Sankhyā: The Indian Journal of Statistics
1998, Volume 60, Series B, Pt. 2, pp. 349-376

A NONPARAMETRIC EXPLORATORY ANALYSIS OF THE FORM OF ENGEL CURVES AND RANK OF THE DEMAND SYSTEM - THE CASE OF RURAL MAHARASHTRA

By AMITA MAJUMDER
and
MANISHA CHAKRABARTY
Indian Statistical Institute, Calcutta

SUMMARY. In this paper nonparametric Kernel smoothing technique is used to determine the functional form of the engel curves of some selected commodities separately for household groups with different household occupations, social classes and household compositions. The performances of these are compared with those of some popular parametric engel curves using the NSS 38th round household level data of rural Maharashtra. The rank of the demand system consisting of these commodities is determined using the rank test of Lewbel (1991) and Gill and Lewbel(1992) and the likelihood ratio test.

1. **Introduction**

Household expenditure patterns differ across households according to family size, age-sex composition, household occupation and other household characteristics. In modelling household demand one should, therefore, relate item specific expenditures not only to prices and income or total expenditure, but also to these household characteristics. The various econometric approaches to estimating household demand functions have a major drawback in common, namely, that the functional form of the demand equations are specified in advance with no a priori knowledge of the underlying true functional form. This may lead to misspecification of the functional form which may have serious consequences on the econometric results. Nonparametric Kernel method enables the functional form to be determined by data and allows consistent estimation of a regression

---

Paper received. October 1997; revised March 1998.
AMS (1991) *subject classification.* 62G07, 62P20.
Key words and phrases. Kernel regression, engel curves, rank test.

model without specifying a functional form a priori.[1]

In this paper we nonparametrically determine the shape of engel curves for some selected items using the household level central sample consumption data of rural Maharashtra for the NSS 38th round. Note that given prices and household characteristics, the demand of a commodity is a function of total expenditure, which is the engel curve for that commodity. In this cross-section based study, where prices may be taken to be fixed, in order to place emphasis on the shape of engel curves we use homogeneous subsamples of households and perform the analysis separately for groups of households with different household characteristics. A comparison between the nonparametric and some parametric engel curves is also made.

Having determined the approximate shapes of the engel curves, the next question one would like to ask is what the *rank* would be [Gorman(1981), Lewbel(1989, 1991)] of the demand system with these commodities. The *rank* of any demand system is the maximum dimension of the function space spanned by the engel curves of the demand system. In most empirical demand analyses the demand system is estimated without knowledge of the actual rank of the system, which may distort the ultimate result. Gill and Lewbel (1992) proposed a *rank test* which is a prespecification test, that indicates the rank, which in turn provides information about the degree of separability, aggregate structure and cost function structure consistent with a given data set. This test has been applied by Lewbel (1991) to U.S. and U.K. consumer survey data and by Banks, Blundell and Lewbel (1997) to U.K. data. We apply the rank test of Gill and Lewbel (1992) to determine the rank of the demand system consisting of the commodities considered in the engel curve analysis.

The plan of this paper is as follows: section 2 briefly describes the methodology used in this paper; section 3 presents the data and results, and finally, section 4 draws the conclusions.

2. Methodology

To explore the engel curve form, we approximate the mean response curve m in the regression relationship

$$w_{ih} = m_i[\ln y_h] + \epsilon_{ih},\ h = 1, ..., H,\ i = 1, 2, ..., n \qquad \ldots(2.1)$$

where w_{ih}, the budget share of the i-th commodity for the h-th household, is a function of logarithm of total expenditure, $\ln y$, n is the number of commodities

[1]Nonparametric Kernel method has been applied in a number of recent consumer demand studies. These include the works of Deaton (1989), Deaton and Subramanian (1993), Bierens and Pott-Butter (1990), Nicol (1993) and Banks, Blundell and Lewbel (1997).

under study for each homogeneous demographic group and H is the number of households in the group[2].

We apply the nonparametric Kernel smoothing technique using the Kernel weights proposed by Nadaraya(1964) and Watson(1964). The optimal bandwidth is chosen using the 'cross-validation leave-one out' method which minimises a quadratic error measure and optimally trades-off bias and variance. For computational simplicity we choose the quartic Kernel [see Härdle (1990) and Silverman (1986)].

The rank test applied here is outlined below. Let

$$w_h = A.G(Z_h) + \epsilon_h, \ h = 1, 2, ..., H \qquad \ldots (2.2)$$

where $Z_h = \ln y_h$, ϵ_h is a n-vector of mean zero errors that are assumed to be independent of Z_h and $\epsilon' l = 0$, where l is the n-vector of ones (in view of $w' l = 1$).

To determine r, the rank of the demand system (i.e, rank of A), the equation is postmultiplied by $Q(Z_h)'$, where $Q(Z_h)$ is the vector of n functions having finite mean. Taking expectation yields

$$E(w_h Q(Z_h)') = A.E(G(Z_h)Q(Z_h)') + E(\epsilon_h Q(Z_h)') \qquad \ldots (2.3)$$

Since ϵ is independent of Z,

$$Y = E(w_h Q(Z_h)') = AE(G(Z_h).Q(Z_h)')$$

so $rank(Y) = r$, unless some component of G is orthogonal to all elements of Q or the price regime has locally rank less than r.

Under the null hypothesis

$$H_0 : rank(Y) = r$$

$$H\hat{d_2}'\hat{W}^{-1}\hat{d_2} \rightsquigarrow {\chi^2}_{n-r}.$$

where $\hat{d_2}$ and $\hat{W}$ are functions of the elements of the matrices from the Lower-Diagonal-Upper-Triangular (LDU) decomposition of $\hat{Y} = \sum_h [w_h Q(Z_h)']/H$ [for a detailed description of the test see Gill and Lewbel (1992), Lewbel (1991)]. It may be noted that the test is consistent against the alternative that $rank(Y) > r$.[3]

[2]Instead of running a separate regression of the form (2.1) for each demographic group one could run only one nonparametric regression incorporating the household characteristic variable along with total expenditure in equation (2.1), using a mixed continuous-discrete Kernel regression approach. However this is computationally very heavy. Also, these estimates and the estimates from the present approach turn out to be asymptotically the same [Bierens and Pott-Butter(1990)].

[3]In a recent paper Cragg and Donald (1995) have modified the test. However, due to non-availability of this paper we have not been able to incorporate the modification.

3. Data and Results

3.1 *Data.* The data used in this analysis are the NSS 38th round(1983) central sample data for the rural sector of Maharashtra. We use the detailed consumption expenditure and demographic data. Excluding single-person households the final sample size becomes 4823. The delimitation of the geographical coverage of the study is determined largely by our access to NSS data and computational tractability.

We first group the households according to the occupational group of the household head, viz., (a) nonagricultural self employed households (NASE) (397),[4] (b) agricultural labourer households (AL)(1875), (c) other labourer households (OL) (335), (d) agricultural self employed households (ASE) (1868), and (e) other households (OH)(348). We also group the households according to the social caste of the household, viz., (a) scheduled caste and scheduled tribe combined (SC-ST) (1157), and (b) others (OTHER) (3631) to examine the consumption pattern across social castes. In the demographic category we consider the age group of 15 years and above as the adult age group and below 15 years as the child age group. To place emphasis on the sensitivity of the shape of the engel curves to demographic characteristics, for each occupation group/social class we divide the sample into subgroups homogeneous in household composition. Thus, for each occupation group/social class the first group consists of only 2 adults (denoted as A), the second one consists of 2 adults and 1 child (denoted as B), the third group consists of 2 adults and 2 children (denoted as C), and in the fourth and fifth groups we have 2 adults and 3 children (denoted as D) and 2 adults and 4 children (denoted as E), respectively. This grouping within the classes NASE, OL and OH renders sample sizes that are not sufficient to apply nonparametric method, and therefore we confine our attention only to the two occupational groups, AL and ASE and to the two social classes SC-ST and OTHER mentioned above.

