Discussion Papers No. 234, October 1998
Statistics Norway, Research Department

John K. Dagsvik and Leif Brubakk

Price Indexes for Elementary Aggregates Derived from Behavioral Assumptions

Abstract:
This paper discusses the properties of price- and Cost-of-Living indexes that follow from specific assumptions about the structure of consumer preferences. Of particular interest are indexes for elementary aggregates. In the first part of the paper we show how particular indexes for elementary aggregates emerge from a micro model with heterogeneous consumers and unobservable choice sets of product variants. Subsequently, we demonstrate that these indexes also follow from a particular preference structure of a representative consumer. Indexes that are currently used in many countries emerge as special cases of the ones proposed in this paper.

Keywords: Elementary aggregates, Price indexes, Cost-of-Living indexes, Price aggregation

JEL classification: C43

Acknowledgement: We are grateful for comments by Olav Ljones, Lasse Sandberg and Ådne Cappelen

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

1. Introduction

The problem of computing price indexes is an important and troublesome aspect of the construction of national accounts data and of macro- as well as microeconomic analyses in general. Part of the problem is related to the fact that products are differentiated and the number of variants may be very large. Also many of the variants that appear in the market today represent dramatic improvements over their counterparts a few years back. Moreover, since products are sold in retail stores with different locations with different characteristics and costs, and consumers have preferences over stores, product prices may vary across stores for a given variant. For example, Carlson and Pescatrice (1980) have found that prices of "identical products" tend to be dispersed.[1] The traditional way of accounting for differentiated products in demand analyses is either simply to increase the number of (observable) variant categories or to apply Hicks aggregation. Although many variants can in principle be classified in observable categories, there will, in practice, be a limit to how many variants one can treat as separate goods in a demand system. To aggregate goods into composite ones is also problematic. If consumers have heterogeneous preferences the corresponding price indexes will be individual specific and can therefore not readily be implemented. As a result, it becomes a forbiddening task to estimate for example empirical demand systems and price indexes without some sort of aggregation of commodity variants.

The traditional approach to the construction of price indexes is to apply some sort of Laspeyres or fixed quantity index, which can be used to obtain a first order approximation to a Cost-of-Living index. Most Consumer Price Indexes (CPI) are based on the Laspeyres formulae. The point of departure for computing the CPI is a classification of items into successively higher item group levels. At the lowest level (elementary level) price observations are collected. According to the Laspeyres formulae, the CPI can be constructed on the basis of price (indexes) from any item group level using the corresponding group budget shares as weights. However, for items at the elementary level (the first level at which price observations are combined) information on budget shares is not readily available. Hence, at the elementary level, most countries rely on less relevant weighting information or simply unweighted measures. This has led to the search for an appropriate alternative that can be justified by theoretical arguments. One strand of theory (see Eichhorn and Voeller (1983) for a survey), which dates back to Fisher (1922), advocates the view that the price indexes should pass certain tests (or axioms) such as monotonicity, proportionality, etc. From a theoretical point of view it is, however, not always clear to

---

[1] Carlson and Pescatrice obtained prices of 34 identical products from different stores in New Orleans. Thus, it is possible that some of this price dispersion may be explained by the locations of the stores.

what extent indexes that pass these tests are consistent with consumer theory. Thus, to the extent that a Cost-of-Living index should serve as a basis for the CPI (for which there are strong arguments that it should), it would be desirable to derive indexes for elementary aggregates that are consistent with behavioral assumptions. In this paper we shall discuss the choice of index for elementary aggregates based on exact aggregation that follows from particular assumptions about consumer preferences. In the context of demand analysis the setting we draw on here has been discussed in Dagsvik (1996), Dagsvik et al. (1998), and Brubakk and Dagsvik (1998). However, since these papers were primarily concerned with demand analysis we shall in the present paper discuss the relevance to index construction, and in particular the construction of indexes for elementary aggregates.

Other authors that have discussed the problem of price and commodity aggregation include for example Anderson (1979) and Feenstra (1995). Feenstra assumes a finite number of variants within a single commodity group and an “outside” numeraire commodity. Furthermore, he discusses the so-called hedonic index problem which arises when nonpecuniary attributes associated with the variants are observed. In contrast, the framework developed in this paper is designed to deal with random sets of feasible product variants that are unobservable by the analyst. This randomness can be interpreted as stemming from variations in for example the set of feasible stores across consumers. Alternatively, one may attribute random choice sets to consumers being boundedly rational in the sense that they only take into account a subset of alternatives within their respective "objective" choice sets in the decision making process. However, we do not consider the issue of hedonic regression nor do we explicitly discuss how the distribution of prices are determined in market equilibrium. Feenstra shows that the inddex derived from a discrete choice type of micro model also can be interpreted within a representative consumer setting. Similarly to Feenstra we also demonstrate that the index referred to above is consistent with a representative approach.

While the present approach, and those of Anderson (1979) and Feenstra (1995) assume that the consumers are perfectly informed about the distribution of variants and their prices, a few authors have assumed that the consumers are not fully informed about the price distribution they face, and consequently they search to obtain acceptable prices. These authors include Baye (1985), Anglin and Baye (1987), and Reinsdorf (1994).

The paper is organized as follows: In Section 2 we present the modelling framework, and in Section 3 we discuss Cost-of-Living indexes. In Section 4 we discuss estimation and computational issues. In Section 4 we demonstrate that the results of Section 3 can also be obtained from a representative consumer analogue to the model introduced in Section 2.

2. The model

The commodity space is supposed to consist of n different types of products (goods), where each product consists of a set of different variants/locations characterized by price and quality attributes. The n goods refer to the observed commodity categories while the product variants refer to the items in the lowest level of grouping, i.e. the individual retail stores and unobservable variants. 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 quality/location attribute associated with variant z. For example, let the commodity type be bread, available in two stores as the variants wheat bread and rye bread. Let $z = 1$ represent store A and wheat bread, $z = 2$ store A and rye bread, $z = 3$ store B and wheat bread, and finally $z = 4$ store B and rye bread. These are all possible combinations of locations and variants in the example. The T^*-attributes are consumer specific in the sense that they are subjectively perceived. The setup above is similar to the approach of Lancaster (1966), where the T^*-attributes represent the characteristics dimension.

Next we state the assumption about the distribution of consumers preferences and the quality attributes. Evidently, we can represent the vector of product variants and their attributes as the Cartesian product

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

The consumer is assumed to be perfectly informed about the distribution of product locations, variants and prices. He is assumed to have preferences over product variants and associated quantities, represented by a utility function U($\mathbf{Q}$, $\mathbf{T}^*$).

Assumption 1

The utility function U($\mathbf{Q}$, $\mathbf{T}^*$) has the structure

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

where

$$S_j(z) = Q_j(z) T_j^*(z),$$

and u is a mapping $u: R_+ \to R_+$, *that is increasing and quasiconcave.*

Assumption (2.1) 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. This setup is therefore a version of the "Ideal Variety Approach", proposed by Lancaster (1979). The realism of (2.1) depends of course on how detailed the observable commodity types are defined. It also depend on the time unit because the consumer specific attributes $\{T_j^*(z)\}$ 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.[2]

The budget constraint is given by

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

where y is income.

Let

(2.3) $$R_j(z) = P_j(z) / T_j^*(z).$$

The consumers optimization problem is equivalent to maximizing the utility function (2.1) with respect to $\{S_j(z), z = 1,2,\ldots, j = 1,2,\ldots, n\}$ subject to the "budget" constraint

(2.4) $$\sum_{j=1}^{n} \sum_{z} S_j(z) R_j(z) \leq y.$$

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 $\{R_j(z)\}$ 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

(2.5) $$R_j(\hat{z}_j) = \min_z R_j(z),$$

[2] The price observations in the Consumer Price Index in Norway are supposed to be valid for one day (the 15th.) each month.

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 $y_j(\mathbf{r}, y),\ j = 1,2,\ldots,m,$ be the function that yields expenditure on good of type j that follows from maximizing $u(s_1, s_2, \ldots, s_m)$ subject to $\sum_{j=1}^{n} r_j s_j \le y$, where $\mathbf{r} = (r_1, r_2, \ldots, r_m)$. We realize immediately that the purchased quantity of good j, $\hat{Q}_j$, is given by

(2.6)
$$\hat{Q}_j = \frac{\hat{S}_j \hat{R}_j}{\hat{P}_j} = \frac{y_j(\hat{\mathbf{R}}, y)}{\hat{P}_j}$$

where $\hat{\mathbf{R}} = (\hat{R}_1, \hat{R}_2, \ldots, \hat{R}_n)$. Thus, we have expressed the expenditures that correspond to the chosen quantities by means of an ordinary and deterministic Marshallian demand system where $\hat{\mathbf{R}}$ represents the vector of prices. 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 prices, $\{\hat{P}_j\}$, are of course endogeneous because they are associated with the respective chosen product variants/locations, while prices are exogenous to the consumers. Note that the virtual prices are not observable. They can be interpreted as 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.

