Discussion Papers No. 223, August 1998
Statistics Norway, Research Department

Leif Brubakk and John K. Dagsvik

Consumer Demand and Unobservable Product Attributes

Abstract:
Traditional approaches to consumer demand modelling ignores the problem associated with product heterogeneity where important product characteristics are latent. The point of departure in the present study is a particular framework developed in Dagsvik (1996a,b) and Dagsvik et al. (1998). In this approach the consumer is assumed to make his choice from a discrete set of product variants. The resulting model has the form of a modified conventional demand system, where the modification consists in replacing the standard price indexes by a indexes which are derived from underlying behavioral assumptions. The empirical application is based on a sample of Norwegian micro data market prices and household consumption.

Keywords: Consumer demand, Differentiated products, Unobserved product heterogeneity

JEL classification: C31, C43, D12

Address: John K. Dagsvik, Statistics Norway, Research Department. E-mail: john.dagsvik@ssb.no

Discussion Papers comprises research papers intended for international journals or books. As a preprint a Discussion Paper can be longer and more elaborated than a usual article by including intermediate calculation and background material etc.

Abstracts with downloadable postscript files of
Discussion Papers are available on the Internet: http://www.ssb.no

For printed Discussion Papers contact:

Statistics Norway
Sales- and subscription service
P.O. Box 1260
N-2201 Kongsvinger

Telephone: +47 62 88 55 00
Telefax: +47 62 88 55 95
E-mail: Salg-abonnement@ssb.no

1. Introduction

A rather typical feature of modern markets for consumer goods is the large variety of product characteristics. In addition, the trading of goods also depends on characteristics only indirectly linked to the respective goods, namely variables that represent the quality and service of the stores that offer the products for sale. This is due to the fact that the consumers often may be quite perceptive to marketing, location and quality of the services offered by the stores. As a consequence, the prices of a given product variant may vary across stores. In principle, it is possible to operate with a demand system where the number of goods correspond to every possible attribute category. To do this in practice, however, is rather problematic. There is limit to how many product variants one can treat as separate goods in a demand system. A major problem is that "quality" is an important aspect of the goods and it is rather difficult to construct variables that are able to capture quality in empirical analyses. The textbook material on consumer demand "solve" this problem by Hicks aggregation. However, an empirically useful application of Hicks aggregation requires strong assumptions about relative prices which may seem severely restrictive in many cases.

In this paper we estimate demand models based on a particular approach which was proposed by Dagsvik, see Dagsvik (1996a,b), and Dagsvik et al. (1998). A crucial feature of this approach is that the choice setting is viewed as a discrete/continuous one, rather than the usual setting with infinitely divisible goods. This we believe, is more realistic in todays markets, since the choice between product variants is typically a discrete one. Furthermore, the product variants are characterized by price and latent product attributes, such as quality. This point of departure has the advantage that under suitable assumptions, it is possible to derive convenient expressions for aggregate relations that correspond to observable quantities and unit values. In particular, it follows that one can apply the unit values to obtain price indexes. We apply these relations to derive an empirical model for consumer expenditure on food, which we subsequently estimate from data on consumer expenditure in Norway.

In Dagsvik et al. (1998) the methodological approach advocated in Dagsvik (1996a) was applied to estimate a particular demand model based on micro data on household expenditure and unit values. However, data on market prices were not available and consequently we were unable to estimate key structural parameters. In the present application, however, we apply a different micro data set that also contains information about market prices, and we therefore are able to recover important structural parameters.

The paper is organized as follows: In Section 2 we review the general choice setting and some basic results obtained in the papers mentioned above are discussed. In Section three we discuss the data and in Section four the empirical model and the estimation strategies are described. In this section we also report different estimation results based on several alternative methods. In Section five we consider policy simulation issues.

2. A summary of the modelling framework

In this section we summarize the theoretical and econometric approach. The point of departure is the approach developed in Dagsvik (1996a), which is further developed and adapted in Dagsvik et al. (1998). A consumer's choice set consists of a wide variety of different products and stores (locations). We assume that we can represent the commodity space by d different types of products (goods), where each product consists of an infinite set of different variants/locations characterized by price and quality attributes. In this paper the d goods refer to the observed commodity categories while the product variants and stores are unobservable to the econometrician. Since the variants and stores are unobservable we may, without loss of generality, treat stores and variants symmetrically in the formalism. Let $Q_j(z)$ be the quantity of observable good j and unobservable location and variant z, and let $T_j^*(z) > 0$, be an unobservable, or partly observable, quality/location attribute associated with variant z within commodity category j, $j \leq d$. The attributes $\left\{T_j^*(z)\right\}$ are consumer specific in the sense that they are subjectively perceived. Let $P_j(z)$ be the price of variant/location z of type j. Evidently, we can represent the vector of product variants and their attributes as

$$\left(\mathbf{Q},\mathbf{T}^*\right) = \underset{z}{\times}\left(Q_1(z), T_1^*(z), Q_2(z), T_2^*(z), \ldots, Q_d(z), T_d^*(z)\right).$$

The consumer is assumed to be perfectly informed about the distribution of product locations, variants and prices. He is assumed to have preferences over ($\mathbf{Q},\mathbf{T}^*$), with utility function that has the structure,

$$U\left(\mathbf{Q},\mathbf{T}^*\right) = u\left(\sum_z S_1(z), \sum_z S_2(z), \ldots, \sum_z S_d(z)\right), \tag{2.1}$$

where

$$S_j(z) = Q_j(z)\, T_j^*(z), \tag{2.2}$$

and $u(\cdot)$ is increasing and quasi-concave. The structure of the utility function implies that within a specific type of good, the different variants are perfect substitutes. This implies that the consumer will only buy one variant of each type of good at a time, i.e., for a given set of consumer specific attributes $\left\{T_j^*(z)\right\}$, only one variant of each type of good will be chosen. This formulation is thus a version of the "Ideal Variety Approach", proposed by Lancaster (1979). Note that the realism of assumption (2.1) depends of course on how detailed the observable types are defined. It also depend on the time unit because the taste-shifters may change from one instant of time to another. If the purchases are made on a daily basis then the perfect substitute assumption might seem rather plausible, while this assumption is quite strong if one assumes that "month" is the proper time unit. The budget constraint is given by

(2.3) $$\sum_{j=1}^{d} \sum_{z} Q_j(z) P_j(z) \leq y$$

where y is income. Let

(2.4) $$R_j(z) = P_j(z) \big/ T_j^*(z) \; .$$

Then it follows that the consumer's optimization problem is equivalent to maximizing (2.1) subject to the budget constraint

(2.5) $$\sum_{j=1}^{d} \sum_{z} S_j(z) R_j(z) \leq y \, .$$

We realize immediately that the problem above is formally equivalent to a conventional consumer optimation problem where $S_j(z), z = 1,2,\ldots,$ are perfect substitutes that enter symmetrically in the model, and $\left\{R_j(z)\right\}$ represent "prices". As mentioned above we realize easily that the consumer will choose only one variant within each observable type of good. Specifically, variant $\hat{z}_j$ will be chosen if

(2.6) $$R_j\left(\hat{z}_j\right) = \min_z R_j(z),$$

which means that $\hat{z}_j$ is the variant with the lowest taste and quality adjusted "price".

