National Institute for Regional and Spatial Analysis

NIRSA

Working Paper Series
8-May02

Does the Stochastic Specification of the Linear Expenditure System Matter?

by

DENIS CONNIFFE[*]

NIRSA and Department of Economics
NUI Maynooth

and

JOHN EAKINS

The Economic and Social Research Institute

National University of Ireland, Maynooth
Maynooth
Co. Kildare
Ireland
nirsa@may.ie
www.may.ie/nirsa

[*] Address for correspondence: Denis Conniffe, National Institute for Regional and Spatial Analysis, Economics Department, NUI Maynooth, Maynooth, Co. Kildare, Ireland.

NIRSA is a research institute dedicated to the inter-disciplinary and comparative study of the impact of global processes on regional and spatial development. It involves several third-level institutions in Ireland: NUI Maynooth, Galway-Mayo Institute of Technology, Institute of Technology, Sligo, Mary Immaculate College - University of Limerick, Waterford Institute of Technology.

ABSTRACT

When 'income' in a system of demand equations is defined as total expenditure, actual expenditure on any commodity must lie between zero and income, or equivalently, budget shares must lie between zero and one. But models for expenditures or shares are often the sum of deterministic components (predicted values), which are functions of prices and income, and disturbances, usually assumed multivariate normal. The predicted values ought to satisfy the same bounds as the dependent variables and will do so if the demand system is 'regular'. But even then, the situation is theoretically inconsistent with unbounded disturbances and it has been proposed (Fry, et al, 1996) that analysis be appropriately modified. In assessing how much practical difference this makes, the linear expenditure system (LES) is, for reasons described in the paper, the crucial case. We compare estimation methods for the LES, using Irish data from 1979-99 on some broadly defined commodities, and find that the differences are not of practical concern.

I INTRODUCTION

When income, y, in a system of demand equations is defined as total expenditure $\Sigma p_j q_j$, where q_i and p_i are the quantity and price of the i th commodity, it is obvious that the actual expenditure on any commodity must lie between zero and y, or equivalently, budget shares must lie between zero and one. However, such systems are usually modelled by the sets of equations

$$p_i q_i = f_i(\mathbf{p}, \mathrm{y}) + \mathrm{e_i}, \tag{1}$$

or

$$w_i = g_i(\mathbf{p}, \mathrm{y}) + \mathrm{u_i}, \tag{2}$$

where w_i is the i th budget share. With n commodities, each set contains n-1 equations as the adding up condition of $\Sigma p_j q_j = y$ for (1), or $\Sigma w_j = 1$ for (2), ensures the n th equation is deducible by difference. Clearly, the deterministic components f_i and g_i of these models ought to conform to the constraints and will if the demand system is globally *regular*[1], but even then the usual assumptions made about the stochastic disturbances e_i and u_i − that they are randomly drawn from a multivariate normal distribution – are evidently not precisely appropriate. For example, w_i in (2) cannot exceed unity, but even with a g_i below unity, a u_i drawn from a normal distribution, with its infinite range, could possibly result in a sum greater than unity and hence inconsistency between the left and right hand sides of (2). Of course, in practice, the fitted multivariate normal might well have variances so small that the probability of a u_i being so large might be negligible and it is probably on this presumption that authors have usually ignored the problem.

If the deterministic components do not automatically conform to the constraints, the likelihood of difficulties is far greater. For example, an equation for a single good of the form[2]

$$w = a + b_y \log \frac{y}{d_1} + b_p \log \frac{p}{d_2} + u, \tag{3}$$

---

[1] Systems are *globally regular* if they meet the demand theory conditions implied by utility maximisation (subject to a budget constraint) for all prices and income, although for practical purposes 'all' can be relaxed to 'all relevant'. Regularity implies constraints on the parameters of the utility function, but even so, few demand systems are regular for all relevant prices and incomes.

[2] This is the Working (1943) or Leser (1963) form, which becomes Deaton and Muellbauers' (1980) AIDS model when extended to multiple commodities.

where d_1 and d_2 are price deflators, must, as y increases, inevitably either exceed unity or become less than zero, depending on whether b is positive or negative. For this and other reasons, Conniffe (1993) argued that a logistic transformation of the budget share should replace w in (3) giving

$$\log \frac{w}{1-w} = a + b_y \log \frac{y}{d_1} + b_p \log \frac{p}{d_2} + u \ . \qquad (4)$$

Now the dependent variable can take any positive or negative values like the deterministic part of the right hand side. The model is also far more compatible with a normality assumption for u, since the dependent variable can, theoretically, range from $-\infty$ to $+\infty$, but the motivation for (4) was principally[3] the incompatibility of the dependent variable and the deterministic term in (3).

