European Economic Review 14 (1980) 145–157. © North-Holland Publishing Company

ESTIMATING CONTINUOUS CONSUMER EQUIVALENCE SCALES IN AN EXPENDITURE MODEL WITH LABOUR SUPPLY

R.W. BLUNDELL*

University of Manchester, Manchester M13 9PL, UK

Received November 1977, final version received January 1980

The cubic spline technique is used in conjunction with a utility based approach to measuring household composition effects in demand analysis, in order to derive estimates of continuous consumer equivalence scales in the context of a quasi-homothetic expenditure model that incorporates the labour supply decision.

1. Introduction

When estimating equivalence scales, or more general household composition effects, we often find that the parameters of the composition variables are not identified in a single cross-section, see Muellbauer (1974). This problem is due to the constancy of commodity prices (not always an acceptable assumption) in the cross-section. However, the household does not only choose commodity demands, it also jointly chooses leisure demand and therefore labour supply. Incorporating this decision into the analysis introduces a price (the marginal wage rate) which naturally varies over the cross-section and therefore aids the identification of the underlying parameters. In this study we assume a rather simple form for the labour supply decision and do not consider the supply decision for any secondary family worker [for this development see Heckman (1974)]. The adding-up condition across goods and leisure allows us to estimate the commodity expenditure system alone and leads to the introduction of the marginal wage in an otherwise conventional expenditure system.

The recent availability of the U.K. Family Expenditure Survey for individual households with detailed age data allows a close analysis of age effects in household expenditure and in turn accurate estimation of equivalence scales. To tackle the equivalence scale estimation in such a way

*I would like to thank Ian Walker for many helpful comments and also the Social Science Research Council Survey Archive at the University of Essex for providing the necessary data for this study.

as to allow flexibility without over-parameterisation we have used the cubic spline technique. This turns out to be a rather convenient method of imposing continuity on age effects in expenditure systems.

The layout of the paper is as follows. Section 2 introduces the consumer problem and bears heavily on the original work of Barten (1964) and the further development by Muellbauer (1974) and Gorman (1976). This section then goes on to generalise the model to include the labour supply decision and discusses various properties of the model including its use in welfare analysis. Section 3 considers the incorporation of the cubic spline into the system, and section 4 considers the resulting estimation and data problems. In section 5 the empirical results are reported, and section 6 draws some conclusions.

2. The consumer problem

Barten (1964) proposed that to compare households with different composition one should define the preferences of each household h, by a strictly quasi-concave utility function,

$$U[x_{1h}/m_{1h}, x_{2h}/m_{2h}, \ldots, x_{rh}/m_{rh}], \tag{1}$$

where x_{ih} is the quantity of good i consumed, and each deflator m_{ih} measures the corresponding specific effect on utility of household composition.

Letting p_i denote the price of good i and y_h the income of household h the cost function, dual to (1), has the form

$$c_h(\mathbf{p}, U_h) = c_h(p_1 m_{1h}, p_2 m_{2h}, \ldots, p_r m_{rh}, U_h). \tag{2}$$

The similarity between price and household composition effects in (2) noted by Barten (1964), imply very simple generalisations of the traditional individual demand models.

In this study the cost function is specified directly as this leads immediately to both the demand equations and the true cost-of-living index useful to a study of equivalence scales. To begin with let us specify the following quasi-homothetic form due to Gorman (1976):

$$c_h(p, U_h) = a_h(\mathbf{p}) + b(\mathbf{p})U_h, \tag{3}$$

where $a_h(p)$ and $b(p)$ are concave, linear homogeneous functions in p. In our model we shall let the fixed cost $a_h(p)$ depend on household composition, in particular we shall write

$$a_h(p) = \sum_j p_j \gamma_j m_{jh}.$$

The form (3) allows aggregation across household members and corresponds to the translation method of incorporating composition effects developed in Pollak and Wales (1976). Choosing a Cobb–Douglas form for $b(p)$ will lead to the linear expenditure system and equivalence between the Barten and translation methods.