For notational convenience, let $\hat{R}_j = R_j(\hat{z}_j)$, $\hat{Q}_j = Q_j(\hat{z}_j)$, $\hat{S}_j = S_j(\hat{z}_j)$ and $\hat{P}_j = P_j(\hat{z}_j)$. Let $y_j(\mathbf{r}, y), j = 1,2,\ldots,d,$ be the expenditure on good of type j that follows from maximizing $u(s_1, s_2, \ldots, s_d)$ subject to $\sum_{j=1}^{d} r_j s_j \le y$, where $\mathbf{r} = (r_1, r_2, \ldots, r_d)$. We realize immediately that the purchased quantity of good j, $\hat{Q}_j$, is given by

(2.7)
$$\hat{Q}_j = \frac{\hat{S}_j \hat{R}_j}{\hat{P}_j} = \frac{y_j(\hat{\mathbf{R}}, y)}{\hat{P}_j}$$

where $\hat{\mathbf{R}} = (\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_d)$. Thus, we have expressed the chosen quantities by means of an ordinary and deterministic demand system and $\hat{\mathbf{R}}$. We shall call $\{\hat{R}_j\}$ virtual prices. The effect of unobserved heterogeneity in quality and preferences is thus entirely captured by the virtual prices. The virtual prices as well as the unit values, $\{\hat{P}_j\}$, are endogeneous because they are associated with the respective chosen product variants/locations. Note that the virtual prices are not observable. They are taste-and-quality-adjusted-prices in the sense that if the virtual prices were known, consumer behavior could be represented by an ordinary deterministic demand system that does not depend on the consumer (within suitable defined population groups) nor on the unobservable product variants. This is so because the "quantities" $S_j(z)$ enter symmetrically in the utility function within each commodity type. Due to this property the virtual prices are in fact latent stochastic price indexes.

To obtain aggregate relations it is necessary to make distributional assumptions. Assume first that there are only a finite number of feasible variants in the market. Write $T_j^*(z) = T_j(z)\,\xi(z)$ where $T_j(z)$ is the mean attribute value of variant z, commodity j in the population and $\{\xi_j(z)\}$ are taste-shifters that represent the heterogeneity in consumer tastes. Following Lancaster (1966) the attributes $\{T_j(z)\}$ correspond to the notion of *vertical* product differentiation, while the taste-shifters $\{\xi_j(z)\}$ correspond to the notion of *horizontal* product differentiation. We shall subsequently call $T_j(z)$ the quality attribute associated with variant z. Let $B_j(p,t)$ be the set of variants of type j with $T_j(z) = t$ and $P_j(z) = p$, and let $b_j(p,t)$ be the number of variants in $B_j(p,t)$. Furthermore, we assume that $\xi_j(z), z = 1,2,\ldots, j = 1,2,\ldots,d,$ are i.i.d. with

(2.8) $$P\left(\xi_j(z) \le y\right) = \exp\left(-y^{-\alpha_j}\right)$$

for $y > 0$, where $\alpha_j > 0$ is a constant. A useful interpretation of α_j is as

(2.9) $$\alpha_j^2 = \frac{\pi^2}{6\,\mathrm{Var}\left(\log \xi_j(z)\right)}.$$

A possible justification for assumption (2.8) is that it is consistent with the notion of independence from irrelevant alternatives which we shall discuss further below. Given the number of variants of each category, $\{b_j(p,t)\}$, the probability, $\hat{g}_j(p,t)$, that a consumer shall choose $\hat{z}_j$ such that $\left(\hat{P}_j = p, \hat{T}_j = t\right)$, is formally defined by

(2.10) $$\begin{aligned}\hat{g}_j(p,t) &= P\left(\min_{z \in B_j(p,t)} R_j(z) = \min_{k,r}\ \min_{z \in B_j(r,k)} R_j(z)\right)\\ &= P\left(\max_{z \in B_j(p,t)}\left(\frac{T_j(z)}{P_j(z)} \cdot \xi_j(z)\right) = \max_{k,r}\ \max_{z \in B_j(r,k)}\left(\frac{T_j(z)}{P_j(z)} \cdot \xi_j(z)\right)\right).\end{aligned}$$

From (2.8) and (2.10) it follows readily that

(2.11) $$\hat{g}_j(p,t) = \frac{\left(\frac{t}{p}\right)^{\alpha_j} g_j(p,t)}{\sum_{(x,y) \in D_j}\left(\frac{y}{x}\right)^{\alpha_j} g_j(x,y)}$$

where D_j is the set of potential variants (combinations of price and quality attributes of type j) and

(2.12) $$g_j(p,t) = \frac{b_j(p,t)}{\sum_{(x,y) \in D_j} b_j(x,y)}.$$

The empirical counterpart of $\hat{g}_j(p,t)$ is the number of consumers that purchase variants within $B_j(p,t)$ to the number of consumers that purchase a variant of type j. The empirical counterpart of $g_j(p,t)$ is the fraction of variants of type j with list price p and quality attribute t that appear in the stores that are accessible to the consumer. How $g_j(p,t)$ is determined is discussed briefly in Dagsvik et al. (1998). Due to the random taste-shifters, $\{\xi_j(z)\}$, a selection effect arises and the distribution of prices (unit values) and quality attributes of the purchased variants will differ from the corresponding distribution of list prices and quality attributes offered in the market. Note that the selection effect decreases when the variance of $\log \xi_j(z)$ increases, and disappears when the variance approaches infinity, which means that the distributions of unit values and market values coincide in the limit.

It follows directly from (2.8) that the distribution of $\hat{R}_j$ has the structure

$$P\left(\hat{R}_j \le r\right) = 1 - \exp\left(-r^{\alpha_j} K_j\right) \tag{2.13}$$

for $r \ge 0$, where

$$K_j = \sum_{(x,y)\in D_j} \left(\frac{y}{x}\right)^{\alpha_j} b_j(x,y) = b_j \sum_{(x,y)\in D_j} \left(\frac{y}{x}\right)^{\alpha_j} g_j(x,y) \tag{2.14}$$

and

$$b_j = \sum_{(x,y)\in D_j} b_j(x,y). \tag{2.15}$$

The virtual prices $\left\{\hat{R}_j\right\}$, have the surprising property that they are stochastically independent of the set $\left\{\left(\hat{P}_k, \hat{T}_k\right), k = 1,2,\ldots,d\right\}$, i.e., the virtual prices are independent of the unit values and quality attributes of the purchased variants. However, as is seen from (2.11) and (2.14) the distribution of virtual prices and the distribution of unit values and chosen attributes are functionally related. This thus means that the "noisy" structure of the preferences imply that the unit values and virtual prices are uncorrelated across purchases for a single consumer, while the corresponding aggregates within, say, $B_j(p,t)$, are dependent.