Fry, Fry and McLaren (1996) discuss the treatment of stochastic terms in the estimation of *regular* demand systems, when the g_i in (2) are sure to be between zero and unity, and argue for estimation of the n-1 equation model

$$\log \frac{w_i}{w_n} = \log \frac{g_i}{g_n} + u_i \ , \qquad (5)$$

instead of (2), assuming multivariate normality of the u_i. (The choice of the n th good for the denominator is arbitrary). From a data analysis viewpoint, it is undeniable that multivariate normality is a more plausible operational assumption if choosing model (5) rather than model (2), for the reasons already stated in the case of (4). More theoretically, if we visualise (5) as the true model generating the w_i, it is clear that they will lie between zero and unity. So it is appealing to work with the form (5) and to suspect there could have been errors introduced by failure to do so in the past. That need not mean that research with the forms (1) or (2) has to be have been seriously incorrect, but it would seem well worth checking out.

In fact, many of the commonly employed demand systems do not satisfy (near) global regularity and for them the form (5) could be quite unsuitable, as Fry et al appreciated, because the g_i might not be appropriately bounded. The linear expenditure system (LES) and the indirect addilog system are the only (near) globally regular systems (given essential constraints on the parameters) that have been frequently employed in applications. There have been considerable theoretical efforts to find other systems[4] with good regularity

---

[3] Utility theory justification for (4) follows from considering it a two equation case (a commodity and all other commodities, so that $w_2 = 1 - w_1$) of Houthakker's (1960) indirect addilog system.

[4] Fry et al (1996) mention the MAIDS system of Cooper and McClaren (1992), but, other than the application by Boyle (1996), this has not featured in the applied literature.

properties, but it is unclear how much progress of practical importance has been made. As regards the indirect addilog system, it has always been estimated in the form (5) anyway, not (at least explicitly) because of concern about the formulation of the stochastic terms, but because it was is computationally convenient to do so[5]. So interest must centre on how LES estimation is affected by the choice of (5) rather than (1) or (2). The LES is certainly an important system, in spite of the limitation that its assumptions are strictly only appropriate for broadly defined commodities of a non-durable nature. It has been popular with Irish researchers since the seventies (Casey, 1973; O'Riordan, 1976; McCarthy, 1977) and is still employed. For example, the ESRI (Duffy et al, 2001) review and forecast of the Irish economy was based on methodology incorporating an LES for the household consumption sector.

[5] The indirect addilog equations have the form $w_i = \gamma_i (\frac{y}{p_i})^{\beta_i} / \sum_j \gamma_j (\frac{y}{p_j})^{\beta_j}$ and it is clear that dividing w_i by w_n and taking logs cancels the denominators and leaves linear equations in the logs of income and prices.

II ESTIMATING THE LES

The LES is usually considered in expenditure form

$$p_i q_i = \gamma_i p_i + \beta_i (y - \Sigma\gamma_j p_j) + u_i \,, \tag{6}$$

where regularity is assured if γ_i are positive, β_i positive and adding to unity over the n commodities and $y > \Sigma\gamma_j p_j$. Sometimes the budget share form

$$w_i = \frac{\gamma_i p_i}{y} + \beta_i (1 - \frac{\Sigma\gamma_j p_j}{y}) + u_i \,, \tag{7}$$

is employed. As regards estimation in either case the n th equation is omitted (and deduced by difference) to avoid singularity resulting from the adding-up constraint. However this is not the only way to proceed. Working with the n-1 equations

$$\frac{w_i}{w_n} = \frac{p_i q_i}{p_n q_n} = \frac{\gamma_i p_i + \beta_i (y - \Sigma\gamma_j p_j)}{\gamma_n p_n + (1 - \Sigma^* \beta_j)(y - \Sigma\gamma_n p_n)} + u_i \,, \tag{8}$$

where Σ^* denotes summation excluding j = n, or the equations

$$\log \frac{w_i}{w_n} = \log \left\{ \frac{\gamma_i p_i + \beta_i (y - \Sigma\gamma_j p_j)}{\gamma_n p_n + (1 - \Sigma^* \beta_j)(y - \Sigma\gamma_n p_n)} \right\} + u_i \tag{9}$$

are also effective ways of accounting for the adding up constraint. Of course, (9), where the dependent variable can range from $-\infty$ to $+\infty$, is the preferred form for Fry et al (1996). For (6) and (7) the dependent variables are bounded above and below, while in (8) the dependent variable can range from 0 to $+\infty$.

Maximum likelihood has been, and remains, the dominant estimation method in applied economics and by far the most frequent assumption about the likelihood is that it is multivariate normal. Indeed, many econometric packages do not permit any other assumption when providing estimation routines for *non-linear systems* of equations. Systems (6), (7), (8) and (9) are really identical as regards deterministic components, but as they differ in how the stochastic and deterministic components combine, estimation involves the maximisation of rather different likelihoods for each case. So it is reasonable to think that estimates of coefficients could be affected to some degree, in terms of bias or precision or both, by the choice of model, and it is interesting to see if this will matter in practice.

