Journal of Quantitative Economics, Vol. 4, No. 2 (July 1988), 329-339

ESTIMATION OF ADDITIVE ENGEL CURVES IN THE PRESENCE OF HETEROSCEDASTICITY*

N. SREENIVASA IYENGER

and

SHAILAJA RAGHUPRASAD

Indian Statistical Institute, Bangalore-560 059

This study is mainly concerned with the empirical determination of additive Engel curves from disaggregated National Sample Survey (NSS) data for the Indian State of Karnataka, using an extended Working-Leser model suitably adjusted for heteroscedasticity.

1. INTRODUCTION

Various functional forms of Engel curves have been used from time to time in the literature on consumer behaviour. Engel curves of the 'price-independent generalised log-linear (PIGLOG)' form have been popularised recently by Deaton and Muellbauer (1980), who used them to construct what is widely

*This study is part of a research project on regional demand patterns carried out at the Economic Analysis Unit, Indian Statistical Institute, Bangalore Centre. The authors wish to thank the authorities of the NSS Organisation for allowing free access to their unpublished data. Grateful acknowledgements are due to an anonymous referee who made valuable suggestions for improvement of an earlier version, presented at the 24th Indian Econometric Conference, Ahmedabad, January 8-10, 1987. Thanks are also due to B. V. Srikantiah for providing typographical assistance.

known as the 'almost ideal demand system (AIDS)'. The Working-Leser model appears as a special case of the PIGLOG family. Ray (1980, 1982, 1985), Murty (1984) and Majumder (1986) have all used the AIDS to describe the consumer behaviour in India. However, in all these studies the problem of heteroscedasticity has not received adequate attention. But, as is well known, the presence of heteroscedasticity in data can have serious implications for inferences about the Engel parameters.

Also, studies on consumer behaviour in India have been, for the most part, based on grouped data (Iyengar, 1963, 1964; Coondoo, 1969; Bhattacharya and Maitra, 1969; Jain and Tendulkar, 1972; Murthy, 1980; Murty, 1984, Ray, 1985; Majumder, 1986).

In this paper we propose to examine the implications of ignoring the heteroscedasticity problem in the estimation of the Working-Leser model using disaggregated household data on consumption. In Section 2 a brief description of the data is given. The outline of the methodology used is given in Section 3. Our main results are briefly discussed in Section 4. Finally, Section 5 sums up the conclusions.

2. THE DATA

The study uses disaggregated household expenditure data for the State of Karnataka collected by the NSS Organisation (NSSO) in its 32nd Round (1977-78).[1] Four regions, namely Inland Northern, Inland Eastern, Inland Southern, Coastal and Ghats, are distinguished and following ten broad commodity categories are considered : (1) cereals and pulses, (2) milk and milk products, (3) vegetables, fruits and nuts, (4) sugar and gur, (5) pan, tobacco, and intoxicants, (6) other foods, (7) fuel and light, (8) clothing, including footwear and bedding, (9) durables, and (10) miscellaneous.

3. METHODOLOGY

The simplest Working-Leser form regresses the share of household's expenditure on a specific item of consumption on its total outlay :

$$w_i = \alpha_i + \beta_i \ln E + e, \quad (i = 1, 2, \ldots, 10) \tag{1}$$

where E is total outlay and w_i is the budget share of the ith item in total outlay. Of course, it is assumed that e is i.i.d. normal, with zero mean and variance σ^2.

1. For a detailed description of the data, concepts and definitions, see *National Sample Survey Report* (1985).

Since $\Sigma w_i = 1$, we should have $\Sigma\alpha_i = 1$ and $\Sigma\beta_i = 0$. It is well known that if (1) is estimated equation by equation by ordinary least squares, then the above adding-up criterion is automatically satisfied. The Engel curves in this form are easily derivable from the generalised indirect utility function,

$$u = \psi_1 \times \frac{\psi_2}{\ln E - \psi_3} \tag{2}$$

for appropriately chosen functions ψ_1, ψ_2 and ψ_3 (see, Gorman, 1981).

The price vector p may also be introduced into equation (2), in which case the parameters α_i and β_i become functions of p. Consequently, ψ_1, ψ_2 and ψ_3 also become functions of p (Muellbauer, 1980). In cross-section based studies, prices are assumed constant across households. The PIGLOG model (1) has no price variables in it. It does not also explicitly include the demographic variables such as household size and composition which appear important in a cross-section study.

