ON THE FIRST STEP IN THE CALCULATION OF A CONSUMER PRICE INDEX

Bert M. Balk*)
Statistics Netherlands, Voorburg

1. Introduction

Most official Consumer Price Indices (CPI's) are, at least in the short run, defined as fixed weight indices. Usually for the estimation the weighted-average-of-price-relatives form is used. As a matter of fact the weights are only available down to a certain level of commodity aggregation. Below that level there may be available some rough stratification, but as Szulc (1987) remarks "there is always the first step, when a price index for a certain commodity is directly derived from the sample price data." These indices are called micro-indices. Extensive discussion about the proper method for calculating these micro-indices can be found in Carruthers, Sellwood and Ward (1980), Morgan (1981), Szulc (1987), Turvey (1989, pp. 87-92), Dalén (1992), Sabag and Finkel (1994), Silver (1994) and Diewert (1995).

Typically, the discussion centers around the following issues:

* Should a ratio of average prices or an average of price relatives be used?
* Which type of average (arithmetic, geometric, harmonic, etc.) should be used?

---

*) The views expressed in this paper are those of the author and do not necessarily reflect the policies of Statistics Netherlands. The initial writing was stimulated by discussions with Jörgen Dalén. The first version has been presented at the Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, 18-21 November 1991. The second version was submitted to the Journal of Official Statistics and obtained a number of very instructive comments. Also the discussions in the Eurostat Harmonisation of CPI Task Force III were helpful.

One has compared the various formulas with the help of Bortkiewicz-type relationships, (super)population models, or neo-Fisherian axiomatic (or test) theory, but an unanimous conclusion has not emerged. Also the Fourteenth International Conference of Labour Statisticians 1987 recommendation that "In the calculation of elementary aggregate indices, consideration should be given to the possible use of geometric means." seems to have had no impact. The official practices show considerable differences.

Recently the topic has gained in importance. This has to do with the European project of developing harmonized CPI's which can be used within the framework of a communitarian monetary policy. As first fruit of this project we now have more insight into the numerical consequences of choosing this or that formula for the first step (see Dalén 1994). Also Schultz (1994) provides evidence that the formula question is no minor issue.

Sellwood (1994) considered the reasons why it has proved so difficult to resolve the problem of which micro-index formula to use in compiling a CPI. He suggests "that there is a lack of an agreed point of reference on what is to be measured."

The present paper tries to add to the discussion by carefully specifying the object of estimation. In my view a CPI can best be considered as a group price index, that is a price index which is an average of household-specific price indices. This implies that the form of the household-specific indices determines the form of the CPI. As will be shown, this has certain consequences for the form of the price indices at the lowest level of commodity aggregation. In the majority of cases, however, the available information is not sufficient to calculate these indices. They have to be approximated. Various assumptions and/or sampling considerations come into play and give rise to various micro-index formulas. It seems that the search for a micro-index formula which can be applied universally is misdirected. As a by-product we obtain the conclusion that it is hard to justify formally the use of a geometric mean.

2. The Laspeyres CPI and its estimation

We start with some definitions. Let the scope of the CPI be given by the commodity groups A_k, k=1,...,K. Each commodity group consists of a finite, in principle (although generally not in practice) known, number of commodities; $g \in A_k$ means that commodity g belongs to group A_k. Before proceeding, we have to make clear what we understand by the expression 'commodity'.

One frequently distinguishes between 'homogeneous' and 'heterogeneous' commodities. This distinction however overlooks the fact that the fundamental primitive economic concept is that of the individual *transaction*.[1] No two transactions, even if they are concerned with the same physically specified good or service, are the same. However, for intertemporal comparability some aggregation of these individual transactions is necessary.

It is important to emphasize that these aggregates are not in any way 'naturally' given. They depend on the availability of basic data but also on our knowledge of the distinctions that matter. The set of all transactions involving a specific good or service can be large or small, more or less homogeneous, depending on its description For example, the commodity 'man's haircut' gives a large set of transactions, 'man's haircut at Harry's barber saloon' (where Harry's saloon has a specific location) gives a small set, and 'men's haircut at Harry's barber saloon without waiting time' gives an even smaller set. Whether we use one of these descriptions depends on what we consider to be the economically relevant features of a set of transactions. Ultimately, this depends on whereof, in our perception, the average household derives its utility.