In order to identify all the parameters of the cost function we generalise this model to include the labour supply decision. The usefulness of this generalisation when using a single cross-section was first noticed by Betancourt (1971) and we develop his work to analyse composition effects and to consider some interesting relationships between goods and leisure. We shall suppose that there is a single male worker in each family facing a marginal wage rate w and a linear budget constraint

$$p'x + wl = wT + y' = y, \tag{4}$$

where l is leisure time, T is the maximum time available, y is full income and y' is unearned income actually defined by identity (4). The supply of labour hours is then given by $T - l = h$.

The cost function we choose is a generalisation, suggested by Muellbauer (1980), of the form given in (3), and is given by

$$c_h(w_h, p, U_h) = a_h(p) + w_h d(p) + b(p)^{1-\theta} w_h^{\theta} U_h, \tag{4}$$

where $0 < \theta < 1$ and $d(p)$ is homogeneous of degree zero in prices. Function (4) preserves quasi-homotheticity and therefore implies parallel Engle curves in full income. For simplicity we specify

$$d(p) = \gamma_l \prod_i p_i^{\delta_i}, \qquad \sum_i \delta_i = 0, \quad \gamma_l > 0,$$

and

$$b(p) = \prod_i p_i^{\beta_i}, \qquad \sum_i \beta_i = 1.$$

Taking derivatives of c_h with respect to price we derive the following expenditure equations for all i:

$$p_i x_{ih} = p_i \gamma_i m_{ih} + \gamma_l \delta_i w_h + (1-\theta)\beta_i (y_h' + (T - d(p)) w_h - a_h(p)). \tag{5}$$

Separability of goods from leisure, a sufficient condition for an optimal uniform commodity tax is only obtained when $\delta_i = 0$ for all i, in other words $d(p) = \gamma_l$. These $r-1$ restrictions simplify the system considerably and we test them empirically in section 5. With these restrictions imposed all goods are substitutes for leisure however system (5) itself allows for complementarity between some goods and leisure.

The labour supply equation corresponding to system (5) is given by

$$h=\frac{\theta}{w}(a(p)-y')+(1-\theta)(T-d(p)).$$

which has the interesting property of being backward sloping for $y'<a(p)$. For the sample of households considered in our example this condition is typically satisfied.

To develop the model let us suppose that there are D age groups and associated with each age group, a_i, there are n_{ah} members in household h. We then write

$$m_{ih}=\sum_{a=a_1}^{a_D} e_{ia}n_{ah}=e_i'\mathbf{n}_h \quad \text{for} \quad i=1,\ldots,r, \quad h=1,\ldots,H,$$

where the e_{ia}, after normalisation with respect to an adult (of age 40 years say), are the specific equivalence scales.

Apart from these specific effects we may be interested in the construction of a true cost of living index which could be used as a total equivalence scale. Consider the comparison of a household J, with a base household I, which has the same marginal valuation of time. The constant utility cost of living index is given by

$$c_J(\mathbf{p},w,u)/c_I(\mathbf{p},w,u).$$

From (4) we see that this may be written as

$$(c_I+a_J-a_I)/c_I.$$

Given a value $y_I=c_I$, we simply require knowledge of the additional necessary cost a_J-a_I. Noticing that in our model this is

$$\sum_k p_k\gamma_k(m_{kJ}-m_{kI}),$$

it can be seen that this extra cost can be calculated from knowledge of the specific effects. Due to the form of m_{ih}, $\Sigma_k e_{ka}p_k\gamma_k$ gives the cost of an additional person of age, a. That is the extra cost that is required to keep the household, as a whole, at the same level of well being measured by (1).

