DEMAND SYSTEM CHOICE BASED ON TESTING THE ENGEL CURVE SPECIFICATION

by

CLAUDIO SOREGAROLI*
(claudio.soregaroli@cr.unicatt.it)

KAREN HUFF**
(KHUFF@KW.IGS.NET)

KARL MEILKE**
(MEILKE@AGEC.UOGUELPH.CA)

WORKING PAPER 02/09

Department of Agricultural Economics and Business
University of Guelph
Guelph, Ontario

March 2002

*Istituto di Economia Agro-alimentare, Università Cattolica del Sacro Cuore, Piacenza, Italy

**Department of Agricultural Economics and Business, University of Guelph, Guelph, Ontario, Canada

ABSTRACT

It is common to use a demand systems approach in estimating the key parameters from household consumption data. In conducting these studies the researcher is faced with selecting a functional form. In turn, each functional form implies a particular shape for the Engel curves. This analysis highlights the importance of testing the shape of Engel curves, especially if the researcher is interested in elasticity estimates well away from the sample mean. Using consumption data for selected households in Italy it is shown that many popular functional forms are rejected by the data.

JEL classification: *C52, B12*

Keywords: *demand, Engel curves, consumption*

Demand System Choice Based on Testing the Engel Curve Specification

1. *Introduction*

Given the dramatic reduction in the cost of computation over the past several decades, it is now commonplace for empirical studies of household consumption patterns to employ a demand systems approach, such as the Almost Ideal (AI) demand system (Deaton and Muellbauer, 1980a), the Linear Expenditure System (Stone, 1954), the Rotterdam model (Theil, 1965), or the Translog model (Christensen *et al*., 1975). While the purpose of many of these studies was to provide numerical measures of theoretical parameters like demand elasticities, many of the studies also conduct tests of the underlying economic theory. However, this "falsificationist" approach to demand systems research is often not applied in a consistent manner given that researchers frequently test only a subset of the refutable hypotheses of demand theory (Cozzarin and Gilmour, 1998). Knowing that the objective of most empirical studies is to produce elasticity estimates, a more practical test – but one that is rarely considered – would be to test the suitability of the Engel curves underlying the chosen demand system, given the data employed.

The behaviour of the income elasticities of a given demand system is determined by its underlying Engel function, and an Engel curve that better represents the observed expenditure patterns at different income levels will reduce the bias in the computed values of the elasticities. The problem of choosing a demand system based on the shape of its underlying Engel curves has been considered previously by Aasness and Rodseth (1983) using Norwegian household data, and by Banks *et al.* (1997) using U.K. data. However, most studies ignore the problem or refer to the empirical evidence, on the shape of Engel curves, available in the literature. Hence, the suitability of the Engel curves arising from an applied demand system is often not tested against real data.

This problem also holds for studies that analyse household consumption in Italy. Most of the studies use well-known demand systems like the AI demand system (Patrizii and Rossi, 1991; Moschini and Rizzi, 1997) without considering the implicit assumptions being made regarding the shape of the underlying Engel curves. The purpose of this study is to test the shape of the Engel curves underlying a number of popular applied demand systems using Italian household expenditure data, and to demonstrate that the choice of demand system and its underlying Engel specification have important implications for the estimated expenditure elasticities.

2. *Model Selection*

Numerous studies have proposed different functional forms for Engel curves. The concern of the early literature was to provide a set of equations that fit the data well, rather than satisfying the underlying economic assumptions. A classic example is the work by Prais and Houthakker (1955). As Deaton argues (1986, p.1799): "the Prais-Houthakker methodology is unashamedly pragmatic, choosing functional forms on the grounds of fit, with an attempt to classify particular forms as suitable for particular types of goods". Indeed, only one of the functional forms proposed by Prais and Houthakker is consistent with economic theory in that it satisfies the budget constraint.

Besides goodness-of-fit considerations, the theoretical plausibility of a model is an important choice criterion for Engel curves. Gorman (1981) characterises the possible set of Engel curves, derived from a general Engel form, which satisfy economic theory, while Deaton and Muellbauer (1980b) focus on the aggregation problem and the related properties of Engel curves.[1]

Empirical studies that focus on selecting functional forms for systems of demand equations can be divided into two main categories based on the use of either parametric or non-parametric regression techniques. Parametric studies include Prais and Houthakker (1955), Leser (1963), Bewley (1982), and Giles and Hampton (1985). All of these studies use non-nested tests to select among specific functional forms. Other parametric studies such as Aasness and Rodseth (1983), Laitinen *et al.* (1983), De Witte and Cramer (1986), and Haque (1988) use nested tests such as the likelihood ratio test to select the best functional form. The non-parametric approach includes more recent studies that test parametric Engel functions against nonparametric alternatives including Delgado and Miles

---

[1] Characterization theorems define and delimit what forms of models can be devised (Lewbel, 1987).

(1997), and Blundell and Duncan (1998), or that test for the rank of the equation system (Banks *et al.*, 1997; Perali, 1999).[2]

To compare the shapes of Engel curves with the underlying assumptions of an applied demand system, it is important that they are evaluated as a system of equations. Non-parametric studies are limited to testing the functional form for single equations, for a given expenditure class, or of providing general information on the system rank. As such, this approach does not allow for performance-based comparisons among a system of specific functional forms of Engel curves.

Only a few studies have estimated Engel curves, as a system, in a parametric framework (Bewley, 1982; Aasness and Rodseth, 1983; Giles and Hampton, 1985). The model applied in this study is based on Aasness and Rodseth (1983). This model is of particular interest since it nests and therefore allows testing most of the Engel curves that underlie commonly applied demand systems. The selection framework includes the general model

$$w_i = \frac{a_i + b_i y^{(-\lambda)} + \rho c_i y^{(\lambda)} y^{(-\lambda)}}{1 + d y^{(-\lambda)}}, \tag{1}$$

where $y^{(\lambda)}$ is the Box-Cox transformation of total expenditure y:

$y^{(\lambda)} = (y^{\lambda} - 1)/\lambda$ for $\lambda \neq 0$,

and $y^{(\lambda)} = \ln y$ for $\lambda = 0$.