We have aggregated the detailed item level consumption expenditure data to a commodity groupwise data. The commodity groups we consider are (1)*cereals*, (2) *other food*, (3) *adult good* (4) *fuel and light* and (7)*clothing* (excluding the adult clothing items) [5].

3.2 *Results.* We have first estimated the budget share curve for the different commodities nonparametrically. Two most commonly used parametric models, namely, the Working-Leser (WL) form and the extended Working-Leser or Quadratic Logarithmic (QL) form given by

$$w = \alpha + \beta \ln y + \epsilon \qquad \ldots(3.4)$$

and

$$w = \alpha + \beta \ln y + \gamma (\ln y)^2 + \epsilon \qquad \ldots(3.5)$$

[4]Figures in parentheses are the sample sizes.

[5]See appendix for the list of items included in each group.

respectively, have also been estimated on the same data set. We have plotted the budget share against $\ln(y)$ for all commodities for both household occupation groups and social classes. Figures 1-5 present the nonparametric Kernel regression for the five commodities for the occupation groups ASE and AL and for the social classes SC-ST and OTHER by demographic groups[6]. As an illustration the nonparametric fit (along with the 95% confidence bands), the WL and the QL fit (evaluated at decile points) of the budget share curves for the five commodities for the demographic group C in the occupation group ASE are presented in figures A - C[7]. In most cases the parametric curves are quite close to the nonparametric one and lie within the confidence band for all cases.

Table 1. CORRELATION COEFFICIENT BETWEEN OBSERVED AND ESTIMATED BUDGET SHARES, ESTIMATED PARAMETRICALLY AND NON-PARAMETRICALLY FOR CEREALS, OTHER FOOD, ADULT GOOD, FUEL AND LIGHT AND CLOTHING BY DEMOGRAPHIC GROUPS WITHIN OCCUPATIONAL CATEGORIES

Demo-graphic group	Sam-ple size	Cereals			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Agricultural Labourer(AL)					
A	138	.10	.10	.10	.72
B	174	.18	.04	.08	.57
C	226	.05	.05	.06	0.00
D	185	.24	.19	.22	0.00
E	110	.22	.22	.22	0.00
Agricultural Self Employed(ASE)					
A	110	.27	.18	.22	0.00
B	78	.16	.14	.17	0.00
C	107	.40	.36	.38	.93
D	118	.34	.27	.29	0.00
E	67	.42	.37	.37	0.00

Demo-graphic group	Sam-ple size	Other food			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Agricultural Labourer(AL)					
A	138	.42	.40	.41	0.00
B	174	.29	.13	.20	0.00
C	226	.29	.04	.11	0.00
D	185	.43	.40	.40	0.00
E	110	.36	.36	.36	0.00
Agricultural Self Employed(ASE)					
A	110	.55	.48	.50	0.00
B	78	.22	.22	.22	0.00
C	107	.50	.47	.50	0.00
D	118	.48	.46	.47	0.00
E	67	.53	.45	.45	0.00

[6]The flatness of the nonparametric curves may seem to be a result of oversmoothing due to large bandwidth. However, in our case, the choice of optimum bandwidth is governed by minimisation of quadratic error measure which optimally trades-off variance and bias.

[7]The behaviour of the commodities with respect to the other groups turns out to be similar (figures have not been presented here due to lack of space).

Table 1. (CONTINUED)

Demo-graphic group	Sam-ple size	Adult good			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Agricultural Labourer(AL)					
A	138	.36	.30	.31	0.00
B	174	.17	.14	.16	0.00
C	226	.36	.18	.18	0.00
D	185	.31	.16	.26	0.00
E	110	.40	.12	.22	0.00
Agricultural Self Employed(ASE)					
A	110	.15	.12	.12	0.00
B	78	.19	.19	.19	0.00
C	107	.22	.09	.11	0.00
D	118	.08	.07	.12	0.00
E	67	.40	.12	.19	0.00

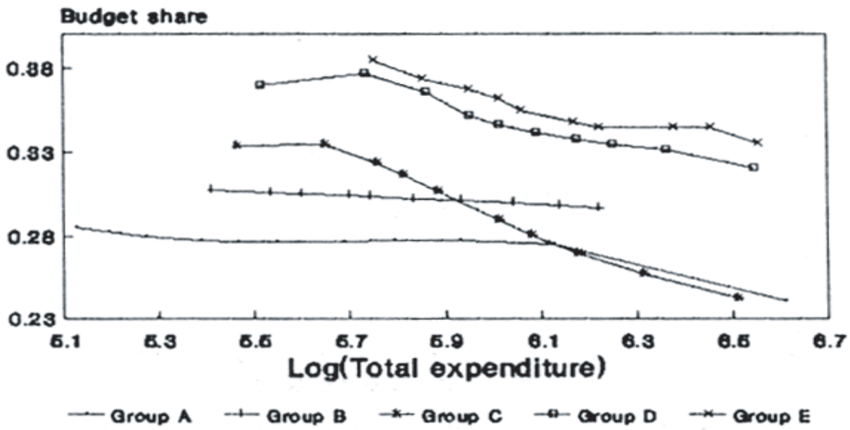
Demo-graphic group	Sam-ple size	Fuel and light			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Agricultural Labourer(AL)					
A	138	.25	.25	.25	0.00
B	174	.17	.14	.17	0.00
C	226	.07	.02	.09	0.00
D	185	.17	.12	.14	0.00
E	110	.16	.14	.15	.91
Agricultural Self Employed(ASE)					
A	110	.36	.36	.36	0.00
B	78	.26	.21	.21	0.00
C	107	.35	.34	.36	0.00
D	118	.21	.20	.22	0.00
E	67	.21	.21	.21	0.00

Demo-graphic group	Sam-ple size	Clothing			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Agricultural Labourer(AL)					
A	138	.31	.12	.16	21.60
B	174	.07	.08	.11	14.90
C	226	.07	.07	.11	8.40
D	185	.23	.13	.15	10.80
E	110	.05	.03	.11	15.50
Agricultural Self Employed(ASE)					
A	110	.25	.09	.09	21.80
B	78	.20	.19	.22	12.80
C	107	.13	.11	.14	4.70
D	118	.19	.17	.18	10.20
E	67	.08	.04	.08	9.00

*NP: Non-parametric.

Nonparametric Curves (cereal)

ASE



AL

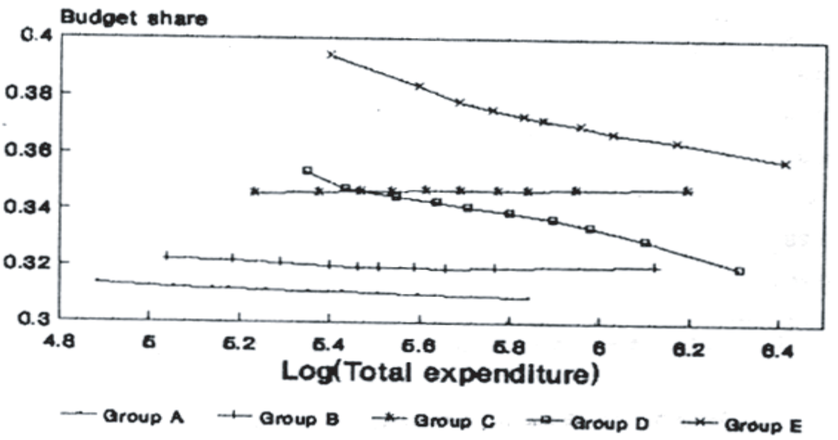


Figure 1 (contd.)

