ECONOMETRICA
JOURNAL OF THE ECONOMETRIC SOCIETY

Forms of Engel Functions
Author(s): C. E. V. Leser
Source: *Econometrica*, Vol. 31, No. 4 (Oct., 1963), pp. 694-703
Published by: The Econometric Society
Stable URL: http://www.jstor.org/stable/1909167
Accessed: 07/01/2011 09:38

Econometrica, Vol. 31, No. 4 (October, 1963)

FORMS OF ENGEL FUNCTIONS[1]

BY C. E. V. LESER

The properties of various forms of Engel functions satisfying the additivity criterion are investigated. It is suggested that a form of relationship used by H. Working, but recently neglected, offers on balance great advantages. A logical generalisation of it yields a flexible function from which a good fit may be expected for most commodity groups.

THE PROBLEM of finding the most appropriate form of an Engel function is an old one in econometrics, but as yet no solution appears to have found general acceptance. Generally speaking, it is perhaps true to say that the specification of the form of relationships has attracted less attention than have methods of estimating parameters for specified equations.

To some extent, of course, the answer depends on the degree of emphasis placed on various properties that one desires the function to possess. Here it is assumed that a set of Engel functions for a number of commodity groups, which include all expenditure, is to be derived from family budget data, with total expenditure as the independent variable. Bias arising from abnormal expenditure is assumed to have been removed, or at any rate reduced, by using income or another instrumental variable for grouping, as has been done by Liviatan [**6**].

It is postulated, furthermore, that the Engel functions for the various commodity groups studied have the same mathematical form and should satisfy the additivity criterion. This is somewhat restrictive but still permits a fair choice among different functions.

The considerations governing the choice may be broadly divided into three categories. First, a close connection with a direct or indirect utility function may appear desirable. Secondly, the function should, ideally, be valid for all positive values of total outlay, or at any rate within a wide range. The variations in the income elasticities of demand entailed by the formula should also be plausible. Finally, there are statistical and computational considerations: Estimation of parameters should be simple and convenient, allowing an assessment of reliability and goodness of fit; the error specification should also be reasonable.

The weight given to the first group of considerations will necessarily be different depending upon whether the problem is approached from the viewpoint of the theoretical economist or from that of the economic statistician. In the latter case, the emphasis will be on the second and third sets of con-

[1] The author is a member of the staff of the Economic Research Institute, Dublin. The author is responsible for the contents of this article, and the views expressed in it do not neccessarily reflect the opinions of the Institute.

siderations. The theoretical desirability of finding a set of Engel functions valid, i.e., giving nonnegative values, for all income levels will, however, not be allowed to override other considerations, since extrapolation of the Engel functions well beyond the observed range of total outlays may in any event give inaccurate results.

A general statistical consideration is that the errors, representing individual variations in tastes as differences between observed and habitual expenditure patterns, may be assumed to be proportional to total outlay on all goods and services. This implies that the residuals have uniform variance if the dependent variables in the regression equations are the expenditure proportions $w_i = v_i/M$, where v_i is expenditure on commodity group i and M is income.

In the light of these considerations, some possible additive Engel functions containing two parameters will now be discussed. This number of parameters should be sufficient where the number of income groups is comparatively small and the income range is not too wide. The following models are familiar or suggest themselves readily:

$$w_i = \alpha_i + \beta_i M + \varepsilon_i \, , \tag{1}$$

$$v_i = \alpha_i + \beta_i M + \varepsilon_i, \tag{2a}$$

$$w_i = \alpha_i + \beta_i/M + \varepsilon_i \, , \tag{2b}$$

$$\log w_i = \alpha_i + \beta_i \log M - \log \left(\Sigma e^{\alpha_j + \beta_j \log M}\right) + \varepsilon_i \, , \tag{3}$$

$$w_i = \alpha_i + \beta_i \log M + \varepsilon_i,. \tag{4}$$

Equation (1) has the advantage of greatest simplicity given the error specification adopted here, but otherwise it does not appear very attractive. These functions are not valid for high M; they could be valid for low M but will often fail to be since estimation produces some negative α_i. What is more important, empirical evidence does not suggest a linear relation between total outlay and the expenditure proportion, and an attempt to force the function into this form may well result in negative figures for outlays on a commodity group at income levels within the observed range. Furthermore, these functions imply that demand elasticities for moderate luxuries rise with income, which is hardly plausible.