Household size (N) can easily be brought into the model either in implicit form or explicitly, so that an extended Working-Leser version of (1) can be written as

$$w_i = \alpha_i' + \beta_i' \ln \frac{E}{N} + u_i' \tag{2a}$$

$$w_i = \alpha_i^* + \beta_i^* \ln E + \gamma_i^* \ln N + u_i^* \tag{2b}$$

Equations (2a) and (2b) could be further enlarged by including a quadratic term as follows:

$$w_i = \alpha_i^* + \beta_i^* \ln \frac{E}{N} + \gamma_i^* \left(\ln \frac{E}{N}\right)^2 + u_i^* \tag{3a}$$

$$w_i = \alpha_i^{**} + \beta_i^{**} \ln E + \gamma_i^{**} \ln N + \delta_i^{**} (\ln E)^2 + \pi_i^{**} (\ln N)^2 + u_i^{**} \tag{3b}$$

It must be noted that (3a) uses ln (per capita expenditure) and ln (per capita expenditure) squared as the explanatory variables.[2] This model implicitly

2. However, a slightly different version of (3a) is used by Deaton *et al.* (1984). They assert on the basis of their empirical work that the Engel curves (2a) or (3b) when fitted to different groups of households that are demographically homogeneous within the groups are well approximated across the groups by replacing total household expenditure by *per capita* household expenditure. According to these authors, the residual effects of demographic structure can be approximated by including the household composition variables linearly, without affecting the additivity properties.

forces the coefficient of ln N to be equal in absolute magnitude to the coefficient of ln E, i.e., $|\beta_i^*|$. On the other hand, model (3b) allows the coefficients of ln E and ln N, β_i^{**} and γ_i^{**}, to be completely different. Also, it provides for the explanation of the separate effects of total outlay (E) and household size (N) on the shares of specific commodities (w_i), whereas (3a) presents the combined effect of total outlay and household size on budget shares. The separate effects are confounded in this model because of the use of per capita expenditure. When the quadratic demographic effect is not significant, we may further simplify (3b) and write

$$w_i = \alpha_i + \beta_i \ln E + \gamma_i \ln N + \delta_i (\ln E)^2 + u_i \quad (4)$$

In what follows we shall be mainly concerned with the estimation of parameters in model (4).

In cross section analysis the problem of heteroscedasticity could arise due to grouping of observations; it could also arise from variations in the relationship between the conditional variance of the error term with one or more of the explanatory variables. For ungrouped data the latter is more readily applicable. This problem is not new, but has generally been neglected in the analysis of NSS data, resulting in possible under- or over-estimation of the regression parameters as well as their sampling variances. It could even lead to erroneous rejection of valid specifications, if not detected early and corrected for (see, Witte and Cramer, 1986).

Since variance is dependent upon the unit of measurement and the expenditure shares (which is what we are trying to analyse here) are much smaller in magnitude compared to absolute expenditure levels, the problem of heteroscedasticity may not be serious.

For purposes of estimation, model (4) can be rewritten as

$$w_{ij} = \alpha_i + \beta_i \ln E_j + \gamma_i \ln N_j + \delta_i (\ln E_j)^2 + u_{ij} \quad (4a)$$

where w_{ij} is the share of commodity i in the budget of the jth household in the sample; E_j is the monthly total expenditure of household j; N_j is the size of the jth household (i.e., number of heads); and u_{ij} is the usual error term, $i = 1, 2, \ldots, 9$.[3]

3. Data on durables contain many zero observations, so that model (4) or (4a) is not entirely appropriate. In cross sections the zero observations could arise in two ways: in one case those who do not purchase have no use for the item, i.e., the marginal utility is less than the price for any positive amount and in the second case the reported zero expenditure may be because of infrequency of purchase due to seasonality. It is also plausible that most households are just too poor to buy a durable. For a detailed treatment of zero expenditures see Deaton *et al.* (1985) and Nayak (1987).

The Goldfeld-Quandt test when applied to model (4a) indicated the presence of heteroscedasticity. As a first step, therefore, it was surmised that per capita expenditure (PCE) is one of the key variables accounting for the observed heteroscedasticity. If, for example, we assume that heteroscedasticity is of the form

$$\sigma_j^2 = \sigma^2 z_j^b \tag{5}$$

where $z_j = \ln(E_j/N_j)$, i.e., $z_j = \ln(PCE)_j$, then each commodity involves two additional parameters, σ^2 and b. The latter can be interpreted as the measure of the strength of heteroscedasticity depending upon its magnitude and sign. A negative b implies that for lower magnitudes of PCE the variance in expenditure share on the particular item is more than that for higher magnitudes of PCE. Similarly, a positive b implies that for lower magnitude of PCE the variation in expenditure share on the particular item is less than that for higher magnitudes of PCE. In particular, when $b = 0$, the model is homoscedastic. The two parameters σ^2 and b are both unknown and have to be estimated along with the regression coefficients, using the method of maximum likelihood. The estimating equations turn out to be highly non-linear and their solution intractable. However, a practical method would be to compute the ln likelihood function $l(b)$ for a range of values of b and pick up that value of b for which $l(b)$ is a maximum. In this case, we have a set of five equations in five unknowns for each value of b, leading to unique estimates (see, Kmenta, 1971). In our exercise the values of b range from -6 to $+6$, with finer intervals in the neighbourhood of the maximum. The results are reported in Tables 1(R) and 1(U) separately for the rural and urban sectors.