The parameters a_i, b_i, and c_i are equation specific, where the subscript i represents the different classes of goods, and d and λ have the same value across all of the equations in the system. The symbol ρ represents a dichotomous (0, 1) variable that helps in specifying the various “lower-level” models in the framework. Finally, function (1) satisfies the adding-up requirement if

$$\sum_{i=1}^{n} a_i = 1, \sum_{i=1}^{n} b_i = d, \text{ and } \sum_{i=1}^{n} c_i = 0.$$

---

[2] The definition of rank of Engel curves derives from Gorman (1981). Gorman characterized Engel curves of the general form $w_i = \sum_{r \in R} a_{ir}(p)\phi_r(\ln y)$, where R is a finite set, i is a given good, and $\phi_r(.)$ is a series of functions in the natural log of income. He concluded that the rank of the Engel function, defined by the rank of the matrix of coefficients $a_{ir}(p)$, is consistent with the theory when $a_{ir}(p)$ is at most of rank three.

Figure 1 shows the hierarchical order among the various Box-Cox functional forms. The dotted lines in the figure mean that two specifications are related by other properties, such as symmetry conditions. When ρ is equal to zero, the Engel curves belong to the class of Generalised Quadratic Indirect Utility functions which includes as subclasses Engel curves from well known demand systems such as the Translog model (Christensen *et al*., 1975) and the Generalised Leontief (Diewert, 1971). For $d = 0$ the Generalised Quadratic Expenditure class of Engel curves is derived. This class can be considered a special case of Gorman's (1981) forms. For example, the Quadratic Expenditure curve is obtained by setting $\lambda = 1$ and the Quadratic-Log form by setting $\lambda = 0$. Moreover, the Quadratic Expenditure curve corresponds to the Linear Expenditure function when ρ is set equal to zero. When ρ and d are both equal to zero the Engel curves are of the price independent generalized linear (PIGL) form. The PIGLOG or Working-Leser form is a special case of the PIGL specification that occurs when λ is set equal to zero.

When dealing with a multi-commodity demand system, the Engel curves are estimated as systems of equations with the value of the power coefficient λ constrained to be equal across all of the equations included in the system. This ensures that all of the Engel curves for a given system will have the same functional form. In testing the appropriateness of the various Engel curves, for the Italian household data employed in this study, each of the various Box-Cox functional forms presented in figure 1 are first tested using a multi-commodity systems approach. The legitimacy of forcing the value of λ be the same in all of the equations is then tested by obtaining single equation estimates of λ for each commodity and given functional form.

3. *Data and Estimation Issues*

The data employed in this study come from the 1995 Family Expenditure Survey of the Italian Institute of Statistics (ISTAT). The database includes information about household characteristics and their expenditures on different categories of goods. The number of households surveyed each month is about 3,250. Each family records its expenditures for ten days of a given month and participates in the survey only once making it impossible to track households over time. Also, no price information is provided in the dataset.

For the sample employed in this study, observations were selected from families with two adults, whose head is employed, that live in the North of Italy, and that recorded expenditures in the first quarter of 1995. While it is possible to further divide the sample into more homogeneous households, the number of observations would decrease dramatically. From the original data, the total number of households conforming to the above characteristics is 406. Other sample choices were possible, but the actual choice appears to strike an appropriate balance between homogeneity of preferences and number of observations. Homogeneity of preferences is particularly important for the empirical tests, since the model does not account for all of the differences in purchasing behaviour resulting from demographics. This choice was made to better evaluate the shape of the functional form, without any influence from the way other variables, such as demographic effects, are included in the specification.

Care must be taken in estimating Aasness and Rodseth's model. The use of different starting values for the parameters helps to ensure that the maximum likelihood method finds the global maximum. Also the tree shape of the model (Figure 1) is useful in checking for and avoiding local maximums. For example, if the value of the likelihood function for the general equation is lower than the values of the restricted equations it means that its estimate did not converge to the global maximum.

Another problem in estimating the general model (1) is that the value of the denominator must not equal zero or the maximum likelihood value will not make any sense. A practical solution is to restrict the domain of d and λ in such a way that, for a given sample, the denominator is not zero in the interval between the two extreme observations of y (Aasness and Rodseth, 1983). Appendix 1 discusses in more detail the empirical model and some additional estimation issues.

When using data from budget surveys two econometric problems are of particular importance, namely, the errors-in-variables problem and the presence of zero expenditures.

The errors-in-variables problem arises when an independent variable is correlated with the error term of the estimated equation leading to inconsistent estimates of the regression parameters. The problem can be caused either by errors in measurement or the endogeneity of an independent variable.

Econometricians have proposed different solutions to the errors-in-variable problem for the linear regression model. A popular method is the instrumental variables approach (IV). However, this method has two important problems in this context. First, cross-section household data does not provide "good" instruments given that “lagged” observations of current income are not available. Second, the application of IV methods to non-linear models might lead to inconsistent estimates (Hausman *et al*., 1995).

The errors-in-variables problem has frequently been neglected in model selection studies. Aasness and Rodseth (1983) recognise the problem, but select Engel curves under the assumption that total expenditure is independent of the error term. Banks *et al*. (1997) find that for their data set, the dependence between the independent variable and the Engel curve estimation errors are not responsible for the observed curvature in the Engel curves. In light of this result, and the data and theoretical constraints on the application of IV methods in model (1), this study treats total expenditure as an independent variable in the selection framework.

Besides the possible presence of errors-of-measurement it is often the case that family survey data contains zero expenditures for some classes of goods. The reasons for the presence of zeros can be divided into two categories: 1) infrequency of purchase, and 2) corner solutions. In the first case the family consumes the commodity but does not buy it during the survey period. This is especially true for durable goods or when the survey period is relatively short. In the second case, the good is not consumed because of household preferences and/or market conditions.