Equation (2a) is, of course, the form used by Allen and Bowley [**1**], and (2b) is a version of it with a different error specification. These are generally valid for high M (unless applied to an inferior good) but not for low M. They are very convenient for computation, but the fit is not always good, and they imply that demand elasticities for necessities increase with income. The estimate for β_i in (2b) will generally differ from the estimate for α_i in (2a) and vice versa. (2b) is less likely to yield negative values within the observed income range than (2a), which is a point in its favour.

Equation (3) is the constant elasticity function adjusted for additivity, which has been used recently, e.g., by Houthakker [**3**], the Netherlands Central Bureau of Statistics [**7**] and also by this writer [**5**]. It can also be written as a set of equations for log v_i. The model has the advantage of being valid for all M; and the implied elasticity behaviour is a slow decline with differences between elasticities remaining constant, which is quite plausible. The estimation of parameters is not difficult but presents at least two awkward features. First, the computation partly depends on the grouping adopted; if the grouping is changed, the results for all income elasticities at average level have to be modified. Secondly, zero observations cannot be utilised, and data close to 0 unduly influence the estimation of parameters.

Equation (4) has been used by H. Working [**11**]. It is not be confused with the nonadditive function $v_i = \alpha_i + \beta_i \log M + \varepsilon_i$ which may give a good fit for some groups but is not universally applicable. (4) is not valid for extreme values of M; but, in the neighbourhood of the average, it gives a good approximation of a constant elasticity function and of (3). The model was rejected by Prais [**8**] and appears to have been neglected in recent work, not being discussed at all by Prais and Houthakker [**9**] or by Goreux [**2**]. It implies a decline in demand elasticities with rising income, which is more marked, the more the elasticity differs from 1, and which is absent for groups with unit elasticity—a behaviour which would seem acceptable. The use of this model avoids the mentioned difficulties of (3) and also has the advantage of greater simplicity.

As an illustration of differences in results and fit obtained by the various models, the data given below have been used as a basis. They have been derived from expenditure data grouped by size of household and per capita income for Ireland [**4**]. Whilst the full set of 4×4 size-income groups was utilised for the demand analysis in [**5**], a simplified procedure was applied here. By computing weighted averages of the four size groups with a set of standard weights based on the correct proportions, an attempt was made to eliminate, as far as the nature of the material permitted it, the effects of differences in household size from the four expenditure groups. It is realised that the data contain recording errors, particularly for alcoholic drinks, included in group 7.

Since the number of households in each income group is approximately equal, unweighted regressions have been calculated. The goodness of fit obtained for each commodity group has been measured by the explained proportion of $\Sigma(w_i - \bar{w}_i)^2$, which is equal to R^2 in models (1), (2b), and (4). It is suggested that this is preferable to using the explained proportion of $\Sigma(v_i - \bar{v}_i)^2$, which may be fairly high even when the naive model $v_i = \alpha_i M$ is used.

It may be noted that the explained proportion of the variance is generally low for "Housing" and "Tobacco" because there is little to explain. Varia-

TABLE I

EXPENDITURE DISTRIBUTION FOR 4 INCOME GROUPS, IRELAND 1951–52

M (s. per week)	100.8	166.4	257.7	408.7
w_i:				
1. Food	.5162	.4400	.3652	.2882
2. Clothing	.0784	.1129	.1410	.1498
3. Fuel and light	.1059	.0784	.0686	.0528
4. Housing	.0734	.0654	.0656	.0668
5. Tobacco	.0527	.0557	.0526	.0397
6. Household durables	.0084	.0165	.0301	.0379
7. Miscellaneous goods	.0411	.0450	.0464	.0505
8. Travel and holidays	.0131	.0277	.0533	.0920
9. Entertainment	.0155	.0252	.0293	.0299
10. Miscellaneous services	.0953	.1332	.1479	.1924
Total	1.0000	1.0000	1.0000	1.0000