Table 1(R)

Maximum Likelihood Estimates of b for Selected Commodities, Rural Karnataka (1977-78)

Commodity group	*Region*			
	I	*II*	*III*	*IV*
Cereals and pulses	0.0	0.0	0.0	0.0
Milk and milk products	0.0	2.0	2.0	2.0
Sugar	0.0	0.0	−4.00	0.0
Pan, tobacco, intoxicants	0.0	−2.0	0.0	−1.5
Fuel and light	0.0	−3.0	2.0	1.5
Vegetables, fruits, nuts	0.0	3.0	0.0	0.0
Other foods	5.0	4.0	4.0	2.0
Clothing	5.0	0.0	0.0	0.0
Miscellaneous items	5.0	0.0	3.0	3.0

Table 1(U)

Maximum Likelihood Estimates of b for Selected Commodities, Urban Karnataka (1977-78)

Commodity group	Region			
	I	II	III	IV
Cereals and pulses	0.0	0.0	0.0	3.0
Milk and milk products	3.0	0 0	0.0	0.0
Sugar	0 0	0.0	0.0	0.0
Pan, tobacco, intoxicants	0.0	0.0	−3.0	3.0
Fuel and light	0.0	0.0	3.0	1.5
Vegetables, fruits, nuts	0.0	0 0	0.0	0.0
Other foods	4.0	0.0	0.0	0.0
Clothing	0.0	0.0	0.0	0.0
Miscellaneous items	0.0	3.0	−1.5	−2.0

4. EMPIRICAL RESULTS

From our analysis it turns out that heteroscedasticity is more marked in the rural sector as indicated by significantly large negative values of b for three commodity groups, viz., pan, tobacco, intoxicants; sugar and gur; and clothing, in all the regions. Fuel and light has a negative b in region II only (see, Table 1(R)). On the other hand, in urban Karnataka, there are fewer instances of heteroscedasticity. A negative b is encountered only in some regions for items like pan, tobacco and intoxicants, and miscellaneous items (see, Table 1(U)).

For necessities and most items of luxury the value of b is positive, which indicates that as households move up in the income scale, their expenditure pattern tends to become more and more divergent. This is as expected. On the other hand, the existence of a negative b in some cases indicates that for those commodity groups the divergence may actually decline with increasing income. This is an interesting hypothesis which needs further examination.[4]

4. An approximate test for the significance of b is provided by the statistic, Q, defined as

$$Q = 2\,[l\,(b) - l(b_0)] \tag{7}$$

where b is the maximum likelihood estimate of b and b_0 is zero, i.e., $b_0 = 0$. The variable, Q, in large samples, follows a chi-square distribution with one degree of freedom. The computed values of Q are shown in the Appendix table. The table values of the chi-square distribution at 5% and 1% levels of significance are, respectively, 3.84 and 6.63, for one degree of freedom. Therefore, any value of Q exceeding these critical values should be considered significant (see, Rao (1974 : 417).

After determining the value of b for each commodity group, in each region and sector, the variables in model (4a) were divided by $z_j^{b/2}$, so that the transformed errors become homoscedastic. The adjusted model takes the form

$$\frac{w_j}{z_j^{b/2}} = \frac{\alpha_i}{z_j^{b/2}} + \beta_i \frac{\ln E_j}{z_j^{b/2}} + \gamma_i \frac{\ln N_j}{z_j^{b/2}} + \delta_i \frac{(\ln E_j)^2}{z_j^{b/2}} + \frac{u_j}{z_j^{b/2}} \quad (6)$$

The various parameters in model (6) were estimated for all selected groups of commodities in all regions and sectors of the State. The variables of interest, viz., the budget shares and the elasticities, were then estimated at the mean levels of the explanatory variables. These are shown in Table 2.