The discrete setting considered above is somewhat unsatisfactory for several reasons. First, it appears to be a rather wide variety in product quality and location and service of the stores which makes it difficult to classify variants and stores in a few groups. As a result, the distribution of prices—which may be observed—seem to vary nearly continuously across variants and stores. Finally, the set of feasible variants may vary across consumers, due for example to spatial variations in the location of stores. The set of feasible variants may also be genuinly random due to bounded rationality of the consumers. By this it is understood that a consumer in fact only considers a subset of the whole set of feasible alternatives in his decision-making process. In each decision "experiment", these subsets may appear to be random although they will, on average, be more or less closely linked to the underlying "objective" choice set. It is therefore desirable to extend the setting discussed above so as to allow for continuous distributions of prices and quality attributes. This issue is discussed in

Dagsvik (1996a) and Dagsvik et al. (1998). Specifically, under assumptions proposed in Dagsvik (1996a) it follows that (2.11), (2.13) and (2.14) hold, with "sum" replaced by "integral", i.e.

$$(2.16) \qquad K_j = b_j \int_0^\infty \int_0^\infty \left(\frac{y}{x}\right)^{\alpha_j} g_j(x,y)\,dx\,dy$$

and

$$(2.17) \qquad \hat{g}_j(p,t) = \frac{\left(\frac{t}{p}\right)^{\alpha_j} g_j(p,t)}{\int_0^\infty \int_0^\infty \left(\frac{y}{x}\right)^{\alpha_j} g_j(x,y)\,dx\,dy}$$

which of course is completely analogous to (2.11). Since we do not observe the chosen attributes $\left\{T_j\left(\hat{z}_j\right)\right\}$ we need to derive the correspondance between the marginal distributions of $\left\{\hat{P}_j\right\}$ and $\left\{P_j(z)\right\}$, which we denote by $\hat{g}_j(p)$ and g(p). From (2.17) it follows immediately that

$$(2.18) \qquad \hat{g}_j(p) = \frac{p^{-\alpha_j} \lambda_j(p) g_j(p)}{\int_0^\infty x^{-\alpha_j} \lambda_j(x) g_j(x)\,dx}$$

where

$$(2.19) \qquad \lambda_j(p) \equiv E\left(T_j(z)^{\alpha_j} \,\middle|\, P_j(z) = p\right).$$

The interpretation of $\lambda_j(p)^{1/\alpha_j}$ is as the conditional mean quality level (adjusted for the dispersion in taste-shifters) across variants of type j, given price level p. We realize that if $\lambda_j(p) = w_j\, p^{\alpha_j}$, where $w_j > 0$ is a constant, then the distribution of unit values will coincide with the price distribution. In Dagsvik (1996a) it is demonstrated that once $\lambda_j(\cdot)$ and α_j are known, one can express the parameter K_j in the distribution of virtual prices either as a function of the price distribution, or alternatively, as a function of the distribution of unit values. The argument goes as follows: From (2.16), (2.18) and (2.19) we have

$$(2.20) \qquad K_j = c_j \int_0^\infty x^{-\alpha_j} \lambda_j(x) g_j(x)\,dx = b_j\, E\left(P_j(z)^{-\alpha_j} \lambda_j\left(P_j(z)\right)\right).$$

Furthermore (2.18) and (2.20) imply that

$$K_j \hat{g}_j(p) p^{\alpha_j} = b_j g_j(p) \lambda_j(p)$$

which, by (2.19) implies that

$$K_j \int_0^\infty x^{\alpha_j} \hat{g}_j(x) dx \equiv K_j E\left(\hat{P}_j^{\alpha_j}\right) = b_j \int \lambda_j(x) g_j(x) dx \equiv c_j E T_j(z)^{\alpha_j}.$$

Consequently, we have

(2.21) $$K_j = b_j E T_j(z)^{\alpha_j} \left(E \hat{P}_j^{\alpha_j}\right)^{-1}.$$

Unfortunately, neither α_j nor $\lambda_j(\cdot)$ are known. To be able to make inference it is necessary to make further assumptions. In Dagsvik et al. (1998) it is argued that it is plausible to assume that $\lambda_j(p)$ is a power function, i.e.,

(2.22) $$\lambda_j(p)^{1/\alpha_j} = A_j^{1/\alpha_j} p_j^{\kappa_j}$$

where $A_j > 0$ and $\kappa_j > 0$ are constants. From (2.19) it follows that A_j has the interpretation

(2.23) $$A_j = \frac{E\left(T_j(z)^{\alpha_j}\right)}{E\left(P_j(z)^{\alpha_j \kappa_j}\right)}.$$

From (2.22) we realize that $\lambda_j(\cdot)^{1/\alpha_j}$ is convex when $\kappa_j > 1$ and concave when $\kappa_j < 1$. This means that when $\kappa_j > 1$, increasing prices do not reduce the attractiveness of the product variants as much as when $\kappa_j < 1$, due to the fact that high prices are perceived as an indication of high quality, and vice versa. When $\kappa_j > 1$, for example, this relationship between prices and quality is strengthened as the price level increases. When (2.22) is inserted into (2.18) we obtain

(2.24) $$\hat{g}_j(p) = \frac{p^{\alpha_j \kappa_j - \alpha_j} g_j(p)}{\int_0^\infty x^{\alpha_j \kappa_j - \alpha_j} g_j(x) dx}.$$

From (2.22) and (2.20) we can obtain an alternative expression for K_j, namely

(2.25) $$K_j = b_j \, E\, T_j(z)^{\alpha_j} \frac{E\, P_j(z)^{\alpha_j \kappa_j - \alpha_j}}{E\, P_j(z)^{\alpha_j \kappa_j}}.$$

Thus, we may express the parameters $\{K_j\}$ in the distribution of virtual prices (apart from $\left\{E\left(T_j(z)^{\alpha_j}\right)\right\}$) in terms of the distribution of unit values by (2.21), or in terms of the distribution of list prices by (2.25). From (2.13) it follows that

(2.26) $$E\,\hat{R}_{ij} = \Gamma\left(1+\frac{1}{\alpha_j}\right) K_j^{-1/\alpha_j}$$

which indicates that K_j^{-1/α_j} can be interpreted as a so called "elementary" price index that account for product heterogeneity within commodity group j. Note also that K_j^{-1/α_j} depends on the whole price distribution, i.e. it is in this sense a functional. This is of considerable interest in policy contexts because it enables the researchers to carry out policy simulation experiments with respect to nonstandard changes of the price distribution, provided one is willing to assume that $\left\{b_j \, E\, T_j(z)^{\alpha_j}\right\}$ remain unchanged. We shall discuss this further below.

3. Data

The consumption data used in this study are taken from the annual Norwegian household expenditure survey for the years 1989-1994 (see NOS (1993)), where households report their consumption expenditure over a fourteen days period. However, all data on expenditure and quantity are translated into annual figures by multiplying the reported fourteen days figures by twenty-six. Quantity is only reported for food and beverages, which implies that unit values can only be calculated for these goods. Although the accounting period is fourteen days, we have used "month" as time unit. Thus, all households with observed consumption within the same month belong to the same cross-section. This was done to increase the number of observations in each cross-section. The total sample consists of 7 690 households, with roughly 107 households in each cross-section. We have divided household consumption of food and beverages into 11 separate commodity groups. The remaining goods are grouped into two different groups according to their degree of "durability". However, expenditure on the two commodity groups "housing and maintenance" and "transportation equipment" are excluded. The chosen grouping is as follows: "Flour", "meal", "Bakery products" and "Sugar" (1); "Meat and meat products" (2); "Fish and fish products" (3); "Milk", "Cream", "Cheese" and "Eggs" (4); "Edible oils and fat" (5); "Vegetables", "Fruits and berries" (6); "Potatoes and potato products" (7); "Coffee", "Tea and cocoa" (8); "Other foods" (9); "Mineral waters and soft drinks" (10); "Alcohol" (11); "All other non-durables" (12), (including tobacco, cloths, miscellaneous household goods, restaurant visits etc.); "Durables" (13), (including furniture, electrical appliances, etc.).

