Logistic Transformation of the Budget Share in Engel Curves and Demand Functions

DENIS CONNIFFE
The Economic and Social Research Institute

Abstract: Models relating the budget share of a commodity to (the logarithms) of income and prices are popular in the analysis of household budget survey data and of time series data on commodity expenditures. This paper shows that both the behavioural economic properties and the econometric, or statistical, properties of the models can be substantially improved by first employing a logistic transformation of the budget share. The paper mainly concerns estimation in a single equation context, but related issues of demand systems and utility theory are examined. The methodology is illustrated by analysis of some Irish household budget survey data.

I INTRODUCTION

In the analysis of household expenditure data, one popular functional form relates the budget share, w, of a commodity to the logarithm of total expenditure, y, by the equation

$$w = a + b \ln y \tag{1}$$

There is an Irish connection here. Although (1) originated with Working (1943), it was revived by Leser (1963) when he was working at the ESRI and was employed by him in analyses of the CSO's 1951-52 Household Budget Enquiry (Leser, 1964). The equation is obviously easily extended by adding the logarithms of other possibly relevant explanatory variables to the right hand side of (1). Leser (1974) did this himself in the case of a household size variable. For analysis of time series data the obvious extension of (1) is

Paper presented at the Seventh Annual Conference of the Irish Economic Association.

$$w = a + b \ln (y/D) + c \ln (p/\bar{p}) \qquad (2)$$

where p is the price of the commodity and D and $\bar{p}$ are deflators, the former based on a general price index and the latter on an index based on the prices of all the other competing commodities. However, unless the commodity is very broad so that its expenditure share is substantial, D can be taken as $\bar{p}$.

When a set of Engel curves are being estimated, shares must sum to unity, suggesting that the set of equations

$$w_i = a_i + b_i \ln y,$$

where the subscript i refers to commodity, should be estimated subject to the constraints $\Sigma a_i = 1$ and $\Sigma b_i = 0$. The corresponding multiequation extension of (2) is the AIDS model of Deaton and Muellbauer (1980)

$$w_i = a_i + b_i \ln (y/D) + \Sigma_j c_{ij} \ln p_j, \qquad (3)$$

with D usually calculated as the weighted geometric index over all commodity prices

$$\ln D = \Sigma_j w_j \ln p_j. \qquad (4)$$

When estimating the set of equations (3) the extra constraints of

$$\Sigma_i c_{ij} = \Sigma_j c_{ij} = 0 \text{ and } c_{ij} = c_{ji}$$

are required if the demand theory conditions of homogeneity and symmetry are to hold.

Most of this paper will be concerned with the case where only a single equation like (1) or (2) is of interest, rather than that where a complete demand system is being considered. However, it should be noted that an AIDS model for a system consisting of just two commodities reduces to Equation (2) with p and $\bar{p}$ the commodity prices. The second equation of the system is given by subtracting (2) from unity. This point will be returned to subsequently.

This paper will argue that (1) and (2) can be greatly improved upon by taking a logistic transformation of the budget share. The improvement applies not only to the statistical properties of the relationships, but also to their interpretation in terms of plausible economic behaviour.

II STATISTICAL ISSUES AND PROPERTIES

In spite of the antiquity and frequency of (1) and (2), there are funda-mental implausibilities about the models from a statistical viewpoint if the

explanatory variables are to be permitted to vary beyond a limited range. The dependent variable w must take values between 0 and 1, and probably is much more limited within these bounds, and yet it is being related to the logarithm of income which can theoretically vary from 0 to ∞. One consequence is that if (1) was first estimated from data with a limited range of income variation and later re-estimated from data with a far greater range, the second estimate of b would be less than the first. There are various other difficulties too, including predicted ws lying outside the 0 to 1 range and the inevitable heteroscedasticity associated with proportions.

This is a basic dimensionality fault with the models and obviously cannot be corrected by just multiplying both sides of the equations by y. In demand studies with time series data, consecutive changes in (usually aggregate) income may be small so that model (2) may not seem inadequate over a short range of years. Household budget surveys, however, usually exhibit large income differences, often by design, so that model (1) applied to cross-sectional data, or model (2) applied to a time series of cross-sections, could be very suspect from a statistical viewpoint. Matters would be made even worse by the inclusion of other explanatory variables that also have unlimited ranges, something that is not at all uncommon in the applications literature.

