effect because it allows one to translate the per-capita to the household consumption of each good. The size of the household is variable because it depends on the 'adult equivalency weights' of each age cohort. The weights vary by both age and commodity because, for some goods, each age group will count differently. An additional infant in a household will not significantly increase the expenditures on alcohol by the household but will increase the expenditures on baby food and, perhaps, the household will need a bigger car. On the other hand, an elderly person will increase the medical expenses while he or she will not contribute to an increase in purchases of household's durables. The weight of each age group, w_{jt}, depends upon the relative importance of the age group in determining consumption expenditures for commodity i. When determining consumption expenditures on alcohol, households with two adults and two children should have smaller 'sizes' than households with four drinking age adults. This adult equivalency weights approach becomes more attractive when one considers the time-series analysis that will be done. Given this scheme, one can obtain a unique weighted population for each good over time, thus handling the age distribution by converting a distribution into a scalar measure of population which is different for each commodity. Therefore, one can incorporate this information into the time-series estimation so that the increase in the number of infants in the population associated with a baby boom will be an important variable in explaining changes in expenditure pattern and, as the baby boomers become adult and then old, we should see the consumption of different commodities changing because of these variations in the age structure. It is necessary to pre-determine one of the weights because equation (1) is not uniquely identified. By normalizing the weights to be one for a certain age group, for example the individuals between the ages of 30 and 40, the equation can be estimated. The technique of using an age-weighted population assumes no cohort effects. In other words, our approach assumes that the age effect on consumption estimated by the adult equivalency weights is 'pure' and that it represents the characteristic life-cycle profile of consumption. In fact, this age effect might also include a cohort effect because individual preferences change not only across ages but also across generations. Therefore, it is not just age that matters, but also when someone was born. Consumption patterns of the elderly in the year 1970 are not the same as those in the year 2000 by individuals of the same age. A distinction of these effects, a 'pure' age effect and a cohort effect, is necessary as a further development of the adult equivalency weights scheme. A preliminary analysis to assess the issue and its relevance in Italian microeconomic data is presented in Bardazzi (2000).

The time-series system of consumer demand equations used in this study is the model developed by Almon (1996) and called PADS (Perhaps Adequate Demand System).[13]

The PAD system, in its basic form, specifies the demand of a good to depend, linearly, on income, a cyclical variable and a time trend,[14] and, non-linearly, upon the prices of all other goods. However, the price effects take also into account that consumption items can be combined into economically relevant groups. Through this technique, a commodity can be a strong complement/substitute for other items in its own group while interacting less strongly with the prices of goods in other groups. Similarly, the functional form allows subgroups within which we suppose even greater sensitivity of the demand for one product to the price of others in the same subgroup. For instance, various food products form a natural group within which a subgroup was designed, including the protein-rich foods (Meat, Fish and Dairy Products). This specification of groups and subgroups serves two purposes. It economizes on the number of parameters, making this an empirically estimable system; and it divides consumption up into 'natural' functional categories of human needs. This method recalls the implementation of two-stage budgeting which is, in Deaton's words, 'almost the only sensible way to deal with very large systems' (Deaton, 1986, p. 1816). This procedure is based on a two-stage allocation process. In the first stage, total expenditure is allocated among different groups of commodities; in the second stage, total expenditure of each group is allocated among individual goods inside the group. The usefulness of this approach is that it reduces the number of own-price and cross-price elasticities of demand to be estimated.[15]

The analytical form of the PAD system is the following:

$$\frac{q_{it}}{\text{Pop}_t} = (a_i(t) + b_i y_t) \prod_{k=1}^{n} p_k^{c_{ik}} \tag{2}$$

where:

q_{it}/Pop_t = consumption per-capita in constant prices of product i,
a_i = a constant term for other non-income, non-price factors,
y_t = income (total expenditure) per-capita in constant prices,
p_k = the price index for product k, equal to 1 in the base year,
b, c = parameters to be estimated.

The system satisfies the following constraints:

$$\sum_{i=1}^{n} b_i = 1 \tag{3}$$

$$\sum_{i=1}^{n} a_i = 0 \tag{4}$$

$$\sum_{k=1}^{n} c_{ik} = 0 \tag{5}$$

Equations (3) and (4) ensure constant-price adding up: as we move away from the base year, a spreader is employed. The spreader adjusts expenditure for each

commodity by allocating the difference between total expenditures (y) and the sum of expenditures in proportion to the marginal propensities to consume with respect to y at the current prices. Equation (5) imposes homogeneity of degree zero in all prices and income on the system. This system has a lot of price parameters to be estimated. To reduce this number, Almon imposed Slutsky symmetry at some base year prices. Therefore, we assume that:

$$c_{ij} \cdot q_i/p_j = c_{ji} \cdot q_j/p_i$$

Multiplying both sides by $(p_i p_j/y)$, we obtain

$$c_{ij}/s_j^{\circ} = c_{ji}/s_i^{\circ}$$

where s_i° is the base year share of total expenditures for commodity i. If we now define $\lambda_{ij} = c_{ij}/s_j^{\circ}$, we have the symmetry condition

$$\lambda_{ij} = \lambda_{ji} \tag{6}$$

This reduces the number of price parameters by half. A further restriction is imposed by combining commodities into groups and sub-groups. This grouping technique was introduced by Almon (1979) and improved in Almon (1996) with a specification where every product has its own-price elasticity. We will have as many price-exponent parameters as there are commodities plus groups plus subgroups. Almon (1996) then derives the equation to be estimated as

$$\frac{q_{it}}{\text{Pop}_t} = (a_i + b_i y_t + c_i \Delta y_t + d_i \,\text{time}) \left(\frac{p_i}{P}\right)^{-\lambda_i} \prod_{k=1}^{n} \left(\frac{p_i}{p_k}\right)^{-\lambda_k s_k} \left(\frac{p_i}{P_G}\right)^{-\mu_G} \left(\frac{p_i}{P_g}\right)^{-\nu_g} \tag{7}$$

where

q_{it}/Pop_t = consumption per-capita in constant prices of product i,
y_t = income (total expenditure) per-capita in constant prices,
$\Delta y_t = y_t - y_{t-1}$,
time = time trend,
s_k = the budget share of product k in the base year,
p_k = the price index for product k, equal to 1 in the base year,
a, b, c, d, λ, μ, ν = parameters to be estimated,
P, P_G, P_g = overall, group, and subgroup price indexes given by:

$$P = \prod_{\text{all } k} p_k^{s_k} \qquad P_G = \left(\prod_{k \in G} p_k^{s_k}\right)^{1/\sum_{k \in G} s_k} \qquad \text{and} \qquad P_g = \left(\prod_{k \in g} p_k^{s_k}\right)^{1/\sum_{k \in g} s_k}$$

Then, with respect to the price effect, the functional form allows: one price parameter for each commodity, (i.e. the λs); one price parameter for each group, (i.e. the μs); and one price parameter for each subgroups (i.e. the νs). A positive value of these parameters indicates substitutability relative to the other commodities (λ), to the commodities in the group (μ) and within the subgroup (ν). A negative value indicates complementarity.

In the first parenthesis of equation (7) we have constant real income per-capita, the first difference of real income per-capita, and a linear time trend. Furthermore, consumption per-capita for all items is computed with reference to total population. In fact, when household data are available, results from the cross-section work may be incorporated within a modified-version of the demand system described above.

3. The Aggregation Procedure: The Linkage

In the cross-section, we estimated to what extent household consumption was affected by demographic composition, age structure and the distribution of income. The task now is to combine these cross-section results with information on the structure of the population as a whole, to create variables for use in the time-series analysis described above.

The estimated cross-section parameters allow us to introduce into the system of demand:

- the marginal propensities to consume specific to each good in each of the income brackets;
- the impact of demographic variables;
- the effect of age composition on consumption of each item.