TABLE II

PROPORTION OF VARIANCE EXPLAINED BY VARIOUS REGRESSION MODELS, IRELAND 1951–52

Commodity group	$1 - \Sigma(w_i - w_{ie})^2/\Sigma(w_i - \bar{w}_i)^2$ for Engel function				
	(1)	(2a)	(2b)	(3)	(4)
1. Food	.966	.903	.949	.999	.999
2. Clothing	.828	.990	.991	.961	.950
3. Fuel and light	.880	.980	.986	.964	.968
4. Housing	.286	.415	.667	—.656	.471
5. Tobacco	.757	—.267	.365	.777	.564
6. Household durables	.944	.897	.932	.949	.983
7. Miscellaneous goods	.952	.873	.929	.818	.975
8. Travel and holidays	.999	.666	.829	.999	.951
9. Entertainment	.686	.931	.966	.873	.864
10. Miscellaneous services	.963	.837	.914	.964	.972

tions in the share of outlay are small and irregular, or, at any rate, not monotonic.

Equation (1) gives a relatively good fit for "Tobacco," "Miscellaneous goods," "Travel and holidays," and "Miscellaneous services," though it gives a superior fit, jointly with (4), only for "Travel and holidays." The fit is, however, notably inferior to all other models with respect to "Clothing," "Fuel and light," and "Entertainment," and inferior to all but (3) for "Housing." This seems sufficient to condemn its use.

Equation (2a) is, by definition, less efficient than (2b) when the present criterion is adopted. For "Tobacco" the minimisation of $\Sigma(v_i - \bar{v}_i)^2$, which is the basis of estimation for (2a), implies that the observation for the highest income group strongly influences the estimates and makes $\Sigma(w_i - w_{ic})^2$ larger than $\Sigma(w_i - \bar{w}_i)^2$. For "Clothing" and "Fuel and light," where the linear function fits the data well, the loss of efficiency is slight. An objective criterion in favour of (2b) as opposed to (2a) is that with (2a), v_{ic} and w_{ic} become negative in the group "Travel and holidays" for $M = 100.8$, the level of total outlay in the lowest income group. This is not the case with (2b).

Equation (3) gives a surprisingly good fit for "Tobacco" but a very poor one for "Housing." This is due to the peculiar nature of the function which may show a rise in w_{ic} followed by a decline, but not the reverse. For similar reasons, the fit for "Miscellaneous goods" is relatively poor. It is perhaps surprising that (3) gives a better result by the criterion used here than does (2b) in the case of five commodity groups, even though the usual least-square procedures minimises $\Sigma(w_i - \bar{w}_i)^2$ with (2b), but is not exactly designed to do so in (3). All in all, however, (4) seems slightly preferable to (3).

The choice between (2b) and (4) may appear more difficult. Equation (4) has the edge on (2b) in six groups, and its positive differentials over the corresponding value for (2b) are, on the whole, larger than the negative differentials. The goodness of fit criterion thus tends to reinforce, in this case, the theoretical considerations in favour of using (4).

It is also of interest to see how estimation of the income elasticity of demand η is affected by the choice of the function. To simplify matters, assume