There is nothing particularly original in the comments just made about the unpalatable statistical properties of taking budget shares as dependent variables. The same issues arise in many situations where ratios, relative frequencies, or estimated probabilities occur as dependent variables and there is a substantial statistical literature relating to them. It is usually considered to be statistical good practice to transform such variables before analysis and arc-sine-root, probit and logit have been advocated in various situations. The transformation

$$\ln \{w/(1-w)\} \qquad (5)$$

seems specially appropriate for budget shares. The dependent variable now has a potential range from $-\infty$ to $+\infty$ and relating (5) to a linear function of the logarithm of income is reasonable. The conventional assumptions of an additive disturbance term satisfying most of the usual OLS conditions are also much more plausible with transformed than with the untransformed shares.

III BEHAVIOURAL PROPERTIES OF THE MODELS

The fact that one model may be preferable to another on the basis of statistical properties is not the only consideration of interest. The models should comply with plausible economic behaviour. The Engel curve

corresponding to the transformation (5) is

$$\ln \frac{w}{1-w} = a^* + b^* \ln y \qquad (6)$$

How do (1) and (6) compare in terms of economic behaviour? The income elasticity corresponding to (1) is

$$1 + \frac{b}{w} \qquad (7)$$

Clearly a commodity is a necessity only if b is negative. But since budget share decreases with income for a necessity, all necessities become inferior goods as income increases. This may not matter too much if incomes are restricted sufficiently to prevent this occurring, but otherwise it is a most disconcerting property.

The model (6) gives an income elasticity of

$$1 + b^* (1 - w) \qquad (8)$$

and once again the commodity is a necessity if b^* is negative. But it does not become an inferior good no matter how large y becomes, provided b^* is greater than minus one.

Turning to Equation (2) and the transformed version of that model

$$\ln \frac{w}{1-w} = a^* + b^* \ln (y/D) + c^* \ln (p/\bar{p}) \qquad (9)$$

the same comments apply as regards income elasticities. For own price elasticity model (2) gives

$$\frac{c}{w} - 1 - b \qquad (10)$$

This is also very unappealing for a necessity, since its magnitude tends to infinity for large incomes. For model (9), the elasticity is the much more acceptable

$$c^* (1 - w) - 1 - b^* w (1 - w). \qquad (11)$$

In deriving the price elasticities, D in Equations (2) and (9) has been taken as given by (4), but it should be said that this choice has nothing to do with behavioural plausibility or implausibility; replacing D by $\bar{p}$ would just remove the third terms in (9) and (10).

So it seems a logistic transformation of the budget share not only has the potential to improve statistical properties, but to improve behavioural economic properties too.

IV DEMAND SYSTEMS AND UTILITY

The justification for certain demand systems can follow from arguments based on utility maximisation, or the dual formulation of cost minimisation, and the AIDS model has been argued to be derivable from optimisation of a viable approximation to an arbitrary cost function. It was remarked in the introduction that model (2) is the special two-commodity case of the AIDS model. So, does utility theory lend more support to untransformed than to transformed budget shares? In fact, it is easy to show that the transformed equations follow directly from a known demand system. The additive indirect utility function (Houthakker, 1960)

$$U = \Sigma_j f_j \left(\frac{y}{p_j} \right)^{h_j}$$

leads to the demand equations

$$w_i = \frac{g_i (y / p_i)^{h_i}}{\Sigma_j g_j (y / p_j)^{h_j}} \tag{12}$$

where $g_i = f_i h_i$. For a two commodity system with budget shares w and $1 - w$, and prices p and $\bar{p}$, dividing the equations and taking logarithms gives

$$\ln \frac{w}{1 - w} = a + h_1 \ln \frac{y}{p} - h_2 \ln \frac{y}{\bar{p}}$$

where

$$a = \ln (g_1/g_2)$$

This equation can be rewritten as

$$\ln \frac{w}{1-w} = a + b^* \ln \frac{y}{\bar{p}} + c^* \ln \frac{p}{\bar{p}} \tag{13}$$