To proceed towards this task, a 'prediction' of consumption of each good in year t, $C^{\star}_{it}$, may be constructed by using the estimated $\hat{\beta}_{jt}$ and $\hat{\delta}_{jt}$ parameters summing the linear term of equation (1) over all households as follows

$$C^{\star}_{it} = \sum_{h=1}^{N}\left(\sum_{j=1}^{k} x_{hjt}\hat{\beta}_{jt} + \sum_{j=1}^{m} d_{hjt}\hat{\delta}_{jt}\right) \qquad (8)$$

Equation (8) can be viewed as the result of evaluating the cross-section equation for each individual in the population and then averaging the results.[16] The $C^{\star}_{it}$ variable is a 'prediction' of what the expenditure per adult equivalent would be if it were determined only from income and demographic variables. This $C^{\star}_{it}$ takes into account changes in total income, its distribution and demographic variables. This information is interpreted for each item through the use of the cross-section coefficients; that is, the demographic and income variables are summed with the weights being the cross-section parameters.

A commodity specific weighted population is constructed through the estimated adult equivalency weights, $\hat{w}_{jt}$, so that a more relevant population size for each good is provided. A weighted population time-series for a commodity, WP_{it}, is defined by:

$$\mathrm{WP}_{it} = \sum_{h=1}^{N} \hat{n}_{ht} = \sum_{h=1}^{N}\sum_{j=1}^{g} n_{hjt}\hat{w}_{jt} = \sum_{j=1}^{g} N_{jt}\hat{w}_{jt} \qquad (9)$$

where N_{jt} is the number of individuals in age group j in year t. The advantages of using the commodity specific weighted population sizes in the analysis of time-series consumption behaviour are many. If we were to choose to use only total population in the examination of all commodities, we would be implicitly assuming either that the age structure of the population is unimportant to the determination of consumption or that the shifts through time in the relative sizes of age groups are minor. The cross-section results clearly contradict the first assumption and the second can be refuted with casual observation of recent data.

To summarize, the variable $C^{\star}_{it}$ captures the effects of demographic and income variables over time while the WP_{it} variable incorporates the growth and changing age structure of the population, as it affects demand for each product. These two variables allow us to perform the linkage between the cross-section and the time-series equations. As for the linear term in equation (7), instead of the real income per-capita, y_t, and the cyclical variable, Δy_t, we use the prediction $C^{\star}_{it}$ and its first difference $\Delta C^{\star}_{it}$. We recall that $C^{\star}_{it}$ is not really income but rather consumption per-capita predicted only from changes in income and demographic variables, with no change in base period prices. It provides a path for slowly moving demographic variables as well as changes in income to affect the value of consumption of each good over time. The elasticity of time-series per-capita consumption with respect to this variable is expected to be positive but the size of this coefficient does not have to be equal to the income elasticity. Then, the dependent variable of the time-series equation, the consumption per-capita, is computed by using a different population for each commodity, that is the variable WP_{it}, as in (9). Therefore, after the linkage with the cross-section work, equation (7) becomes:

$$\frac{q_{it}}{\mathrm{WP}_{it}} = (a_i + b_i C^{\star}_{it} + c_i \Delta C^{\star}_{it} + d_i \,\mathrm{time})\left(\frac{p_i}{P}\right)^{-\lambda_i} \prod_{k=1}^{n}\left(\frac{p_i}{p_k}\right)^{-\lambda_k s_k}\left(\frac{p_i}{P_G}\right)^{-\mu_G}\left(\frac{p_i}{P_g}\right)^{-\nu_g} \qquad (10)$$

where the subscript i of the WP and $C^{\star}$ variables emphasizes that they are product-specific. Equation (10) is the time-series demand system to be estimated.

4. The Data

The data used for the cross-sectional analysis are obtained from the 'Household Expenditure Survey', conducted by the Italian Institute of Statistics (ISTAT) from 1985 to 1996 and they are classified in 76 expenditure categories.[17] The time-series data come from the National Accounts produced by ISTAT. The reason for using different sources of data is because the demand system we are describing is part of a modern interindustry multisectoral model, where the statistical information concerning the whole set of economic variables, including personal consumption expenditure, come from the National Accounts. This accounting consistency is a constraint we must satisfy.

For practical purposes of linking the cross-sectional analysis with the time-series work, we have aggregated the cross-section expenditure categories into 40 items listed in the first column of Table 1.[18] Inspection of survey data immediately revealed the presence of many zero purchases for most of the consumption items considered in the classification. The problem of zero expenditures is severe at least in more than a half of expenditure categories—the percentage of zero observations is above 30%—as may be seen in the second column of Table 1, where the percentages of zeros for the year 1996 are presented. Therefore, this problem must be treated with an appropriate model. The model used in this study will be briefly described in the following paragraph.

The sample represents about 2‰ of the population. In the Italian data, each household is assigned a weight (grossing-up factor) representing the number of households in the whole population corresponding to that particular unit. Weights are given for each quarter of observation; a reweighting procedure is applied in order to use yearly-based observations.[19] In our study, to compute the $C^{\star}_{it}$ and WP_{it} variables (equations (8) and (9)), a time-series of the average per-capita total expenditure, the population proportions of each demographic group, and the number of individuals in each age group, are needed. The problem is that these data either are not available in the Italian statistics or they are not properly classified for our purposes. Therefore, we have used the grossing-up factor to produce this data set.

Table 1. Expenditure categories

Consumption categories	Percentages of zeros—1996
1. Cereals and bakery products	1.0
2. Meat	2.7
3. Fish	22.1
4. Dairy products	1.6
5. Fats & oils	19.4
6. Fruit	1.9
7. Potatoes	1.9
8. Sugar	33.6
9. Coffee, tea and cocoa	26.0
10. Other food	26.1
11. Non-alcoholic beverages	13.0
12. Alcoholic beverages	12.0
13. Tobacco	60.6
14. Clothing	20.0
15. Footwear and repair	53.0
16. Tenant occupied rent	0.0
17. Electricity, oil, gas	0.0
18. Furniture	42.0
19. Household linen	78.5
20. Kitchen and household appliances	42.0
21. China, glassware and tableware	80.9
22. Other non-durables and services	97.3
23. Domestic services	15.5
24. Drug Preparation and sundries	40.1
25. Orthopedic equipment	80.9
26. Physicians, dentists, other medical professionals	84.5
27. Hospitals, nursing homes	82.0
28. Vehicles	53.0
29. Operation of motor vehicles	23.0
30. Public transportation	72.4
31. Communication	7.7
32. TV, radio, musical instruments	40.1
33. Books, magazines and newspapers	24.6
34. Textbooks	92.5
35. Recreational services	47.0
36. Personal care	15.8
37. Hotels and motels, restaurants	34.9
38. Other goods	92.4
39. Financial services and insurance	50.7
40. Other services	55.7

5. The Estimation Procedure

The cross-section/time-series system has required an estimation procedure based both on a cross-section analysis dealing with the problem of zero expenditures, identifying the effect of income distribution and demographic factors, and on a time-series analysis.

As for the cross-section analysis, 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 it is currently observed is total purchases over the survey period. 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 where it can assume zero values.

Several models have been proposed to treat zero expenditures due to various causes such as infrequency of purchase, economic reasons, conscientious abstention and misreporting. We must stress that these models are independent from the analytical form of cross-section function specified: consumption could be measured in quantity or in budget shares, the underlying Engel curve could be different as well as the specification of demographic variables. The problem of zero expenditures is a general one, and the functional form of the consumption model adopted is not particularly important in how it is handled.

To tackle the problem of zero expenditures, in this work we have decided to adopt a non-linear probability model which may be seen as a sequential choice model based on the following steps.

(i) Estimation of the probability to purchase. Suppose we let the binary choice of whether to consume at time t be represented by a dichotomous variable I_h which takes the value 1 when the hth family consumes, and 0 when it doesn't. That is,

$$I_h = \begin{cases} 1 & \text{if } c_h > 0 \\ 0 & \text{if } c_h \leqslant 0 \end{cases}$$

where we have suppressed the index t for convenience. Let us write the model (1) at time t in summary notation as $c_h = (\mathbf{x}_h\boldsymbol{\beta} + \mathbf{d}_h\boldsymbol{\delta})\mathbf{n}_h\mathbf{w} + u_h$ where $\mathbf{x}_h$, $\mathbf{d}_h$, $\mathbf{n}_h$ are row vectors respectively of income, demographic and size variables; $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, $\mathbf{w}$ the column vectors of parameters to be estimated; u_h the random component. Then, the probability to purchase is given by

$$P(I_h = 1) = P[u_h > -(\mathbf{x}_h\boldsymbol{\beta} + \mathbf{d}_h\boldsymbol{\delta})\mathbf{n}_h\mathbf{w}] = \Phi_h[(\mathbf{x}_h\boldsymbol{\beta} + \mathbf{d}_h\boldsymbol{\delta})\mathbf{n}_h\mathbf{w}]$$

where Φ_h is assumed to be a normal distribution function.