One obvious approach to comparing (6), (7), (8) and (9) would be through a simulation study, generating the data from exact LES equations for the deterministic components and combining samples from an exact multivariate normal for the stochastic components, and then comparing the distributions of estimates of the parameters. To at least some extent it is intuitively clear what would result. If the variance matrix of the multivariate normal is 'small' (in the sense that the diagonal terms are) so that deterministic components greatly outweigh the stochastic components, there will be no difference, while if the reverse holds, there will be. But this is not satisfactory as a practical assessment. No one believes that consumer demand is precisely represented, even as regards deterministic components, by the LES – at best it is a reasonable approximation for some broad commodities. Nor would anyone believe that with real-world data, exact multinormality is at all plausible. What matters for practical purposes is whether choice of (6), (7), (8) and (9) makes any difference with the sort of data set typically analysed by applied economists.

III DATA AND ANALYSES

Time series of domestic expenditures on commodities at current and constant prices are available from the Irish Central Statistics Office's *National Income and Expenditure Accounts*. Dividing current by constant series gives price indices for commodities. Five broad commodities – food, alcohol, clothing, energy (domestic fuels) and other non-durable goods – were chosen for the 21 years 1979 to 1999. Other commodities could have been added and the time scale extended back, although there could have been corresponding weakening of the plausibility of the LES framework[6]. In terms of composition and number of observations the data set is quite typical of those to which Irish researchers have applied the LES.

All the models 6, 7, 8 and 9 are non-linear in the parameters and so maximum likelihood estimation requires an iterative approach. We employed the SHAZAM (2001) package, which iterates from initial 'guesstimates' to some maximum of the likelihood function. From a computational viewpoint, the models differ in their complexity and it is more difficult to find the maximum likelihood estimate for 9 than for 6 or 7. This is not a matter of number of iterations, which is not a concern with modern computing power, but because convergence to local maxima, rather than to the global maximum, can occur with non-linear estimation routines and it is important to either start from an estimate known to be close to the global maximum or to take many starting points and compare likelihood values at convergence. With model 9 it seems particularly important to have a good initial estimate. It is also worth noting that the problem of trying to take the logarithm of a negative could possibly arise in the course of iterative solution of model 9. Although the LES is regular, given the requirements for positive parameters, SHAZAM does not constrain estimates of parameters to remain positive through all iterations. Nor should it, because it is always possible that consumers are not behaving (or are not appearing to) in accordance with utility maximisation, which could be signalled by a negative β_i in the maximum likelihood solution. Under such circumstances, a negative predicted value could arise. Packages differ in how an undefined operation like log of a negative is handled – often with a warning and the setting of the 'result' to zero, but continued iteration.

[6] These are well known considerations: a fine division of commodities would be incompatible with the LES's inability to reflect specific substitution and complimentary effects; inclusion of durable goods might necessitate extra terms in demand equations; long time series risk structural change or instability of parameters; etc.

However, we had no such problem with our data. Having obtained and carefully checked the locations of the global maxima, we found, somewhat to our surprise, that they were remarkably similar for all models.

Table 1 shows the estimates of the β parameters. β_5 was not actually estimated, but obtained by difference from unity and is just included for completeness.

Table 1: *Estimates of β Parameters with standard errors*

Model	β_1	β_2	β_3	β_4	β_5
6	.1928 (.0109)	.2866 (.0159)	.2415 (.0108)	.0500 (.0055)	.2291
7	.1923 (.0129)	.2768 (.0136)	.2389 (.0085)	.0616 (.0062)	.2304
8	.1941 (.0127)	.2775 (.0125)	.2375 (.0086)	.0626 (.0052)	.2283
9	.1945 (.0106)	.2763 (.0122)	.2395 (.0082)	.0584 (.0054)	.2313

The β estimates are almost all equal across models to the second place of decimals. Of course, real interest in demand studies usually focuses on elasticities rather than model parameters, but these are functions of the parameters, with the LES income elasticities, for example, equal to the β_i divided by budget shares. For example, the income elasticity of food (at the 1999 end-point) calculated from model 6 is .582 and calculated from model 9 it is .587. Standard errors obtained from maximum likelihood solutions of non-linear models are obtained from formulae that are only asymptotically valid and may differ from true finite sample standard errors. However, they should still be useful for relative comparisons and again there are no appreciable differences.

Continuing to the γ parameters, estimates are shown in Table 2.