Similarly, it follows that the Hicksian (compensated) demands can be expressed as

(2.7)
$$\hat{Q}_j = \frac{c_j(\hat{\mathbf{R}}, u)}{\hat{P}_j}$$

where $c_j(\mathbf{r}, u),\ j = 1,2,\ldots,m,$ is the function that yields minimum expenditure on good j given that the utility level is equal to u. From (2.7) it follows that the expenditure function $c(\cdot)$ equals

(2.8)
$$c(\hat{\mathbf{R}}, u) = \sum_{j=1}^{n} c_j(\hat{R}, u).$$

From (2.8) we realize that the expenditure function has the crucial property that it depends on prices solely through virtual prices. This property is useful in the context of price indexes and cost of living indexes.

To obtain aggregate relations that apply to empirical settings, it is necessary to make further assumptions. Without loss of generality we can write $T_j^*(z) = T_j(z)\xi_j(z)$, where $T_j(z)$ represents the mean attribute value of variant z of type j in the population, and $\xi_j(z)$ are taste-shifters that represent the heterogeneity in consumers tastes. According to 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 in the sequel call $T_j(z)$ the quality attribute associated with variant z.

Assumption 2

The taste-shifters $\{\xi_j(z)\}$ *are assumed to be i.i.d. random variables with*

$$P\left(\xi_j(z) \le y\right) = exp\left(-y^{-\alpha_j}\right) \tag{2.9}$$

for $y > 0$, *where* $\alpha_j > 0$ *is a constant.*

A useful interpretation of α_j is as

$$\alpha_j^2 = \frac{\pi^2}{6\,\mathrm{Var}\left(\log \xi_j(z)\right)}. \tag{2.10}$$

A possible justification for (2.9) is that it is consistent with the notion of "Independence from Irrelevant Alternatives", which is discussed in Dagsvik et al. (1998).

Let $g_j(p,t)$ be the density of prices and quality attributes of the variants in the market within commodity group j, and let D_j be the support of $g_j(\cdot)$. Let $\$_j(p,t)$ be the probability that a consumer shall purchase a variant with price and quality (p,t) given that a variant of type j is purchased. The empirical counterpart to $g_j(p,t)$ is the fraction of variants of type j with list price p and quality attribute t that appear in the stores. The empirical counterpart to $\$_j(p,t)$ is the number of consumers that purchase a variant with price and quality (p,t) to the number of consumers that purchase a variant of type j. From the assumptions above it follows readily from the theory of discrete choice that the relationship between $\hat{g}_j(p,t)$ and $g_j(p,t)$ is given by

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

The interpretation of (2.11) is as follows: Due to the consumer's 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. Eq. (2.11) expresses the structure of this selection effect. Note that according to (2.10) the selection effect decreases when the variance of $\log \xi_j(z)$ increases, and disappears when the variance approaches infinity, which means that the distribution of unit values and market values coincide in the limit.

It follows from (2.9) that the distribution of $\hat{R}_j$ has the structure

$$(2.12) \qquad P\left(\hat{R}_j \le r\right) = 1 - \exp\left(-r^{\alpha_j} K_j\right)$$

for $r \ge 0$, where

$$(2.13) \qquad K_j = b_j \sum_{(x,y)\in D_j}\left(\frac{y}{x}\right)^{\alpha_j} g_j(x,y),$$

and where b_j represents the number of variants of type j that is supplied to the market.

In empirical analyses, (2.11) is not readily applicable due to the fact that the quality attributes are unobservable. We shall therefore derive aggregate relations that correspond to observations of prices and unit values. To this end let

$$(2.14) \qquad \lambda_j(p) \equiv E\left(T_j(z)^{\alpha_j} \,\middle|\, P_j(z) = p\right) \equiv \frac{\sum_t t^{\alpha_j} g_j(p,t)}{g_j(p)}$$

where $g_j(p) \equiv \sum_t g_j(p,t)$. The interpretation of $\lambda_j\left(p^{1/\alpha_j}\right)$ is as the conditional mean of $T_j(z)^{\alpha_j}$ across variants of type j, given $P_j(z)^{\alpha_j} = p$. Thus, $\lambda_j(p)$ represents the mean level of perceived quality across variants with price level p. (See Stiglitz (1987) for a discussion on the relationship between price and quality.) It follows immediately from (2.11) and (2.14) that the relationship between the marginal densities of prices, $g_j(p)$, and unit values, $\tilde{g}_j(p)$, is given by

(2.15)
$$\tilde{g}_j(p) \equiv \sum_t \tilde{g}_j(p,t) = \frac{p^{-\alpha_j}\lambda_j(p)\,g_j(p)}{\sum_{x>0} x^{-\alpha_j}\lambda_j(x)\,g_j(x)\,dx}.$$

We realize that if $\lambda_j\left(p^{1/\alpha_j}\right) = w_j\,p$, where $w_j > 0$ is a constant, then the distribution of unit values will coincide with the price distribution.

The setting considered above may seem somewhat unsatisfactory for several reasons. First, there appears to be a rather large variety in product quality, 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—seems to be nearly continuous. Also the sets of feasible variants may vary across consumers, due for example to spatial variations in the location of stores. Finally, it is desirable where $\hat{\mathbf{R}}_0$ denotes the vector of virtual prices in the basis period. The index given in (3.1) is household specific and random due to the fact that the virtual prices are household specific. The corresponding aggregate index $E\,I\left(\hat{\mathbf{R}}_0, \hat{\mathbf{R}}_t, u\right)$ can, by first order Taylor approximation, be expressed as

$$E\,I\left(\hat{\mathbf{R}}_0, \hat{\mathbf{R}}_t, u\right) \approx \frac{c\left(E\,\hat{\mathbf{R}}_t, u\right)}{c\left(E\,\hat{\mathbf{R}}_0, u\right)} \equiv I\left(E\,\hat{\mathbf{R}}_0, E\,\hat{\mathbf{R}}_t, u\right).$$

Thus, to a first approximation we can interpret $I\left(E\,\hat{\mathbf{R}}_0, E\,\hat{\mathbf{R}}_t, u\right)$ as an (aggregate) Cost-of-Living index. As a result, the corresponding Laspeyres and Paasche indexes follow from the usual expression by substituting the prices by the respective mean virtual prices.

Let us next discuss the issue of commodity group-specific indexes (elementary indexes). To this end we consider the cost function conditional on group j. Due to assumption (2.1) it follows that

$$\sum_z S_j(z)$$

is equivalent to a utility function when only consumption allocation within group j is considered. This implies that the "subutility"

$$u_j \equiv \hat{S}_j \tag{3.2}$$

represents the "indirect" utility, due to the fact that only one variant within group j is purchased. Therefore, the group-specific cost function at time t equals

$$c_j\left(\hat{R}_{jt}, u_j\right) = u_j\,\hat{R}_{jt} \tag{3.3}$$

and consequently, the group-specific cost of living index equals

$$I_j^*\left(\hat{R}_{j0}, \hat{R}_{jt}, u_j\right) \equiv \frac{c_j\left(\hat{R}_{jt}, u_j\right)}{c_j\left(\hat{R}_{j0}, u_j\right)} = \frac{\hat{R}_{jt}}{\hat{R}_{j0}}. \tag{3.4}$$

This index is reference free, i.e., it does not depend on the level of the subutility. As in the unconditional case, the index given in (3.4) is household specific and random. We shall therefore be interested in the corresponding aggregate index

(3.5) $$E\, I_j^*\left(\hat{R}_{j0}, \hat{R}_{jt}, u_j\right) = E\left(\frac{\hat{R}_{jt}}{\hat{R}_{j0}}\right).$$

In contrast to the treatment in the general unconditional case it is in fact possible to calculate an exact formulae for the right hand side of (3.5) due to a bivariate extension of (2.12), provided the correlation between $\hat{R}_{jt}$ and $\hat{R}_{j0}$ is sufficiently strong. Specifically, we prove in the Appendix B that

(3.6) $$E\left(\frac{\hat{R}_{jt}}{\hat{R}_{j0}}\right) = \frac{E\,\hat{R}_{jt}}{E\,\hat{R}_{j0}} \cdot \frac{\rho_j \pi}{\alpha_j \sin\left(\frac{\rho_j \pi}{\alpha_j}\right)}$$

where ρ_j is a constant such that $\rho_j < \alpha_j$, with the interpretation as

(3.7) $$1 - \rho_j^2 = \mathrm{Corr}\left(\log \hat{R}_{jt}, \log \hat{R}_{j0}\right).$$