Let B_g denote the set of outlets (or other selling places) in which commodity g can be bought. Each B_g is finite and its members are assumed to be known, at least in principle; $b \in B_g$ means that commodity g can be bought in outlet b. In general a certain outlet belongs to more than one B_g. Some sets B_g contain only one element. Thus in these cases there is only one outlet selling commodity g, e.g. a 'man's haircut at Harry's barber saloon' can only be obtained at Harry's barber saloon. Initially, we assume that A_k (k=1,...,K) and B_g ($g \in A_k$; k=1,...,K) are fixed during time.

Consider firstly an individual household. Its base period expenditures are

1) This point should be familiar from economic theory. It was stressed recently by Triplett (1990) and Winkler (1993).

(1) $\Sigma_k \Sigma_{g \in Ak} \Sigma_{b \in Bg} p^0_{gbh} x^0_{gbh}$,

where x^0_{gbh} is the quantity of commodity g bought at outlet b by household h during period 0 (the base period) and p^0_{gbh} is the corresponding unit value.[2] The set $\{x^0_{gbh}$; all k, all $g \in A_k$, all $b \in B_g\}$ constitutes the base period consumption pattern of household h and is a reflection of household h's base period standard of living.

Notice that most of the quantities x^0_{gbh} are zero, because (i) for most A_k an individual household does not buy all commodities $g \in A_k$ but only a limited number of them, (ii) for these commodities the household does not buy in all possible outlets, but only in a limited number of them, generally lying within the neighborhood of the household's domicile. Although in practice the x^0_{gbh} are not available they could be observed if we wished. We assume that $p^s_{gbh} = p^s_{gb}$ for all h and all periods s considered. Thus the unit values are the same for all households.

The Laspeyres price index for household h is then

$$(2)\ P^t_h \equiv \frac{\Sigma_k \Sigma_{g \in Ak} \Sigma_{b \in Bg} p^t_{gb} x^0_{gbh}}{\Sigma_k \Sigma_{g \in Ak} \Sigma_{b \in Bg} p^0_{gb} x^0_{gbh}} .$$

This is the ratio of the base period consumption pattern valued at comparison period t prices and at base period 0 prices, and is an approximation of the cost-of-living index with the base period standard of living as reference.[3]

The price index for the group of households H is defined as the so-called plutocratic average of the individual household indexes P^t_h, i.e. the weighted average where the base period expenditures (1) serve as weights. Defining $x^0_{gb} \equiv \Sigma_{h \in H} x^0_{gbh}$, we obtain

2) As Diewert (1995) demonstrates, this is the appropriate concept to be used at the lowest level of aggregation. See also Balk (1995).

3) The Laspeyres index is chosen because of its linear aggregation properties.

$$(3)\ P^t \equiv \frac{\Sigma_k \Sigma_{g\in Ak} \Sigma_{b\in Bg} p_{gb}^t x_{gb}^0}{\Sigma_k \Sigma_{g\in Ak} \Sigma_{b\in Bg} p_{gb}^0 x_{gb}^0} .$$

This can also be interpreted as the Laspeyres price index for the average household of the group H. Expression (3) is the *object of the estimation.*

The group price index (3) can be written as a weighted average of commodity group price indices,

$$(4)\ P^t = \sum_k w_k^0 P_k^t \ ,$$

where

$$(5)\ P_k^t \equiv \Sigma_{g\in Ak} \Sigma_{b\in Bg} p_{gb}^t x_{gb}^0 / \Sigma_{g\in Ak} \Sigma_{b\in Bg} p_{gb}^0 x_{gb}^0 \quad (k=1,\ldots,K),$$

is a commodity group price index and

$$(6)\ w_k^0 \equiv \Sigma_{g\in Ak} \Sigma_{b\in Bg} p_{gb}^0 x_{gb}^0 / \Sigma_k \Sigma_{g\in Ak} \Sigma_{b\in Bg} p_{gb}^0 x_{gb}^0 \quad (k=1,\ldots,K)$$

is the aggregate (over all households belonging to H) base period expenditure share of commodity group A_k. We assume that the weights w_k^0 $(k=1,\ldots,K)$ can be estimated from a household expenditure survey, and that the commodity groups $A_1,\ldots,A_K$ constitute the lowest level for which the weights can be found. The remaining problem now is the estimation of the commodity group price indexes P_k^t.