Market prices are taken from the data base containing the price information used to construct the Norwegian consumer price index (see Statistics Norway (1991)). Each month, about 50,000 prices are collected for roughly 900 "representative" goods, which means that the sample of goods is rather limited. A majority of these prices relates to food and beverages. In this collection process a possible selection bias may arise due to the fact that stores are instructed to report the price of the product variant that is most frequently purchased whenever there is more than one variant that fits the good description. Thus the distribution of the reported prices provide a more or less crude approximation of the true price distribution for the different commodity groups used in this study.

Table 3.1. Summary statistics of expenditure and budget shares for the sample period 1989-1994

Variable	Number of obs.	Sample mean	Standard deviation	Minimum	Maximum
Total expenditure[1]	7 687	163 647.2	106 936.3	8 655.4	1 173 491.0
Expenditure (excl. zero expenditure):					
Flour, meal, bakery products, sugar	7 609	3 936.1	2 902.2	59.2	39 919.1
Meat and meat products	7 420	8 451.9	9 271.0	76.4	210 836.6
Fish and fish products	6 472	2 599.0	3 104.2	12.6	71 343.2
Milk, cream, cheese, eggs	7 646	6 314.8	3 873.2	65.0	33 996.7
Edible oils and fats	6 623	1 007.5	787.6	6.3	7 601.1
Vegetables, fruits and berries	7 559	4 731.4	3 722.4	34.58	83 835.4
Potatoes and potato products	6 415	1 173.1	1 135.0	6.5	19 500.0
Coffee, tea and coccoa	6 023	1 249.2	1 115.2	3.1	34 106.0
Other foods	7 414	4 643.8	5 535.7	6.9	255 970.9
Mineral waters, etc.	6 076	2 295.8	2 167.2	6.9	25 818.0
Alcohol	4 037	5 666.5	7 423.4	3.1	128 596.0
Other non-durables	7 687	86 087.9	61 280.4	725.0	947 794.0
Durables	7 381	38 219.0	51 017.2	1.0	736 539.6
Budget shares (excl. zero expenditure):					
Flour, meal, bakery products, sugar	7 609	0.029	0.023	$5.2 \cdot 10^{-4}$	0.534
Meat and meat products	7 420	0.058	0.052	$6.4 \cdot 10^{-4}$	0.678
Fish and fish products	6 472	0.020	0.024	$6.75 \cdot 10^{-5}$	0.423
Milk, cream, cheese, eggs	7 646	0.047	0.031	$3.1 \cdot 10^{-4}$	0.366
Edible oils and fats	6 623	0.008	0.008	$4.33 \cdot 10^{-5}$	0.147
Vegetables, fruits and berries	7 559	0.033	0.026	$4.6 \cdot 10^{-4}$	0.347
Potatoes and potato products	6 415	0.008	0.009	$4.33 \cdot 10^{-5}$	0.207
Coffee, tea and coccoa	6 023	0.010	0.011	$2.2 \cdot 10^{-5}$	0.235
Other foods	7 414	0.030	0.026	$1.1 \cdot 10^{-4}$	0.736
Mineral waters, etc.	6 076	0.015	0.013	$3.38 \cdot 10^{-5}$	0.136
Alcohol	4 037	0.033	0.040	$1.61 \cdot 10^{-5}$	0.535
Other non-durables	7 687	0.532	0.152	0.0364	1.000
Durables	7 381	0.195	0.148	$6.99 \cdot 10^{-6}$	0.889

[1] Total expenditure = total living expenditure - "housing and maintainance" - "transport equipment".

Table 3.2. Summary statistics of unit values and market prices for the sample period 1989-1994

Variable	Number of obs.	Sample mean	Standard deviation	Minimum	Maximum
Unit values:					
Flour, meal, bakery products, sugar	26 677	30.21	26.47	0.94	418.67
Meat and meat products	21 936	85.96	47.51	2.86	498.00
Fish and fish products	14 300	54.79	31.28	1.67	400.00
Milk, cream, cheese, eggs	24 441	35.22	24.66	1.46	182.00
Edible oils and fats	7 909	25.85	11.44	1.25	264.00
Vegetables, fruits and berries	37 177	20.05	16.20	0.51	589.00
Potatoes and potato products	9 203	29.99	31.36	0.04	250.00
Coffee, tea and coccoa	8 245	94.04	106.48	2.95	598.89
Other foods	16 025	102.63	70.82	3.88	598.78
Mineral waters, etc.	6 074	12.45	4.09	0.90	50.00
Alcohol	5 136	65.68	81.27	1.86	540.00
Market prices:					
Flour, meal, bakery products, sugar	239 602	35.73	38.42	2.20	366.67
Meat and meat products	245 295	98.00	56.31	8.90	295.90
Fish and fish products	152 134	68.77	48.57	9.90	278.00
Milk, cream, cheese, eggs	138 705	41.90	33.44	4.30	151.20
Edible oils and fats	43 546	32.80	12.50	11.00	66.36
Vegetables, fruits and berries	337 124	21.93	13.59	1.00	117.86
Potatoes and potato products	46 548	20.13	21.44	0.76	98.56
Coffee, tea and coccoa	56 273	180.79	152.28	9.90	599.40
Other foods	174 163	68.87	67.70	2.50	500.00
Mineral waters, etc.	47 691	13.93	5.11	0.14	32.43
Alcohol	26 397	39.42	68.00	8.69	531.43

4. Empirical specification and estimation based on a modified Linear Expenditure System

4.1. The empirical specification

In this section, we apply the framework discussed in Section 2 to estimate a consumer demand system consisting of thirteen different commodity groups. The point of departure is the LES model, given by

$$(4.1) \qquad y_{ijt} = \gamma_j \hat{R}_{ijt} + \beta_j \left(y^*_{it} - \sum_{k=1}^{13} \gamma_k \hat{R}_{ikt} \right), \quad j = 1, \ldots, 13,$$

where y_{ijt} denotes the expenditure for household i on commodity j in month t, and $\hat{R}_{ijt}$ is the corresponding virtual price. Total expenditure for household i in month t is denoted by y^*_{it}, and $\{\gamma_j\}$ and $\{\beta_j\}$ are unknown parameters. The adding-up restriction implies that

$$\sum_{k=1}^{13} \beta_k = 1.$$

As mentioned above, the virtual prices are unobservable and must be related to some observables to make identification possible. From (2.24) and (2.26) it follows that

$$E\,\hat{R}_{ijt} = \Gamma\left(1 + 1/\alpha_j\right) K_{jt}^{-1/\alpha_j} = \Gamma\left(1 + \frac{1}{\alpha_j}\right)^{1/\alpha_j} b_j^{-1/\alpha_j} \left(E\,T_{jt}(z)^{\alpha_j}\right)^{-1/\alpha_j} \left(E\,\hat{P}_{ijt}^{\alpha_j}\right)^{1/\alpha_j}. \tag{4.2}$$

If we assume that $E\,T_{jt}(z)^{\alpha_j}$ is constant over time we can, without loss of generality, choose b_j such that

$$b_j\, E\,T_{jt}(z)^{\alpha_j} = \Gamma\left(1 + \frac{1}{\alpha_j}\right)^{\alpha_j}. \tag{4.3}$$

Total expenditure per month, y_{it}^*, is not observed. We therefore use total expenditure per year divided by 12, y_{it}, as an instrument for y_{it}^*. From (4.1) to (4.3) we now get

$$y_{ijt} = \gamma_j \left(E\,\hat{P}_{ijt}^{\alpha_j}\right)^{1/\alpha_j} + \beta_j \left(y_{it} - \sum_{k=1}^{13} \gamma_k \left(E\hat{P}_{ikt}^{\alpha_j}\right)^{1/\alpha_j}\right) + \varepsilon_{ijt} \tag{4.4}$$

where

$$\varepsilon_{ijt} = \gamma_j \left(\hat{R}_{ijt} - E\,\hat{R}_{ijt}\right) - \beta_j \sum_{k=1}^{13} \gamma_k \left(\hat{R}_{ikt} - E\,\hat{R}_{ikt}\right) + \beta_j \left(y_{it}^* - y_{it}\right). \tag{4.5}$$

It follows that the random terms ε_{ijt} will have zero mean. The system of equations (4.4) is the basis for the estimation strategies to be discussed next.