If the only reason for zero observations is the infrequency of purchase, the problem can be modeled by confining attention only to those households with positive observations for all of the aggregates (Deaton, 1987), or by the use of IV methods (Keen, 1986). A more complex situation is when market conditions or household preferences influence the consumption decision. In this case, complex estimation techniques (Lee and Pitt, 1986), or two-step models (Shonkwiler and Yen, 1999) are required.

The goods included in this study are grouped into broad classes of expenditure such as food, housing, and recreation. Since the level of aggregation is fairly high and expenditures are recorded monthly, it is reasonable to assume that all households consume some goods in each class. It is therefore concluded that the presence of zero expenditures is the result of the short period of the survey.[3] As such, and to avoid further complications in the model that could influence the selection process, only those observations with positive expenditures in all of the categories are employed in the study. This choice is supported by the results discussed in Appendix 2 that show the distribution of zero consumption across products, and the results of a Probit analysis used to evaluate sample selection bias. The final sample used for estimation contains 354 observations.[4]

4. ***Results***

The Engel functional forms represented in figure 1 are estimated as a system of equations including the following expenditure aggregates: 1) food; 2) beverages and tobacco; 3) housing; 4) transportation; 5) recreation; and 6) other goods and services. The choice of the commodity aggregates is driven by the fact that most of the empirical work on Italian consumption uses a similar aggregation (Patrizii and Rossi, 1991; Moschini and Rizzi, 1997). This makes it possible to compare the results of this study with those obtained previously.

Table 1 presents descriptive statistics for the above classes of expenditure. The class with the largest expenditure share is housing. The expenditure shares are asymmetrically distributed with positive skewness. The range of values for the measured expenditure shares is quite large.

Table 2 presents the estimates obtained for the parameters, d and λ, as well as the value of the log-likelihood function for all of the Engel equations in figure 1. However, before proceeding with the

---

[3] The Italian consumption data report the values for monthly expenditures, but it is computed from a survey period of ten days.

[4] From the original sample the following observations were deleted: 1) those households recording values of total expenditure more than five standard deviations above the mean, which removed six observations, and 2) households reporting zero consumption for at least one aggregate, a total of 46 observations.

Likelihood Ratio (LR) test, it is important to explain how the test is performed. As emphasised by Aasness and Rodseth (1983) there are two methods for carrying out the LR tests.

The first method follows the structure of Figure 1. From the General model (1), downward conditional tests are made at different levels. For example, the Generalised Quadratic Expenditure model (3) is tested against the General model (1). If (3) is rejected, it means that all of its special cases are rejected. If, on the other hand, the test fails to reject model (3), its special cases (4), (5), and (6) can be rejected or accepted only after each is tested against function (3). Note that if at the top-level test - (3) against (1) - the chosen level of significance is 0.01, then the tests at the second level - (4), (5), or (6) against (3) - have a different significance level. This will be 0.01 times the number of tests that can reject the particular lower level hypothesis (Aasness and Rodseth, 1983). For example, the test of the Quadratic-Log specification (4) against (3) has a significance level of 0.04 (0.01*4).

The second method of conducting the LR test is to test each special case directly against the General model (1). If the chosen significance level is 0.01, it will be the same for all of the tested hypotheses. These two methods of testing are not alternatives, but complements that might lead to slightly different results.

The results of each type of LR test are summarised in table 3. The first column presents the results of the direct test of each model against the General Box-Cox formulation (1). The abbreviation NR represents the tests that fail to reject the null hypothesis at a significance level of 0.01. Based on this direct test all of the simplest Engel equations – models (11) through (16) – are rejected. Moreover, the direct LR test rejects the Generalised Quadratic Indirect Utility (7) and its restricted equations (8), (9), and (10). All of the rejected functional forms have a common characteristic: they do not allow for a higher order polynomial term for total expenditure ($c_i = 0$). The direct test results fail to reject the Expanded Translog (2), the Generalised Quadratic Expenditure function (3), and each of its special cases (4), (5), and (6).

The remaining columns of table 3 present the results of the first method of conducting the LR test in which the special cases of each model are tested in a downward fashion. Looking only at models (4), (5) and (6), which were not rejected under the direct test, the results of the downward conditional

LR tests indicate that the both the Quadratic Expenditure Engel curve (6) and model (5) fail to be rejected against model (3). The Quadratic-Log (4) Engel curve fails the conditional test against the Generalised Quadratic expenditure (3). However, it is not rejected by the conditional test against model (2). Finally, rank-two models (12) to (16) are all rejected by the conditional tests against rank-three models (4) to (6). Hence, the testing illustrates the importance of rank-three specifications, even if among these models there is no clear evidence regarding which Engel curve should be used based on the data employed.[5]

This last result indicates that the value of the parameter λ is not particularly important for model choice, since its value varies from 0 to 1 across equations (4), (5), and (6). However, this could be influenced by the commodity aggregation. For example, if each commodity requires a different value of λ, it would be difficult to find a single, significant value for it at the system level. Since the value of λ can influence the results for the single commodities, this issue is investigated in the section presenting the single equation estimates. In this way it is possible to determine whether its value differs significantly across commodities for a given functional form.

4.1 Implications for the Estimated Expenditure Elasticities

It is time to consider whether the differences between the selected and the rejected models have important implications for the values of the expenditure elasticities and whether there is any difference in the behaviour of elasticities among the selected models.

Table 4 presents the values of the expenditure elasticities for a subset of the estimated Engel curves. The elasticities are computed at the average total expenditure of the sample. The first five columns present expenditure elasticities for the General model (1), the Generalised quadratic expenditure model (3), and its special cases (4), (5), and (6). The last two columns present elasticity

---

[5] One possible way to evaluate the potential bias in the selection of functional form using the adopted selection framework is to conduct a simulation study. Generating data from a specified functional form and then following the Aasness and Rodseth selection process would allow for an evaluation of the likelihood of choosing a wrong functional form. However, the ease of finding the correct functional form would be highly dependent on their assumed shapes. Engel curves with lots of curvature would be easier to discover than more linear forms that might be approximated by many potential functions.

estimates for two of the rejected specifications: the Linear Expenditure function (12) and the PIGLOG form (14).