3. Application of the cubic spline technique

For estimation purposes we could set all $e_{ia}=1$, then $m_{ih}=N_h$ the total number of persons in each household. A slightly less restrictive alternative is that $e_{ia}=e_i$ for $a \leqq 15$ years and $e_{ia}=1$ for $a \geqq 15$ years, allowing a separate constant child scale for each good. As a maintained hypothesis neither of these suggestions seem sufficiently flexible. It seems more likely that age has a significant and continuously changing effect on consumption. We could suggest that structural changes in behaviour occur, in mid-teens and late-twenties for example, but not so as to destroy the overall continuity of behaviour. We wish to impose a continuous, albeit fairly flexible, structure on e_i seen as a function of age. Blokland (1977) has also made a strong case for considering such continuous equivalence scales.

The cubic spline technique satisfies these properties while allowing a significant reduction in the number of parameters to be estimated. We assume that between each point of structural change (knot) the function is at most a cubic, but at these knots a change in behaviour is represented by a jump in the third derivative, preserving continuity.

Suppose we can identify $k-1$ possible points at ages $\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_{k-1}$, then our cubic spline, for each good, is given by

$$e(a)=q_k+q_{k+1}a+q_{k+2}a^2+q_{k+3}a^3+\sum_{j=1}^{k-1} q_j(a-\bar{a}_j)^3_+, \tag{6}$$

where

$$\begin{aligned}(a-\bar{a}_j)^3_+ &=(a-\bar{a}_j)^3 \quad \text{if} \quad a>\bar{a}_j,\\ &=0 \qquad\qquad\quad \text{if} \quad a\leqq\bar{a}_j,\end{aligned}$$

and q_i $(i=1,\ldots,k+1)$ are unknown parameters.

From (6) we know the form of e_i for all a, so given observations at ages $a_1, a_2, \ldots, a_D$, we could substitute $e_i(a_j)$ directly into our model (5). However, the data matrix corresponding to (6) can be extremely ill-conditioned, and so for practical application we follow Poirier (1976) and derive an alternative expression below. In addition we can restrict the form of $e_i(a)$ at the end points $\bar{a}_0$ and $\bar{a}_k$. For this study we restrict the second derivative of e_i at a_0 to be zero and the first derivative to be zero at $\bar{a}_k$, imposing a certain stability of behaviour with age, which seems *a priori* reasonable. These two restrictions reduce the number of parameters to be estimated to $k+1$. In particular we write

$$\boldsymbol{e}_i=\boldsymbol{W}\boldsymbol{b}_i \quad \text{for} \quad i=1,\ldots,r, \tag{7}$$

where $\boldsymbol{b}_i$ is a $k+1$ vector of ordinates corresponding to $\bar{a}_0, \bar{a}_1, \ldots, \bar{a}_k$ and W is a known $D \times (k+1)$ matrix, details of which can be found in Poirier (1976, ch. 3). Given estimates of $\boldsymbol{b}_i$ we can then use (7) to generate estimates of $\boldsymbol{e}_i$ that satisfy the smoothness conditions outlined above.

Returning to our expenditure model (5) we have

$$p_i x_{ih} = n_h' e_i p_i \gamma_i - (1-\theta)\beta_i \sum_{j=1}^{r} n_h' e_j p_j \gamma_j + (1-\theta)\beta_i y_h' + (1-\theta)\beta_i \phi_i w_h, \quad \text{for all } i \text{ and } h,$$

where

$$\phi_i = ((\gamma_l \delta_i / (1-\theta)\beta_i) + (T - d(p))).$$

From (7) we may write our model as

$$p_i x_{ih} = \boldsymbol{n}_h' W \boldsymbol{b}_i p_i \gamma_i - (1-\theta)\beta_i \sum_j \boldsymbol{n}_h' W \boldsymbol{b}_j p_j \gamma_j + (1-\theta)\beta_i y_h' + (1-\theta)\beta_i \phi_i w_h. \tag{8}$$

4. Estimation

The data used in the following empirical example was the Northwest Regional subsample of the 1974 U.K. Family Expenditure Survey. A particular region was chosen in order to remove, to some extent, possible regional price variation. Our aim in this paper is to concentrate on the estimation of equivalence scales and with this in mind we have chosen the simplest of labour supply considerations. We analyse only those households with a male head in full employment and include all other family earning in our unearned income category, using a linear approximation to the budget constraint for each worker.