4.2. Estimation; Method I

The first approach to estimation is the most intuitive one. For simplicity, we chose to proceed in two stages. In the first stage, the marginal budget shares, $\{\beta_j\}$, are estimated from the following transformed system

$$y_{ijt} - \bar{y}_{jt} = \beta_j (y_{it} - \bar{y}_t) + (\varepsilon_{ijt} - \bar{\varepsilon}_{jt}) \tag{4.6}$$

where $\bar{y}_{jt}$ and $\bar{y}_t$ denote the respective population means of $\{y_{ijt}\}$ and $\{y_{it}\}$. This approach yields consistent estimates of the marginal budget shares and the results can be found in Table 4.1. In the second stage the remaining parameters are estimated by non-linear least squares. As mentioned in Section 3, unit values can only be calculated for food and beverages. Thus, for the remaining goods which are summarized in the two commodity groups "other non-durables" and "durables", labelled commodity group 12 and 13, respectively, we constructed two proxies for the mean of the corresponding virtual prices. These proxies are defined as weighted averages of the price indexes, I_{kt}, of the different goods included in the two commodity groups. As weights we used the corresponding budget shares, w_k. If we now derive the aggregate version of (4.4), we obtain the following relation

$$\bar{y}_{jt} = \gamma_j Z_{jt} + \beta_j \left(\bar{y}_t - \sum_{k=1}^{13} \gamma_k Z_{kt} \right) + u_{jt} \tag{4.7}$$

where

$$Z_{jt} = \begin{cases} \left(\dfrac{1}{N_{jt}} \displaystyle\sum_{i=1}^{N_{jt}} \hat{P}_{ijt}^{\alpha_j} \right)^{1/\alpha_j} & \text{if} \quad j = 1, \ldots, 11 \\ \displaystyle\sum_k w_{kj} I_{kjt} & \text{if} \quad j = 12,13 \end{cases} \tag{4.8}$$

and $\{u_{jt}\}$ are error terms that have approximately zero means.

Although all parameters $\{\gamma_j\}$ and $\{\alpha_j\}$ in (4.7) can in principle be estimated by non-linear least squares or maximum likelihood procedures, it turned out that the data did not vary sufficiently to allow estimation of commodity-specific $\{\alpha_j\}$. Thus, we had to impose the restriction $\alpha_j = \alpha$ for all j. The parameter estimates are displayed in Table 4.1 (Model A). For the sake of comparison, we also estimated a system identical to (4.7) except for the prices which are generalized (share weighted) price indexes. This system is referred to as Model B and its estimates can also be found in Table 4.1.

Table 4.1. Parameter estimates of the modified Linear Expenditure System. Method I[1]

Commodity groups	β_j	Model A γ_j	Model A α_j	Model B γ_j^a
Flour, meal, bakery products, sugar	0.017 (30.7)	113.133 (10.3)	0.63 (3.9)	97.362 (10.1)
Meat and meat products	0.034 (20.9)	79.348 (12.5)		80.367 (10.5)
Fish and fish products	0.010 (16.6)	40.395 (13.3)		43.100 (11.4)
Milk, cream, cheese, eggs	0.020 (31.2)	165.330 (13.5)		157.489 (15.2)
Edible oils and fats	0.006 (34.3)	27.148 (8.5)		24.156 (6.5)
Vegetables, fruits and berries	0.020 (32.3)	199.194 (11.7)		203.953 (10.2)
Potatoes and potato products	0.007 (32.7)	31.712 (5.9)		30.728 (6.5)
Coffee, tea and coccoa	0.006 (27.0)	11.779 (8.7)		12.801 (8.8)
Other foods	0.024 (20.1)	35.592 (9.7)		28.149 (8.0)
Mineral waters, etc.	0.012 (29.2)	135.507 (10.1)		102.192 (8.0)
Alcohol	0.023 (17.8)	81.303 (8.8)		69.993 (9.9)
Other non-durables	0.476 (100.7)	234.641 (9.4)		212.214 (7.1)
Durables	0.345 (73.1)	80.359 (3.6)		60.492 (2.3)

[1] t-values in parenthesis.

From Table 4.1, we note that the estimated alpha is 0,63. For both models, the estimated $\{\gamma_j\}$ are positive and highly significant. By comparing the parameter estimates of the two models, we note that they are quite similar.

4.3. Estimation; Method II

We shall now discuss an alternative approach to estimating the parameters $\{\alpha_j\}$ and $\{\gamma_j\}$. Let

$$V_{ijrt} = \frac{y_{ijt}}{\beta_j} - \frac{y_{irt}}{\beta_r}.$$

From (4.1), it follows that

(4.9) $$V_{ijrt} = a_j \hat{R}_{ijt} - a_r \hat{R}_{irt},$$

where $a_j = \gamma_j / \beta_j$. From (4.9) we obtain that

(4.10) $$\mathrm{Cov}\left(V_{ijrt}, V_{ijkt}\right) = a_j^2 \,\mathrm{Var}\,\hat{R}_{ijt},$$

for $j \neq r, j \neq k,$ and $r \neq k$. It follows easily from (2.13) (cf. Dagsvik, 1996a) that

(4.11) $$\mathrm{Var}\,\hat{R}_{ijt} = m_j^{-2}\left(E\,\hat{R}_{ijt}\right)^2$$

where

(4.12) $$m_j^{-2} = \frac{\Gamma\left(1+\frac{2}{\alpha_j}\right)}{\Gamma\left(1+\frac{1}{\alpha_j}\right)^2} - 1$$

and $\Gamma(\cdot)$ as usual denotes the Gamma function. Let

(4.13) $$S_{jt}^2 = \frac{1}{(d-1)(d-2)} \sum_{\substack{j\neq k \\ j\neq r,\, k\neq r}} \frac{1}{N_t} \sum_{i=1}^{N_t} \left(V_{ijkt} - \overline{V}_{jkt}\right)\left(V_{ijrt} - \overline{V}_{jrt}\right)$$

where $\overline{V}_{jkt}$ denotes the mean of V_{ijkt} across households and d is the number of goods. (In our application, $d = 11$.) From (4.10) and (4.11), it now follows that

(4.14) $$E\,S_{jt}^2 = \frac{1}{(d-1)(d-2)} \sum_{\substack{j\neq k \\ j\neq r,\, k\neq r}} \mathrm{cov}\left(V_{ijkt}, V_{ijrt}\right) = a_j^2 \,\mathrm{Var}\,\hat{R}_{ijt} = m_j^{-2} a_j^2 \left(E\,\hat{R}_{ijt}\right)^2$$

which implies that if N_t is large, then

(4.15) $$\gamma_j \, E\hat{R}_{ijt} = \beta_j a_j \, E\hat{R}_{ijt} \cong \beta_j m_j S_{jt} .$$

When (4.15) is inserted into the aggregate counterpart of (4.1) we obtain

(4.16) $$\bar{y}_{jt} - \beta_j \bar{y}_t = m_j \beta_j S_{jt} - \sum_{k=1}^{13} m_k \beta_k S_{kt} + \eta_{jt}$$

where η_{jt} is a random term defined by

(4.17) $$\eta_{jt} = \beta_j \left(\bar{y}_t^* - \bar{y}_t\right) + \beta_j m_j \left(\hat{R}_{jt} - S_{jt}\right) - \beta_j \sum_{k=1}^{13} \beta_k m_k \left(\hat{R}_{kt} - S_{kt}\right),$$

with $\hat{R}_{jt}$ denoting the population mean of the virtual prices in period t for commodity j. From (4.17) we realize that $E\,\eta_{jt} \approx 0$, when the sample is large.