Nonparametric Curves (cereal)

SC-ST

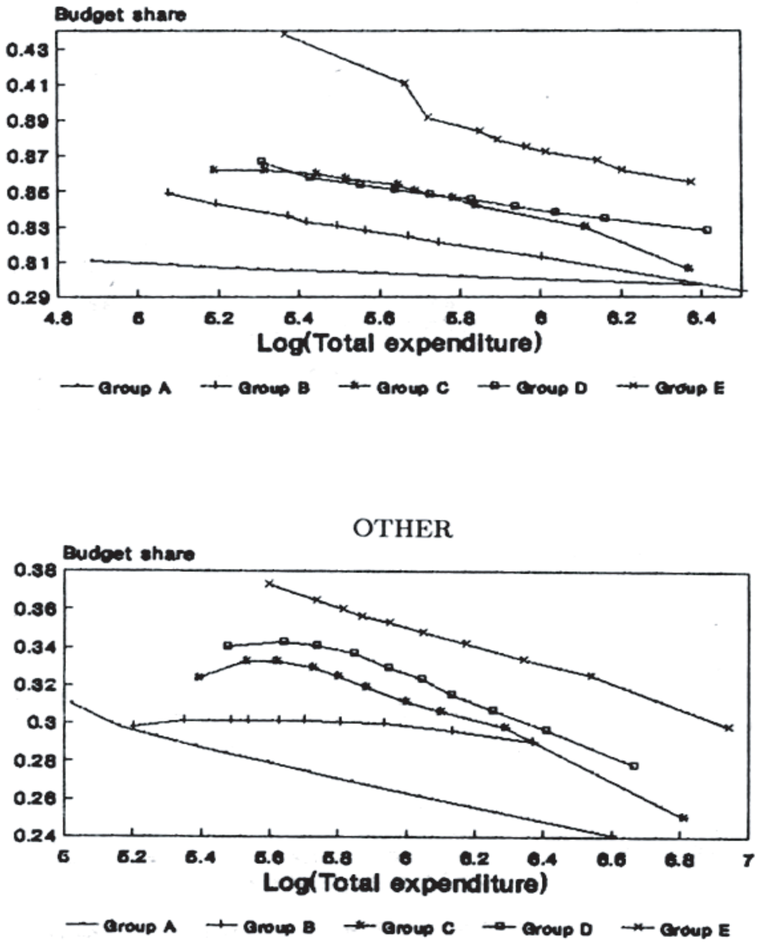
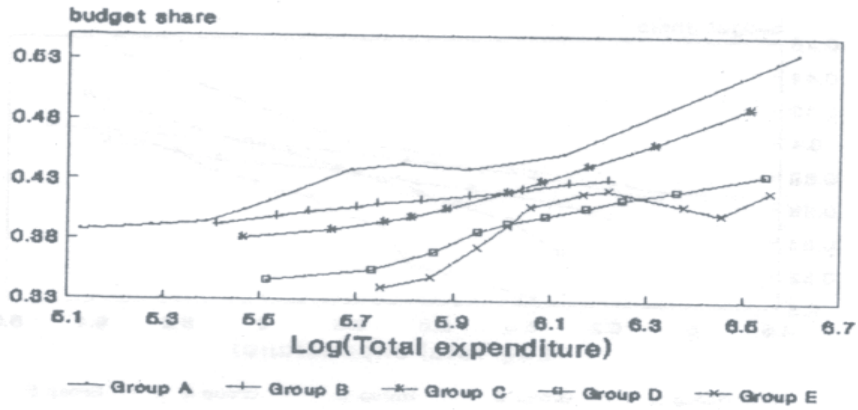


Figure 1

Nonparametric Curves (other food)

ASE



AL

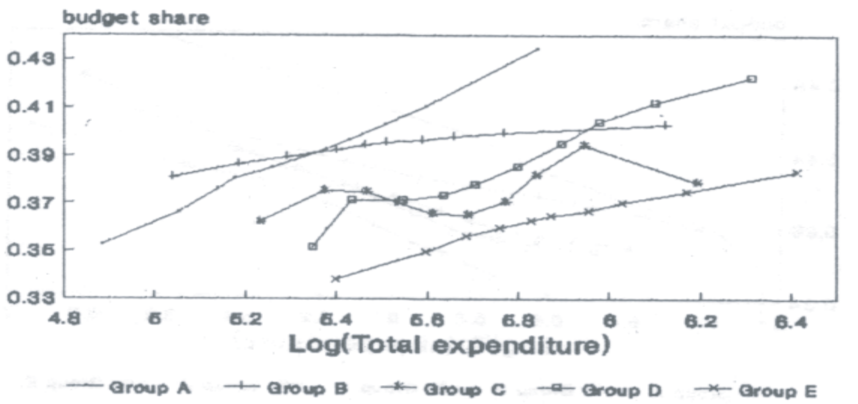
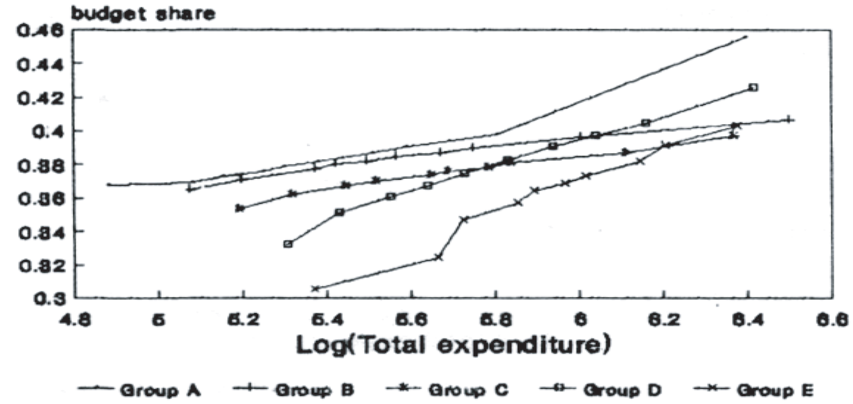


Figure 2 (contd.)

Nonparametric Curves (other food)

SC-ST



OTHER

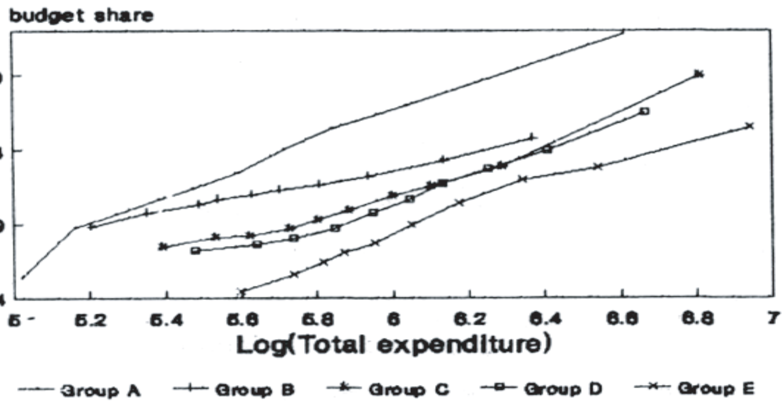
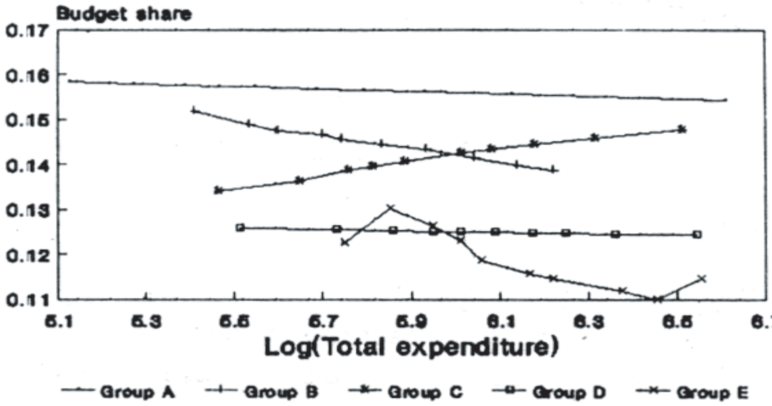