Table 2: *Estimates of γ Parameters with standard errors*

Model	γ_1	γ_2	γ_3	γ_4	γ_5
6	835.8 (39.60)	286.2 (79.77)	153.7 (35.89)	176.0 (10.27)	228.0 (46.40)
7	821.9 (25.97)	273.9 (52.80)	137.9 (21.55)	159.0 (8.04)	207.2 (30.55)
8	815.6 (23.96)	264.7 (48.00)	133.4 (18.94)	156.5 (7.74)	202.9 (26.88)
9	818.3 (23.99)	271.9 (46.15)	135.7 (19.00)	161.8 (7.75)	204.5 (27.08)

Again differences are small with almost all estimates across models 7, 8 and 9 equal to two significant digits. For model 6 – the LES in expenditure form – estimates do seem slightly larger than for the other three models, but the magnitudes make no practical difference. For example, the own-price elasticity[7] of food (at the 1999 end-point) calculated from model 6 is -.44 and calculated from model 9 it is -.45. For standard errors the dominating difference is between model 6, where standard errors do seem larger and models 7, 8 and 9, within which differences are not appreciable. This contradicts the idea that the issue of bounds and multinormality can matter much, because 7 is just as suspect as 6 in that regard. Probably the data conditioning by scaling involved in all models except 6 is responsible.

[7] The formula being $-1+\gamma_i(1-\beta_i)/q$.

IV CONCLUDING REMARKS

As we have already indicated, we do agree with Fry et al (1996) that a theoretical case can be made for logistic transformation of shares and, in particular, for estimating the LES in the form (9), when we are confident of the utility maximisation context. While we did find it intuitively plausible that estimates of parameters should be affected by the treatment of the stochastic terms, the actual magnitudes of differences between parameter estimates and standard errors just do not appear to be of appreciable practical importance. But if this is disappointing in terms of return to increased sophistication of analysis, it is reassuring about the content and quality of past research findings.

Should (9) always be estimated on the grounds that, even if there are no practically important differences in estimates, it is still the theoretically preferred model? Not on its own though, partly because good starting estimates are needed with (9) to find the global maximum likelihood solution quickly, and also there is the possibility that the solution is not compatible with utility maximisation. If some true β_i are negative, iterative solution of (9) could become meaningless. So it would seem (7), taking account of the possibly less precise estimates of the γ_i by (6), ought always be solved before (9).

We analysed annual time series data aggregated over households. While this accords with the practice of all other researchers who have applied the LES to Irish data, demand equations and sometimes complete systems are also estimated on more disaggregated data. It is probably always true that much greater variation in commodity consumptions exist at household or individual levels with even zero consumptions possible with finely disaggregated goods. Then variances of distributions are much larger and stochastic assumptions can matter greatly and, of course, there is a considerable literature on some issues, even where the focus is on a single commodity, rather than on a systems context. But the deterministic components of models capable of adequately representing household variation will not be the LES's simple linear functions of prices and income. Issues connected to regularity reappear and the objectives of dependent variable transformation can be as much or more about improving the plausibility of the deterministic term as about the appropriateness of the stochastic formulation.

REFERENCES

BOYLE, G. E. (1996) "A MAIDS Model of Meat Demand", *The Economic and Social Review*, Vol. 27: pp 309-319.

CASEY, M. G. (1973) "An Application of the Samuelson-Stone Linear Expenditure System to Food Consumption in Ireland", *The Economic and Social Review*, Vol. 4: pp 309-342.

CONNIFFE, D. (1993) "Logistic Transformation of the Budget Share in Engel Curves and Demand Functions", *The Economic and Social Review*, Vol. 25: pp 49-56.

COOPER R. J. and K. R. McLAREN, (1992) "An Empirically Oriented Demand System with Improved Regularity Properties", *Canadian Journal of Economics*, Vol. 25: pp 653-668.

DEATON, A. and J. MUELLBAUER, (1980) "An Almost Ideal Demand System", *American Economic Review*, Vol. 70: pp 312-316.

DUFFY, D., J. FitzGERALD, I. KEARNEY, J. HORE and C. MacCOILLE, (2001) *Medium Term Review: 2001-2007*, Dublin: ESRI.

FRY, J. M., T. R. L. FRY and K. R. McLAREN, (1996) "The stochastic specification of demand share equations: restricting budget shares to the unit simplex", *Journal of Econometrics*, Vol. 73: pp 377-385.

HOUTHAKKER, H. S. (1960) "Additive Preferences", *Econometrica*, Vol. 28: pp 244-256.

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

McCARTHY, C. (1977) "Estimates of a System of Demand Equations using Alternative Commodity Classifications of Irish Data, 1953-1974", *The Economic and Social Review*, Vol. 8: pp 201-211.

O'RIORDAN, W. (1976) "Consumer Response to Price and Income Change", *Journal of the Statistical and Social Inquiry Society of Ireland*, Vol. 23: pp 65-89.

SHAZAM, (2001) *User's Reference Manual Version 9.0,* Vancouver: Northwest Econometrics.

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