(ii) Fitting the non-linear regression model $c_h = (\mathbf{x}_h\boldsymbol{\beta} + \mathbf{d}_h\boldsymbol{\delta})\mathbf{n}_h\mathbf{w} + u_h$ for the subsample for which the consumption is positive. To solve the problem we applied ordinary least squares to the Gauss–Newton regression obtained through a first-order Taylor-series approximation around some parameter vector. As known, cross-section data could produce heteroscedasticity: higher income, size and composition of the family will, presumably, enlarge the expenditure. However, we tried to catch these aspects acting not just on the random error component, but on the specification of the model where classes of income and structure of the family are considered. In fact, we also tried to estimate the model with a generalized least square approach, taking into account the problem of heteroscedasticity without obtaining significative changes of the results.

(iii) Computing expected expenditure for all families using estimates of Step (ii).

(iv) Computing the expected purchases, or consumption, for all households multiplying the probabilities obtained in Step (i) by the expected expenditure of Step (iii).[20]

Although we recognize the necessity of using different models in the explanations of zeros for different goods, for the purpose of the present study we decided to use only one model for the whole set of consumption items. The choice of the non-linear probability model is due to its flexibility and capacity to justify different sources of zero expenditures in a wide set of consumption items. This allows us to economize on the number of models that would be required for 40 expenditure categories over 11 years. Besides, although non-linear, the functional form of the cross-section model becomes linear when we divide the dependent variable by the family size. This allows a relatively simple way to aggregate (by summation) the cross-section results and to use them in a time-series analysis. In fact the computation of what we called C-star, $C^\star_{it}$, is straightforward.

As described above, $C^\star_{it}$ is the vehicle used to transfer the effects of demographic and income variables to the time-series model by inserting it in place of the real income per-capita and the cyclical variable. Then, the estimation of the full cross-section/time-series system proceeds by estimating the time-series model regressing consumption on prices, time trend and the 'new' variable, $C^\star_{it}$. This approach has a strict analogy with the famous two-stage least squares. In fact, in stage 1 we obtain the ordinary least square predictions from the regression of expenditure on income, demographic variable and family size; in stage 2 we estimate the parameters of the time-series model regressing the consumption on prices, time trend and $C^\star_{it}$.

A brief description of the estimation procedure applied to the time-series model is now necessary.

Let $z_{it} = q_{it}/\text{Pop}_t$ $i = 1, \ldots, n$; $t = 1, \ldots, T$, be the consumption *per adult equivalent* of the tth observation of product i; $\xi_{it}(\boldsymbol{\alpha}_i, \boldsymbol{\beta})$ the tth observation on the regression function that determines the conditional mean of the dependent variable; $\boldsymbol{\alpha}_i$ denotes the vector of parameters of the 'income-and-time term', the term in the first round bracket in equation (2) for product i; $\boldsymbol{\beta}$ denotes the vector of parameters of the 'price term', the rest of the formula (2). Thus, $\boldsymbol{\beta}$ consists of all values of λ, μ, v. Note that $\boldsymbol{\beta}$ is the same for all products, although a particular μ or v may not enter the equation of a given commodity. Then, we can write the multivariate non-linear regression model as

$$z_{it} = \xi_{it}(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) + \omega_{it} \quad i = 1, \ldots, n; \quad t = 1, \ldots, T$$

where u_{it} is an error term, presumably with an autocorrelated structure. In vector form, the non-linear regression model for the product i may be written as

$$\mathbf{z}_i = \boldsymbol{\xi}_i(\boldsymbol{\alpha}_i, \boldsymbol{\beta}) + \mathbf{u}_i \quad i = 1, \ldots, n \quad \mathbf{u}_i \sim ID(0, \boldsymbol{\Sigma}_i) \tag{11}$$

where $\mathbf{z}_i = [z_{i1}, \ldots, z_{iT}]$, $\boldsymbol{\xi}_i(\cdot) = [\xi_{i1}(\cdot), \ldots, \xi_{iT}(\cdot)]$. The estimation procedure is carried out on the Gauss–Newton regression, which is based on a first-order Taylor-series approximation to (11) around some parameter vector $\boldsymbol{\alpha}_i^\star$, $\boldsymbol{\beta}^\star$. This yields

$$\mathbf{z}_i = \boldsymbol{\xi}_i(\boldsymbol{\alpha}_i^\star, \boldsymbol{\beta}^\star) + \mathbf{A}_i(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_i^\star) + \mathbf{B}_i(\boldsymbol{\beta} - \boldsymbol{\beta}^\star) + \text{higher order terms} + \mathbf{u}_i \tag{12}$$

where $\mathbf{A}_i$ is the matrix of partial derivatives of $\xi_{it}(\cdot)$ with respect to the $\boldsymbol{\alpha}_i$, calculated in $\boldsymbol{\alpha}_i^\star$; $\mathbf{B}_i$ is the matrix of partial derivatives of $\xi_{it}(\cdot)$ with respect to $\boldsymbol{\beta}$, calculated in $\boldsymbol{\beta}^\star$.

For estimation purposes, it may be convenient to stack all regressions given by (12) in a unique regression

$$\begin{bmatrix} \mathbf{y}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{y}_n \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 & 0 & & \cdots & 0 & \mathbf{B}_1 \\ 0 & \mathbf{A}_2 & 0 & \cdots & 0 & \mathbf{B}_2 \\ \cdot & 0 & \cdot & & & \\ \cdot & & & & 0 & \cdot \\ 0 & & \cdots & 0 & \mathbf{A}_n & \mathbf{B}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \cdot \\ \cdot \\ \boldsymbol{\alpha}_n \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{r}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{r}_n \end{bmatrix}$$

or, in summary notation,

$$\mathbf{Y}^c = \mathbf{W}\boldsymbol{\theta} + \mathbf{R}^c$$

where $\mathbf{y}_i = \mathbf{z}_i - \boldsymbol{\xi}_i(\boldsymbol{\alpha}_i^\star, \boldsymbol{\beta}^\star) + \mathbf{A}_i\boldsymbol{\alpha}_i^\star + \mathbf{B}_i\boldsymbol{\beta}^\star$ and $\mathbf{r}_i$ is for residuals and it is given by the error term $\mathbf{u}_i$ plus higher-order terms of the Taylor-series approximation.

Then, $\boldsymbol{\theta} = (\mathbf{W}'\mathbf{W})^{-1}\mathbf{W}'\mathbf{Y}^c$ is a first attempt of a solution obtained through an iterative procedure, following the Marquardt approach (Fletcher, 1987).

Although the fits calculated by the mean squared percentage error are fairly good, the results obtained showed several problems both economically and statistically.

(a) About half of the own price elasticities are positive. Since they are 'compensated', all should be negative.
(b) The income elasticities should be positive or zero, but for about 20% of products it is negative.
(c) Many parameters show a non-significant t-statistic.
(d) The autocorrelation coefficients of the residuals are not clear.

To overcome problems (a)–(d), we introduced into the model a first-order serial correlation structure and imposed some stochastic constraints (Theil & Goldberger, 1960) both on price and income parameters.