The development above, which is due to Dagsvik et al. (1998), suggests an estimation procedure in several stages.

Stage one: Estimate $\{\beta_j\}$ as in Method I and use these estimates $\{\hat{\beta}_j\}$ to compute $\{S_j\}$ given by (4.13).

Stage two: Estimate $\{m_j\}$ by using (4.16) with $\{\hat{\beta}_j S_{jt}\}$ as explanatory variables and $\{\bar{y}_{jt} - \hat{\beta}_j \bar{y}_t\}$ as dependent variables.

Stage three: From the estimates of $\{m_j\}$, compute estimates of $\{\alpha_j\}$ from (4.12).

Stage four: Use the estimates of $\{\alpha_j\}$ to compute the price indexes given by (4.8) and estimate $\{\gamma_j\}$ by using (4.7).

Estimation results are displayed in Table 4.2.

Table 4.2. Estimates of the modified Linear Expenditure System. Method II[*]

Commodity groups	m_j	γ_j	α_j
Flour, meal, bakery products, sugar	0.677 (23.1)	111.409 (13.0)	0.69
Meat and meat products	0.464 (23.4)	81.058 (13.5)	0.51
Fish and fish products	0.362 (19.2)	41.766 (14.5)	0.43
Milk, cream, cheese, eggs	1.142 (39.4)	144.777 (19.1)	1.14
Edible oils and fats	0.062 (1.5)	28.062 (8.8)	0.20
Vegetables, fruits and berries	0.643 (22.2)	198.983 (13.1)	0.66
Potatoes and potato products	0.104 (3.1)	42.497 (8.2)	0.24
Coffee, tea and coccoa	0.212 (6.2)	13.411 (11.2)	0.32
Other foods	0.216 (9.9)	37.878 (10.5)	0.32
Mineral waters, etc.	0.298 (11.7)	138.183 (10.4)	0.38
Alcohol	0.350 (20.0)	90.184 (13.1)	0.42
Other non-durables	0.434 (15.9)	236.523 (9.6)	0.49
Durables	-0.144 (5.9)	82.042 (3.8)	0.27

[*] t-values in parenthesis.

Note that the coefficient of variation of $\hat{R}_j$ is inversely related to α_j. Thus, large α_j corresponds to a small coefficient of variation, which means that a large α_j implies little heterogeneity in consumer tastes. From Table 4.2 we find that the smallest α is obtained for commodity group 5, which contains "edible oils and fat". This would imply that there exists a relatively high degree of heterogeneity in taste for this group. The largest alpha is found for commodity group 4 ("milk, cream, cheese and eggs"). However, the main impression is that the α-parameters are quite similar across commodity groups, indicating that heterogeneity in taste does not vary substantially across commodity groups.

4.4. The relation between price and quality

Based on the marginal density of the unit values, given by (2.24), Dagsvik (1996b) considers a (modified) maximum likelihood procedure to estimate $n_j \equiv \alpha_j \kappa_j - \alpha_j$. He shows that the first order condition for maximum is given by

(4.18)
$$\frac{1}{N_{jt}} \sum_{i=1}^{N_{jt}} \log \hat{P}_{ijt} = \frac{\sum_{k=1}^{M_{jt}} P_{kjt}^{n_j} \log\left(P_{kjt}\right)}{\sum_{k=1}^{M_{jt}} P_{kjt}^{n_j}}$$

where M_{jt} is the number of variants/locations. From (4.18), it is clear that we would obtain the same estimate of n_j by applying non-linear least squares to estimate the following regression

(4.19)
$$\log \hat{P}_{ijt} = \frac{\sum_{k=1}^{M_{jt}} P_{kjt}^{n_j} \log\left(P_{kjt}\right)}{\sum_{k=1}^{M_{jt}} P_{kjt}^{n_j}} + e_{ijt}$$

where e_{ijt} is a zero-mean random error term. In the expressions (4.18) and (4.19), n_j is constant over time. The estimates of $\left\{n_j\right\}$ are displayed in Table 4.3. In addition, we report the corresponding κ_j in the case where $\left\{\alpha_j\right\}$ are given in Table 4.2.

Table 4.3. Estimates of n_j and κ_j[1]

Commodity groups	n_j	κ_j
Flour, meal, bakery products, sugar	-0.10 (15.5)	0.86
Meat and meat products	-0.25 (27.7)	0.51
Fish and fish products	-0.49 (37.9)	-0.14
Milk, cream, cheese, eggs	-0.05 (8.8)	0.96
Edible oils and fats	-1.45 (52.1)	-6.25
Vegetables, fruits and berries	-0.30 (33.5)	0.55
Potatoes and potato products	0.14 (9.6)	1.58
Coffee, tea and coccoa	-0.91 (46.0)	-1.84
Other foods	0.83 (109.6)	3.59
Mineral waters, etc.	-0.59 (17.4)	-0.55
Alcohol	0.58 (53.9)	2.38

[1] t-values in parenthesis.

Recall the interpretation of κ_j discussed in Section 2. If $\kappa_j > 1$, or equivalently $n_j > 0$, expensive variants are more attractive than cheaper variants, because in this case expensive variants are perceived to have a higher quality/price ratio. The opposite is true when $\kappa_j < 1$. A negative value of κ_j does not make much sense since it means that quality is decreasing as we move from cheaper to more expensive variants. As can be seen from Table 4.3, this is the case for commodity group three (Fish and fish products), group five (Fat and butter), group eight (Coffee, tea and cocoa), and group ten (Mineral waters). Furthermore, the results indicate a strong relation between quality and price for commodity group 7 (Potatoes and potato products), group 9 (Other foods) and group 11 (Alcohol). For the remaining commodity groups, there seems to exist a positive, but somewhat weaker relation between quality and price. However, several of the estimates of n_j do not seem reasonable. This is particularly the case for the groups "Edible oils and fats" and "Coffee, tea and cocoa". The relationship between the distribution of unit values and list prices are influenced by $\{\kappa_j\}$ in the following way. Without loss of generality, assume that the probability mass for prices less than one is negligible. We then realise from (2.24) that $\kappa_j < 1$ implies that $\hat{g}_j(p)$ is more skew to the left than $g_j(p)$, whereas the opposite is true when $\kappa_j > 1$.