We have therefore demonstrated that the index I_j defined by

(3.8) $$I_j = \frac{E\,\hat{R}_{jt}}{E\,\hat{R}_{j0}}$$

can (apart from a multiplicative constant) be interpreted as an exact aggregate cost of living index for commodity group j.[3]

4. Calculation of mean virtual prices

We shall in this section consider the problem of calculating (estimating) the virtual prices. We shall demonstrate that the assumptions introduced in Section 2 imply rather convenient expressions for $E\,\hat{R}_j$. For simplicity we drop the indexation of time in the notation. From (2.12) and (2.14) it follows, with the normalization,

(4.1) $$b_j^{1/\alpha_j} = \Gamma\left(1 + \frac{1}{\alpha_j}\right)$$

---

[3] It is easy to verify that the result (3.6) does not essentially depend on the Weibull distribution (2.12). It is in fact sufficient that the virtual prices have the structure $\hat{R}_{jt} = E\,\hat{R}_{jt} \cdot \eta_{jt}$, where $\{\eta_{jt}\}$ are positive random variable with distributions that do not depend on the mean virtual prices, and with the property that $E\left(\eta_{jt}/\eta_{j0}\right)$ is finite for all j.

that one gets

$$(4.2)\qquad E\,\hat{R}_j = K_j^{-1/\alpha_j} = \left(\sum_p p^{-\alpha_j}\,\lambda_j(p)\,g_j(p)\right)^{-1/\alpha_j} \equiv \left(E\,P_j(z)^{-\alpha_j}\,\lambda_j\left(P_j(z)\right)\right)^{-1/\alpha_j}.$$

One can also express the mean virtual price $E\,\hat{R}_j$ by means of the distribution of unit values. Specifically, it can be demonstrated that

$$(4.3)\qquad E\,\hat{R}_j = \left(\frac{\sum_p p^{\alpha_j}\,\hat{g}_j(p)}{\sum_t t^{\alpha_j}\,\tilde{g}_j(t)}\right)^{1/\alpha_j} \equiv \left(E\,T_j(z)^{\alpha_j}\right)^{-1/\alpha_j}\left(E\,\hat{P}_j^{\alpha_j}\right)^{1/\alpha_j}$$

where $\tilde{g}_j(t) \equiv \sum_p g_j(p,t)$, is the marginal density of $T_j(z)$ across variants (see Dagsvik et al. (1998)).

We shall next introduce an additional assumption which implies a useful restriction on the functional form of $\lambda_j(\cdot)$.

Assumption 3

The conditional distribution of unit values within each commodity group given that a variant is purchased, is not affected by a scale transform of the prices of the variants.

Assumption 3 seems reasonable since only changes in relative prices matter due to the fact that $\hat{g}_j(\cdot)$ is independent of income.

In Appendix A we demonstrate that Assumption 3 implies that $\lambda_j(\cdot)$ is a power function. Thus, under Assumption 3

$$(4.4)\qquad \lambda_j\left(p^{1/\alpha_j}\right) \equiv E\left(T_j(z)^{\alpha_j}\,\middle|\,P_j(z)^{\alpha_j} = p\right) = A_j\,p^{\kappa_j}$$

where $A_j > 0$ and $\kappa_j > 0$ are constants. From (4.4) we obtain that A_j has the interpretation

$$(4.5)\qquad A_j = \frac{E\left(T_j(z)^{\alpha_j}\right)}{E\left(P_j(z)^{\alpha_j\kappa_j}\right)}.$$

From (4.4) we realize that $\lambda_j\left(p^{1/\alpha_j}\right)$ 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 perceived attractiveness of the product variants as much as when $\kappa_j < 1$, because high prices are perceived as an indication of high quality, and vice versa. When $\kappa_j > 1$, for example, the relationship between prices and quality is strengthened as the price level increases.

From (4.2) and (4.4) it follows that we can express $E\hat{R}_j$ as

(4.6) $$E\hat{R}_j = \left(E\,T_j(z)^{\alpha_j}\right)^{-1/\alpha_j}\left(\frac{E\,P_j(z)^{\alpha_j\kappa_j}}{E\,P_j(z)^{\alpha_j\kappa_j-\alpha_j}}\right)^{1/\alpha_j}$$

or, alternatively, $E\hat{R}_j$ can be expressed by (4.3).

To gain some intuition about the properties of the index formulae (4.6), we will discuss a few particular cases below. If we are willing to assume that $E\,T_j(z)^{\alpha_j}$ is constant through time we can without loss of generality normalize such that $E\,T_j(z)^{\alpha_j} = 1$. Consider first the case with $\kappa_j \approx 0$. In this case expression (4.6) reduces to the generalized harmonic mean

$$\left(E\,P_j(z)^{-\alpha_j}\right)^{-1/\alpha_j}.$$

Note that this expression is little affected by the right tail of price distribution. This means that since quality in this case is not correlated with price, high prices will have a small effect on the price index simply because consumers will not buy from stores with high prices (or variants with high prices). In the "reference case" with $\kappa_j = 1$, the index above reduces to the generalized mean

$$\left(E\,P_j(z)^{\alpha_j}\right)^{1/\alpha_j}.$$

This reference case means that relative changes in prices yield the same relative changes in mean perceived quality. In this case we realize that high prices will be much more important that in the previous case, unless α_j is very small. Recall that a small α_j means large heterogeneity in tastes, and consequently the effect of the price dispersion will be reduced. This conforms with the intuition that since consumers value the product variants differently, the influence on demand of a specific price distribution will to some extent vary across consumers in an unpredictable manner. Consider finally the case when $\kappa_j = 2$. Then the index above has the form

$$\left(\frac{E\, P_j(z)^{2\alpha_j}}{E\, P_j(z)^{\alpha_j}} \right)^{1/\alpha_j}.$$

Since $\left(E\, P_j(z)^{2\alpha_j} \right)^{1/\alpha_j}$ is a factor in the formulae above the effect of the right tail of the price distribution will be larger than the previous cases. This is intuitively plausible since in this case high prices are perceived as signals of high quality. Note finally that expression (4.3) is convenient for the

(4.8) $$E\hat{R}_j \cong \left(\frac{1}{N}\sum_{i=1}^{N} \hat{P}_{ij}^{\alpha_j}\right)^{1/\alpha_j}$$

where $\hat{P}_{ij}, i = 1,2,\ldots,N,$ is a random sample of unit values. It is interesting that while it is necessary to know α_j and κ_j to apply (4.7), only α_j is needed to compute the index formulae in (4.8). In practice, however, it will usually not be possible to apply (4.8) due to the fact that the samples used in consumer expenditure surveys are too small.

In Dagsvik (1996), part II, it is discussed how $\alpha_j\kappa_j - \alpha_j$ can be estimated provided one has a sample of unit values and list prices. To estimate α_j, however, one needs to make further assumptions about the structure of the function $y_j(\mathbf{r}, y)$ introduced in Section 2. This issue is discussed in Dagsvik et al. (1998).

Finally, let us consider the case with very large population heterogeneity in tastes, i.e., when $\alpha_j \to 0$, cf. (2.10). By using l'H pital's rule we get from (4.6) that

(4.9) $$\lim_{\alpha_j \to 0} E\hat{R}_j = \exp\left(E \log P_j(z)\right) \cong \left(\prod_{s=1}^{M} P_j(z_s)\right)^{1/M}$$

which we recognize as the geometric mean of the prices. Notice that in this case the parameter κ_j vanishes in the index formulae. From (4.3) we also obtain

(4.10) $$\lim_{\alpha_j \to 0} E\hat{R}_j = \exp\left(E \log \hat{P}_j\right) \cong \left(\prod_{i=1}^{N} \hat{P}_{ij}\right)^{1/N}.$$

In other words, when α_j is close to zero we can estimate the mean virtual prices by a geometric mean of list prices or, alternatively, by the geometric mean of unit values. The geometric mean alternative (4.9) has been recommended by the so called CPI Commision, see Boskin et al. (1997). See also Dalén (1992).

We conclude this section by a discussion on the differences between the generalized mean given by (4.7) when $\kappa_j = 1$, and the geometric mean. From Hölder's inequality (see for example Berck and Sydsæter, 1993) it follows easily that

(4.11) $$\left(\frac{1}{M}\sum_{s=1}^{M} P_j(z_s)^{\alpha_j}\right)^{1/\alpha_j} \geq \left(\prod_{s=1}^{M} P_j(z_s)\right)^{1/M}$$

where equality holds only when all prices are equal. Moreover, the difference between the right and the left hand side will increase as the variance of the logarithm of prices increases, provided the central moments of order higher than two are not too "large". To see this we note that the asymptotic counterpart to (4.11) is

$$\left(E\, P_j(z)^{\alpha_j}\right)^{1/\alpha_j} \geq \exp\left(E \log P_j(z)\right)$$

which is equivalent to

$$\left[E \exp\left(\alpha_j \left(\log P_j(z) - E \log P_j(z)\right)\right)\right]^{1/\alpha_j} \geq 1. \tag{4.12}$$

By a second order Taylor expansion it follows that the left hand side of (4.12) is approximately equal to

$$\left(1 + \frac{\alpha_j^2}{2} \operatorname{Var} \log P_j(z)\right)^{1/\alpha_j}$$

which shows that the left hand side of (4.12) is increasing in $\operatorname{Var} \log P_j(z)$ provided that

$$\frac{1}{n!} E\left(\log P_j(z) - E \log P_j(z)\right)^n$$

is small for $n \geq 3$.