In our system, with 40 consumption categories there are over 250 parameters involved in the simultaneous estimation. The autocorrelation introduces a large number of additional parameters, which makes it difficult to obtain reliable estimates of the parameters. In order to reduce the number of parameters to be estimated, we allowed u_{it} to depend only on $u_{i,t-1}$ and not on $u_{s,t-1}$ for $s \neq i$. As for the stochastic constraints, we introduced formally the following restrictions

$$\begin{bmatrix} \mathbf{r}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{r}_n \end{bmatrix} = \begin{bmatrix} \mathbf{R}_1 & 0 & & \cdots & 0 & \mathbf{S}_1 \\ 0 & \mathbf{R}_2 & 0 & \cdots & 0 & \mathbf{S}_2 \\ \cdot & 0 & \cdot & & & \\ \cdot & & & & 0 & \cdot \\ 0 & & \cdots & 0 & \mathbf{R}_n & \mathbf{S}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \cdot \\ \cdot \\ \boldsymbol{\alpha}_n \\ \boldsymbol{\beta} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_1 \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{v}_n \end{bmatrix}$$

or, in summary notation,

$$\mathbf{r}^c = \mathbf{R}\boldsymbol{\theta} + \mathbf{V}^c$$

where $\mathbf{r}_i$ is a column vector with known elements suggested by our *a priori* beliefs on the value of the parameters of the product i $(= 1, \ldots, n)$ that are, nevertheless, approximate and, to some extent, subjective; $\mathbf{R}_i$ and $\mathbf{S}_i$ are 'zero-one' matrices with one in the position where we want to pick up, respectively, the parameter of the 'income-and-time-term' and/or of the 'price-term' of the product i; $\mathbf{v}_i$ is the stochastic component with zero mean and variance equal to the diagonal matrix $\mathbf{\Lambda}_i$ which is determined on an empirical basis as will be explained later.

The introduction into the model of a first-order serial correlation and of a stochastic constraint leads to solving the following optimization problem

$$\min_{\boldsymbol{\theta}} (\mathbf{Y}^c - \mathbf{W}\boldsymbol{\theta})'\mathbf{\Sigma}_d^{-1} (\mathbf{Y}^c - \mathbf{W}\boldsymbol{\theta}) - (\mathbf{r}^c - \mathbf{R}\boldsymbol{\theta})'\mathbf{\Lambda}_d^{-1}(\mathbf{r}^c - \mathbf{R}\boldsymbol{\theta}) \tag{13}$$

where $\mathbf{\Lambda}_d$ and $\mathbf{\Sigma}_d$ are block diagonal matrices with elements, respectively, $\mathbf{\Lambda}_i$ and $\mathbf{\Sigma}_i$. The analytical solution to (13) is given by

$$\boldsymbol{\theta} = (\mathbf{W}'\mathbf{\Sigma}_d^{-1}\mathbf{W} + \mathbf{R}'\mathbf{\Lambda}_d^{-1}\mathbf{R})^{-1}(\mathbf{W}'\mathbf{\Sigma}_d^{-1}\mathbf{Y}^c + \mathbf{R}'\mathbf{\Lambda}_d^{-1}\mathbf{r}^c) \tag{14}$$

which is the multivariate version of the well known solution given by Theil & Goldberger (1960) for multiple linear regression models.

Because of the complexity of the estimation of equation (14) especially in light of observations (a)–(d) previously mentioned, we decided to follow a specific-to-general estimation approach, which begins with the estimates of equation (14) for each product i. That is,

$$\boldsymbol{\delta}_h = (\mathbf{Q}_h'\mathbf{\Sigma}_h^{-1}\mathbf{Q}_h + \mathbf{P}_h'\mathbf{\Lambda}_h^{-1}\mathbf{P}_h)^{-1}(\mathbf{Q}_h'\mathbf{\Sigma}_h^{-1}\mathbf{y}_h + \mathbf{P}_h'\mathbf{\Lambda}_h^{-1}\mathbf{v}_h) \tag{15}$$

where $\boldsymbol{\delta}_i = [\boldsymbol{\alpha}_i' \quad \boldsymbol{\beta}']$; $\mathbf{Q}_i = [\mathbf{A}_i \quad \mathbf{B}]$; $\mathbf{P}_i = [\mathbf{R}_i \quad \mathbf{S}_i]$ and then passes the results into the general solution estimating simultaneous equation (15).

That is, the iterative process of estimation may be described as follows.

First stage. Specific for each product i.

(i) Set initially $\mathbf{\Lambda}_i = \mathbf{0}$.
(ii) Estimate the matrices $\mathbf{\Sigma}_i^{-1}$ adopting the Hildreth–Lu approach, call it $\hat{\mathbf{\Sigma}}_i^{-1}$. This also leads to obtaining a first estimate of the parameters.
(iii) With $\hat{\mathbf{\Sigma}}_i^{-1}$ and the first estimate of the parameters obtained in Step (ii) enter equation (15). Initially, set the elements of $\mathbf{\Lambda}_i^{-1}$ equal to the corresponding elements of $\hat{\mathbf{\Sigma}}_i^{-1}$ so as to have an order of magnitude of the variance for the stochastic constraints.
(iv) Find iteratively the solution of (15).
(v) Evaluate the results both from economic and statistical viewpoints.
(vi) If results are satisfactory, the estimation procedure stops. Otherwise
(vii) Act on $\mathbf{\Lambda}_i$ pursuing the following rule of thumb. Increase or reduce the λs looking for the best compromise between the loss in terms of the fitting of the model and the economic and/or statistical significance of the parameters.
(viii) Return to Step (iv).

Second stage. General for the multivariate model.

(i) With the estimates of $\boldsymbol{\theta}$, $\mathbf{\Sigma}_d^{-1}$ and $\mathbf{\Lambda}_d^{-1}$ obtained following stage one, apply Steps (iv)–(viii) to the equation (15).

Because of the number of the equations and, then, of the parameters, Step (vii) may result in a difficult and daunting task. However, this may not be true and the

modification of the $\mathbf{\Lambda}$ matrices could require minor changes if the work conducted in the first stage was made with sufficient care.

6. The Cross-section Results

The cross-section function is the result of three components: per-capita household income, demographic effects and the specific size of the household. Therefore we will discuss the effects of each in turn. Because of the large number of results obtained at this stage, 40 equations for 11 years, we have chosen to present here only a very short selection of estimates. For the most recent year, 1996, we have selected seven consumption categories just to give an example of our findings. The categories are Bakery Products, Alcoholic Beverages, Clothing, Footwear, Furniture, House Appliances, and Vehicles. Some of these are very representative as far as the problem of zero expenditures is concerned (which applies especially to Furniture, Household Appliances and Vehicles).

The estimation results are shown in Tables 2 and 3 and in Figures 1(a)–1(d). The tables report, for each consumption category, the coefficients of demographic variables (Table 2) and the adult equivalency weights (Table 3).[21]

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 1(a)–1(d)). These plots 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 household head aged between 35 and 55 working as an employee. The consumption expenditure per adult equivalent is plotted against per-capita household income.

The coefficients 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 us to underline specific patterns for certain groups of consumption goods. For instance, Bakery products (Figure 1(a)) and Alcoholic Beverages (Figure 1(b)) show a definite necessity pattern of consumption, while Clothing (Figure 1(c)) has a relatively constant slope of the Engel curve throughout all the income brackets. Furniture (Figure 1(d)) is an example of luxuries showing a steeper slope in the upper levels of income as Durable Household Appliances and Vehicles.

The results of demographic variables shown in the tables should be interpreted as deviations with respect to the base category omitted from the regression.[22] A negative dummy variable coefficient—for example, in Table 2 the parameter of Northwest for Clothing (−65.76) and Footwear (−37.23)—means that the consumption of these items of a household living in this part of the country is lower than a family living in the Centre, and vice-versa if the coefficient is positive.

The Region is significant for at least one region category and for all goods except for Vehicles. Also, the Family Size has a significant impact on the household consumption, particularly for durables (Table 2). Smaller households have a lower consumption of Furniture and Household Appliances and the expenditure for Vehicles increases with the number of family members.