There is little difference in the elasticity estimates obtained from the various models, and the greatest variability occurs for food; and beverages and tobacco. Moreover, the ranking of the expenditure groups, from the most to the least elastic, is preserved across models. Ranking product categories from the highest to the lowest elasticities results in the following list: transportation; other goods and services; recreation; housing; beverages and tobacco; and food.

Among the selected Engel models, the Quadratic Log (4) and Quadratic Expenditure (6) specifications correspond, respectively, to the commonly used demand systems QUAIDS of Banks *et al*. (1997) and Quadratic Expenditure (QES) of Pollak and Wales (1978). Their ability to provide expenditure elasticity estimates over a range of expenditures can be compared, together with the PIGLOG (14) function corresponding to the widely used AI demand system of Deaton and Muellbauer (1980). When these models are compared, elasticities computed from the PIGLOG model do not always provide estimates similar to the other two models. For food; beverages and tobacco; housing; and other goods and services, the elasticity estimates of all three models behave in a similar fashion as the level of expenditure increases. However, for transportation, and recreation the PIGLOG (14) model gives elasticity estimates at expenditures away from the mean that are quite different from the Quadratic Log (4) and Quadratic Expenditure (6) models. In particular, Figure 2 shows the behaviour of the expenditure elasticity for transportation under each of the three models as the level of total expenditures varies. The elasticity estimates of the rejected PIGLOG model (14) vary significantly from those obtained with models (4) and (6) at expenditure levels above and below the mean.

Given that the underlying Engel curves of the popular AI Demand System are of the PIGLOG form, its questionable performance in providing elasticity estimates over a range of expenditures for this data set is of interest. It is clear that the choice of the underlying Engel model has important consequences for inferences made far away from the average level of total expenditure. The Quadratic Log (4) and Quadratic Expenditure (6) models, which were not rejected against the general model (1) provide estimated expenditure elasticities that behave similarly at various expenditure levels. Demand system choice based on these Engel formulations should result in more reliable elasticity estimates.

4.2 Single equation estimates of Engel curves

In this section the various Box-Cox functional forms are tested for each commodity. This gives some indication of the expenditure classes that determine the overall choice of an appropriate Engel system. Moreover, since a common value of λ for the equations in any specific system is a maintained hypothesis, it is of interest to test this hypothesis using single equation estimates.

Table 5 shows the results of the individual commodity likelihood ratio tests. From the LR tests it is evident that transportation and recreation are the two expenditure classes driving the choice of the Engel model in the system estimation. Beverages and tobacco; housing; and other goods and services fail to reject any of the specifications. The food commodity rejects the Linear Expenditure function (12) as well as models (15) and (16). Finally, note that models (4), (5), and (6) are the only specifications that are not rejected by any of the commodities. These models also were not rejected under the direct test for the maximum likelihood system estimation implying that the choice of λ is not influenced by the commodity grouping given its value varies from 0 to 1 across these particular models.

4.3 Comparison with Other Studies

The results of this study suggest that for the available data set, the Engel curves perform better when a higher order polynomial term for total expenditure is included. This means that the Engel curves are curvilinear when the dependent variable is the expenditure share (rank-three Engel curves). This is especially true for the expenditure classes transportation, and recreation.

Comparing these results with those obtained by Aasness and Rodseth (1983) is of particular interest since their model is applied in this study. Moreover, they used classes of expenditure similar to those considered here. Their results from a Norwegian family survey are quite similar to those obtained in this study for Italy. The data reject the Generalised Quadratic Indirect Utility formulation (7) and all of its special cases – (8), (9), (10), and (11). Also the Linear Expenditure (12) and the PIGLOG (14) Engel curves are rejected. The evidence favours the Generalised Quadratic Expenditure formulation (3) and its restricted equations, the Quadratic-Log (4) and Quadratic Expenditure (6) Engel curves. These are the same conclusions reached using the available data for Italy. Hence, for an aggregation featuring non-food commodities, the need for a rank-three Engel model is confirmed. Aasness and Rodseth's single equation results also suggest the need for a rank-three Engel curve for four expenditure classes: food; transportation; recreation and education; and other goods and services. There is a clear similarity with the results obtained in this study for transportation, and recreation.

Turning to two recent applied demand studies employing Italian household data, Patrizii and Rossi (1991) and Moschini and Rizzi (1997) both use demand systems based on the AI Demand System (Deaton and Muellbauer, 1980a). The Engel curve assumed by this demand system has a PIGLOG functional form. Both of these studies use a commodity aggregation similar to this study that includes categories such as food, housing, and transportation. Moschini and Rizzi (1997) also include an expenditure category for recreation. The results obtained in this study suggest that, for these expenditure categories, the AI demand system model may not be consistent with the observed expenditure patterns, and a rank-three demand system should be considered. Hence, a preliminary evaluation of Engel curves for any data set would help in selecting the appropriate model.

As highlighted earlier, the incorrect choice of an Engel curve is particularly important when the expenditure elasticities are computed far from the average expenditure level. Indeed, for low and high expenditures the values of the expenditure elasticities for transportation and recreation vary considerably depending on the model chosen, and this can have important implications in some types of policy analysis.[6]

5. *Conclusions*

---

[6] Banks *et al.* (1997, p.536) give an example of the biases introduced when using an incorrect Engel specification. They computed the welfare losses from a tax increase (sales tax on clothing) using both the AI demand system and the QUAIDS model. They found that “the AI demand system understates the welfare losses for the majority of the distribution and overstates the welfare losses for the richer households”.

The purpose of this study was to select the set of Engel curves that best represent the consumption behaviour of different commodities in Italy and to highlight the importance of a demand system choice based on the selected curves. The results indicate that for macro-aggregates such as food; transportation; recreation; housing; and beverages and tobacco, a rank-three model should be used. In particular, the evidence favours the Quadratic-Log (4) or the Quadratic Expenditure (6) models. The transportation, and recreation commodities show the most need for rank-three Engel curves.