Figure 2

Nonparametric Curves (adult good)

ASE



AL

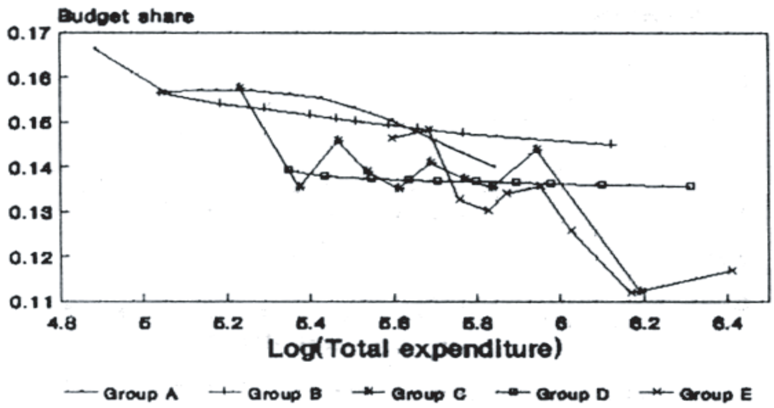
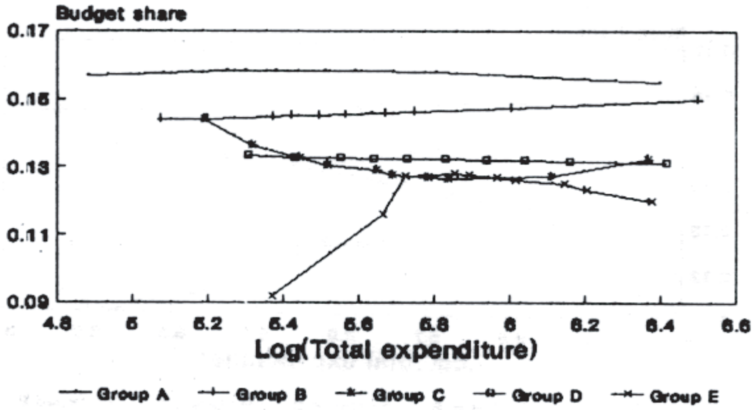


Figure 3 (contd.)

Nonparametric Curves (adult good)

SC-ST



OTHER

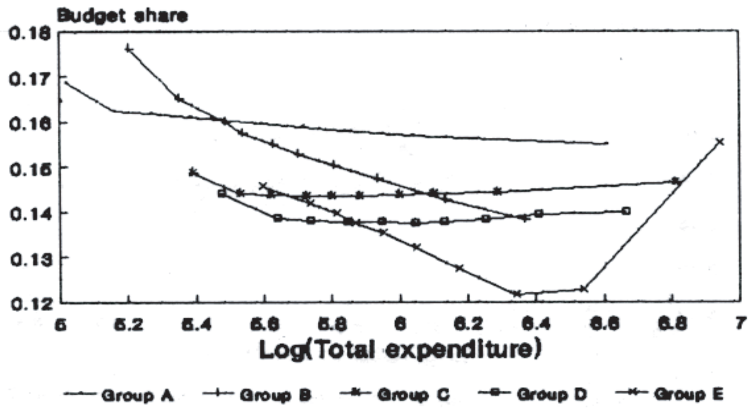
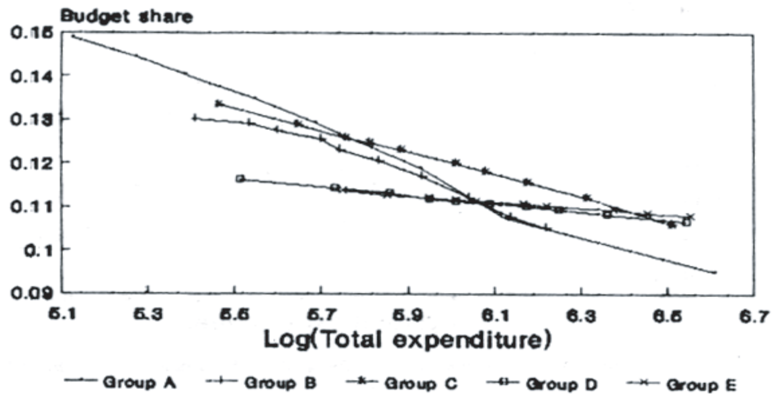


Figure 3

Nonparametric Curves (fuel and light)

ASE



AL

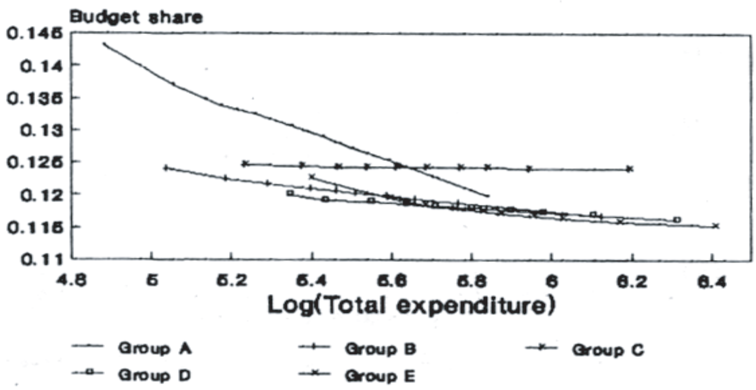
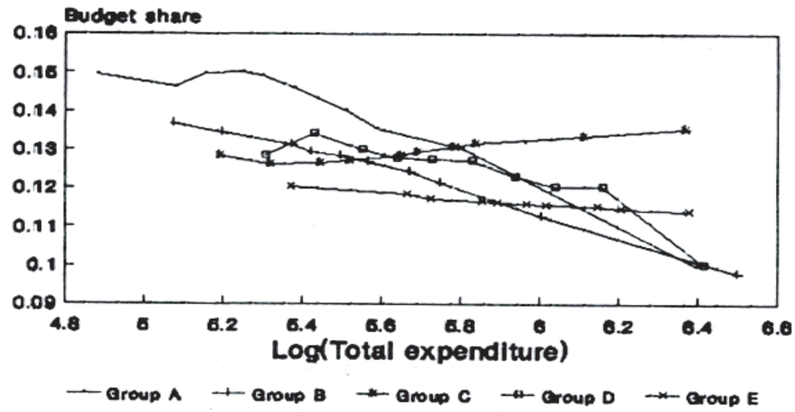


Figure 4 (contd.)

Nonparametric Curves (fuel and light)

SC-ST



OTHER

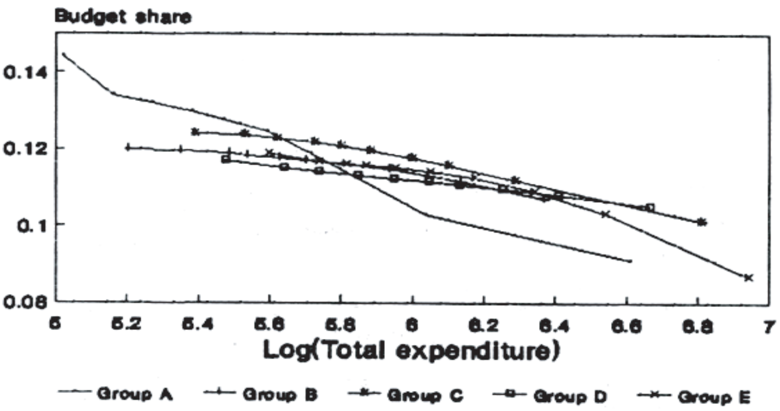
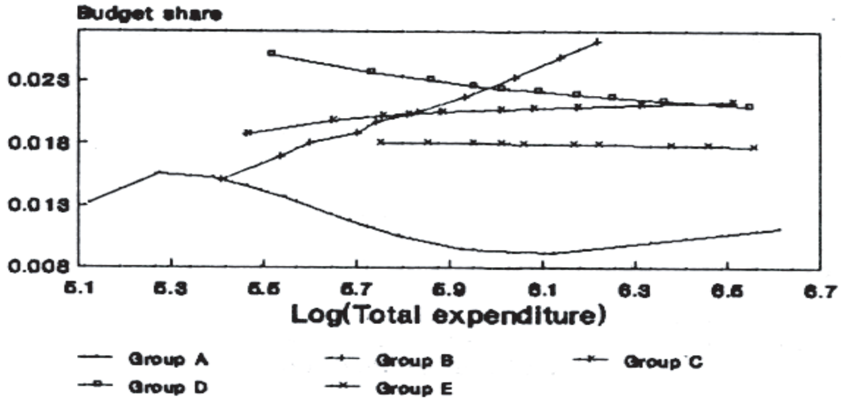