5. A representative consumer analogue

Anderson et al. (1992) and Feenstra (1995), among others, have discussed how discrete choice behavior can be interpreted within a representative consumer setting. In this section we shall demonstrate that the indexes derived above can also be derived from a representative consumer approach.

To this end we assume that the representative consumer has utility function given by

$$U(\mathbf{Q}, \mathbf{T}) = u\left(V_1, V_2, \ldots, V_n\right) \tag{5.1}$$

where

$$V_j = \left(\sum_z \left(T_j(z)\, Q_j(z)\right)^{\frac{\alpha_j}{1+\alpha_j}}\right)^{\frac{1+\alpha_j}{\alpha_j}}. \tag{5.2}$$

The representative consumer's problem is to maximize (5.1) subject to (2.2). This problem can be formulated as a two stage budgeting problem as follows: First, maximize utility with respect to variants within commodity groups subject to expenditure on each commodity type. Second, maximize utility with respect to consumption allocation between commodity groups.

Consider first allocation within commodities. As above, let c_j denote the (conditional) expenditure function for commodity group j. It follows readily that

(5.3) $$Y_j = \left(\sum_z \left(\frac{T_j(z)}{P_j(z)} \right)^{\alpha_j} \right)^{-1/\alpha_j} u_j.$$

The interpretation of (5.3) is as the expenditure Y_j which is required to achieve utility level u_j within group j, given prices and quality attributes. With the same notation as in Section 2 it follows that we can write (5.3) as

(5.4) $$Y_j = \left(b_j \sum_{(x,y)\in D_j} \left(\frac{y}{x} \right)^{\alpha_j} g_j(x,y) \right)^{-1/\alpha_j} u_j = \left(b_j\, E \left(\frac{T_j(z)}{P_j(z)} \right)^{\alpha_j} \right)^{-1/\alpha_j} u_j .$$

From (2.12) we have that (5.4) also can be expressed as

(5.5) $$Y_j = u_j\, K_j^{-1/\alpha_j}.$$

If now K_j^{-1/α_j} is linear homogeneous in prices, it is clear from (5.5) that u_j can be interpreted as composite consumption of type j while K_j^{-1/α_j} is the corresponding "price" (price index). Thus the total expenditure function can therefore be expressed as a function of the "price indexes", K_j^{-1/α_j}, $j = 1,2,\ldots,n$. But from (4.2) we realize that the index formulae for $E\hat{R}_j$ and K_j^{-1/α_j} are the same. In other words, the representative consumer approach presented above yields the same price index as the micro-approach outlined in Section 2.

In the representative consumer setting the interpretation of the parameters $\{\alpha_j\}$ is different from the case with a population of consumers. In that case α_j is associated with the dispersion of the random taste-shifters, cf. (2.10), and is constrained to be positive. In the representative consumer approach $1+\alpha_j$ can be interpreted as the elasticity of substitution between variants within commodity group j, and α_j can take any real value except zero.

6. Empirical results

To gain some insight on the importance of the value of the parameter κ_j in the formulae (4.7) one can compute (4.7) for different values of κ_j and α_j. We have computed estimates by means of (4.7), the arithmetic mean $(\kappa_j = \alpha_j = 1)$, and the geometric mean given by (4.9), for selected commodities based on data from January 1989 to December 1994. Plots of the resulting indexes are displayed in Figures 1 to 19. For the sake of comparison we have also displayed the index currently in use by Statistics Norway. The three indexes shown are: (i) The actual elementary index used by Statistics Norway in the construction of the Consumer Price Index (CPI) (See Koht and Sandberg (1997)), which is based on regional ratios of mean prices, weighted together using appropriate regional weights.[4] This index will be referred to as k_j, where j denotes the commodity group. (ii) The index given by (4.7), with $\kappa_j = 1$ and $\alpha_j = 0.63$ (the chosen value of α is taken from Brubakk and Dagsvik (1998)), is referred to as d_j and, finally, (iii) the geometric mean given by (4.9), is referred to as g_j. In Figures 10 to 19 we display the respective arithmetic and geometric means, where the arithmetic mean for group j is referred to as a_j. All the indexes are normalized to 100 in January 1989.

From Figures 10 through 19 we realize that the geometric and the arithmetic means yield very similar result except for the commodity groups "Bread" and "Fish products". From the discussion in Section 4 we realize that this may be due to the fact that the variance of the logarithm of prices increases for these particular commodity groups. From Figures 1 to 9 we note that the elementary index currently in use in Statistics Norway differs from the other indexes for some goods in some months. Since the difference between this index and the geometric mean evidently is due to the regional weighting, we realize that it may be of some importance how the weights are selected.

[4] Thus, for a given commodity group j, we have that the elementary index for the time period 0 to t can be expressed as

$$k_j = \sum_r w_r \frac{\frac{1}{n_r}\sum_s P_{jt}(z_{sr})}{\frac{1}{n_r}\sum_s P_{j0}(z_{sr})}$$

where $\{w_r\}$ denote regional weights. For each region r, the summation is made across the set of stores, indexed z_{sr}. The number of stores in each region is denoted n_r.