There are important implications for the behaviour of the expenditure elasticities depending on the choice of an Engel model. While the values of the elasticities are similar across the various specifications when computed at the average expenditure level of the sample, important differences arise at low and high values of total expenditure. Failure to consider the appropriate Engel model when selecting a demand system could introduce important biases in the results. Hence, when applying a particular demand system on a given data set, this problem needs to be considered and preferably tested.

REFERENCES

AASNESS J. and RODSETH A., "Engel Curves and Systems of Demand Functions", *European Economic Review*, 1983, 20, pp. 95-121.

BANKS J., BLUNDELL R., and LEWBEL A., "Quadratic Engel Curves and Consumer Demand", *The Review of Economics and Statistics*, 1997, 79, pp. 527-39.

BEWLEY R.A., "On the Functional Form of Engel Curves: the Australian Household Expenditure Survey 1975-76", *The Economic Record*, 1982, 58, pp. 82-91.

BLUNDELL R. and DUNCAN A., "Kernel Regression in Empirical Microeconomics", *Journal of Human Resources*, 1998, 23, pp. 62-87.

CHRISTENSEN L.R., JORGENSON D.W., and LAU L.J., "Transcendental Logarithmic Utility Functions", *American Economic Review*, 1975, 65, pp. 367-83.

COZZARIN B.P. and GILMOUR B.W., "A Methodological Evaluation of Empirical Demand Systems Research", *Canadian Journal of Agricultural Economics*, 1998, 46, pp. 297-316.

DE WITTE M.A.C. and CRAMER J.S., "Functional Form of Engel Curves for Foodstuffs", *European Economic Review*, 1986, 30, pp. 909-13.

DEATON A., "Demand Analysis", *Handbook of Econometrics*, New York: Elsevier Science Publishers, 1986.

-------, "Estimation of Own- and Cross-Price Elasticities from Household Survey Data", *Journal of Econometrics*, 1987, 36, pp. 7-30.

DEATON A. and MUELLBAUER J., "An Almost Ideal Demand System", *American Economic Review*, 1980a, 70, pp. 312-36.

------ and ------, *Economics and Consumer Behaviour*, Cambridge: Cambridge University Press, 1980b.

DELGADO M.A. and MILES D., "Household Characteristics and Consumption Behaviour: a Nonparametric Approach", *Empirical Economics*, 1997, 22, pp. 409-29.

DIEWERT E.W., "An Application of the Shephard Duality Theorem: a Generalized Leontief Production Function", *Journal of Political Economy*, 1971, 79, pp. 481-507.

GILES D. and HAMPTON P., "An Engel Curve Analysis of Households' Expenditures in New Zealand", *The Economic Record*, 1985, 61, pp. 450-62.

GORMAN W.M., "Some Engel Curves", *Essays in the Theory of Measurement of Consumer Behaviour*, Cambridge: Cambridge University Press, 1981.

HAQUE M.O., "Estimation of Engel Elasticities from the Box-Cox Engel Function", *Metroeconomica*, 1988, 39, 317-35.

HAUSMAN J.A., NEWEY W.K., and POWELL J.L., "Nonlinear Errors-in-Variables: Estimation of some Engel Curves", *Journal of Econometrics*, 1995, 65, pp. 205-33.

KEEN M.J., "Zero expenditures and the estimation of Engel curves", *Journal of Applied Econometrics*, 1986, 1, 277-86.

LAITINEN K., THEIL H., and RAPARLA T., "A Generalization of Working's Model", *Economic Letters*, 1983, 13, pp. 97-100.

LEE L.F. and PITT M.M., "Microeconometric demand systems with binding nonnegativity constraints: the dual approach", *Econometrica*, 1986, 54, 1237-42.

LESER C.E.V., "Forms of Engel Functions", *Econometrica*, 1963, 31, pp. 694-703.

LEWBEL A., “Characterizing some Gorman Engel Curves”, *Econometrica*, 1987, 55, pp. 1451-59.

MOSCHINI G.C. and RIZZI P.L., “La Struttura dei Consumi delle Famiglie in Italia”, *Rivista Internazionale di Scienze Sociali*, 1997, 105, pp. 449-71.

PATRIZII V. and ROSSI N., *Preferenze, Prezzi Relativi e Redistribuzione*, Bologna: Il Mulino, 1991.

F. PERALI, “Stima delle Scale di Equivalenza Utilizzando i Bilanci Familiari ISTAT 1985-1994”, *Rivista Internazionale di Scienze Sociali*, 1999, 107, pp. 481-541.

POLLAK R.A. and WALES T.J, “Estimation of Complete Demand Systems from Household Budget Data”, *American Economic Review*, 1978, 68, pp. 348-59.

PRAIS S.J. and HOUTHAKKER H.S., *The Analysis of Family Budgets*, Cambridge: Cambridge University Press, 1955.

SHONKWILER J.S. and YEN S.T., “Two-step estimation of a censored system of equations”, *American Journal of Agricultural Economics*, 1999, 81, 972-82.

STONE R, “Linear Expenditure System and Demand Analysis: an Application to the British Pattern of Demand”, *Economic Journal*, 1954, 64, pp. 511-27.

THEIL H, “The Information Approach to Demand Analysis”, *Econometrica*, 1965, 33, pp. 67-87.