Although the quasi-homothetic form of (4) looks attractive for cross-section estimation it is unlikely to provide a reasonable approximation over the whole income range. In this study we see (4) as an approximation over a narrow income range towards the lower end of the income distribution. Selection of households is done according to expenditure as full income is an unknown before estimation and we choose a range from £25.00 to £45.00. This data set provides yearly age groups and referring to section 3 above we choose k to be 4 which taking knots at 15, 30, 50 and 65, seemed to allow sufficient flexibility in the form of the equivalence scales. Finally we chose to aggregate goods into the following six categories:

(i) Food.

(ii) Alcohol and Tobacco,

(iii) Fuel, Light and Power
(iv) Clothing and Footwear,
(v) Services, Other Goods, Transport and Vehicles and Miscellaneous,
(vi) Durable Goods and Housing.

A more detailed disaggregation is of course possible but this level was chosen in order to concentrate on necessary expenditures and provide an illustration of a fairly low dimension.

Assuming as we have done that prices are constant in cross-section we may write (8) as

$$p_i x_{ih} = \mu_i' z_h - (1-\theta)\beta_i \sum_j \mu_j' z_h + (1-\theta)\beta_i y_h' + (1-\theta)\beta_i \phi_i w_h \quad \text{for all} \quad i = 1, \ldots, r, \tag{9}$$

where

$$\mu_i = b_i \gamma_i p_i \quad \text{and} \quad z_h = W' n_h.$$

In order to reduce the possibility of heteroscedasticity in a stochastic specification of (9) we work with budget shares as the dependent variable and assume an additive disturbance distributed independently across households but not goods. Due to the budget constraint we do not need to estimate the parameters of the labour supply equation directly and can identify all the underlying parameters of the cost function from the r expenditure share equations. The model is a non-linear in parameters seemingly unrelated system and was estimated by a Newton–Raphson maximum likelihood program details of which can be found in Wymer (1973). This program uses analytic derivatives to calculate the Hessian matrix at each iteration and from a variety of quite different starting values, for the models considered above, it converged in no more than six iterations (approx. 6 C.P.U. seconds on the CDC 7600).

5. Empirical results

Table 1a presents the results for the general non-separable model that allows for complementarity between goods and leisure. In terms of the estimated parameters the cross-substitution effects are given by

$$S_{il} = \delta_{il}\gamma_l/p_i + (1-\theta)(\beta_i/p_i)(T-d-h)$$
$$= (1-\theta)(\beta_i/p_i)(\phi_i - h),$$

where h is the number of hours worked. The mean number of hours worked in the sample is 43 and therefore a comparatively low value of ϕ_i implies

Table 1a

General model: Parameter estimates, asymptotic t-ratios in parentheses; 96 observations.

Commodity group	$(1-\theta)\beta_i$	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}	ϕ_i
Food	0.1468 (5.29)	0.2350 (0.37)	2.8342 (5.95)	4.1000 (4.53)	4.5001 (5.56)	3.1433 (3.81)	34.0471 (4.50)
Alcohol and Tobacco	0.0032 (0.08)	−0.3340 (0.63)	0.7352 (1.87)	0.8712 (1.16)	0.6318 (0.94)	0.9370 (2.01)	765.9883 (0.08)
Fuel, Light and Power	0.0350 (2.48)	0.4576 (1.77)	0.2588 (1.45)	0.9944 (2.94)	1.1049 (3.65)	0.9371 (3.04)	19.5963 (1.30)
Clothing and Footwear	0.1513 (2.94)	0.6512 (1.20)	0.4032 (1.00)	0.3733 (0.50)	0.4804 (0.70)	0.4278 (0.61)	40.9821 (2.81)
Services and Other Goods	0.2008 (5.51)	0.1539 (0.18)	1.0097 (1.60)	1.1003 (0.92)	1.3565 (1.27)	0.3277 (0.30)	45.3878 (5.53)
Durable Goods and Housing	0.2293 (4.86)	0.0102 (0.02)	1.0201 (1.37)	3.7082 (2.61)	3.6894 (2.91)	2.3597 (1.83)	50.2587 (5.07)

Table 1b

Separable model: Parameter estimates, asymptotic t-ratios in parentheses; 96 observations.