Figure 1. Price indexes for a particular type of bread

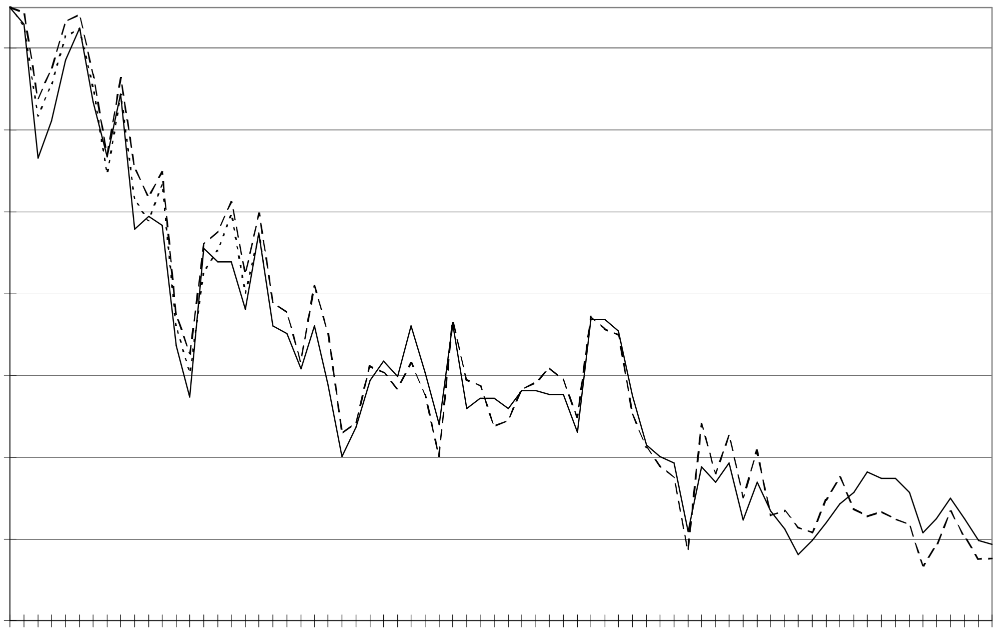


Figure 3. Price indexes for particular fish products

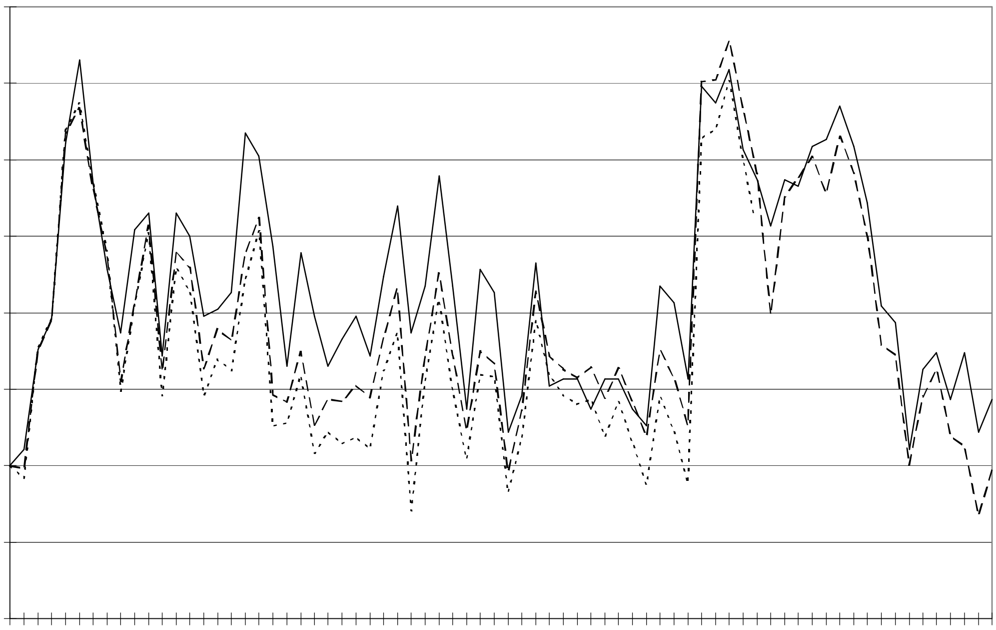


Figure 5. Price indexes for margarine

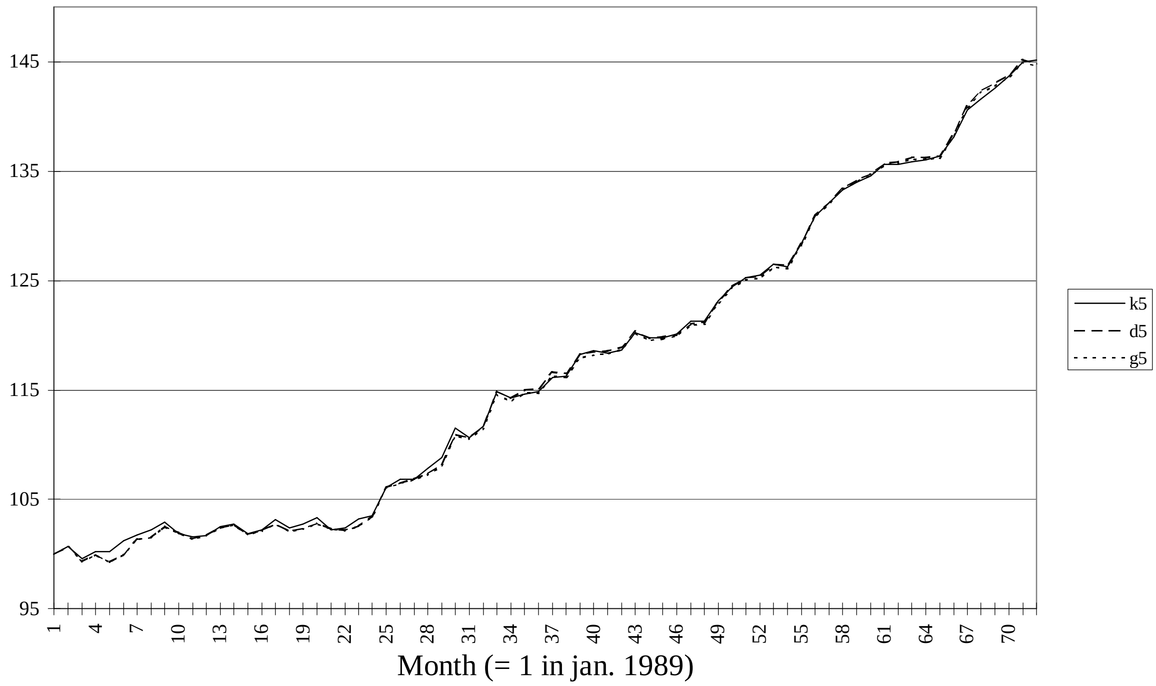


Figure 6. Price indexes for oranges

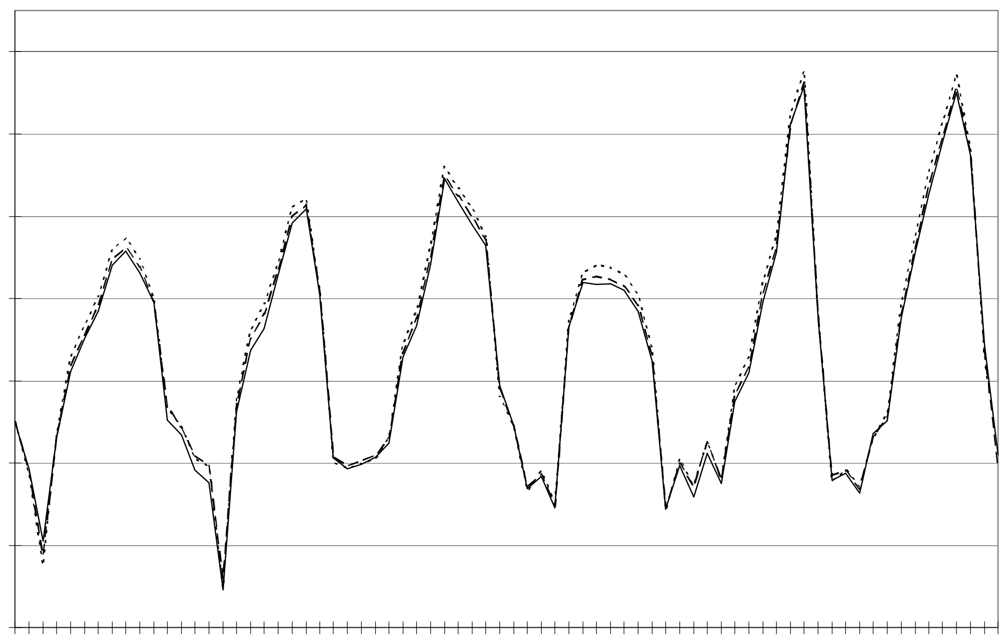