In the case of the other demographic variables, with respect to Household Appliances, Furniture and Vehicles, the model shows a significant impact only for Education, although the coefficients present a sign not consistent with our *a priori* expectations: household heads with a lower degree of education tend to spend more for these luxury goods than the college educated household heads. A possible

Table 2. Non-linear probability model—Results for 1996

Demographic variables	1–Bakery prod. (1%)*	12–Alcohol (12%)*	14–Clothing (20%)*	15–Footwear (53%)*	18–Furniture (42%)*	20–Hous. appliances (42%)*	28–Vehicles (53%)
Region (Central = base):							
Northwest	44.13	11.73	−65.76	−37.23	−36.87	−3.12	26.23
	(3.24)	(2.39)	(−2.31)	(−4.72)	(−2.49)	(−0.43)	(0.39)
Northeast	20.52	6.17	24.11	−13.75	6.35	14.33	5.91
	(1.51)	(1.24)	(0.84)	(−1.76)	(0.43)	(1.92)	(0.09)
South and Islands	13.03	−2.88	88.92	20.70	18.24	25.79	64.82
	(1.09)	(−0.60)	(3.48)	(2.98)	(1.44)	(3.89)	(1.12)
Family size (3–4 = base):							
single	41.86	−15.13	−78.61	13.35	−211.42	−72.18	−1009.81
	(1.56)	(−1.66)	(−1.33)	(0.86)	(−5.44)	(−4.78)	(−5.97)
2 components	15.95	1.03	−42.46	−15.32	−10.06	−6.67	−339.41
	(1.04)	(0.19)	(−1.32)	(−1.79)	(−0.54)	(−0.83)	(−4.07)
5 or more components	−14.75	−2.02	28.65	7.82	14.18	9.34	102.15
	(−1.36)	(−0.48)	(1.25)	(1.22)	(1.33)	(1.59)	(2.08)
Age of householder (35–55 = base):							
less than 35	7.74	11.53	23.330	1.02	−10.47	7.49	3.00
	(0.46)	(1.73)	(0.75)	(0.11)	(−0.93)	(0.99)	(0.05)
more than 55	−3.77	−1.03	17.89	−4.56	−24.32	−8.31	−20.07
	(−0.26)	(−0.20)	(0.64)	(−0.59)	(−1.74)	(−1.19)	(−0.34)
Education (high school = base):							
lower than high school	11.80	−7.36	−27.447	−6.51	33.65	21.51	100.85
	(1.06)	(1.75)	(−1.18)	(−1.02)	(2.96)	(3.50)	(1.92)
bachelor degree or higher	−14.26	−7.63	−8.45	−7.24	−45.59	−16.12	−208.65
	(−0.80)	(−1.14)	(−0.22)	(−0.70)	(−2.57)	(−1.66)	(−2.47)
Occupation (Employee = base):							
unemployed, retired, unoccupied	6.71	8.30	−31.37	−17.40	5.49	7.69	75.90
	(0.49)	(1.65)	(−1.11)	(−2.23)	(0.38)	(1.05)	(1.18)
professional class	4.20	7.46	41.47	5.15	−40.13	−3.86	−28.84
	(0.35)	(1.63)	(1.65)	(0.74)	(−3.32)	(−0.59)	(−0.52)
working class	12.35	11.09	−12.50	2.47	1.30	−2.66	22.40
	(0.98)	(2.32)	(−0.48)	(0.34)	(0.10)	(−0.39)	(0.39)
Workers other than the householder (1 = base):							
no workers	10.23	4.57	−22.19	0.6	−22.58	1.28	−27.86
	(1.05)	(1.25)	(−1.11)	(0.11)	(−2.31)	(0.25)	(−0.62)
more than one	2.65	0.15	25.31	0.86	6.68	−4.72	62.92
	(0.19)	(0.03)	(0.85)	(0.11)	(0.41)	(−0.60)	(0.90)

(*)Percentage of zero observations. *t*-values in parentheses.

Table 3. Adult equivalency weights—results for 1996

Age groups	1–Bakery prod.	12–Alcohol	14–Clothing	15–Footwear	18–Furniture	20–Household appliances	28–Vehicles
0–4	0.963	0.704	0.787	0.627	3.867	1.234	2.353
	(−0.21)	(−1.90)	(−1.49)	(−2.78)	(9.20)	(1.33)	(5.55)
5–14	1.079	0.721	0.891	0.682	2.018	1.158	1.591
	(0.55)	(−2.37)	(−1.01)	(−3.23)	(5.59)	(1.22)	(3.69)
15–19	1.014	0.621	0.804	0.741	0.654	0.673	1.334
	(0.10)	(−3.11)	(−1.75)	(−2.42)	(−2.45)	(−2.70)	(2.22)
20–29	0.961	0.746	0.871	0.906	1.077	0.990	1.302
	(−0.43)	(−3.41)	(−1.88)	(−1.32)	(0.99)	(−0.13)	(3.78)
30–39	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
40–49	1.100	1.239	0.968	1.019	0.932	0.991	0.958
	(1.02)	(2.46)	(−0.45)	(0.26)	(−0.97)	(−0.13)	(−0.63)
50–64	1.145	1.264	0.899	0.880	1.150	0.998	0.982
	(1.42)	(2.66)	(−1.53)	(−1.73)	(2.17)	(−0.03)	(−0.31)
> 64	1.159	1.163	0.780	0.837	0.629	1.065	0.652
	(1.45)	(1.60)	(−3.17)	(−2.14)	(−6.01)	(0.83)	(−5.98)

explanation is that in Italy a high educational level is not always connected with higher income.

As to the adult equivalency weights (Table 3), we can observe that for Cereals and Bakery Products the age structure of the family does not make very much difference in the household consumption: the coefficients vary approximately between 0.96 and 1.15.

In the other cases, the patterns vary across goods. The age groups from 0 to 30 years old do not contribute very much to the household consumption of Alcoholic Beverages while their weight in the expenditure for Furniture and Vehicles is at least double the reference age group (almost four times in the case of Furniture). This result is consistent with what we have found studying the demographic effects: there, we have observed that the consumption of durables increases with the number of family components; that is, we may add now, the number of children. The presence of children aged from 0 to 14 years old also has an impact on the consumption of Household Appliances, even though their weight is only 20% above the reference adult.

In the case of Clothing and Footwear, the profile of weights is almost constant across the age groups, but we observe a lower weight for the oldest and the youngest classes.

7. Results of the PAD System in the Italian Model: Some Comments

The results of the time-series estimation of the PAD system equations are summarized in Table 4 and in Figures 2(a)–2(h).

As a general comment on these results, we may observe that the system reproduced quite well the historical values and almost all coefficients had the desired signs and sensible magnitudes.

The groups designed for the Italian consumption categories are as follows. Group 1: Food; Group 2: Dress; Group 3: House furnishing and operation;