Commodity group	$(1-\theta)\beta_i$	μ_{i1}	μ_{i2}	μ_{i3}	μ_{i4}	μ_{i5}
Food	0.1268 (5.31)	0.1673 (0.28)	2.7582 (6.22)	3.7223 (4.85)	4.0897 (6.28)	2.9330 (3.94)
Alcohol and Tobacco	0.0313 (0.95)	−0.2173 (0.38)	0.8659 (2.09)	1.5218 (2.50)	1.3385 (3.04)	1.5332 (2.58)
Fuel, Light and Power	0.0240 (2.02)	0.3837 (1.73)	0.2118 (1.31)	0.7606 (3.08)	0.8509 (4.53)	0.8068 (3.06)
Clothing and Footwear	0.0903 (3.32)	0.6409 (1.20)	0.3917 (1.00)	0.3158 (0.51)	0.41797 (0.80)	0.3598 (0.79)
Services and Other Goods	0.2015 (6.41)	0.1771 (0.21)	1.0358 (1.63)	1.2301 (1.08)	1.4975 (1.52)	1.1012 (1.37)
Durable Goods and Housing	0.2437 (6.01)	0.0605 (0.06)	1.1136 (1.44)	4.1733 (3.02)	4.1946 (3.50)	2.6186 (2.00)
$(T-d)=45.1425$ (13.21)						

complementarity whereas a higher value implies substitution. The two ends of the spectrum are illustrated by Fuel, Light and Power an obvious complement and Durable Goods a typical substitute, see Barnett (1979). The implied value of $(T-d)$, the maximum amount of work time available is 45.97. Although probably a little on the low side due to the linearity of the budget constraint, it is still an intuitively sensible value. The large value of ϕ_i for Alcohol and Tobacco is complemented by a very small value for $(1-\theta)\beta_2$, and implies quite a reasonable point estimate for $\gamma_l\delta_2$ of 2.28.

Given this maintained model we may go on to test the separability of goods from leisure. Such separability implies that all the ϕ_i's are equal and the results for this restricted model are given in table 1b. On the null hypothesis, twice the difference of the loglikelihood values should follow a χ^2 distribution asymptotically, with five degrees of freedom. The calculated value of this statistic is 6.327 and therefore we can accept the null hypothesis at the 5% level, in this sample. In this restricted model we see that most of the parameter estimates are statistically well determined with intuitively reasonable values. The income effect on alcohol and tobacco consumption is still insignificantly different from zero but apart from this all other income effects are significantly positive and summing the coefficients of the income variables we find the implied estimate of $(1-\theta)$ is 0.7104. The μ_{ij} parameters are the estimates of $p_j\gamma_j b_{ij}$ and allow us to analyse family composition effects on expenditure. First we turn to the implied equivalence scales.

Fig. 1 provides the estimated scales where we have normalised so that they peak at unity. It is interesting to note that not all scales start close to zero and neither do all scales for adults dominate those for children. This is especially true for such goods as Fuel, Light and Power, and Clothing and Footwear. The additional cost of an extra individual may now be calculated by summation of the estimated non-normalised specific scales,

$$\sum_{j=1}^{r} W\boldsymbol{b}_j p_j \gamma_j.$$

This cost is presented in fig. 1 ('Total') and is particularly interesting for children as it measures the additional cost of a child in order to keep the household at the same level of utility. It therefore provides precisely the information required to construct the true cost of living index and as a result for considerations of horizontal equity across households.