For most commodity groups this involves sampling, both of specific commodities and of outlets. Several sampling designs are possible. We consider a two-stage procedure: in the first stage a fixed sample (panel) of commodities is taken (the set of so-called representative commodities) and in the second stage for each of the sampled commodities a sample of outlets is taken. Let $\hat{A}_k$ denote the sample[4] drawn from A_k $(k=1,\ldots,K)$. The first stage yields

$$(7)\ \underline{P}_k^t \equiv \Sigma_{g\in \hat{A}k} \Sigma_{b\in Bg} p_{gb}^t x_{gb}^0 / \Sigma_{g\in \hat{A}k} \Sigma_{b\in Bg} p_{gb}^0 x_{gb}^0$$

4) Ideally this should be a simple random sample. In practice it is usually a judicious sample.

as the estimator of P_k^t. Expression (7) can again be written as a weighted average

$$(8)\quad \underline{P}_k^t = \sum_{g\in \hat{A}k} \underline{w}_g^0 P_g^t,$$

where

$$(9)\quad P_g^t \equiv \Sigma_{b\in B g} p_{gb}^t x_{gb}^0 / \Sigma_{b\in B g} p_{gb}^0 x_{gb}^0$$

for $g\in\hat{A}_k$ $(k=1,\ldots,K)$ is a commodity price index, and

$$(10)\quad \underline{w}_g^0 \equiv \Sigma_{b\in B g} p_{gb}^0 x_{gb}^0 / \Sigma_{g\in \hat{A}k} \Sigma_{b\in B g} p_{gb}^0 x_{gb}^0$$

for $g\in\hat{A}_k$. Expression (10) gives the base period expenditure share of the sampled commodity g with respect to the entire sample. Usually for the weights $\underline{w}_g^0$ $(g\in\hat{A}_k)$ rough estimates are available, for instance from turnover statistics or market information. In the absence of any information one frequently replaces the $\underline{w}_g^0$ by $1/n(\hat{A}_k)$, where $n(\hat{A}_k)$ is the size of the sample $\hat{A}_k$.

In the second stage we must estimate the commodity price indices P_g^t, which are again Laspeyres price indices. Recall that x_{gb}^0 denotes the total quantity of commodity g sold by outlet b to the group of households $\ddot{H}$ in the base period and that p_{gb}^0 denotes the corresponding unit value; p_{gb}^t is the comparison period unit value of commodity g in outlet b.

It is important to notice that if B_g contains only one element, P_g^t simplifies to the relative p_{gb}^t/p_{gb}^0, where b is the only outlet in which g can be bought. In the next section we consider the case where the size of B_g is larger than 1.

3. The estimation of a commodity price index

The estimation of (9) usually proceeds with help of a stratification of the set of outlets B_g e.g. according to region, type and/or size-class, that is a decomposition

$$(11)\quad B_g = \cup_{i=1}^{I} B_{gi}\ ,\ B_{gi}\cap B_{gi'} = \emptyset\ (i\neq i')\ .$$

Based on this stratification we can decompose (9) as follows

$$(12)\quad P^t_g = \sum_{i=1}^{I} \frac{\Sigma_{b\in B_{gi}} p^0_{gb} x^0_{gb}}{\Sigma_{b\in B_g} p^0_{gb} x^0_{gb}} \frac{\Sigma_{b\in B_{gi}} p^t_{gb} x^0_{gb}}{\Sigma_{b\in B_{gi}} p^0_{gb} x^0_{gb}} \equiv \sum_{i=1}^{I} w^0_{gi} P^t_{gi}$$

with obvious definitions. The stratum weights w^0_{gi} $(i=1,\ldots,I)$ are the relative base period sales. It is assumed that these weights can be obtained from retail trade turnover statistics. They usually do not apply to specific commodities but to broader commodity groups. The stratum price indices P^t_{gi} $(i=1,\ldots,I)$ are the population analogs of Szulc's (1987) *micro-indices*. They can be estimated in various ways, dependent on the information available and the assumptions one is prepared to make.