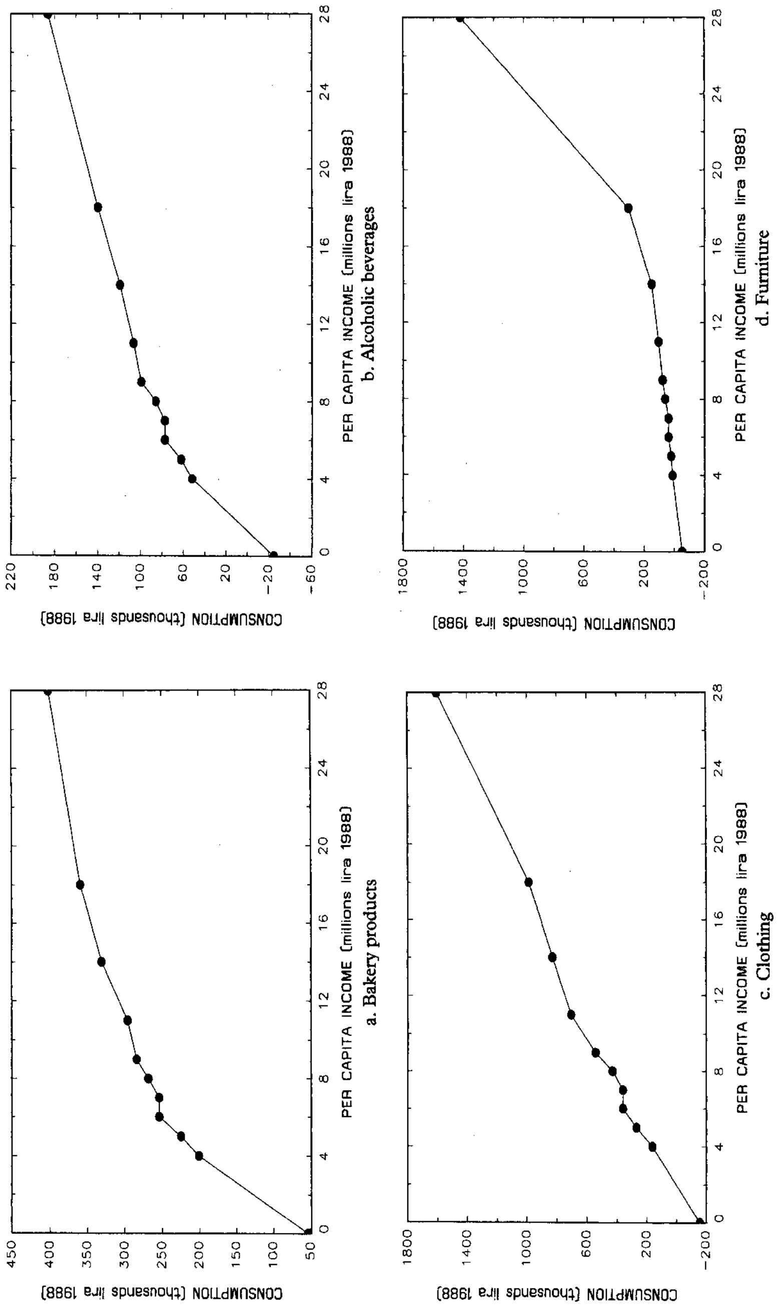

Figure 1. Piecewise Linear Engel curves, 1996.

Table 4. Regression results of Italian private consumption expenditures

μ: −0.11 −0.11 −0.26 −0.19 0.10
ν: 0.08 −0.07

	NSEC Title	G	S	I	Lamb	Share	C*El	DC*	Time%	PrEl	Err%	rho
1	Cereals and Bakery Products	1	0	1	0.09	0.025	0.14	−0.73	0.11	−0.15	0.21	−0.02
2	Meat	1	1	1	0.09	0.056	0.13	−0.95	−0.91	−0.20	0.63	0.25
3	Fish	1	1	1	0.07	0.013	0.44	−0.43	−0.67	−0.20	1.53	0.16
4	Dairy Products	1	1	1	0.10	0.029	0.17	0.38	0.17	−0.22	0.88	0.52
5	Fats & Oils	1	0	1	0.12	0.008	0.16	−1.06	−1.30	−0.18	0.58	0.27
6	Fruit	1	0	1	0.11	0.044	0.34	−0.52	−0.50	−0.18	0.68	0.22
7	Potatoes	1	0	1	0.18	0.002	0.11	0.07	−0.30	−0.24	1.24	−0.06
8	Sugar	0	0	0	0.00	0.003	0.16	−0.87	−0.41	0.00	0.63	−0.05
9	Coffee, Tea and Cocoa	1	0	1	0.09	0.005	0.02	0.33	−0.41	−0.15	1.11	0.56
10	Other Food	1	0	1	0.09	0.006	0.35	−0.27	1.27	−0.15	0.73	0.06
11	Non Alcoholic Beverages	1	0	1	0.09	0.003	1.03	−0.52	3.99	−0.15	1.15	0.21
12	Alcoholic Beverages	1	0	1	0.09	0.011	0.14	−0.18	−2.09	−0.15	0.42	−0.15
13	Tobacco	0	0	0	0.00	0.014	0.02	−0.03	−1.41	0.00	2.06	0.05
14	Clothing	2	0	1	0.10	0.083	0.84	0.05	−0.01	−0.23	1.72	−0.08
15	Footwear and Repair	2	0	1	0.35	0.021	0.36	−0.32	−0.01	−0.41	2.70	0.49
16	Tenant Occupied Rent	0	0	1	0.10	0.115	0.05	1.76	2.59	−0.24	0.32	−0.56
17	Electricity, Oil, Gas	0	0	1	0.03	0.032	0.61	−0.29	0.88	−0.20	2.76	−0.06
18	Furniture	3	0	1	0.12	0.028	0.86	−1.33	0.07	−0.09	5.78	0.28
19	Household Linen	3	0	1	0.79	0.011	1.05	0.33	2.55	−0.71	2.98	0.17
20	Kitchen and Household Appliances	3	0	1	0.08	0.011	1.23	−0.01	0.73	−0.01	1.88	0.05
21	China, Glassware and Tableware	3	0	1	0.23	0.006	1.29	−0.66	−0.99	−0.15	1.33	−0.23
22	Other Non-Durables and Services	0	0	1	−0.03	0.011	0.13	−2.10	0.18	−0.13	6.89	0.83
23	Domestic Services	0	0	1	0.24	0.025	0.38	−1.97	0.03	−0.40	4.76	0.51
24	Drug Preparation and Sundries	0	0	0	0.00	0.022	1.64	−0.63	4.21	0.00	2.04	−0.31
25	Orthopedic Equipment	0	0	0	0.00	0.003	0.77	−0.59	1.86	0.00	1.75	0.30
26	Physicians, Dentists, Other Medical Professionals	4	0	1	0.13	0.025	0.76	−0.56	2.02	−0.23	1.67	−0.05
27	Hospitals, Nursing Homes	4	0	1	0.22	0.013	0.25	−1.55	2.01	−0.25	2.20	0.13
28	Vehicles	5	2	1	0.80	0.039	1.11	−0.20	0.00	−0.93	4.63	0.06
29	Operation of Motor Vehicles	5	2	1	−0.20	0.051	0.03	0.29	2.57	−0.01	0.71	−0.07
30	Public Transportation	5	0	1	0.09	0.017	0.04	1.80	2.66	−0.33	0.65	−0.31
31	Communications	0	0	1	2.36	0.012	0.01	−1.60	0.24	−2.47	5.26	0.52
32	TV, Radio, Musical Instruments	3	0	1	0.10	0.039	0.91	−0.07	3.36	−0.10	1.60	−0.28
33	Books, Magazines and Newspapers	0	0	1	0.10	0.016	0.64	−0.44	0.65	−0.26	1.67	0.34
34	Textbooks	0	0	1	0.05	0.008	0.32	0.00	2.07	−0.21	1.13	−0.09
35	Recreational Services	0	0	1	0.11	0.024	0.23	−0.32	2.58	−0.27	1.91	−0.06
36	Personal Care	0	0	1	0.09	0.030	0.66	0.16	1.93	−0.25	1.03	0.31
37	Hotels & Motels, Restaurants	0	0	1	0.10	0.097	0.27	1.34	2.33	−0.25	1.27	−0.07
38	Other Goods	0	0	1	0.55	0.030	0.78	−0.86	1.74	−0.68	1.51	−0.29
39	Financial Services and Insurance	0	0	1	0.75	0.005	0.16	−1.69	0.66	−0.91	17.37	0.91
40	Other Services	0	0	1	−0.04	0.007	0.42	−3.66	0.16	−0.12	5.18	0.45

Note: The abbreviations in the heading are: G and S—Group and Subgroup numbers; I—whether or not a sector is included in the price sensitive group (1 = yes); Share—budget shares; C*El—elasticity with respect to C*; PrEl—price elasticities; DC*—the coefficient on the change in C* divided by the C* coefficient; time %—the annual change due to the time trend expressed as a percentage of the base year; Err%—standard error of estimate as a percentage of the base year value; rho—autocorrelation coefficient of the residuals.