Figure 4

Nonparametric Curves (clothing)

ASE



AL

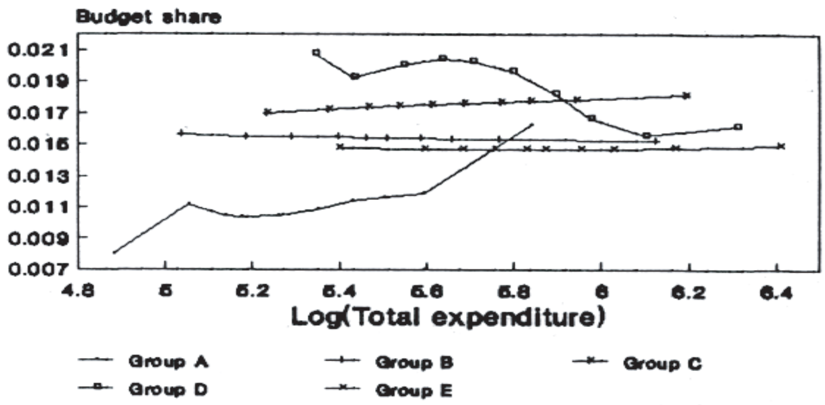
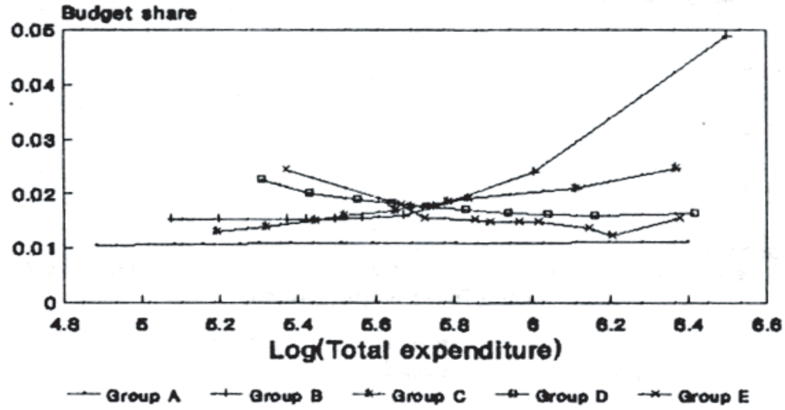


Figure 5 (contd.)

Nonparametric Curves (clothing)

SC-ST



OTHER

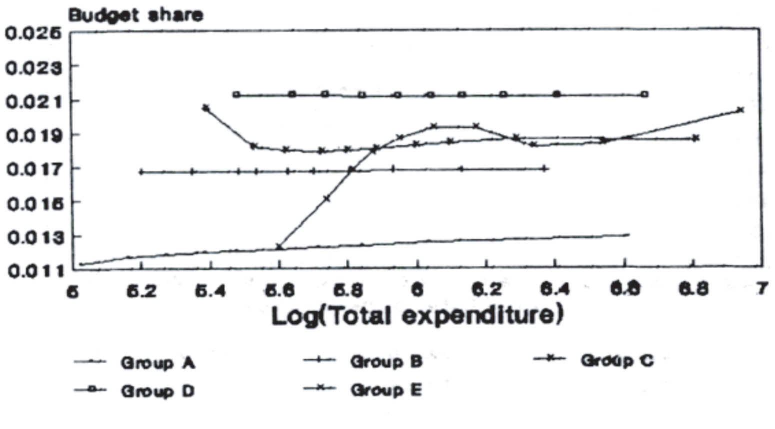
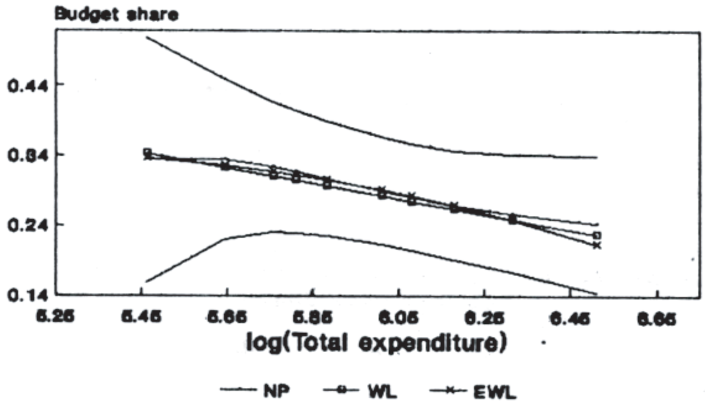


Figure 5

Parametric and Nonparametric Curves with Confidence Bands (ASE-C)

cereal



other food

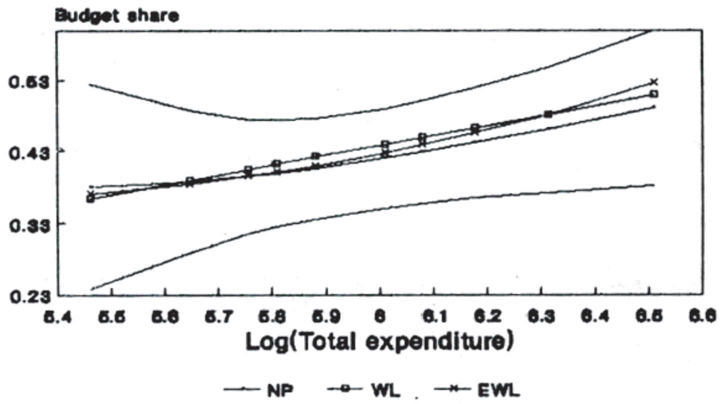
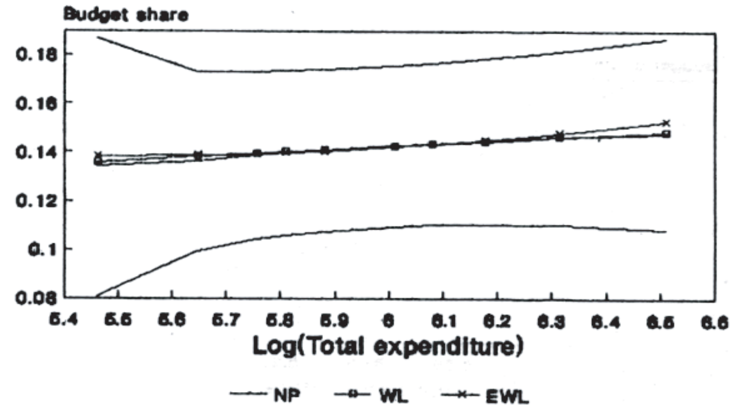