Table 1: Descriptive statistics for the expenditure shares of the sample and total expenditure (average total expenditure = 1)

	Mean	Std. Dev.	Min	Max	Skewness
Food	0.175	0.084	0.026	0.444	0.70
Beverages & tobacco	0.036	0.032	0.001	0.195	1.66
Housing	0.320	0.134	0.073	0.748	0.71
Transportation	0.158	0.130	0.010	0.799	2.26
Recreation	0.078	0.064	0.002	0.353	1.51
Other goods and services	0.232	0.130	0.015	0.773	0.82
Total expenditure	1	0.522	0.299	3.141	1.70

Table 2: Restrictions, estimated parameters, and value of the log-likelihood function for different systems of Engel curve equations.[a]

System of Engel curves[b]		Restrictions			Estimates		Log-Likelihood
		ρ	d	λ	$\hat{d}$	$\hat{\lambda}$	
1	General Model	-	-	-	0.14	-2.30***	2179.04032
2	Expanded Translog	-	-	0	-0.33		2176.50605
3	Gen. Quadratic Expenditure	-	0	-		-1.75***	2178.77440
4	Quadratic Log	-	0	0			2176.09518
5		-	0	0.5			2176.54770
6	Quadratic Expenditure	-	0	1			2177.63332
7	Gen. Quadratic Indirect Utility	0	-	-	-0.01	0.00	2167.79641
8	Quadratic Indirect Utility	0	-	1	-0.53***		2166.83329
9	Generalised Leontief	0	-	0.5	-0.26		2167.55167
10	Translog	0	-	0	-0.01		2167.79641
11	PIGL	0	0	-		-0.01	2167.79628
12	Linear Expenditure	0	0	1			2163.59925
13		0	0	0.5			2166.64837
14	PIGLOG	0	0	0			2167.79583
15		0	0	-0.5			2166.69427
16		0	0	-1			2163.47101

***Different from zero at the 0.01 level of significance (**= 0.05; *=0.10).

[a] The system of equations is estimated for the following categories of expenditure: 1) food; 2) beverages and tobacco; 3) housing; 4) transportation; 5) recreation; and 6) other goods and services.

[b] The model is indicated by a number corresponding to the Box-Cox functional forms reported in figure 1.

Table 3: Likelihood ratio test for alternative hypotheses of Engel curve systems[a] [b]

	Alternative hypothesis										
	1	2	3	4	5	6	7	8	9	10	11
2	**NR**										
3	**NR**										
4	**NR**	**NR**	R								
5	**NR**		**NR**								
6	**NR**		**NR**								
7	R										
8	R						R				
9	R						R				
10	R	R					R				
11	R		R				R				
12	R					R		R			R
13	R				R				R		R
14	R			R						R	R
15	R				R				R		R
16	R					R		R			R

[a] NR = tests that fail to reject the null hypothesis of the given model being equal to the alternative hypothesis at the appropriate level of significance (R =rejected).

[b] The model is indicated with a number that corresponds to the functional forms reported in Figure 1.

Table 4: Expenditure elasticities computed for selected Engel models at the sample average total expenditure level[ab]

	1	3	4	5	6	12	14
Food	0.617	0.583	0.432	0.445	0.483	0.506	0.438
Beverages and tobacco	0.680	0.653	0.544	0.553	0.580	0.601	0.546
Housing	0.943	0.946	0.912	0.915	0.924	0.937	0.912
Transportation	1.410	1.383	1.513	1.502	1.469	1.305	1.451
Recreation	0.949	0.994	1.001	1.000	0.998	1.089	1.025
Other goods and services	1.143	1.163	1.235	1.228	1.209	1.211	1.248

[a] The model is indicated with a number that corresponds to the functional form reported in Figure 1.

[b] All of the elasticities are different from zero at the 0.01 level of significance.

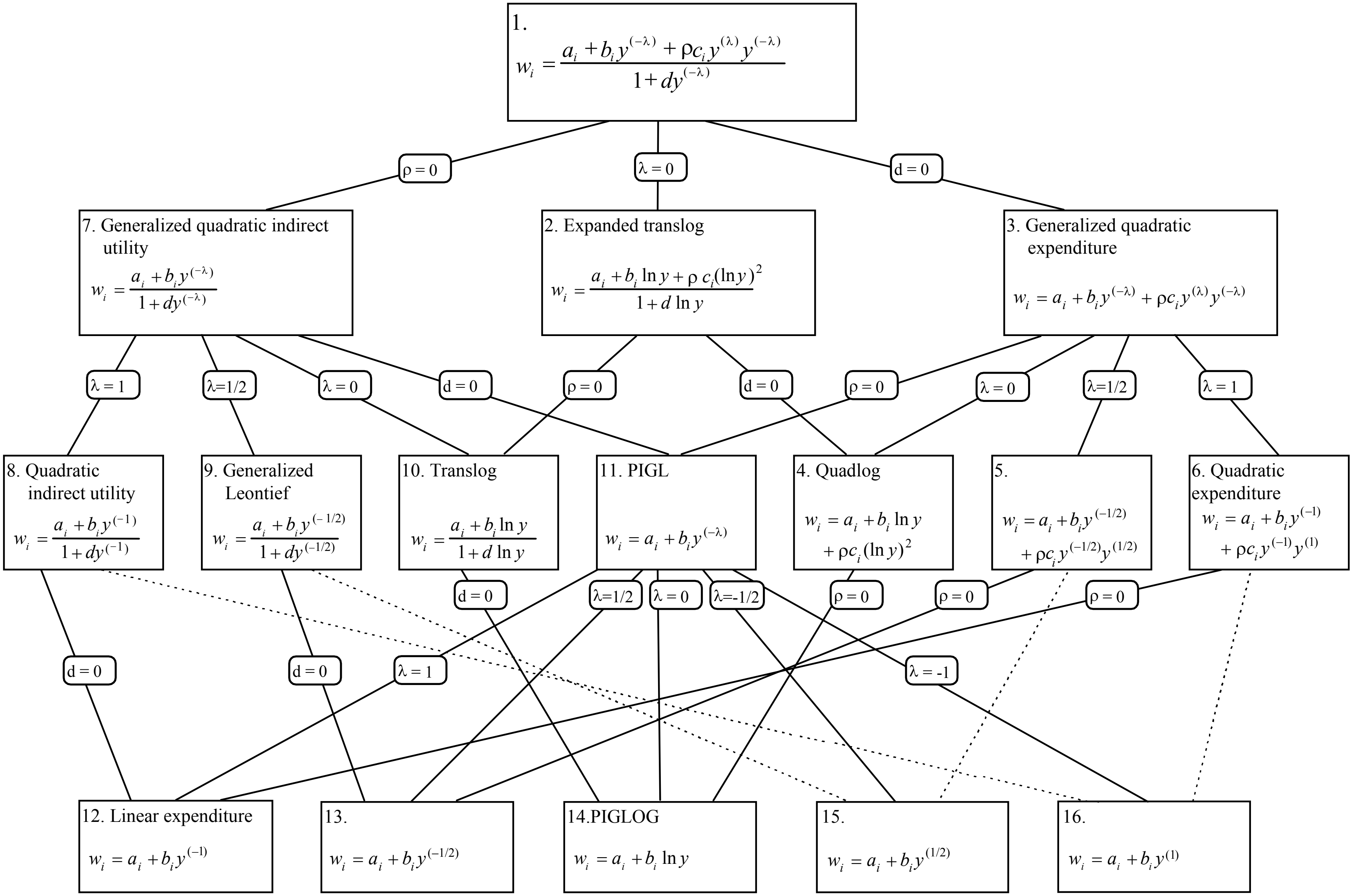
Table 5: Results of testing nested Engel models against the general Box-Cox specification for single categories of expenditure[a]