where

$$b^* = h_1 - h_2 \quad c^* = -h_1$$

At constant prices (13) becomes the Engel curve (6) and otherwise it is the demand equation (9) with D taken as $\bar{p}$. So the logistic transformation of the share can be regarded as a two commodity indirect addilog system, applying to the commodity and the rest of expenditure. Thus, the plausible economic behaviour noted in the previous section is quite compatible with utility theory. The less plausible behaviour of the AIDS model may perhaps be related to its utility justification being dependent on an approximation argument, which may be least accurate when income variation is large.

Indeed, looking at the two-commodity AIDS system more closely, reveals even more problems. It has already been said that the own price elasticity (10) is implausible in tending to infinity for high incomes in the case of a necessity. The cross-price elasticity on $\bar{p}$ displays the same problem for a necessity. The problem does not occur for a luxury, but then the other commodity must be a necessity and the problem reappears for its price elasticities. It is easily verified that no such difficulties arise with the transformed model.

Such breakdowns of total expenditure, or of some separable category of expenditure (with consequent redefinition of y), are not uncommon. For example, Deaton and Muellbauer (1986) and Deaton, Ruiz-Castillo and Thomas (1989) considered such budget breakdowns as food and non-food, or adult and non-adult goods. However, this paper is concerned with the common situation where a particular commodity is the subject of interest, and the demand system in which it is embedded is not explicitly considered. The reason for including this section in this paper is to counter the idea that an untransformed share equation has greater prior support from utility theory than has a transformed share equation.

V AN ENGEL CURVE EXAMPLE

The data are drawn from the 1987 Household Budget Survey and are based on 36 groups of households. The commodity chosen for examination corresponds to the classification "fuel and light" and "income" is measured as total expenditure. Fitting model (1) gives a b value of −.064 with a t value of −13.0. So, using (7), the commodity is a necessity and will become an inferior

good when w becomes small enough. In fact, the mean value of w in the data was .084, while the minimum was .032, so it is clear the commodity becomes an inferior good not very far above average income. Fitting model (6) gave a b^* value of $-.917$ with a t value of -13.6. So, using (8), the commodity is again a necessity, but stays positive even when w becomes small (or incomes large). The comparisons are made clear by Figures 1 and 2. Figure 1 plots the shares as predicted by the two models against income (total expenditure).
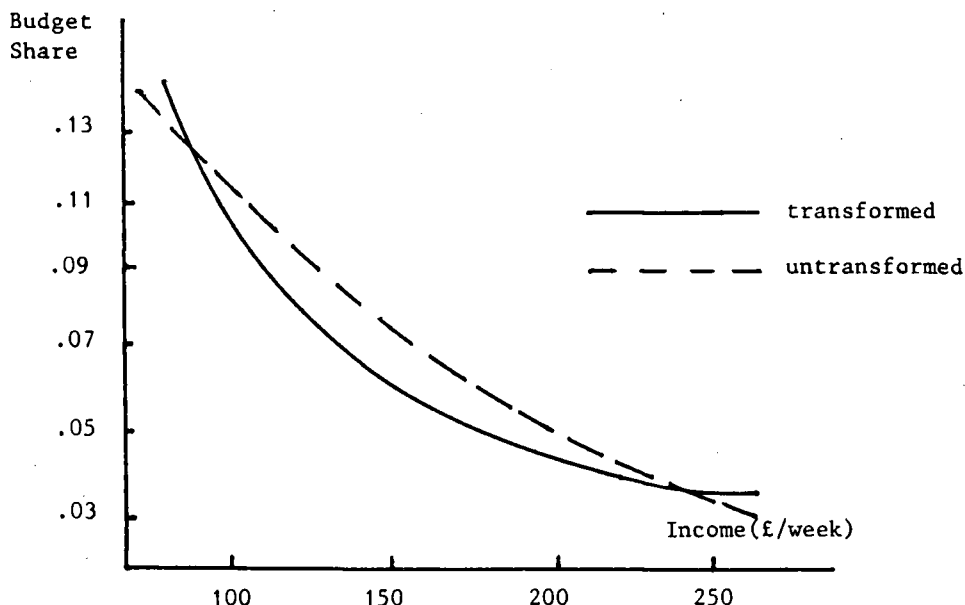