In table 2 we present some more detailed results on (a) the changes in the distribution of expenditures, and (b) the hours of work, on the addition of one child to a household. Household characteristics may also affect the decision of other family members to work and to this extent there could occur additional indirect effects through y'. The assumption that the male working head is not near a kink in his budget constraint is also an important assumption behind the hours of work consideration. However, the detailed effects of household composition on labour supply (given full income) certainly add an interesting new dimension to the more usual model, expenditures do not have to add up as leisure time is variable. However, some expenditures do fall in services and durables to compensate for the increases in more necessary expenditures.

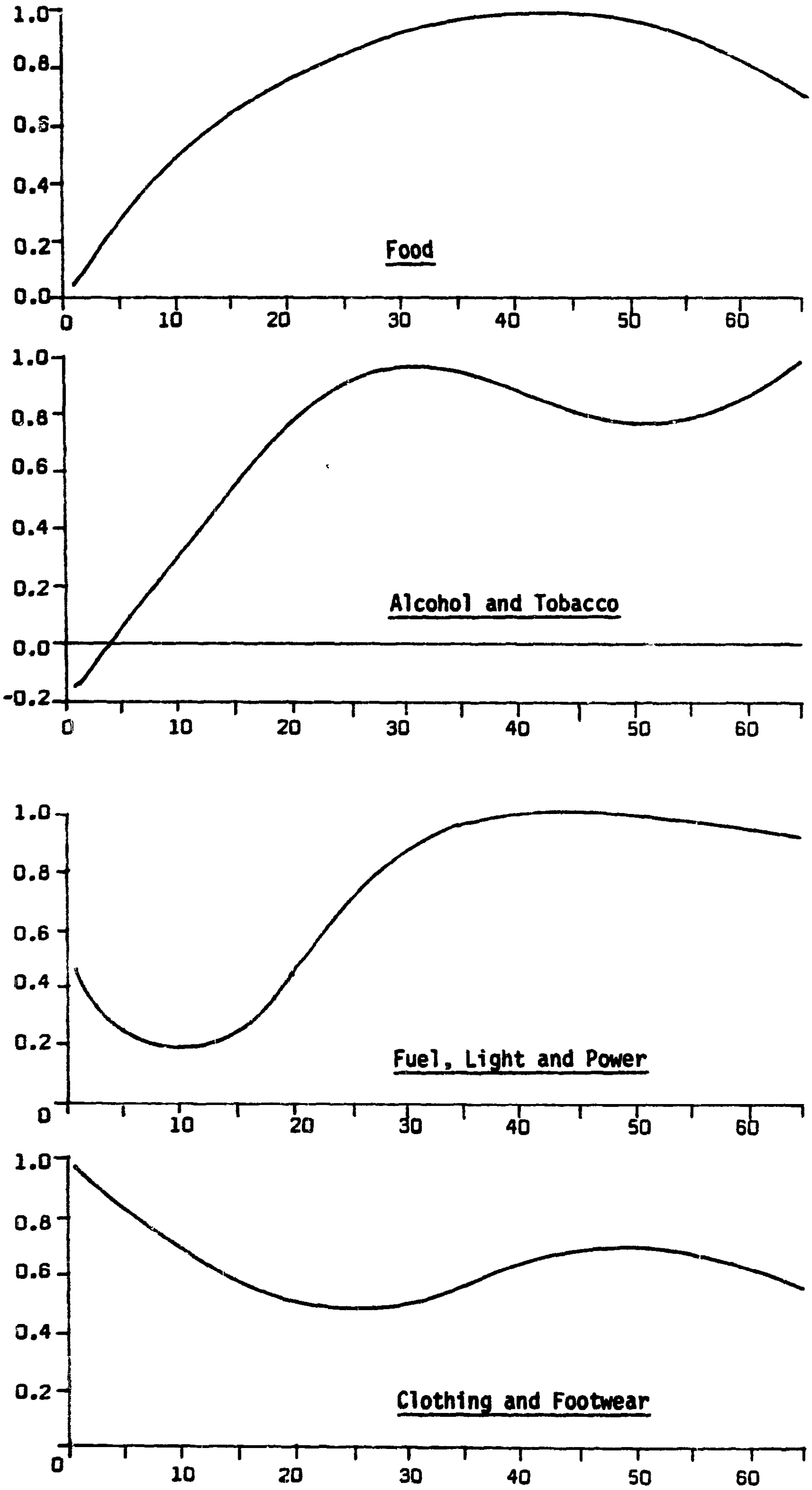

Fig. 1. Equivalence scale estimates; 'Total' in 1974 £'s.