Figure A

Parametric and Nonparametric Curves with Confidence Bands (ASE-C)

adult good



fuel and light

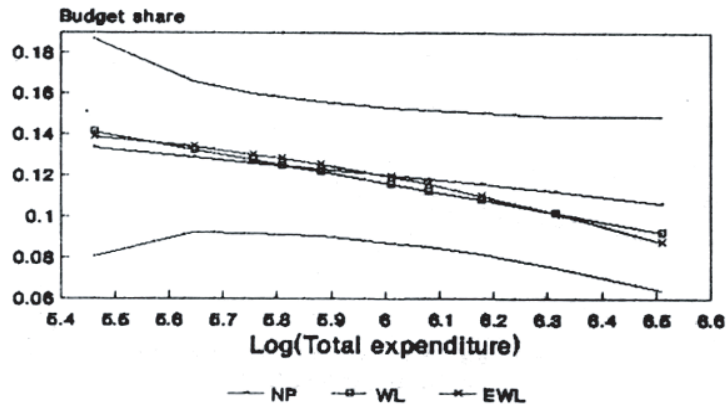


Figure B

Parametric and Nonparametric Curves with Confidence Bands (ASE-C)

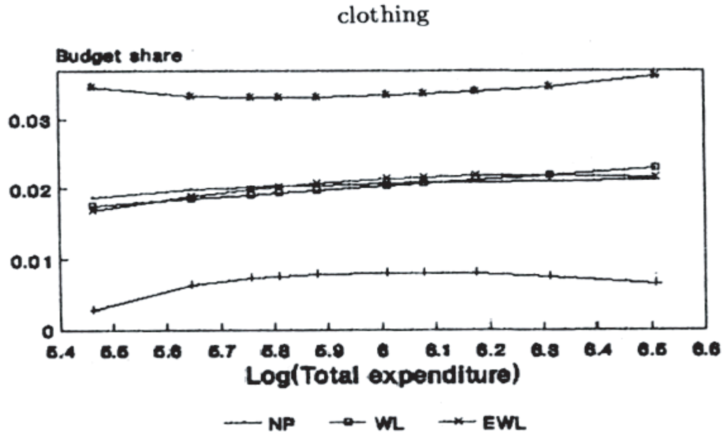


Figure C

Table 1 presents the correlation coefficient between the observed and estimated budget shares for the nonparametric and parametric engel curves for the occupation groups AL and ASE, and household composition groups A,B,C,D and E. It also presents the percentage of households not reporting purchase of these commodities. Table 2 presents the correlation coefficients for the household demographic groups A, B, C, D and E of the two social classes.

The findings from the tables and the plots can be summarised as follows:

For *cereals*, an overall downward sloping budget share curve is observed for all demographic categories and occupation groups although the slopes and intercepts differ. For both occupation groups, in most cases the performance of the QL form is marginally better than that of the WL form. A similar result holds for the two social classes. This is evident from the correlation coefficients in Tables 1 and 2. Generally the budget share curve shifts upwards monotonically with the increase in the number of children. This is possibly indicative of the fact that in the rural sector children are, by and large, *cereals*-consuming units.

Table 2. CORRELATION COEFFICIENT BETWEEN OBSERVED AND ESTIMATED BUDGET SHARES, ESTIMATED PARAMETRICALLY AND NON-PARAMETRICALLY FOR CEREALS, OTHER FOOD, ADULT GOOD, FUEL AND LIGHT AND CLOTHING BY DEMOGRAPHIC GROUPS WITHIN SOCIAL CLASSES

Demo-graphic group	Sam-ple size	Cereals			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Scheduled Caste and Scheduled Tribe(SC-ST)					
A	100	.16	.16	.16	0.00
B	101	.33	.32	.33	0.00
C	111	.32	.23	.23	0.00
D	102	.25	.25	.25	0.00
E	65	.40	.40	.40	0.00
Others(OTHER)					
A	230	.35	.33	.35	.87
B	229	.18	.05	.18	.44
C	316	.34	.24	.32	.32
D	316	.30	.25	.27	0.00
E	181	.31	.30	.30	0.00

Demo-graphic group	Sam-ple size	Other food			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Scheduled Caste and Scheduled Tribe(SC-ST)					
A	100	.36	.31	.34	0.00
B	101	.35	.15	.31	0.00
C	111	.26	.26	.26	0.00
D	102	.45	.44	.44	0.00
E	65	.55	.54	.55	0.00
Others(OTHER)					
A	230	.56	.55	.55	0.00
B	229	.42	.31	.37	0.00
C	316	.45	.41	.43	0.00
D	316	.45	.43	.43	0.00
E	181	.43	.41	.41	0.00

Table 2. (CONTINUED)

Demo-graphic group	Sam-ple size	Adult good			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Scheduled Caste and Scheduled Tribe(SC-ST)					
A	100	.16	.01	.18	0.00
B	101	.51	.27	.49	0.00
C	111	.29	.12	.20	0.00
D	102	.04	.04	.04	0.00
E	65	.32	.09	.29	0.00
Others(OTHER)					
A	230	.29	.16	.20	0.00
B	229	.31	.26	.26	0.00
C	316	.23	.02	.19	0.00
D	316	.31	.08	.19	0.00
E	181	.33	.02	.26	0.00

Demo-graphic group	Sam-ple size	Fuel and light			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Scheduled Caste and Scheduled Tribe(SC-ST)					
A	100	.35	.28	.29	0.00
B	101	.31	.30	.31	0.00
C	111	.16	.02	.06	0.00
D	102	.35	.20	.21	0.00
E	65	. .19	.18	.19	0.00
Others(OTHER)					
A	230	.35	.31	.32	0.00
B	229	.22	.18	.21	0.00
C	316	.25	.23	.25	0.00
D	316	.23	.23	.24	0.00
E	181	.29	.27	.28	.55

Demo-graphic group	Sam-ple size	Clothing			
		Budget share type			% of households not purchasing
		NP*	WL	QL	
Scheduled Caste and Scheduled Tribe(SC-ST)					
A	100	.08	.06	.14	21.00
B	101	.35	.20	.21	12.90
C	111	.23	.19	.19	9.00
D	102	.24	.16	.26	10.00
E	65	.45	.22	.30	10.80
Others(OTHER)					
A	230	.11	.13	.13	22.20
B	229	.08	.01	.05	12.20
C	316	.27	.10	.21	7.60
D	316	.03	.03	.04	10.10
E	181	.27	.16	.16	11.60

*NP: Non-parametric.

But this monotonicity is not strictly observed in case of groups ASE-C, AL-D. In case of SC-ST curves for the two demographic groups C and D almost coincide. In some cases, e.g., ASE-B, AL-A, B and C, the budget share curve is almost flat indicating small variation in budget share for change in $\ln y$.

In case of *other food* an upward sloping curve is observed for all demographic categories in the occupation groups and social classes. A downward shift of the budget share curve with the increase in the number of children is observed for both social groups, but this feature is not clearly observed for the occupation groups. *Other food*, being a luxury item (the budget share curves are upward rising) households are expected to reduce their share of expenditure on *other food* as more and more children appear in the family. Here in most cases the QL form is closer to the nonparametric curve in terms of correlation coefficients presented in tables 1 and 2. In some cases (e.g., AL-C) the nonparametric curves tend to show a shape with break points which may induce one to use a change-point model[8]. In this case too QL is a better approximation as compared to WL as is evident from the correlation coefficients in table 1.

The budget share curve for *adult good* is downward sloping in most of the cases for both occupational groups and social classes indicating that this item can be treated as a necessary item[9]. Exceptions are the cases for ASE-C (upward sloping), SC-ST-B and E and OTHER-E. For the occupation group as well as the social groups generally a downward shift in the curves is observed with increase in the number of children. This shift is self explanatory in view of the item being *adult good*. Here also like *other food* break points occur in the groups AL-C and E. The values of the correlation coefficient in tables 1 and 2 indicate that QL provides a better explanation than the WL model especially for the social classes.

For *fuel and light* a downward slope is observed for all demographic compositions for both occupation groups and social classes excepting SC-ST-C, although the slope and intercept differ. Here in all cases the QL form is marginally better than the WL form (see also tables 1 and 2). The effect of increase in the number of children on the consumption of *fuel and light* is not clear. Given the nature of the commodity the result is not unexpected.