TABLE III

ESTIMATES OF AVERAGE INCOME ELASTICITIES OF DEMAND, IRELAND 1951–52

Commodity group	$\eta_i(G)$ estimated from Engel function				
	(1)	(2a)	(2b)	(3)	(4)
1. Food	.650	.546	.626	.585	.594
2. Clothing	1.394	1.369	1.378	1.470	1.435
3. Fuel and light	.603	.502	.537	.521	.521
4. Housing	.954	.972	.931	.939	.936
5. Tobacco	.815	.702	.869	.809	.822
6. Household durables	1.966	1.882	1.740	2.108	1.945
7. Miscellaneous goods	1.129	1.150	1.120	1.142	1.140
8. Travel and holidays	2.355	2.270	1.888	2.404	2.210
9. Entertainment	1.356	1.304	1.372	1.464	1.412
10. Miscellaneous services	1.452	1.491	1.383	1.480	1.465
Average	1.000	1.000	1.000	1.000	1.000

that the main interest is concentrated on the value of the elasticity which applies at the level $M = G$, where G is the geometric mean of the observations. Since M may be lognormally distributed, G may reasonably be considered a more "typical" value than $\bar{M}$, where $\bar{M}$ is the arithmetic mean of the observations. In the present example, $G = 205.0$ ($\bar{M} = 233.4$).

In contrast to the wide variations in results of various regressions of butter expenditure on income obtained by Prais [8], the elasticities in Table III exhibit fairly small variations. This is explained by the fact they were derived from grouped data rather than individual budgets. Nevertheless, the differences are not negligible. In particular, those between (2a) and (2b) are greater than one might anticipate considering that the mathematical forms of the relationships are the same and only the error specifications are different. (1) gives a higher elasticity for the necessities "Food" and "Fuel and light" than the other functions. The main differences between (2b), (3), and (4) appear in the results for the two groups with the highest demand elasticities—"Household durables" and "Travel and holidays"—for which (3) gives the highest result, (2b) the lowest result, and (4) gives an intermediate result.

In the example given above, the fit could, of course, be further improved if a three-parameter Engel function was used. However, with only two degrees of freedom available for the error variance, one would hardly be inclined to increase the number of parameters. The position is different when a larger number of observations are available. In such a case, one may well ask whether it would not be worthwhile to use a function which is more flexible than those containing two parameters only.

Ideally, one would like the Engel function to be universally valid. One might envisage an initial expenditure pattern, given by dv_i/dM and w_i which are equal in the neighbourhood of $M=0$. Then dv_i/dM and w_i constantly change and will differ, but eventually both tend toward a limiting "ideal" or "final" expenditure pattern. The changes in dv_i/dM and w_i might be of a fairly complex nature. In practice, it would not be easy to find a simple Engel function fitting into this scheme, nor could one expect to obtain reliable estimates of initial and final expenditure patterns.

A more modest approach consists in constructing a set of three-parameter functions which can be expected to describe variations in expenditure proportions other than regular increases or decreases. There are, of course, various possibilities open, which we shall not investigate. A combination of (2b) and (4) seems particularly attractive, yielding

$$w_i = \alpha_i + \beta_i \log M + \gamma_i/M + \varepsilon_i\,. \tag{5}$$

(5) may be considered a logical generalisation of (4) which implies (assuming natural logarithms are used)

$$\frac{dv_i}{dM} = (\alpha_i + \beta_i) + \beta_i \log M$$
$$v_i = 0 \text{ for } M = 0 .$$

But the last condition does not ensure the validity of (4) for low M and may well be dropped, in which case (5) is obtained. Since the estimation of β_i and γ_i is not of primary concern here, multicollinearity should not present a serious problem.

An interesting feature of (5) is that it permits for any one commodity group, the simultaneous testing of the two hypotheses: a) that marginal outlay is constant, and b) that the elasticity of demand is approximately constant. This is simply done by testing the significance of the partial regression coefficients β_i and γ_i. Both hypotheses may, of course, be rejected.

The procedure raises the question of how far it is permissible to use the residual variance in a significance test, when the correct formulation of the relationship is subject to some uncertainty. Also it is doubtful whether one would be inclined to equate nonsignificant regression coefficients with zero in a set of additive equations as this would involve adjustments to the significant coefficients. Nevertheless, a test of this kind should be of some interest.