	Engel models[b]								
	4	5	6	11	12	13	14	15	16
Food	NR	NR	NR	NR	**R**	NR	NR	**R**	**R**
Beverages & tobacco	NR	NR	NR	NR	NR	NR	NR	NR	NR
Housing	NR	NR	NR	NR	NR	NR	NR	NR	NR
Transportation	NR	NR	NR	NR	**R**	**R**	**R**	NR	NR
Recreation	NR	NR	NR	**R**	**R**	**R**	**R**	**R**	**R**
Other goods and services	NR	NR	NR	NR	NR	NR	NR	NR	NR

[a] R =rejected at 0.01; NR =not rejected.

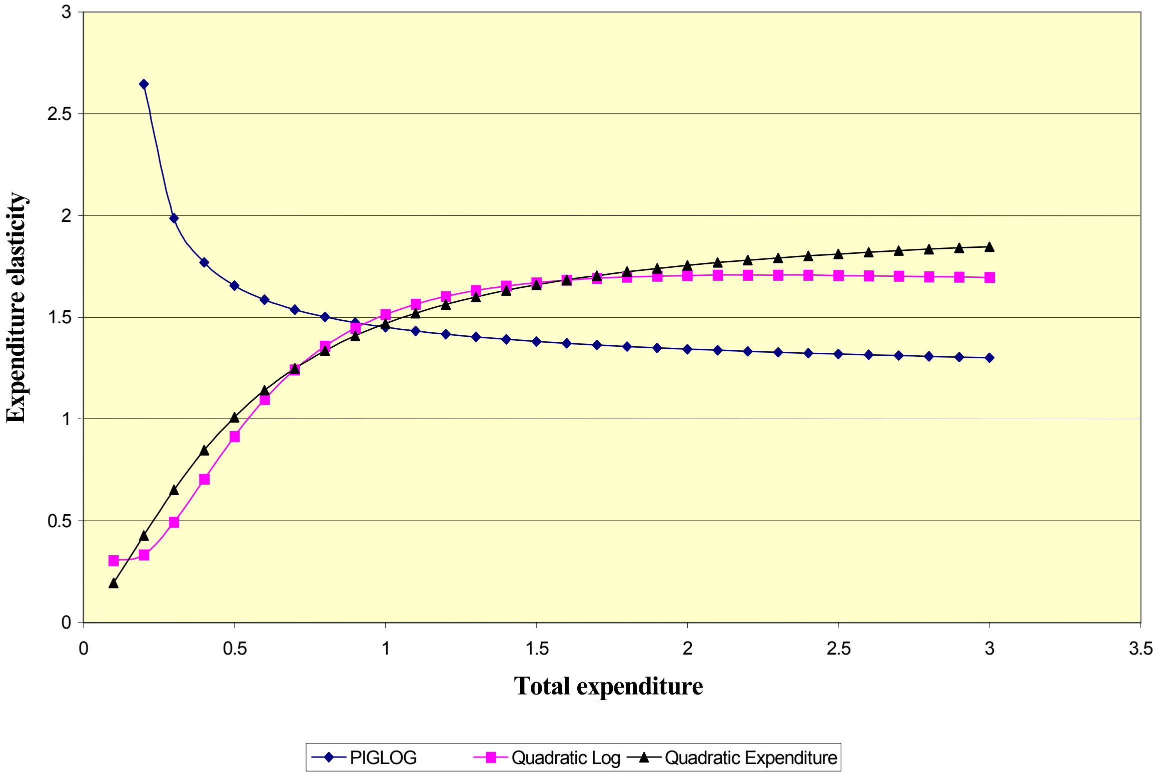
[b] The numbers of the models correspond to those indicated in Figure 1.

Figure 1: Relationships among different systems of Engel curves



Source: Aasness and Rodseth (1983).

Figure 2: Behaviour of the expenditure elasticities for transportation with various Engel curve specifications (average total expenditure of the sample = 1)



APPENDIX 1

Given the sample selection strategy, the empirical model can be specified using total expenditure as the sole explanatory variable. Household preferences are assumed to be sufficiently homogeneous so that the empirical model does not require the inclusion of demographic variables. From the general framework (1), using household data, the general specification can be estimated as

(a) $$w_{ih} = \frac{a_i + b_i y_h^{(-\lambda)} + c_i y_h^{(\lambda)} y_h^{(-\lambda)}}{1 + d y_h^{(-\lambda)}} + \varepsilon_{ih},$$

where:

w_{ih} is the expenditure share for good *i* by family *h*;

y_h is the level of total expenditure of family *h*;

$y_h^{(\lambda)}$ and $y_h^{(-\lambda)}$ are the Box-Cox transformations of total expenditure *y*, specifically:

$y^{(\lambda)} = \frac{(y^{\lambda} - 1)}{\lambda}$ and $y^{(-\lambda)} = \frac{(y^{-\lambda} - 1)}{-\lambda}$;

a_i, b_i, c_i are product specific parameters;

d and λ are common parameter for all of the goods; and,

ε_{ih} is the error term.