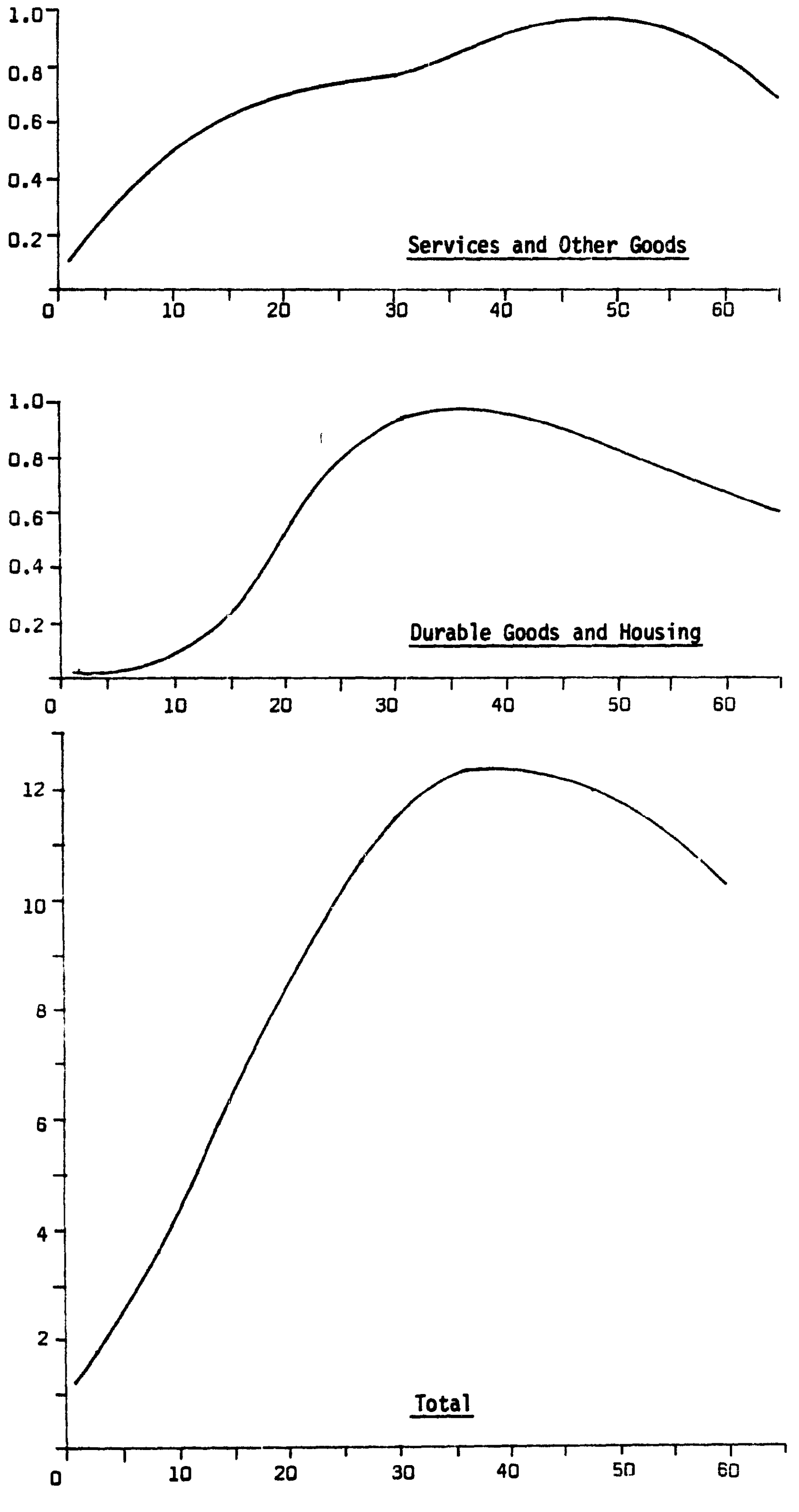

Fig. 1 (continued)

Table 2
(a) The extra expenditure (given full income, all figures in 1974 £'s) and (b) the extra hours worked, both on the addition of one household member.

		Age		
		5	10	15
(a) Food		0.759	1.477	1.951
Alcohol and Tobacco		0.029	0.368	0.662
Fuel, Light and Power		0.164	0.050	0.057
Clothing and Footwear		0.301	0.247	−0.180
Services and Other Goods		−0.035	−0.083	−0.241
Durables and Housing		−0.433	−0.760	−0.445
(b) Marginal	0.5	1.56	2.59	3.60
hourly wage	0.8	0.97	1.62	2.25

6. Conclusions

In this paper we have developed a model of household composition effects in consumer demand which introduces smooth equivalence scales, using the spline technique, into a theoretically consistent model of consumer behaviour. The model is fully identified and straightforward to estimate from a single cross-section of data and an illustrative example using the 1974 Family Expenditure Survey has been given. We have generated continuous consumer equivalence scales that are both useful in considerations of horizontal equality and the analysis of consumer demand. In particular we have incorporated the labour supply decision to an expenditure model in order to identify the underlying parameters and have found that this in itself leads to interesting tests of separability and complementarity.

The quasi-homothetic assumption in the cost function is of course only an approximation and would be a particularly bad one across all income groups. However, by choosing a fairly narrow band of income we hope to have reduced any approximation error. A second criticism of this model arises from the implied backward sloping supply curve of labour, necessary if $a_h(p) > y_h$. A condition which is most likely to hold for households with low marginal wage rates and therefore reverses the usual form of the backward bending supply curve.