The numerical data analysed to provide an example of this kind of analysis are taken from budget data for farm operator families of two or more persons in 1955, published for the United States [**10**]. The average family

TABLE IV

EXPENDITURE DISTRIBUTION FOR SEVEN INCOME GROUPS, U.S. FARMERS, 1955

Weights	1,111	480	469	822	594	663	173
M ($ per year)	1,887	2,077	2,513	2,952	3,395	4,459	6,560
w_i:							
1. Food and beverages	.3242	.3141	.3109	.2903	.2891	.2700	.2427
2. Tobacco	.0201	.0212	.0199	.0176	.0168	.0148	.0107
3. Dwelling upkeep	.0630	.0596	.0569	.0586	.0601	.0727	.0841
4. Housefurnishings and equipment	.0646	.0697	.0728	.0742	.0718	.0740	.0770
5. Fuel, light, etc.	.0720	.0712	.0713	.0681	.0654	.0574	.0489
6. Household operation	.0366	.0380	.0366	.0376	.0362	.0386	.0476
7. Clothing	.1319	.1352	.1469	.1457	.1352	.1413	.1483
8. Transportation	.0959	.1039	.1075	.1253	.1434	.1554	.1607
9. Medical care	.1038	.0943	.0848	.0854	.0807	.0722	.0666
10. Personal care	.0233	.0250	.0251	.0240	.0224	.0240	.0241
11. Recreation	.0339	.0370	.0382	.0427	.0465	.0453	.0492
12. Miscellaneous	.0307	.0308	.0291	.0305	.0324	.0343	.0401
Total	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

size is somewhat larger in the high-income groups than in the low-income groups but not very much, so that this factor is unlikely to introduce any appreciable bias. Combining a few small commodity groups with larger ones, the data are obtained as shown in Table IV.

Using the number of households represented as weights, regression equations (2b), (4), and (5) have been computed. Results for average elasticities of demand (for $M = G = 2,788$) together with coefficients of determination are shown below.

TABLE V

ESTIMATES OF AVERAGE INCOME ELASTICITIES OF DEMAND AND EXPLAINED PROPORTIONS OF VARIANCE, U.S. FARMERS, 1955

Commodity group	$\eta(G)$ estimated from			R^2 obtained from		
	(2b)	(4)	(5)	(2b)	(4)	(5)
1. Food and beverages	.779	.788	.788	.945	.973	.973
2. Tobacco	.595	.603	.635	.851	.930	.956
3. Dwelling upkeep	1.162	1.196	1.134	.275	.431	.941
4. Housefurnishings and equipment	1.149	1.136	1.154	.784	.690	.843
5. Fuel, light etc.	.748	.745	.777	.792	.906	.985
6. Household operation	1.090	1.106	1.074	.267	.394	.751
7. Clothing	1.082	1.073	1.086	.396	.336	.451
8. Transportation	1.540	1.532	1.540	.952	.944	.955
9. Medical care	.598	.631	.594	.963	.925	.964
10. Personal care	.996	.995	.997	.002	.003	.005
11. Recreation	1.357	1.339	1.359	.934	.876	.943
12. Miscellaneous	1.152	1.170	1.136	.491	.652	.955
Average	1.000	1.000	1.000			

The estimates for the average income elasticities are seen to be fairly robust; this does not, of course, apply to elasticities for very high or very low income. It may be worth noting that the estimates obtained by using (5) generally lie outside the range of the estimates provided by (2b) and (4).

The comparison of R^2 is most conveniently carried out through an analysis of variance. The findings are set out in the following.

First, the overall regression is not significant in the case of "Household operation," "Clothing," and "Personal care"; in other words, the direction in which the elasticity of demand differs from 1 is not clearly established, and the hypothesis of a constant expenditure proportion could be maintained. The overall regression is significant at the 5% level only for "Housefurnishings and equipment" and at the 1% level for all other commodity groups.

Secondly, in five of the nine groups for which the overall regression is significant, neither of the partial regressions is significant. The groups are: "Food," "Housefurnishings and equipment," "Transportation," "Medical care," and "Recreation." In these instances, neither the hypothesis of constant marginal outlay nor that of approximately constant demand elasticity can be conclusively rejected.