The equations can be estimated as a system using the maximum likelihood estimator and imposing the restrictions $\sum_{i=1}^{n} a_i = 1, \sum_{i=1}^{n} b_i = d$, and $\sum_{i=1}^{n} c_i = 0$. Hence *i-1* equations are estimated. The estimates were obtained using the Full Information Maximum Likelihood Procedure (FIML) of the TSP software. This procedure computes maximum likelihood estimates from a nonlinear simultaneous equation model using iterative algorithms.

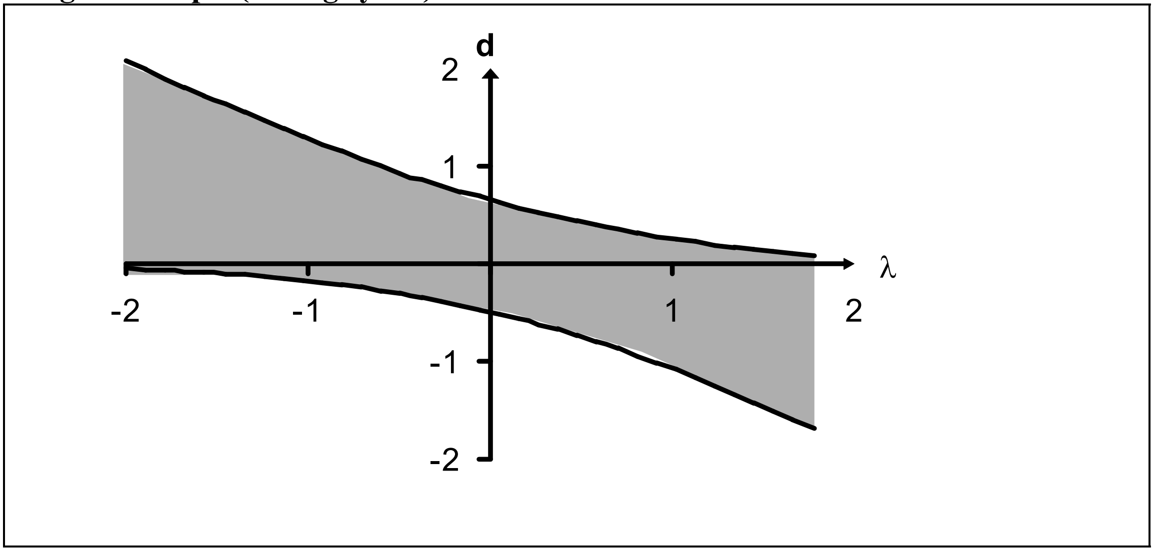
A problem in estimating the empirical model (a) is that the value of the denominator has to be different from zero. As Aasness and Rodseth (1983, p.104) suggest, a practical solution is to restrict the domain of *d* and λ in such a way that, for a given sample, the denominator is not zero in the interval between the two extreme observations of *y* ($y_{\min}$ and $y_{\max}$). The restricted domain of *d* and λ can be defined making the denominator equal to zero and solving for d. This gives the following equation

(b) $$d = \lambda / (y_z^{-\lambda} - 1), \qquad y_z = y_{\min}, y_{\max}.$$

Figure A1 shows the two curves originating from (b). The upper curve is determined by $y_{\min}$ and the lower curve by $y_{\max}$. The shaded area represents all of the combination of *d* and λ where the denominator is not equal to zero. The larger the distribution of total expenditure the more the possible combinations of *d* and λ are restricted (the smaller the shaded area).

Note that the shape of figure A1 is determined by the normalization of the data for total expenditure. This means that the average *y* is equal to one, $y_{\min}<1$, and $y_{\max}>1$. The normalization of the data is also convenient for the computation of expenditure elasticities at the sample mean ($y = 1$). Indeed, when *y* is equal to one, the expenditure elasticities are $E_i = 1 - d + b_i / a_i$.

Figure A1: Possible domain (shaded area) of d and λ for the total expenditure values (y) of a given sample (average y = 1)



APPENDIX 2

The following tables show the importance of zero expenditures for the households and aggregates considered in the analysis.

Table A1 shows that zero observations are rare except for beverages and tobacco, and recreation. For these products, table A2 allows an assessment of the sample selection bias given by the deletion of zero records. The probability of positive household expenditure is evaluated with respect to a family's characteristics including monthly income, age and level of education of the head, and location dummies for families living in the North-East and urban areas. None of the estimated parameters are different form zero at the 0.05 level of significance. Moreover, likelihood ratio tests do not reject the null hypothesis that the explanatory variables are irrelevant in determining the probability of positive expenditure (level of significance 0.05).

Form the above result it is reasonable to conclude that the deletion of the households recording zero expenditures does not lead to a significant sample selection bias.

Table A1: Number of households reporting zero expenditure for each aggregate[a]

Aggregates	Number of households reporting zero expenditure
Food	3
Beverages and tobacco	34
Housing	0
Transportation	1
Recreation	12
Other goods and services	0

[a] Sample includes 400 families with two adults, whose head is employed, that live in the North of Italy, and that recorded expenditures in the first quarter of 1995.

Table A2: Probit estimates of household's non-zero expenditure for beverages and tobacco and recreation as a function of demographic characteristics[a]

	Explanatory variables					
	Constant	Income	North-East	Urban	Age	Education
Beverages and tobacco						
Estimates	1.434	0.000	-0.166	0.195	-0.001	-0.141
t-statistic	1.672	1.900	-0.902	0.340	-0.064	-1.169
P-value	[.095]	[.057]	[.367]	[.734]	[.949]	[.242]
LR (zero slopes)			5.676 [0.339]			
Recreation						
Estimates	0.971	0.000	-0.567	0.956	-0.008	0.098
t-statistic	0.887	0.758	-1.929	1.866	-0.599	0.565
P-value	[.375]	[.448]	[.054]	[.062]	[.549]	[.572]
LR (zero slopes)			10.743 [0.057]			

[a] Sample includes 400 families with two adults, whose head is employed, that live in the North of Italy, and that recorded expenditures in the first quarter of 1995.