The coefficient of determination, R^2, is generally high and significant in most cases, justifying the need for the transformation of the variables to overcome heteroscedasticity. The estimated regression coefficients in model (6) appear highly significant although negative coefficients are observed in the milk and milk products, other foods, clothing, footwear and miscellaneous groups with regard to demographic changes. In all cases, however, the income (i.e., total expenditure) effects are significant and positive. Cereals and pulses appear as necessities whereas milk and milk products, vegetables, fruits and nuts, clothing and footwear as luxuries. On the other hand, items like sugar, pan, tobacco and intoxicants, fuel and light, and other foods appear as semi-luxuries. These results are hardly startling, but the range of variation in the estimated elasticities in Table 2 could be of some interest to students of consumer behaviour in India. A detailed analysis of these results in the context of social classes is found in Raghuprasad (1986).

5. CONCLUSIONS

From an examination of the R^2-coefficients we find that the adjusted model fits well to our data sets. The errors in the estimates of elasticities and budget shares therefrom are not very large, as judged by usual half-sample analyses. It may be noted, however, that the original model would not retain its strict additive property after the correction for heteroscedasticity is applied. This is no doubt a limitation of our study.

As already observed, heteroscedasticity might occur purely as a data problem or because of a specification error. In our model we have not included the prices. This might lead to a specification problem which we have not investigated in this study. In our analysis the assumption of uniform prices across the regions within the State could be justified to some extent.

Even if the prices were to be different, the question of which set of prices to use, is not easy to answer. It might be argued that implicit prices, i.e., value

Table 2

Average Expenditure Elasticities : Regional and Sectoral Estimates : Rural and Urban Karnataka, 1977-78

Region / Commodity	Rural				Urban			
	I	*II*	*III*	*IV*	*I*	*II*	*III*	*IV*
Cereals and Pulses	.64 (.38)	.72 (.42)	.79 (.37)	.76 (.37)	.68 (.30)	.54 (.30)	.74 (.29)	·73 (.27)
Milk and its Products	1.23 (.07)	1.40 (.05)	1.37 (.05)	1.47 (.06)	1.48 (.10)	1.21 (.07)	1.21 (.08)	1.44 (.06)
Sugar	.91 (.03)	1.26 (.02)	.99 (.02)	1.06 (.04)	.95 (.03)	.71 (.02)	.91 (.03)	.98 (.03)
Pan, Tobacco and Intoxicants	.98 (.05)	.86 (.06)	.83 (.05)	.91 (.04)	1.01 (.03)	.56 (.02)	.96 (.02)	.96 (.03)
Fuel and Light	.80 (.07)	.53 (.08)	.67 (.09)	.58 (.07)	1.03 (.08)	.44 (.08)	.84 (.09)	.94 (.08)
Vegetables, Fruits and Nuts	1.03 (.08)	1.17 (.05)	1.01 (.05)	.91 (.05)	.90 (.03)	1.14 (.06)	1.09 (.06)	1.04 (.05)
Other foods	.89 (.16)	1.07 (.17)	.96 (.17)	,77 (.13)	.79 (.28)	1.13 (.21)	1.06 (.20)	.90 (.20)
Clothing	1.80 (.04)	2.11 (.07)	2.64 (.05)	2.43 (.07)	3.02 (.04)	1.79 (.09)	1.69 (.04)	1.94 (.05)
Miscellaneous Goods	1.49 (.16)	1.20 (.10)	1.21 (.11)	1.29 (11)	1.50 (.12)	1.27 (.18)	1.12 (.18)	1.11 (.14)
No. of Households	184	282	595	836	96	91	558	504

of consumption divided by quantity consumed, could have been used. But this is also subject to certain inherent limitations (see, Iyengar, 1965). It is also observed that quantity data would be generally less reliable than value data. In addition, commodity aggregation poses its own problems in the use of implicit prices.

If external information on prices is used to convert quantities into values, we still have the problem of aggregation because in the case of some commodities, only values are available in our data set, and not the quantities. The commodity group 'vegetables, fruits and nuts' is a good example. The problem of prices in this context has been dealt with elaborately in Nayak (1987).

Conventional forms of Engel curves such as the double-log, semi-log, hyperbolic, log-inverse and linear forms were tried on our data set, but the choice of a most satisfactory model among them was not unique. As a viable alternative the Working Leser model was adopted, with suitable modifications incorporating the household size effects. This model was found to be more realistic for Karnataka data after necessary corrections were applied for heteroscedasticity.[5]

REFERENCES

BHATTACHARYA, N. and MAITRA, T. (1969). Analysis of Engel curves based on household budget data for Seventh to Twenty-second Rounds. Presented at the 10th Indian Econometric Conference, Madurai.