It may also be fair to criticise the assumption that household composition only works through the fixed cost element in the cost function. Although this assumption is required for within household aggregation it may not necessarily hold in the data. The overriding consideration here though, is the necessity for a reasonably linear structure in order that estimation of the spline coefficients is computationally tractable. To this end a model that is nonlinear in parameters only is of considerable advantage, and we hope to have struck a reasonable balance between realism and simplicity.

References

Barnett, W. A., 1979, The joint allocation of leisure and goods expenditure, Econometrica, 539–564.

Barten, A.P., 1964, Family composition, prices and expenditure patterns, in: P. Hart, ed., Colston Papers no. 16 (Butterworths, London).

Betancourt, R.R., 1971, The estimation of price elasticities from cross section data under additive preferences, International Economic Review, 283–292.

Blokland, J., 1977, Continuous consumer equivalence scales (Martinus Nijhoff, The Hague).

Gorman, W.M., 1976, Tricks with utility functions, in: M. Artis and P. Nobay, eds., Essays in economic analysis.

Heckman, J., 1974, Shadow prices, market wages and labour supply, Econometrica, 679–694.

Muellbauer, J., 1974, Household composition, Engel curves and welfare comparisons between households, European Economic Review, 103–122.

Muellbauer, J., 1980, Linear aggregation in neoclassical labour supply, Review of Economic Studies 47, forthcoming.

Poirier, D.J., 1976, The econometrics of structural change (North-Holland, Amsterdam).

Pollak, R.A. and T.J. Wales, 1976, Estimation of complete demand system from household budget data, Discussion Paper no. 354 (Department of Economics, University of Pennsylvania, Philadelphia, PA).

Wymer, C.R., 1973, Computer programs: Resimul manual.