Figure 7. Price indexes for potatos

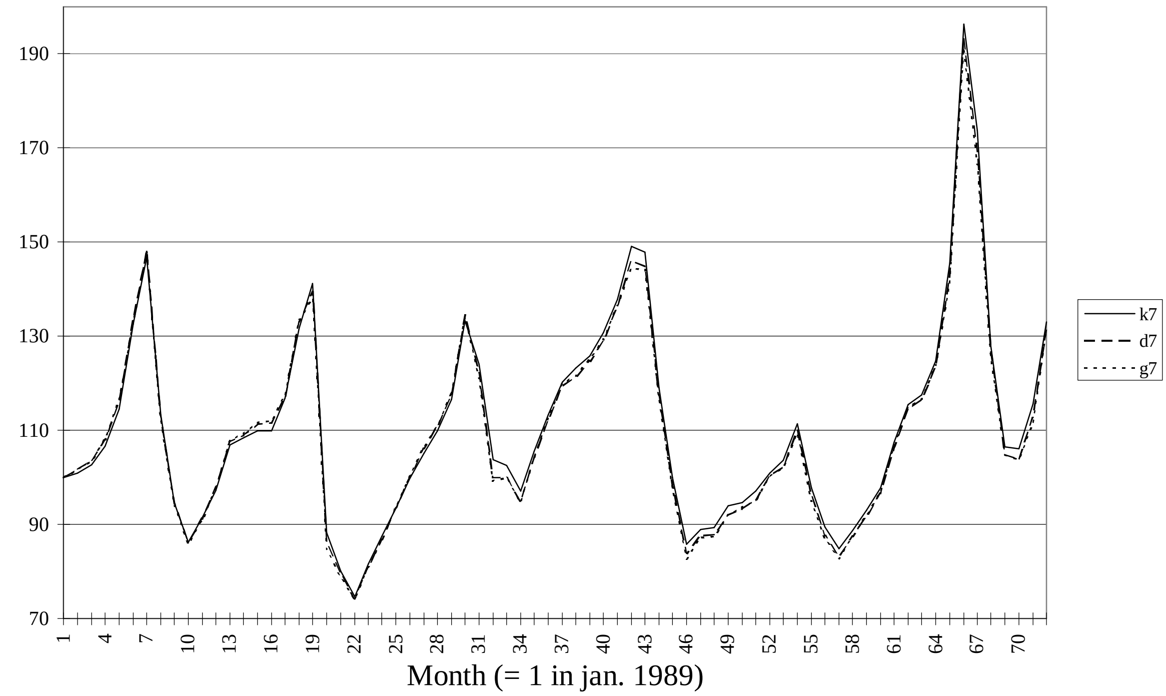


Figure 8. Price indexes for sugar

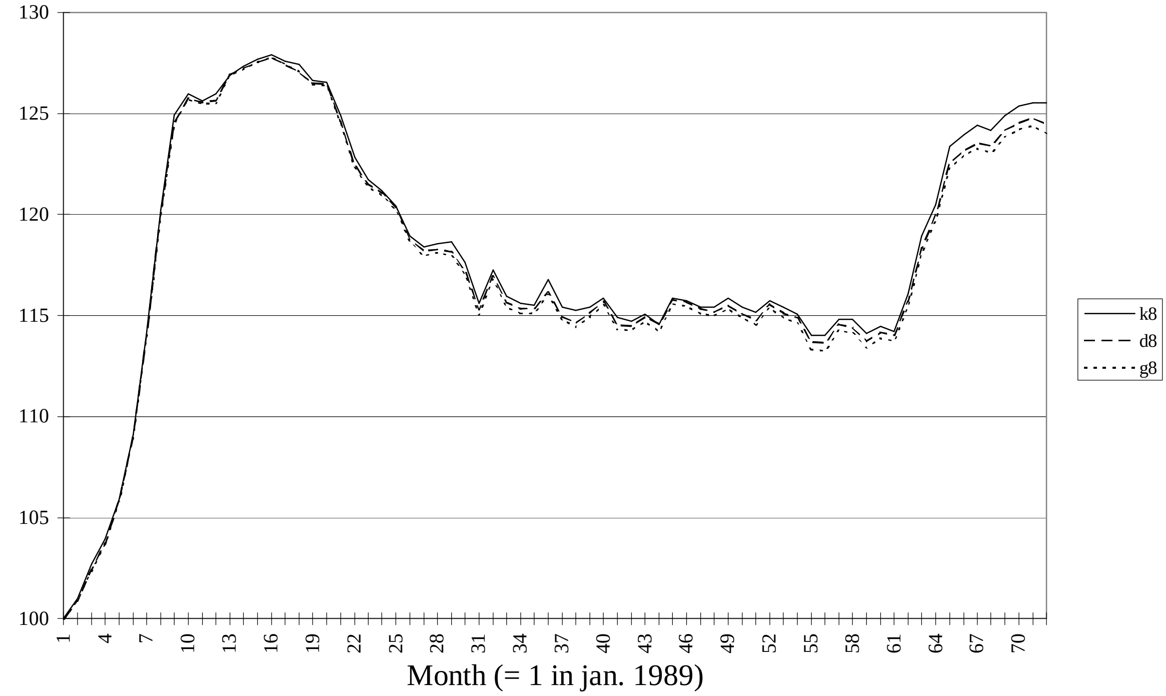


Figure 9. Price indexes for coffee

Figure 11. Price indexes for minced meat

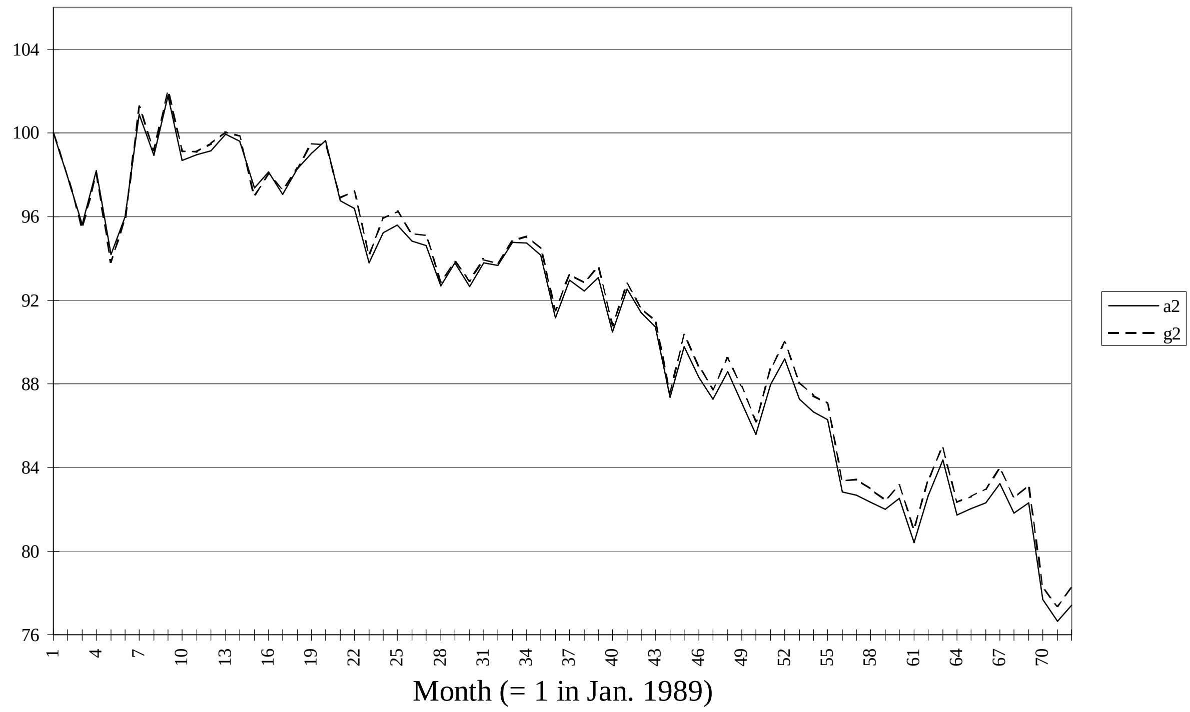


Figure 12. Price indexes for particular fish products