COONDOO, DIPANKAR (1969). The effect of relative prices on Engel elasticities of cereals in India. *Arthaniti* 12.

DEATON, A. and MUELLBAUER, J. (1980). An almost ideal demand system. *American Economic Review* 70(3) : 312-326.

DEATON, A., JAVEIR RUIZ CASTALLE and DUNCAN THOMAS (1985). The influence of household composition on household expenditure patterns : Theory and Spanish evidence. Working Paper No. 122, Research Programme in Development Studies.

IYENGAR, N. S. (1963). Estimation of quality elasticities for certain commodities from National Sample Survey data. *Sankhya* 25 : 15-22.

_____ (1964). A consistent method of estimating the Engel curve from grouped data. *Econometrica* 23 : 591-618.

IYENGAR, N. S., JAIN, L. R. and SRINIVASAN, T. N. (1968). Economies of scale in household consumption : A case study. *Indian Economic Journal* 15 : 465-477.

JAIN, L. R. and Tendulkar, S. D. (1973). Analysis of occupational differences in consumer expenditure patterns in India. *Sankhya* B35 : 140-156.

KMENTA, J. (1971). *Elements of Econometrics*. McMillan, London.

LESER, C. E. V. (1976). Income, household size and price changes, 1953-1973. *Oxford Bulletin of Economics and Statistics* 48 : 1-10.

MAJUMDER, A. (1986). Consumer expenditure in India : A comparison of the AIDS and the LES. *Sankhya*, B48 : 115-143.

5. A half-sample analysis was made using the data on interpenetrating sub-samples collected by the NSS Organisation in reaching this conclusion.

MURTHY, G. V. S. N. (1984). Consumer behaviour : An application of the AIDS. *Anveshak* XIV : 1-47.

MURTY, K. N. (1980). Consumer behaviour in India. *European Economic Review* 14 : 221-235.

NATIONAL SAMPLE SURVEY (1985). *Sarvekshana* VIII(3, 4).

NAYAK, VIJAY (1987). Analysis of household expenditures : Treatment of zero observation in the estimation of price response. *Unpublished Ph.D. thesis*, Indian Statistical Institute, Calcutta.

RAGHUPRASAD, SHAILAJA (1986). A quantitative analysis of poverty, inequality and consumer behaviour : The case of the Scheduled Castes and Tribes of Karnataka. *Unpublished Ph.D. thesis*, Indian Statistical Institute, Calcutta.

RAO, C. R. (1974). *Linear Statistical Inference and its Applications*. Wiley Eastern, New Delhi.

RAY, RANJAN (1980). Analysis of a time series of household expenditure surveys for India. *Review of Economics and Statistics* 62 : 595-602.

____(1982). The testing and estimation of complete demand systems of household budget surveys : An application of AIDS. *European Economic Review* 15 ; 349-369.

____ (1985). A dynamic analysis of expenditure patterns in rural India. *Journal of Development Economics* 19 : 283-298.

WITTE, M. A. C. DE and CRAMER, J. S. (1986). Functional form of Engel curves for foodstuffs. *European Economic Review* 30 ; 909-913.

Appendix Table

Values of Q† for Selected Commodities and Regions : Rural and Urban Karnataka, 1977-78

Commodity group	*Sector*	*Region*			
		I	*II*	*III*	*IV*
Cereals and pulses	Rural	2.02	0.76	5.62*	0.40
	Urban	2.64	0.32	1.21	22.50**
Milk and milk products	Rural	1.92	10.78**	20.11**	27.12**
	Urban	10.76**	1.00	2.10	0.95
Sugar	Rural	0.92	0.66	54.68**	1.00
	Urban	1.96	3.52	3.01	1.00
Pan, tobacco, intoxicants	Rural	1.00	16.00**	2.70	30.90**
	Urban	1.28	3.18	10.56**	24.54**
Fuel and light	Rural	0.14	23.90**	101.21**	21.30**
	Urban	0.76	2 80	12.21**	9.00**
Vegetables, fruits, nuts	Rural	0.58	17.80**	2.56	0.38
	Urban	1.00	0.30	2.86	1.74
Other foods	Rural	31.02**	28.66**	68.26**	30.30**
	Urban	3.00	2.91	0.92	0.46
Clothing	Rural	37.54**	0.91	1.23	1.56
	Urban	2.45	3.20	1.15	1.25
Miscellaneous	Rural	25.48**	2.84	180.68**	89.46**
	Urban	3.11	6.22**	9.89**	82.70**

†$Q = 2\,[l(b) - l(b_0)]$ is the likelihood ratio test statistic as defined in footnote 4.

*Significant at 5% level.

**Significant at 1% level.