Thirdly, for "Tobacco" the partial regression on $\log M$ is significant at the 5% level, but the partial regression on $1/M$ is not significant. Thus, regression model (4), indicating a near-constant demand elasticity, seems adequate in this case, while model (2a), indicating constant marginal outlay, does not.

Finally, for "Dwelling upkeep," "Fuel and light etc.," and "Miscellaneous," both partial regressions are significant, though that on $1/M$ for "Fuel and light etc." only at the 5% level. For these groups, or at any rate for "Dwelling upkeep" and "Miscellaneous," the two-parameter Engel functions considered here do not appear to describe satisfactorily the variations in consumer behaviour over a moderately wide range of incomes.

One may hesitate before going too far in drawing general conclusions from one set of data which, it must be remembered, may contain errors in the everyday meaning as well as in the statistical meaning of the word. Nevertheless, it would not be contrary to common sense to believe that, in many communities, the proportion of total outlay devoted to housing diminishes at low to medium income levels but increases again further up on the scale. Furthermore, one may well imagine the elasticity of demand for fuel and light to be nearly constant at low but rapidly falling at high income levels; one can more readily think of buying expensive food than expensive fuel.

It is this kind of result which is derived from the data used here as an example, and which can only be described by a three-parameter Engel function. If (5) is adopted, then in addition to w_i and η_i for, say $M = G$, a further summary measure is required for description. As such, for example, the "average income elasticity of marginal outlay" could serve, which would be given by

$$\zeta_i(G) = \frac{b_i}{a_i + b_i(1 + \log G)}$$

where a_i, b_i are the estimates of α_i, β_i.

In the present example, the following values for $\zeta(G)$ would be obtained: Food and beverages, $-.27$; Tobacco, -1.21; Dwelling upkeep, $+1.36$; House furnishings and equipment, $-.20$; Fuel, light etc., $-.86$; Household operation, $+.66$; Clothing, $-.16$; Transportation, $+.13$; Medical care, $+.12$; Personal care, $-.03$; Recreation, $-.15$; Miscellaneous, $+.77$. The groups in which the numerically largest values are obtained are, of course, those for which the hypothesis of constant marginal outlay is rejected.

Marginal outlay tends to increase for "Dwelling upkeep" and "Household operation" and to decline for "Tobacco" and "Fuel, light etc."

Economic Research Institute, Dublin

REFERENCES

[1] Allen, R. G. D., and A. L. Bowley: *Family Expenditure*, London: P.S. King and Son, 1935.
[2] Goreux, L. M.: *Income Elasticity of the Demand for Food*, Economic Commission for Europe, Rome, 1959.
[3] Houthakker, H. S.: "The Influence of Prices and Incomes on Household Expenditure," *Bulletin de l'Institut International de Statistique*, Vol. 37 (1960), pp. 9–22.
[4] Ireland — Central Statistics Office: *Household budget enquiry* 1951–52, Dublin, Stationery Office, 1954.
[5] Leser, C. E. V.: *Demand Relationships for Ireland*, Economic Research Institute Paper No. 4, Dublin, 1962.
[6] Liviatan, N.: "Errors in Variables and Engel Curve Analysis," *Econometrica*, Vol. 29 (1961), pp. 336–362.
[7] Netherlands Central Bureau of Statistics: *Statistical studies*, No. 10, Zeist, 1960.
[8] Prais, S. J.: "Non-linear Estimates of the Engel Curve," *Review of Economic Studies*, Vol. 20 (1952–53), pp. 87–104.
[9] Prais, S. J., and H. S. Houthakker: *The Analysis of Family Budgets*, Cambridge University Press, 1955.
[10] United States Bureau of the Census: *Historical Statistics of the United States, Colonial Times to 1957*, Washington, 1960.
[11] Working, H.: "Statistical Laws of Family Expenditure," *Journal of the American Statistical Association*, Vol. 38 (1943), pp. 43–56.