From (2.17), (2.25) and (2.26), we have that

(4.20) $$E\,\hat{R}_{jt} = \Gamma\left(1+\frac{1}{\alpha_j}\right) K_{jt}^{-1/\alpha_j} = \left(b_j\, E\,T_j(z)^{\alpha_j}\right)^{-1/\alpha_j} \Gamma\left(1+\frac{1}{\alpha_j}\right)\left(\frac{E\,P_{jt}(z)^{n_j+\alpha_j}}{E\,P_{jt}(z)^{n_j}}\right)^{1/\alpha_j}.$$

Provided we are willing to assume that $E\,T_j(z)^{\alpha_j}$ remains constant over time we may without loss of generality assume that (4.3) holds. Let

(4.21) $$\tilde{Z}_{jt} = \left(\frac{\sum_z P_{jt}(z)^{n_j+\alpha_j}}{\sum_z P_{jt}(z)^{n_j}}\right)^{1/\alpha_j}.$$

We can now estimate (4.7) based on the estimates of $\{\alpha_j\}$ given in Table 4.2. Alternatively, it is possible to apply Method I to obtain estimates of $\{\gamma_j\}$ and an estimate of a common α with $\{Z_{jt}\}$ replaced by $\{\tilde{Z}_{jt}\}$ given in (4.21) where $\{P_{jt}(z)\}$ are the observed prices. We have only applied Method II. The estimation results are displayed in Table 4.4, column two.

As suggested above, we have serious doubts about the reliability of the estimates of $\{n_j\}$ given in Table 4.3. It may also be reasonable to assume that $\{\alpha_j\}$ have been overestimated due to the fact that corner solutions have been ignored in the present analysis. This is so because the observations with zero expenditure have been ignored and consequently the households in the remaining sample are more homogeneous than the households in the full sample. Since increasing heterogeneity corresponds to decreasing α, the argument is complete. When α_j decreases it follows readily (cf. Dagsvik et al., 1998) that

$$\left(\frac{\sum_z P_{jt}(z)^{n_j+\alpha_j}}{\sum_z P_{jt}(z)^{\alpha_j}}\right)^{1/\alpha_j} \xrightarrow[\alpha_j\to 0]{} \left(\prod_{z=1}^{M_{jt}} P_{jt}(z)\right)^{1/M_{jt}}.$$

Note that the limiting expression is the geometric mean of the market prices, and it is independent of n_j. Hence, when α_j is small one can use price indexes computed as geometric means of market prices.

In column 3 of Table 4.4 we report estimates obtained when geometric means are used as price indexes. From these estimates we realize that the estimates of $\{\gamma_j\}$ seem very robust with respect to the values of $\{\alpha_j\}$. If we compare the estimates in Table 4.4 with those in Table 4.2 we see that some of them are quite different.[1] There may be at least two reasons for this. First the data on prices are not obtained by selecting a random sample of variants and stores, but rather so-called "representative goods". Second, the assumption that variants within each commodity group are perfect substitutes (Eq. (2.2)) may not hold on the relative high aggregation level used in this paper. In either case the index price indexes computed by means of unit values may differ from the corresponding indexes computed from observations on market prices.

[1] Their respective 95 per cent asymptotic confidence intervals do, however, overlap.

Table 4.4. Parameter estimates of the Linear Expenditure System based on market prices[1]

Commodity groups	α_j	Price indexes with estimated $\{\alpha_j\}$ γ_j	Price indexes with $\alpha_j = 0$ γ_j
Flour, meal, bakery products, sugar	0.69	87.490 (10.4)	88.365 (9.2)
Meat and meat products	0.51	71.692 (10.8)	70.771 (9.6)
Fish and fish products	0.43	38.608 (11.9)	37.591 (10.6)
Milk, cream, cheese, eggs	1.14	146.473 (16.0)	141.123 (14.3)
Edible oils and fats	0.20	21.371 (6.4)	21.956 (5.7)
Vegetables, fruits and berries	0.66	174.796 (10.4)	172.406 (9.2)
Potatoes and potato products	0.24	23.949 (6.3)	24.155 (5.5)
Coffee, tea and coccoa	0.32	10.704 (8.7)	8.543 (7.4)
Other foods	0.32	29.172 (8.1)	27.758 (7.1)
Mineral waters, etc.	0.38	89.958 (7.8)	88.068 (6.9)
Alcohol	0.42	65.794 (10.2)	64.677 (9.1)
Other non-durables	0.49	187.445 (7.1)	183.496 (6.2)
Durables	0.27	38.783 (1.7)	35.403 (1.4)

[1] t-values in parenthesis.

We have also used Method I to estimate $\{\gamma_j\}$ and a common α simultaneously. In this case we obtain the estimate, $\hat{\alpha} = 0.94$, with t-value 3.5, which is somewhat higher than the corresponding estimate obtained when using unit values (cf. Table 4.1).

We end this section by illustrating with an example how the modelling framework discussed above can be applied in policy simulation experiments. Suppose that we can identify which market price observations are associated with the respective variants and/or stores. Then it is easy to perform policy simulation experiments in which prices are increased for selected variants or store. Consider as an example the case of prices of all variants z that belong to a set C are changed by a factor f. Then the estimate of the predicted price index in (4.21), $\tilde{Z}^{P}_{jt}$ (say), becomes

(4.22)
$$\tilde{Z}^{P}_{jt} = \left(\frac{\sum_{z \notin C} P_{jt}(z)^{\alpha_j + n_j} + f^{\alpha_j + n_j} \sum_{z \in C} P_{jt}(z)^{\alpha_j + n_j}}{\sum_{z \notin C} P_{jt}(z)^{n_j} + f^{n_j} \sum_{z \in C} P_{jt}(z)^{n_j}} \right)^{1/\alpha_j} .$$

This example provides a nice illustration of the potential of this framework to perform sophisticated policy simulations with respect to particular price policy interventions.

6. Concluding remarks

In this paper we have discussed the notion of product heterogeneity when important product characteristics are unobservable to the econometrician. We have applied and modified a particular methodology developed in Dagsvik (1996a,b) and Dagsvik et al. (1998), and we demonstrated that this methodology works well in practice. In the current application we used a LES-type of framework. We have also experimented with the AIDS demand system, but we found that our data set is too small to provide reliable estimates. Also the observations on market prices are not ideal for our purpose. The data on market prices are based on the notion of so called "representative" goods; i.e., goods that are supposed to represent the variants within respective commodity groups. In other words, the prices of the representative goods are in fact a type of price indexes that are designed to "solve" an aggregation problem. In contrast, the present methodology requires data on a random sample of the "basic" market prices. Thus, as noted above, the estimates of the parameters $\{n_j\}$ are rather questionable. However, the estimates of $\{\gamma_j\}$ in the LES system seem to be quite robust with respect to errors in the estimates of $\{n_j\}$ when data on market prices are used to compute prices indexes.

Ideally, it would be desirable to use a more disaggregated level than the one we have used in this paper. Unfortunately, the sample of households is not large enough to allow a further disaggregation of the commodity groups.

Let us emphasize an important aspect of the present approach. If the analyst only is interested in carrying out policy experiments with proportional changes in market prices, then one only needs data on expenditure and unit values to estimate the demand model. However, if one is interested in assessing the impact of changes in the variance (say) of the market price distribution then one also needs data on prices to recover the structural parameters, $\{\kappa_j\}$.

Finally, it is of interest to note that, as a byproduct of our analysis, we obtain a particular index formula that can be used to construct prices indexes for elementary aggregates (i.e. aggregates at the lowest level). The last issue is discussed in a separate paper, cf. Dagsvik and Brubakk (1998). We have also illustrated the application of this formula by performing a particular policy simulation experiment.