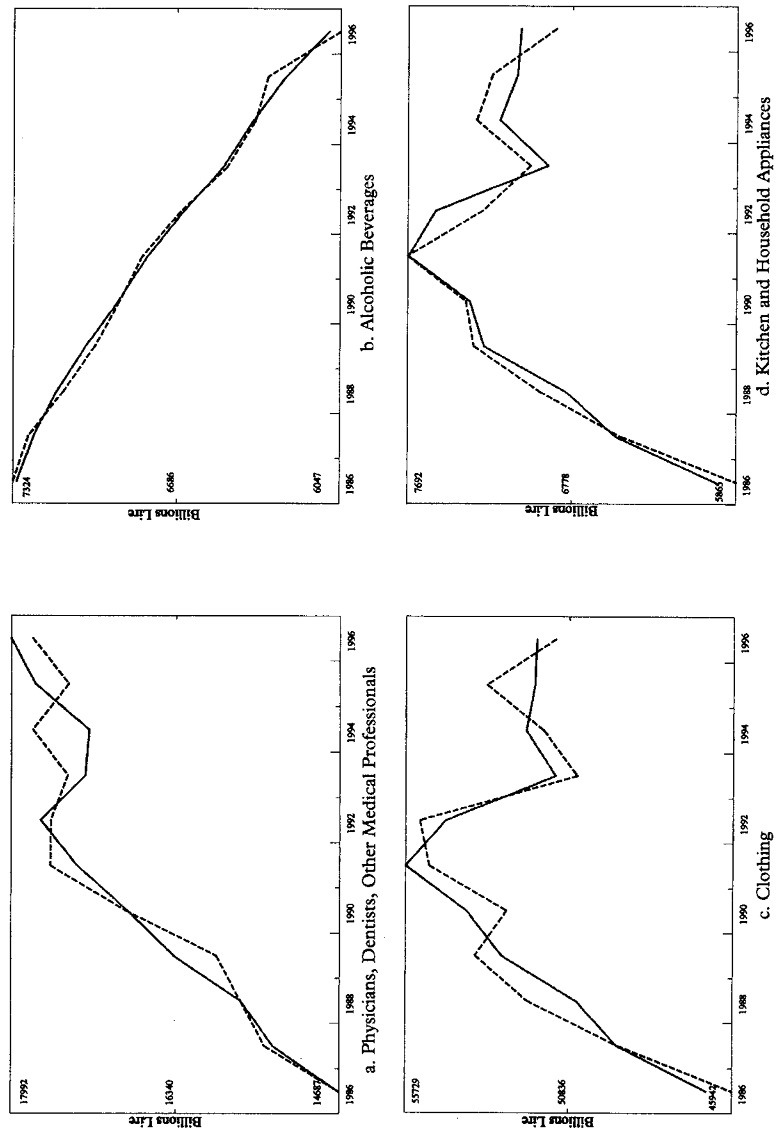

Figure 2. Plots of the time-series regressions for some items. —— predicted values, - - - - - historical data

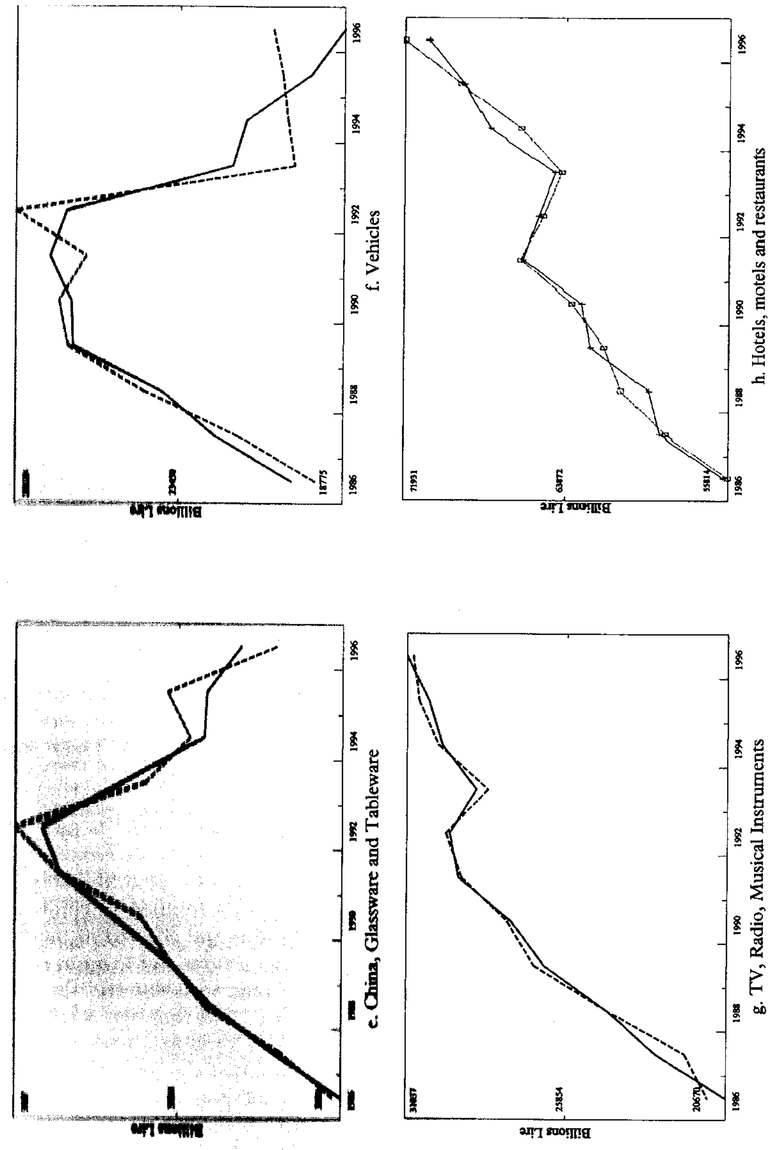

Figure 2. *Continued.* —— predicted values, - - - - - historical data

Group 4: Medical expenditures; and Group 5: Transportation. There are also two subgroups: (1) Protein-rich food, and (2) Vehicles and operation. The categories included in these groups and the ungrouped products can be checked from Table 4, in the columns labelled G and S.

Looking at the first two lines in the table showing the values of the μs and νs, we may observe that the Food group turned out to be complementary while the protein sources were weak substitutes. The interaction inside the groups is in the direction of complementarity also for Clothing and Shoes, House Furnishing and operation, and the medical sectors. Public and Private transportation turned out to be substitutes while buying cars and operating them shows little complementarity.

As to the sectoral results, we may observe that the Cereals and Bakery products are not very sensitive to price and income changes, although they have a relevant budget share if compared with other countries: they may be seen as a necessary part of the Italian diet. On the other hand the expenditure on Fish show a higher income elasticity, as expected, and a lower share. Three food categories—Fats and Oils, Alcoholic Beverages and Non-Alcoholic Beverages—show a very strong time trend in tastes, which has not been removed. The consumption of fats and alcohol is sharply decreasing as a change in habits and, at the same time, all Italians drink more sodas and juice fruits—as we can see from Figure 2(b).

The categories of Household Appliances, China, Glassware and Tableware are in the same group as complements and show the common feature of high income elasticities, low price elasticities but opposite values on the time trend. As for the Transportation group, purchases of vehicles have strong income and price elasticity, while the operation of vehicles has very low income and price elasticities but a strong time trend: this case leaves room for further inspection to improve the equation by limiting the effect of the time trend. This is also the case of other equations where additional analysis is required both in terms of equation form an din terms of soft constraints (item 24—Drug Preparation and Sundries; item 32—TV, Radio and musical instruments, item 37—Hotels, motels and restaurants). In some cases, the data are to blame. Those sectors including goods not elsewhere classified do not have a constant coverage over time and therefore do not respond properly to the variables included in their equations as for Other Non Durables, Financial services and Other services.

8. Conclusions

In this work, a long-run model of personal consumption is constructed. It incorporates both cross-section and time-series data and proposes a solution to combine these two levels of analysis. This model is extremely flexible and allowed us, in the first step, to treat appropriately the problem of zero expenditures and to appreciate the relevance of the income distribution, the demographic variables and the age composition of the household in the consumption patterns. The time-series system, embodying the cross-section results and the time-series variables, enriched the model and confirmed that the basic form of the equation had a good performance for most commodities. In our opinion this analysis of personal consumption expenditures can be improved in several directions. Some of them are very straightforward, in the light of the preceding comments. First of all, commodity groups and subgroups could be revised looking further into the classification of official statistics. For example, in our data, the category of Rent includes not only rental payments but also the consumption of water in the house; carpets are included with furniture; boats along with pets, flowers and toys are included in the category of TV sets and musical instruments. It is obvious that this part of the work will require close cooperation with the Italian Statistical Office. Further developments concern the construction of a demographic model to project forecasts of demographic variables and the study of the cohort effects on personal consumption behaviour. In fact, cohort analysis might be used to get a better measure of the adult equivalency weights. Both issues have been preliminarily investigated in Bardazzi (2000). Furthermore, we should customize the basic equations, introducing additional variables for specific commodities. For instance, special features might be incorporated to deal with durability of certain goods in order to study their speed of adjustment. In future work, we intend to investigate these aspects of the model in order to include these features and improve its interpretation of personal consumption behaviour.