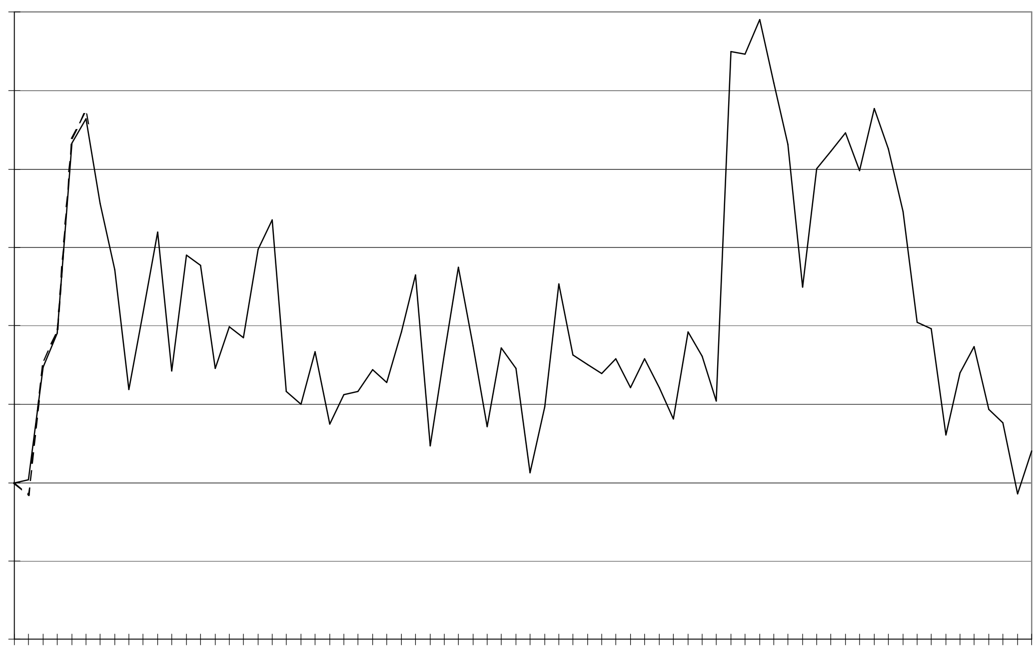


Figure 13. Price indexes for eggs

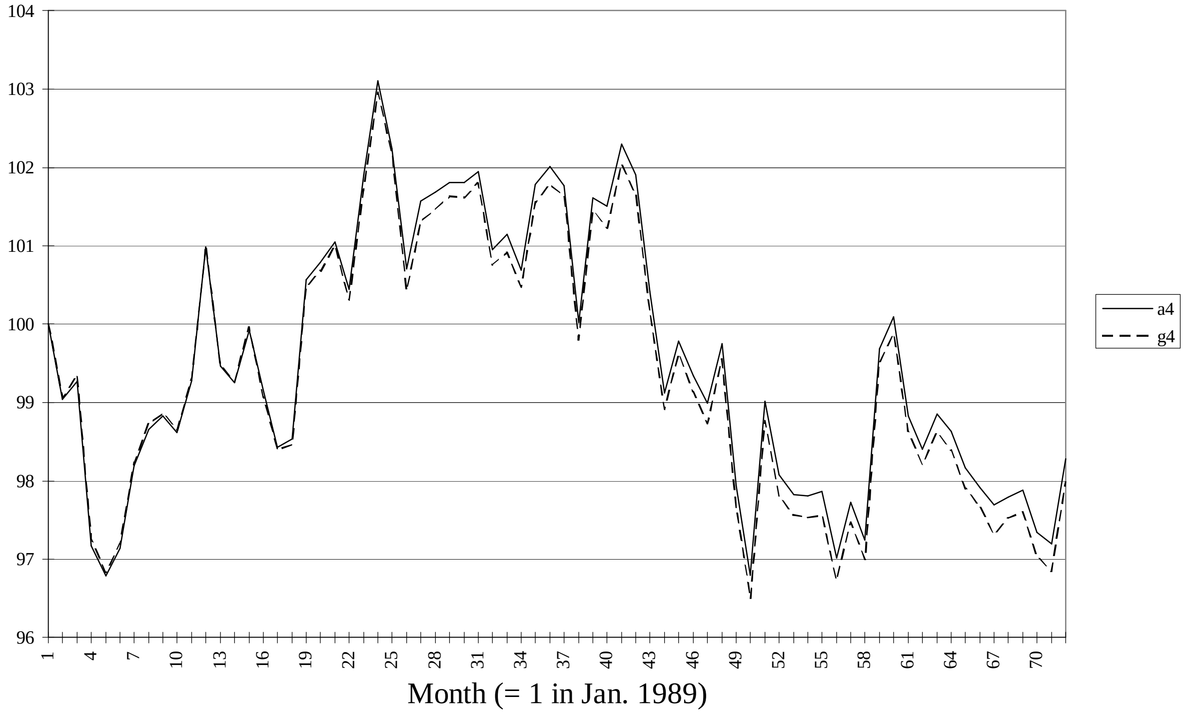


Figure 14. Price indexes for margarine

Figure 17. Price indexes for sugar

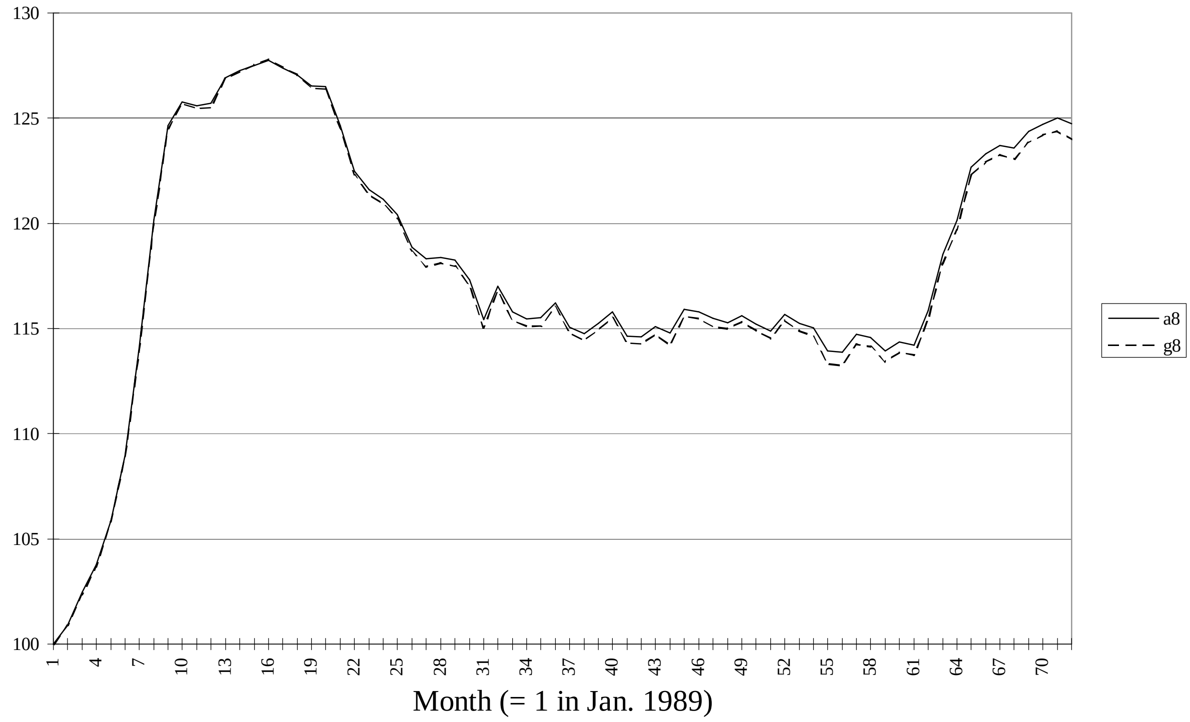


Figure 18. Price indexes for coffee

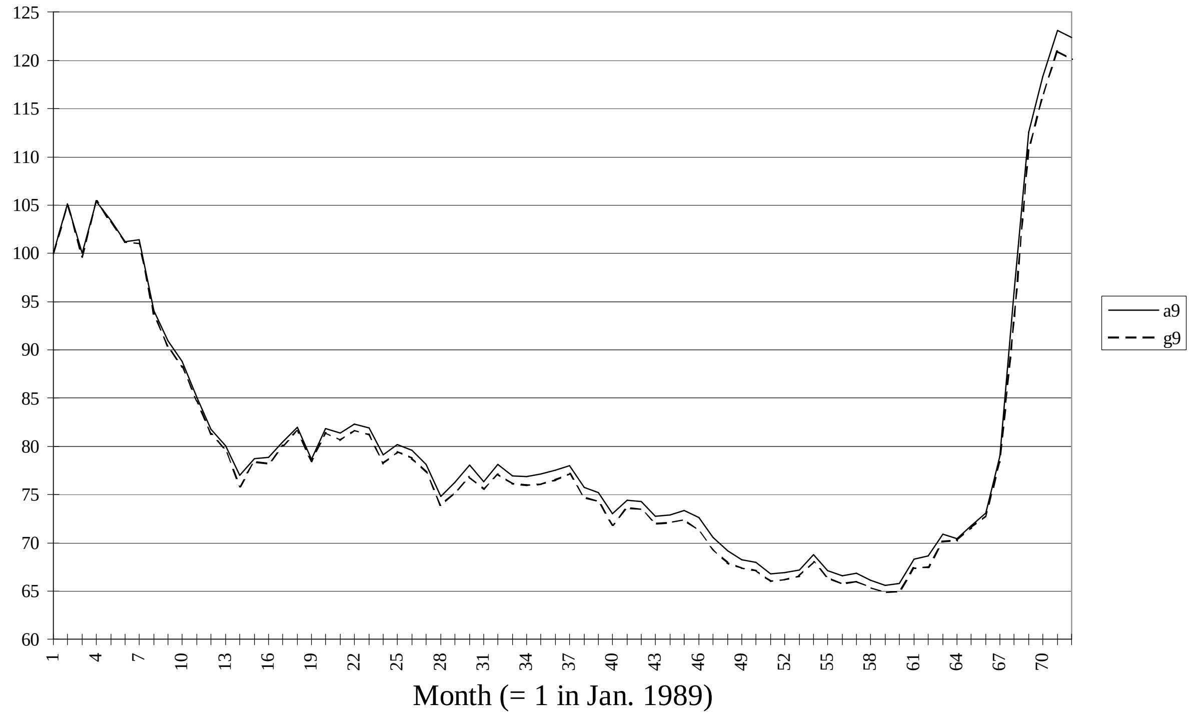
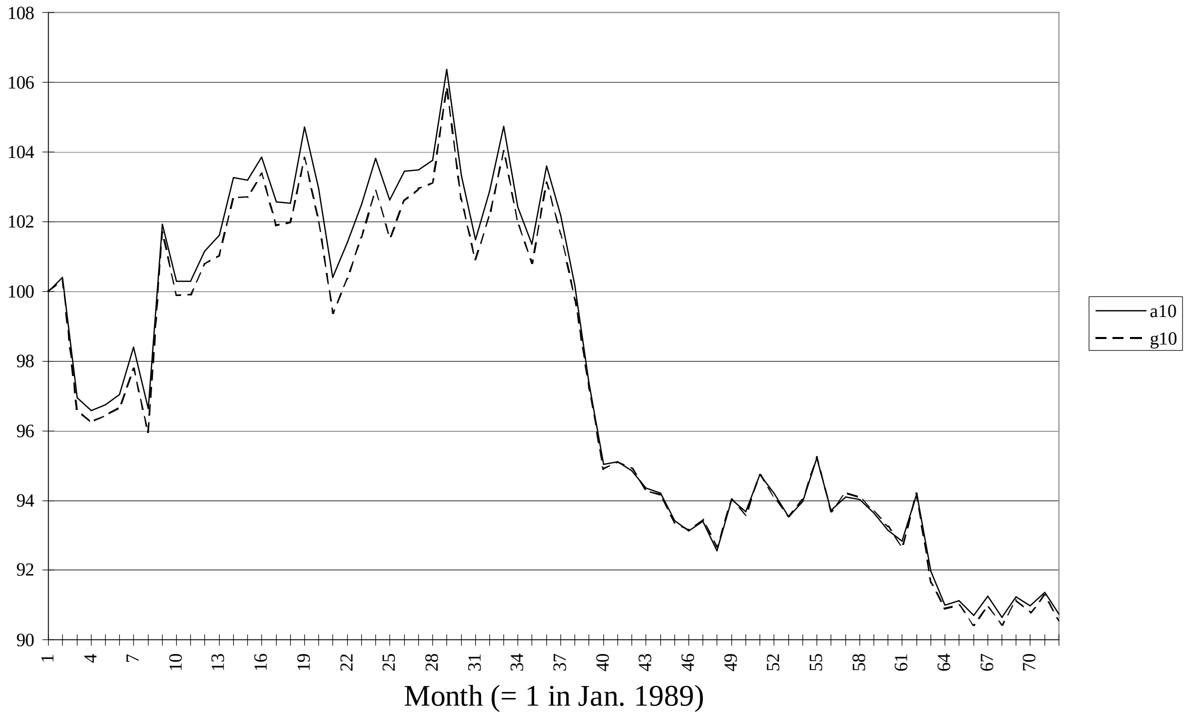