References

Dagsvik, J.K (1996a): Consumer Demand with Unobservable Product Attributes. Part I: Theory, *Discussion Papers* 166, Statistics Norway.

Dagsvik, J.K (1996b): Consumer Demand with Unobservable Product Attributes. Part II: Inference, *Discussion Papers* 167, Statistics Norway.

Dagsvik, J.K , Y. Zhu and R. Aaberge (1998): Consumer Demand of Differentiated Products: An Empirical Analysis Based on Chinese Microdata, *Discussion Papers* ?, Statistics Norway

Dagsvik, J.K. and L. Brubakk (1998): Price Indexes for Elementary Aggregates Derived from Behavioral Assumptions. *Discussion Papers* (to appear), Statistics Norway.

Deaton, A. and J. Muellbauer (1980): *Economics and Consumer Behaviour.* Cambridge: Cambridge University Press.

Lancaster, K.J. (1966): A New Approach to Consumer Theory. *Journal of Political Economy*, **74**, 132-157.

Lancaster, K.J. (1979): *Variety, Equity and Efficiency*. Columbia University Press, New York.

Appendix A

Table A1. Price elasticities for LES Model A based on estimates from Table 4.1

Commodity groups	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13
1	0.754	-0.013	-0.014	-0.014	-0.012	-0.013	-0.011	-0.013	-0.012	-0.012	-0.013	-0.012	-0.008
2	-0.026	0.739	-0.028	-0.029	-0.023	-0.027	-0.022	-0.026	-0.025	-0.025	-0.027	-0.024	-0.016
3	-0.008	-0.008	0.797	-0.008	-0.007	-0.008	-0.007	-0.008	-0.007	-0.007	-0.008	-0.007	-0.005
4	-0.015	-0.015	-0.016	0.812	-0.014	-0.016	-0.013	-0.015	-0.015	-0.014	-0.015	-0.014	-0.009
5	-0.004	-0.004	-0.005	-0.005	0.672	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.004	-0.003
6	-0.016	-0.016	-0.016	-0.017	-0.014	0.766	-0.013	-0.015	-0.015	-0.015	-0.016	-0.014	-0.009
7	-0.005	-0.005	-0.005	-0.006	-0.005	-0.005	0.638	-0.005	-0.005	-0.005	-0.005	-0.005	-0.003
8	-0.005	-0.005	-0.005	-0.005	-0.004	-0.005	-0.004	0.747	-0.005	-0.004	-0.005	-0.004	-0.003
9	-0.018	-0.018	-0.019	-0.020	-0.016	-0.018	-0.015	-0.018	0.715	-0.017	-0.018	-0.017	-0.011
10	-0.009	-0.009	-0.009	-0.010	-0.008	-0.009	-0.007	-0.009	-0.009	0.715	-0.009	-0.008	-0.005
11	-0.018	-0.018	-0.019	-0.019	-0.016	-0.018	-0.015	-0.018	-0.017	-0.017	0.755	-0.017	-0.011
12	-0.365	-0.364	-0.383	-0.394	-0.322	-0.372	-0.306	-0.358	-0.349	-0.344	-0.368	0.372	-0.215
13	-0.265	-0.264	-0.278	-0.286	-0.233	-0.270	-0.222	-0.259	-0.253	-0.249	-0.267	-0.245	0.296

Estimation of a modified Almost Ideal Demand System (AIDS)

Let w_{ijt} denote the budget share of commodity j for consumer i in period t. The AIDS model (cf. Deaton and Muellbauer, 1980) implies that the budget shares in a modified AIDS model can be expressed as follows

(B.1) $$w_{ijt} = h_j + \sum_{k=1}^{13} \delta_{jk} \log \hat{R}_{ijt} + \theta_j \log\left(\frac{y_{it}^*}{q_{it}}\right),$$

where q_{it} is a price index (consumer specific) defined by

(B.2) $$\log q_{it} = h_0 + \sum_{k=1}^{13} h_k \log \hat{R}_{ikt} + \frac{1}{2}\sum_{j=1}^{13}\sum_{k=1}^{13} \delta_{jk} \log \hat{R}_{ijt} \log \hat{R}_{ikt}$$

and h_0, $\{h_j\}$, $\{\beta_j\}$ and $\{\delta_{jk}\}$ are parameters to be estimated. The following restictions apply

$$\sum_{j=1}^{13} h_j = 1, \quad \delta_{jk} = \delta_{kj}$$

and

$$\sum_{j=1}^{13} \theta_j = \sum_{j=1}^{13} \delta_{jk} = \sum_{k=1}^{13} \delta_{jk} = 0.$$

Since $\{\hat{R}_{ijt}\}$ is not observable, (B.1) cannot be estimated directly. However, (B.1) can be written as follows

(B.3) $$w_{ijt} = h_j + \sum_{k=1}^{13} \delta_{jk} E \log \hat{R}_{ijt} + \theta_j \log y_{it} - \theta_j E \log q_{it} + u_{ijt}$$

where

(B.4) $$u_{ijt} = \sum_{k=1}^{13} \delta_{jk} \left(\log \hat{R}_{ikt} - E \log \hat{R}_{ikt}\right) - \theta_j \left(\log q_{it} - E \log q_{it}\right) + \theta_j \left(y_{it}^* - y_{it}\right).$$

Note that one may estimate the marginal budget shares $\{\theta_j\}$ from the following transformed model

(B.5) $$w_{ijt} - \overline{w}_{.jt} = \theta_j \left(\log y_{it} - \overline{\log y_{.t}}\right) + \left(u_{ijt} - \overline{u}_{.jt}\right).$$

From (2.13) it follows readily that

(B.6) $$E \log \hat{R}_{ijt} = -\frac{0.5772}{\alpha_j} - \frac{1}{\alpha_j} \log K_{jt} = -\frac{0.5772}{\alpha_j} - \frac{1}{\alpha_j} \log\left(b_j \frac{E\, T_j^{\alpha_j}}{E\, \hat{P}_j^{\alpha_j}} \right).$$

We can therefore write

(B.7) $$\overline{w}_{jt} - \theta_j \overline{\log y_{it}} + \theta_j \log \tilde{q}_t = \tilde{h}_j + \sum_{k=1}^{13} \delta_{jk} \log Z_{kt} + \nu_{jt}$$

where $\tilde{q}_t$ is a price index (for example a Laspeyres index),

$$\tilde{h}_j = h_j - \sum_{k=1}^{13} \left(\frac{0.5772}{\alpha_k} + \frac{1}{\alpha_k} \log\left(b_k\, E\, T_{kt}(z)^{\alpha_k} \right) \right),$$

$$Z_{kt} = \frac{1}{\alpha_k} \log E\, \hat{P}_{jt}^{\alpha_j},$$

$\overline{w}_{jt}$ and $\overline{\log y_{it}}$ denote population means and ν_{jt} is a random error term which has approximately zero mean. One can estimate this model in two stages using a method similar to Method I with $\alpha_j = \alpha$ for $j \leq 11$. Thus, one first estimates $\{\theta_j\}$ which are used to compute $\overline{w}_{jt} - \theta_j \overline{\log y_{it}} + \theta_j \log \tilde{q}_j$. In the second stage one estimates α, $\{\delta_{jk}\}$ and $\{\tilde{h}_j\}$ by the method of non-linear least squares.