Figure 1: *Budget Share v Income for Transformed and Untransformed Models*

Both curves slope downwards of course, but the curvature is greater for the logistic transformation curve so that it lies above the untransformed curve at low and high incomes. Even if projected to ever higher incomes it will not cross to a negative share, which the untransformed budget share will. Figure 2 plots the expenditures as predicted by the models against income. The untransformed model forces expenditure on fuel and light to fall at higher incomes, making the commodity an inferior good. Mean income for all households was approximately £170 pounds per week and the fall commences between this and £200 per week. In contrast, the logistic transformation model never forces expenditure on the commodity downwards.
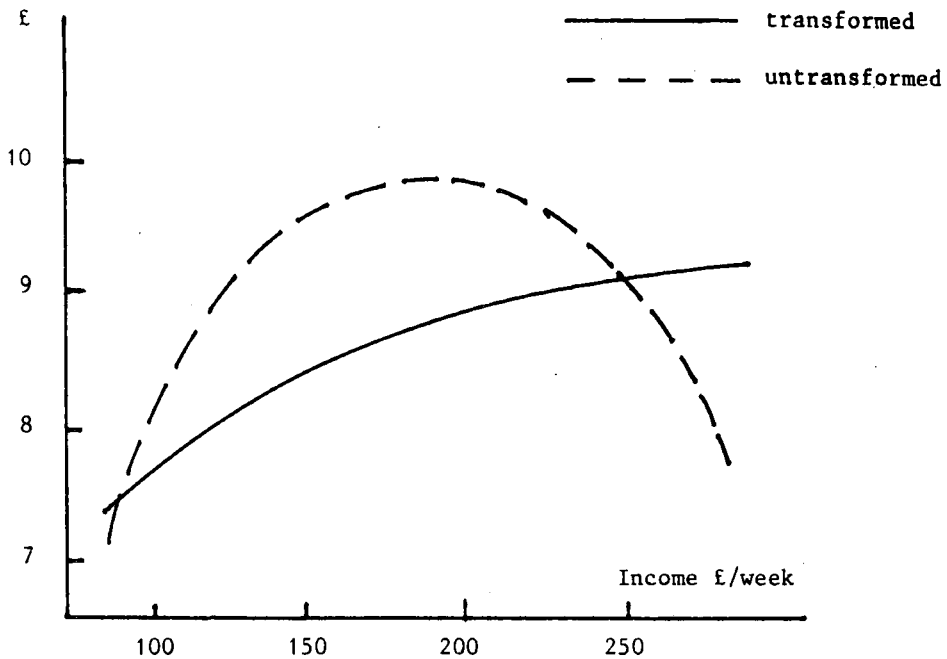
Figure 2: *Expenditure v Income For Transformed and Untransformed Models*

REFERENCES

DEATON, A. and J. MUELLBAUER, 1980. "An Almost Ideal Demand System", *American Economic Review*, Vol. 70, pp. 312-336.

DEATON, A., and J. MUELLBAUER, 1986. "On Measuring Child Costs: With Applications to Poor Countries", *Journal of Political Economy*, Vol. 94, pp. 720-744.

DEATON, A., J. RUIZ-CASTILLO and D. THOMAS, 1989. "The Influence of Household Composition on Household Expenditure Patterns: Theory and Spanish Evidence", *Journal of Political Economy*, Vol. 97, pp. 179-200.

HOUTHAKKER, H.S., 1960. "Additive Preferences", *Econometrica*, Vol. 28, pp. 244-257.

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

LESER, C.E.V., 1964. "A Further Analysis of Irish Household Budget Data, 1951-1952", Dublin: The Economic and Social Research Institute, Paper No. 23.

WORKING, H., 1943. "Statistical Laws of Family Expenditure", *Journal of the American Statistical Association*, Vol. 38, pp. 43-56.