Notes

1. This argument has recently been stressed by Attanasio (1999) in his survey of the most recent contributions to the study of consumption behaviour.
2. A full theoretical description of the time-series demand system is in Almon (1996). This paper extends a previous version of the system suggested by the author (Almon, 1979).
3. The assumption of a representative agent implies that all consumers have identical preferences and identical income. What is wrong with the representative agent approach is well documented in a vast and growing literature surveyed by Kirman (1992). We agree with Stoker (1993) in underlining that this approach ignores the evidence of individual differences in economic behaviour and 'no realistic conditions are known which provide a conceptual foundation for ignoring compositional heterogeneity in aggregate data, let along a foundation for the practice of forcing aggregate data patterns to fit the restrictions of an individual optimization problem' (Stoker, 1993, p. 1829). The constraints derived from the microeconomic consumer theory are also valid for aggregate demand functions if we assume that each consumer has identical preferences and these preferences give rise to individual cost functions that are members of the 'Price Independent Generalized Linear' (PIGLOG) family of cost functions.
4. In fact, the estimation of a demand system with a large group of commodities can be difficult since the number of price parameters without some form of symmetry equals the square of the number of commodities in the system. For example, a demand system with 40 goods (N) and no symmetry would have 1600 price parameters (N^2). Assuming symmetry of some type reduces the price parameters to 860 (that is, $N + (N^2 + N)/2$).
5. The Almon system applied to US data has shown that the size of differences is typically in the range of 2–3% of total expenditures (Almon, 1979; Janoska, 1994). Our study on Italian data has estimated the magnitude of the scaling in the same range.
6. The Italian multisectoral model is member of the INFORUM project (Interindustry FORecasting project University of Maryland). For a description of the INFORUM system, see Almon (1991) and Grassini (2001).
7. Dowd *et al.* (1998) have estimated this cross-section/time-series demand system for the US using the original time-series function by Almon (1979).
8. These differences between sample data and National Accounts are not specific for Italian data. Similar features are shared by British and US data as described by Attanasio (1999).
9. For a detailed description of the different data sources see Bardazzi (2000).
10. For a detailed description of the cross-section consumption function, see Bardazzi & Barnabani (1998a).
11. Application of splines for estimating consumer demand can be found in Diewert & Wales (1992, 1993).
12. With our cross-section data, it is not difficult to construct these interaction effects between the demographic variables but the numbers of parameters in this case would rise sharply and managing the equations system would become very cumbersome.
13. A full theoretical description of the time-series demand system is given in Almon (1996), which extends a previous version of the system suggested by the author (Almon, 1979).

14. It is important to say that, instead of just a linear time trend, a different variable could be used if more appropriate to explain the behaviour of a particular item: the percentage of smokers could explain tobacco expenditure but, for other commodities, the tax treatment, the credit conditions or some other specific variable could be helpful to capture systematic changes in demand that could not be attributed to price, income, age or demographics.
15. For an application of this modelling strategy see Jorgenson *et al.* (1988).
16. The average value of the 0-1 dummy demographic variable is the number of individuals in the category divided by the total number of individuals, that is, the population proportion.
17. In January 2000, the year 1996 was the last one released by the Italian Institute of Statistics.
18. We must stress that the household expenditures are deflated at the sectoral level by the sectoral consumption deflators (base year 1988) available from official statistics. The base year is linked to the intersectoral table for Italy used by the multisectoral model. The deflation is necessary because the parameters of the cross-section estimation are used to compute the C_i^* of equation (8) which is a proxy of income in real terms in the time-series equation.
19. The reweighting procedure is directly provided by Istat.
20. For an analytical description of the non linear probability model, see Bardazzi & Barnabani (1998b).
21. We refer to the consumption categories specified in Table 1. The figures in the parentheses are the *t*-statistics. The level of significance has been assumed 0.05.
22. 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—say the North east—and the coefficient of the base—the Central Region.

References

Almon, C. (1979) A system of consumption functions and its estimation for Belgium, *Southern Economic Journal*, 46, pp. 85–106.

Almon, C. (1991) The INFORUM approach to interindustry modeling, *Economic Systems Research*, 3, pp. 1–7.

Almon, C. (1996) A Perhaps Adequate Demand System, *INFORUM Working Paper*, Series 96(7), University of Maryland.

Attanasio, O. (1999) Consumption, in: J. B. Taylor & M. Woodford (eds) *Handbook of Macroeconomics* (Amsterdam, Elsevier Science), pp. 741–812.

Bardazzi, R. (2000) Effects of the changing structure of population on Italian household consumption, paper presented at the *XIII International Conference on Input–Output Techniques*, University of Macerata (http://policy.rutgers.edu/cupr/iioa/iioa5.htm).

Bardazzi, R. & Barnabani, M. (1998a) Modelling zero expenditures on Italian household consumption, *Economic Notes*, 27, pp. 55–96.

Bardazzi, R. & Barnabani, M. (1998b) A model of affordable and desired household consumption, mimeo.

Cooper, R. J. & McLaren, K. R. (1992) An empirically oriented demand system with improved regularity properties, *Canadian Journal of Economics*, 25, pp. 652–668.

Deaton, A. (1986) Demand analysis, in: Z. Griliches & M. D. Intriligator (eds) *Handbook of Econometrics* (Amsterdam, North Holland), pp. 1767–1839.

Deaton, A. & Muellbauer, J. (1980) An almost ideal demand system, *American Economic Review*, 70, pp. 312–326.

Diewert, W. E. & Wales, T. J. (1992) Quadratic spline models for producer's supply and demand functions, *International Economic Review*, 33, pp. 705–722.

Diewert, W. E. & Wales, T. J. (1993) Linear and quadratic spline models for consumer demand functions, *Canadian Journal of Economics*, 26, pp. 77–106.

Dowd, T. A., Monaco, R. M. & Janoska, J. J. (1998) Effects of future demographic changes on the US economy: evidence from a long-term simulation model, *Economic Systems Research*, 10, pp. 239–262.

Fletcher, R. (1987) *Practical Methods of Optimization* (Chichester, Wiley).

Grassini, M. (1983) A system of demand equation for medium-to-long term forecasting with input-output econometric models, *Economic Notes*, 12, pp. 84–96.

Grassini, M. (2001) The Core of the Multisectoral INFORUM Model, in: Grassini, M (ed.) *Contributions on Multisectoral Modelling* (Florence, Centro Editoriale Toscano), pp. 7–32.

Janoska, J. (1996) Modeling personal consumption and government transfers in a long-run macro-econometric forecasting model. Unpublished PhD Dissertation, University of Maryland, College Park.

Jorgenson, D. W., Slesnick, D. T. & Stoker, T. M. (1988) Two-stage budgeting and exact aggregation, *Journal of Business and Economic Statistics*, 6, pp. 313–325.

Kirman, J. A. P. (1992) Whom or what does the representative individual represent, *Journal of Economic Perspectives*, 6, pp. 117–136.

Muellbauer, J. (1975) Aggregation, income distribution and consumer demand, *Review of Economic Studies*, 42, pp. 525–543.

Muellbauer, J. (1976) Community preferences and the representative consumer, *Econometrica*, 44, pp. 979–999.

Stoker, T. M. (1993) Empirical approaches to the problem of aggregation over individuals, *Journal of Economic Literature*, 31, pp. 1827–1874.

Theil, H. & Goldberger, A. S. (1960) On pure and mixed statistical estimation in Economic, *International Economic Review*, 2, pp. 65–78.