Figure 19. Price indexes for a particular type of bread



6. Conclusions

In this paper we have discussed possible theoretical justifications for a class of price indexes for elementary aggregates. Two different approaches have been discussed. The first one is based on a particular representation of preferences for heterogeneous consumers in which goods are allowed to be differentiated with product variants that are perfect substitutes to the individual consumers. Moreover, this representation allows prices to depend on latent "quality" attributes of the variants. This is of particular interest for product variants where price is perceived by the consumers as a signal of quality. From the stated assumptions, a convenient class of prices indexes for elementary aggregates follows. Second, we demonstrate that the same class of indexes can be derived from a representative consumer approach. Many of the indexes for elementary aggregates proposed in the literature emerge as special cases within this class. An example of a case of particular interest is the geometric mean.

References

Aczél, J. (1966): *Lectures on Functional Equations and Their Applications*. Academic Press, New York.

Anderson, R.W. (1979): Perfect Price Aggregation and Empirical Demand Analysis. *Econometrica*, **47**, 1209-1229.

Anderson, S.P., A. de Palma and J.-F. Thisse (1992): *Discrete Choice Theory of Product Differentiation*. MIT Press, Cambridge, Massachusetts.

Anglin, P.M. and M.R. Baye (1987): Information, Multiple Search, and Cost-of-Living Index Theory. *Journal of Political Economy*, **95**, 1179-1195.

Baye, M.R. (1985): Price Dispersion and Functional Price Indices. *Econometrica*, **53**, 217-223.

Berck, P. and K. Sydsæter (1993): *Economist's Mathematical Manual*. Springer-Verlag, New York.

Boskin, M.J., E.R. Dulberger, R.J. Gordon, Z. Griliches and D.W. Jorgenson (1997): The CPI Commision: Findings and Recommendations. *American Economic Review*, 78-93.

Brubakk, L. and J.K. Dagsvik (1998): Consumer Demand and Unobservable Product Attributes. Discussion Papers 223, Statistics Norway.

Carlson, J. and D. Pescatrice (1980): Persistent Price Distributions. *Journal of Economics and Business*, **33**, 21-27.

Dagsvik, J.K. (1996): Consumer Demand with Unobservable Product Attributes. Part I and Part II. Discussion Papers, 166 and 167, Statistics Norway.

Dagsvik, J.K., Y. Zhu and R. Aaberge (1998): A Framework for Empirical Modelling of Consumer Demand with Latent Quality Attributes. Discussion Papers 229, Statistics Norway.

Dalén, J. (1992): Computing Elementary Aggregates in the Swedish Consumer Price Index. *Journal of Official Statistics*, **8**, 129-147.

Eichhorn, W. and J. Voeller (1983): The Axiomatic Foundations of Price Indexes and Purchasing Power Parities. In *Price Level Measurement: Proceedings from a Conference Sponsored by Statistics Canada.*

Feenstra, R. (1995): Exact Hedonic Price Indexes. *Review of Economics and Statistics*, 634-653.

Fisher, I. (1922): *The Making of Index Numbers*. Houghton-Mifflin, Boston.

Koht, B. and L. Sandberg (1997): Sources of Measurement Error in the Consumer Price Index (In Norwegian). Økonomiske analyser, 22-26, Statistics Norway.

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.

Reinsdorf, M. (1994): The Effect of Price Dispersion on the Cost-of-Living Indexes. *International Economic Review*, **35**, 137-149.

Stiglitz, J.E. (1987): The Causes and Consequences of the Dependence of Quality on Price. *Journal of Economic Literature*, **25**, 1-48.

Proof of the result that Assumption 3 implies that l_j(p) is a power function

For analytic convenience we shall present a proof for the continuous case where $\lambda_j(p)$ is a continuous function. Let $\tilde{g}_{j\theta}(p)$ denote the density of unit values within commodity group j after the prices have been multiplied by a common positive scale θ. The corresponding density of prices equals $g_j(p/\theta)/\theta$. Hence, by (2.15)

(A.1)
$$\tilde{g}_{j\theta}(p) = \frac{p^{-\alpha_j}\lambda_j(p)\,g_j\left(p\theta^{-1}\right)}{\int_0^\infty x^{-\alpha_j}\lambda_j(x)\,g_j\left(x\theta^{-1}\right)dx}.$$

By change of variable in the integral in the denominator of (A.1) we get

(A.2)
$$\tilde{g}_{j\theta}(p\theta) = \frac{p^{-\alpha_j}\lambda_j(\theta p)\,g_j(p)}{\theta\int_0^\infty x^{-\alpha_j}\lambda_j(\theta x)\,g_j(x)\,dx}.$$

Under Assumption 3 it follows readily that

$$\tilde{g}_{j\theta}(p) = \tilde{g}_{j1}\left(p\theta^{-1}\right)\theta^{-1}$$

which implies that for all $p \in (0,\infty)$

(A.3)
$$\frac{\tilde{g}_{j\theta}(p\theta)}{\tilde{g}_{j\theta}(\theta)} = \frac{\tilde{g}_{j1}(p)}{\tilde{g}_{j1}(1)}.$$

When (A.2) and (A.3) are combined we obtain that

(A.4)
$$\frac{\lambda_j(\theta p)}{\lambda_j(\theta)} = \frac{\lambda_j(p)}{\lambda_j(1)}.$$

Let $f_j(p) = \lambda_j(p)/\lambda_j(1)$. Then (A.4) yields

(A.5)
$$f_j(\theta p) = f_j(p)\,f_j(\theta).$$

Eq. (A.5) is a Cauchy type of functional equation which only continuous solution is $f_j(p) = p^{\beta_j}$ (see for example Aczél (1966)). Hence $\lambda_j(p) = a_j p^{\beta_j}$, where $a_j > 0$ and β_j are constants.

Q.E.D.

Derivation of the mean of $\hat{R}_{jt} / \hat{R}_{j0}$

Since $\hat{R}_{jt}$ and $\hat{R}_{j0}$ are Weibull distributed it seems reasonable to assume that $\hat{R}_{jt}$ and $\hat{R}_{j0}$ are bivariate Weibull distributed, i.e., $\hat{R}_{jt} = E\,\hat{R}_{jt} \cdot \eta_{jt}$, where

(B.1) $$F(x,y) \equiv P\left(\eta_{j0} \le x, \eta_{jt} \le y\right) = \exp\left(-k\left(x^{-\alpha_j/\rho_j} + y^{-\alpha_j/\rho_j}\right)^{\rho_j}\right),$$

$x > 0, y > 0, k > 0$ is a constant and $\rho_j \in (0,1]$ is a constant which has the interpretation

(B.2) $$\rho_j^2 = 1 - \operatorname{corr}\left(\log \eta_{jt}, \log \eta_{j0}\right) = 1 - \operatorname{corr}\left(\log \hat{R}_{jt}, \log \hat{R}_{j0}\right)$$

where F_2' denotes the partial derivative with respect to the second variable. From (B.1) it follows by straight forward calculus that

(B.3) $$P\left(\frac{\eta_{jt}}{\eta_{j0}} > v\right) = \int_0^\infty F_2'\left(\frac{y}{v}, y\right) dy = \frac{1}{1 + v^{\alpha_j/\rho_j}}.$$

Accordingly, we obtain

$$E\left(\frac{\eta_{jt}}{\eta_{j0}}\right) = \int_0^\infty P\left(\frac{\eta_{jt}}{\eta_{j0}} > v\right) dv = \int_0^\infty \frac{dv}{1 + v^{\alpha_j/\rho_j}} = \frac{\rho_j \pi}{\alpha_j \sin\left(\frac{\rho_j \pi}{\alpha_j}\right)}$$

provided $\rho_j < \alpha_j$. Hence the proof of (3.6) is complete.

Q.E.D.