In case of *clothing* nothing can be said clearly about the nature of the commodity. The budget share curve shows an upward slope in some groups and a downward slope in some other groups. In many cases the curves tend to be flat, indicating little change with changes in the level of $\ln y$. In most of the cases QL gives a better fit than the WL form. The peculiar behaviour of this commodity can be partly explained by the composition of the commodity (adult clothing

[8]We have, however, not pursued this here.

[9]This is corroborated by the estimates of the coefficients in table 3 (for the range of income we are considering here).

items are not included here). It may also be pointed out that the percentage of people not purchasing this commodity is rather high (see tables 1 and 2) compared to the other commodities.

Table 3. REGRESSION COEFFICIENT OF $\ln y$ AND $(\ln y)^2$ FOR A POOLED REGRESSION OF BOTH WL AND QL FORM OF ENGEL EQUATIONS

Coefficients of	Cereals WL	Cereals QL	Other food WL	Other food QL	Adult good WL	Adult good QL	Clothing WL	Clothing QL	Fuel and light WL	Fuel and light QL
$\ln y$	-.053	.255	.086	-.095	-.010	-.060	.004	-.011	-.026	-.088
	(13.835)	(6.079)	(24.263)	(2.450)	(4.215)	(2.313)	(3.897)	(1.148)	(13.995)	(4.327)
$(\ln y)^2$		-.026		.015		.004		.001		.005
		(7.382)		(4.693)		(1.935)		(1.511)		(3.059)
Chow test statistics* (F value)	54.449		22.034		3.785		2.469		9.367	
Likelihood Ratio test statistic (χ^2 value)**	54.144		21.987		3.784		2.400		9.359	

Figures in parentheses indicate absolute t values.
*Chow test statistic $(F_{1,4639}) = \frac{RSS_{WL} - RSS_{QL}/no.of restriction}{RSS_{QL}/d.f}$.
Critical value at 5% level of significance = 3.84.
**Likelihood ratio test statistic $({\chi^2}_1) = -2(\ln L_{WL} - \ln L_{QL})$.
Critical value at 5% level of significance = 3.84.

Table 3 presents the regression coefficients of $\ln y$ and $(\ln y)^2$ for a pooled regression of WL and QL forms of engel equations incorporating dummies for occupation and social categories. Both the Chow-test and likelihood-ratio test indicate that QL gives a significantly better fit than WL for *cereals*, *other food* and *fuel and light*. For *adult good* and *clothing* the performances are similar. In all cases except for *clothing* the coefficient of the quadratic term are significantly different from zero.

The rank test was applied to all demographic groups within the occupational as well as social classes. The underlying assumption here is that the distribution of preferences is independent of the distribution of income, which may be considered to be valid for homogeneous demographic groups (Lewbel,1991). The "instrument" list Q, in our case, consists of $1, y, \ln y, (\ln y)^2, \frac{1}{y}$. Following Lewbel (1991), each element of Q was divided by its mean in each sample, to ensure that Y is not ill-conditioned as a result of the enormous range of magnitudes in the functions included in Q. Table 4 summarises the results of the rank test for three possible ranks, viz., 1, 2 and 3. In all cases the magnitudes of the χ^2-statistics drops dramatically from $r = 1$ to $r = 2$ indicating that the rank is 2.

Table 4. LDU rank test chi-squared statistics by demographic groups

	Rank= r					
Demographic group	1(4)*	2(3)	3(2)	1(4)	2(3)	3(2)
	Agriculture Labourer(AL)			Agricultural Self Employed(ASE)		
A	18.370	0.020	0.001	10.666	0.734	0.032
B	0.506	0.096	0.002	42.540	2.340	0.001
C	4.257	0.071	0.011	35.960	0.050	0.001
D	19.092	0.233	0.022	23.480	0.010	0.001
E	10.140	0.101	0.001	11.890	0.100	0.001
	Scheduled Caste and Scheduled Tribe(SC-ST)			Others(OTHER)		
A	11.428	0.144	0.017	100.210	0.530	0.010
B	20.010	2.170	0.001	43.960	4.430	0.010
C	5.030	0.050	0.001	50.820	0.340	0.001
D	24.490	0.010	0.001	70.040	0.280	0.001
E	19.970	0.030	0.010	29.400	0.740	0.001

*Figures in parentheses denote the respective degrees of freedom(d.f).
The critical values at 5 percent level of significance are as follows:
χ^2 with 4 d.f : 9.488.
χ^2 with 3 d.f : 7.815.
χ^2 with 2 d.f : 5.991.

The implication of the above result is that given a price regime, for each demographic group within the occupational and social classes the budget share curve can be specified by an extended Working Leser form in view of a better performance of the QL model compared to the WL model. In particular, if the above result holds for multiple price regimes, one could specify a quadratic, yet rank 2 integrable demand system with the budget share curves of the form

$$w_{ih} = \alpha_i + \beta_i(\ln y_h + \rho(\ln y_h)^2) \qquad \ldots(3.6)$$

where ρ is a constant. This formulation basically assumes that one column (that corresponds to $(\ln y)^2$ of the coefficient matrix, denoted as A in section (2)) is a constant multiple of another column, corresponding to $\ln y$.

To determine the rank under different price situations we introduced price variation through regionwise break-up of the data. Price data were constructed using unit values following Yen and Roe (1989).[10] Table 5A summarises the results of the rank test for different price situations under different occupational and social classifications and table 5B gives the results for the overall sample. In few cases, e.g., in regions 1,3 and 4 for AL, in region 1 for ASE and in regions 1,4 and 6 for SC-ST the magnitudes of the χ^2-statistic are non-significant for r=1. In all other cases the rank test indicates a rank of 2.

Given that QL seems to fit quite well, a stronger rank test can be constructed by estimating systems of demand equations with the underlying engel curves

[10]A detailed description of the construction of price data is available in Chakrabarty (1998).

Table 5A. LDU RANK TEST CHI-SQUARED STATISTICS BY REGIONS

Different regions	Rank= r					
	1(4)*	2(3)	3(2)	1(4)	2(3)	3(2)
	Agricultural Labourer(AL)			Agricultural Self Employed(ASE)		
Region 1	6.28	0.20	0.02	4.30	0.62	0.01
Region 2	42.14	0.03	0.00	18.49	0.06	0.00
Region 3	8.69	0.47	0.01	13.87	0.02	0.00
Region 4	6.08	1.01	0.00	15.02	0.40	0.01
Region 5	10.18	0.06	0.01	27.30	0.57	0.01
Region 6	11.54	0.03	0.01	15.97	0.01	0.00
	Schedule Caste and Schedule Tribe(SC-ST)			Others(OTHER)		
Region 1	2.49	0.57	0.05	22.21	0.06	0.01
Region 2	32.86	0.22	0.01	35.95	2.17	0.00
Region 3	15.16	0.52	0.01	19.22	0.14	0.00
Region 4	7.11	0.33	0.10	68.72	2.98	0.01
Region 5	15.12	0.38	0.01	48.94	1.31	0.01
Region 6	8.22	0.74	0.02	20.04	0.04	0.03

Table 5B. LDU RANK TEST CHI-SQUARED STATISTICS FOR OVERALL SAMPLE

Different regions	rank=r 1(4)*	2(3)	3(2)
Region 1	17766920.80	0.01	0.00
Region 2	2995280.06	0.11	0.00
Region 3	653503.51	0.81	0.00
Region 4	51719713.30	1.03	0.00
Region 5	290106741.00	0.11	0.00
Region 6	567285477.00	0.29	0.00

*Figures in parentheses denote the respective degrees of freedom(d.f).
The critical values at 5 percent level of significance are as follows:
χ^2 with 4 d.f : 9.488.
χ^2 with 3 d.f : 7.815.
χ^2 with 2 d.f : 5.991.

of the form (3.4), (3.5) and (3.6) by maximum likelihood method and then performing likelihood ratio tests. The systems we consider are

$$w_i = \begin{cases} \Lambda_i + \beta_i \ln(\frac{y}{a(p)}) & \text{for WL} \\ \Lambda_i + \beta_i \ln \frac{y}{a(p)} + \rho\beta_i(\ln(\frac{y}{a(p)}))^2 & \text{for QL(rank two)} \\ \Lambda_i + \beta_i \ln \frac{y}{a(p)} + (\rho\beta_i + \frac{\gamma_i}{b(p)})(\ln(\frac{y}{a(p)}))^2 & \text{for QL(rank three),} \end{cases} \quad \ldots(3.7)$$

for $i = 1, 2, .., n$; where

$$\ln a(p) = \sum_{i=1}^{n} \Lambda_i \ln p_i + \sum_j \nu_j d_j,$$

d_j's being the occupational/social dummies,

$$b(p) = \Pi_{i=1}^{n} p_i^{\beta_i},\ \sum_i \beta_i = 0$$

and

$$\sum \gamma_i = 0.$$

It may be noted that the QL(rank three) model nests the QL(rank two) model (when $\lambda_i = 0$ for all i) and the WL model (when $\rho = 0$ and $\lambda_i = 0$ for all i). The log-likelihood values are presented in table 6. The log-likelihood values suggest that QL(rank two) is significantly better than the WL model. However, a comparison between the log-likelihood values of QL(rank two) and QL(rank three) shows that QL(rank three) fails to be a significant improvement over the QL(rank two) model.

Table 6. Log-likelihood values for WL, QL(rank two) and QL(rank three) demand systems

Serial number	System	Number of parameters	Log-likelihood values (L_i)
1	WL	39	30064.74
2	QL(rank two)	40	30076.40
3	QL(rank three)	44	30079.92

$-2(L_1 - L_2) = 23.32$ (Critical value of ${\chi^2}_{(1)}$ at 5% level of significance is 3.84).
$-2(L_2 - L_3) = 7.04$ (Critical value of ${\chi^2}_{(4)}$ at 5% level of significance is 9.49).

It may be mentioned that the results obtained here are in sharp contrast with those for developed countries like the U.S., U.K. and Australia, which are found to have rank three demands [Banks, Blundell and Lewbel (1997), Lewbel(1991), Lancaster and Ray(1996)]. The nonlinearity, in these cases are observed for non-basic goods in the higher income range. Here, the possible explanation for the rank two results could be that (i) the budget shares on non basic goods are rather low and their variation is small and (ii) the range of income (total expenditure) is too small to capture the nonlinearity.

4. Conclusion

In this paper we have explored the forms of the budget share curves of some selected commodities for rural Maharashtra using the nonparametric kernel smoothing technique. Dividing the overall sample into homogeneous subsamples in terms of demographic characteristics, occupational groups and social classes, it turns out that the Quadratic Logarithmic form of the budget share curves provides a very close approximation to the nonparametric curves. The results of this analysis may briefly be summarised as follows:

- In majority of the cases *cereals, fuel and light* and *adult good* turn out to be necessary item, while *other food* is a luxury item. The nature of *clothing*, in some cases varies with the groups. Given the composition of this particular commodity, the result is not unexpected.
- Given income, as the number of children increases the budget share increases for *cereals* and decreases for *adult good* and for *other food*. The effect of children on *fuel and light* and *clothing* is not so evident.

Application of LDU rank test of Gill and Lewbel (1992) to the homogeneous groups indicates a rank of 2 for the demand system consisisting of these commodities at the given price regime. Introduction of price variation through regionwise decomposition of the occupation/ social groups yields the same results for the rank test except in few cases. For the overall sample the 'rank two demand' result is retained in all price situations. The likelihood ratio tests indicate that the rank is indeed two with Quadratic Logarithmic engel curves.

Appendix

- CEREALS: This group includes cereals, gram and cereal substitutes.
- OTHER FOOD: This group includes pulse and pulse products, milk, edible oil, meat, egg, fish, other vegetables, fruits, salt, sugar and spices.
- ADULT GOOD: This group includes tea(number of cups and leaf) and coffee(number of cups and powder), pan leaf, pan finished, supari, katha, lime, other ingredients for pan, ganja, bhang, charas, foreign liquor, opium, bear, other drugs and intoxicants and also biri, cigarettes, tobacco, zarda, adult clothing items comprising dhoti, saree, cloth for shirt, pajama, salwar, coat, trousers etc.
- FUEL AND LIGHT: This group includes coke, firewoods, electricity, dungcake, kerosine, matches, coal, coalgas, gas, other oil used for lighting, candles, gobar gas etc.
- CLOTHING: This group includes chaddar, dopatta, wraper, lungi, gamcha, bedcover, mats, knitting wool etc.

Acknowlegement. Thanks are due to the authorities of the National Sample Survey (NSS) Organisation for permitting the use of updated tapes containing the NSS 38th round data for the purpose of the present analysis. The authors would like to thank Professors Nikhilesh Bhattacharya, Dipankor Coondoo, Probal Chaudhuri and an anonymous referee for their comments and helpful suggestions. They are also grateful to Professor Arthur Lewbel for providing them with the computer program for the rank test.

References

BANKS, J., BLUNDELL, R. AND LEWBEL, A. . (1997). Quadratic engel curves, indirect tax reform and welfare measurement; *Review of Economics and Statistics*, **XXIX**, No-4, 527-39.

BIERENS, H. AND POTT-BUTTER, H. (1990). Specification of household engel function by non-parametric regression; *Econometric Reviews*, **9**, 123-210.

CHAKRABARTY, M. (1998). On measuring the cost of children: The case of rural Maharashtra; *Ph.D Dissertation*, Indian Statistical Institute, Calcutta.

CRAGG, J.G. AND DONALD, S.G. (1995). Inferring the rank of a matrix; *Mimeo*, University of British Columbia.

DEATON, A. (1989). Rice prices and income distribution in Thailand: A non-parametric analysis; *Economic Journal*, **99**, 1-37.

GILL, L. AND LEWBEL, A. (1992). Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models; *Jour. Amer. Statist. Assoc.*, **87**, 766-76.

GORMAN, W.M. (1981). Some engel curves; in *Essays in the Theory and Measurement of Consumer Behaviour* in Honour of Sir Richard Stone, Ed. by Angus Deaton, Cambridge University Press, Cambridge.

HÄRDLE, W. (1990). *Applied Nonparametric Regression*; Cambridge University Press, Cambridge, England.

LANCASTER, G. AND RAY, R. (1996). Comparison of alternative methods of estimating household equivalence scales : The Australian evidence on pooled time series of unit record data; *The Economic Record* (forthcoming).

LEWBEL, A. (1989). A demand system rank theorem; *Econometrica*, **57**, 701-705.

LEWBEL, A. (1991). The rank of demand systems: Theory and nonparametric estimation; *Econometrica*, **59**, 711-730.

NADARAYA, E. A. (1964). On estimating regression; *Theory Prob Appl*, **10**, 186-90.

NICOL, J. (1993). An empirical comparison of nonparametric and parametric engel functions; *Empirical Economics*, **18**, 233-249.

SILVERMAN, B. W. (1986). *Density Estimation for Statistics and Data Analysis*; Chapman and Hall, London.

SUBRAMANIAN, S. AND DEATON, A. (1994). The demand for food and calories - Further evidence from India; *Sarvekshana*, **60**, 1-17.

WATSON, G. S. (1964). Smooth regression analysis; *Sankhyā*, Series A, **26**, 359-72.

YEN, T.S. AND ROE, L.T. (1989). Estimation of a two-level demand system with limited dependent variables; *Amer. Jour. Agricultural Economics*, **71**, 85-98.

AMITA MAJUMDER AND MANISHA CHAKRABARTY
INDIAN STATISTICAL INSTITUTE
203 B.T. ROAD
CALCUTTA 700 035
INDIA
e-mail : amita@isical.ac.in and res9206@isical.ac.in