Let $\hat{B}^s_{gi}$ be a simple random sample taken from B_{gi} at period s (s=0,t).[5] Then P^t_{gi} can be estimated approximately unbiased by

$$(13)\quad \frac{\Sigma_{b\in \hat{B}^t_{gi}} p^t_{gb} x^0_{gb}/n(\hat{B}^t_{gi})}{\Sigma_{b\in \hat{B}^0_{gi}} p^0_{gb} x^0_{gb}/n(\hat{B}^0_{gi})} ,$$

where $n(\hat{B}^s_{gi})$ denotes the sample size (s=0,t). This involves observing not only base period and comparison period prices but also base period quantities in the sampled outlets. Usually this is considered as laying too heavy a burden on the respondents.

If we could assume that within each stratum B_{gi} $(i=1,\ldots,I)$ the finite population covariance between x^0_{gb} and p^s_{gb} (s=0,t) is zero[6], then P^t_{gi} becomes equal to

$$(14)\quad \Sigma_{b\in B_{gi}} p^t_{gb}/\Sigma_{b\in B_{gi}} p^0_{gb} ,$$

which can be estimated approximately unbiased by

$$(15)\quad \frac{\Sigma_{b\in \hat{B}^t_{gi}} p^t_{gb}/n(\hat{B}^t_{gi})}{\Sigma_{b\in \hat{B}^0_{gi}} p^0_{gb}/n(\hat{B}^0_{gi})} .$$

The numerator and the denominator of (15) are unbiased estimators of the respective parts of (14). Expression (15) is a *ratio of average prices*.

---

5) Any other sampling design requires more information on B_{gi} than the addresses of the outlets.

6) A specific case is when $x^0_{gb} = x^0_{gb'}$, for all b, b'$\in B_{gi}$.

Alternatively, if we could assume that within each stratum the finite population covariance between $p^0_{gb}x^0_{gb}$ and p^t_{gb}/p^0_{gb} is zero[7], then P^t_{gi} becomes equal to

$$(16)\ \Sigma_{b\in B_{gi}}(p^t_{gb}/p^0_{gb})/n(B_{gi}) \ ,$$

where $n(B_{gi})$ is the size of stratum B_{gi}. This is an *average of price relatives*. It could be estimated unbiasedly by

$$(17a)\ \Sigma_{b\in \hat{B}^0_{gi}}(p^t_{gb}/p^0_{gb})/n(\hat{B}^0_{gi}) \ ,$$

or by

$$(17b)\ \Sigma_{b\in \hat{B}^t_{gi}}(p^t_{gb}/p^0_{gb})/n(\hat{B}^t_{gi}) \ .$$

In both cases, however, we can meet with difficulties. If $b \in \hat{B}^0_{gi}$ but $b \notin \hat{B}^t_{gi}$ we are not able to observe p^t_{gb}, and if $b \in \hat{B}^t_{gi}$ but $b \notin \hat{B}^0_{gi}$ we are in general not able to observe p^0_{gb}.[8] This implies that in estimating (16) we must restrict to the matched sample $\hat{B}^0_{gi}\cap\hat{B}^t_{gi}$. Thus (16) can only be estimated by

$$(17c)\ \Sigma_{b\in \hat{B}^0_{gi}\cap\hat{B}^t_{gi}}(p^t_{gb}/p^0_{gb})/n(\hat{B}^0_{gi}\cap\hat{B}^t_{gi}) \ ,$$

provided that $\hat{B}^0_{gi}\cap\hat{B}^t_{gi} \neq \emptyset$. This is again an unbiased estimator of (16), provided that $\hat{B}^0_{gi}\cap\hat{B}^t_{gi}$ is a simple random sample from B_{gi}.

In the course of time the size of the matched sample $n(\hat{B}^0_{gi}\cap\hat{B}^t_{gi})$ might tend to zero. This is especially the case if a rotating panel of outlets is used. We can, however, also write P^t_{gi} in the following form

$$(18)\ P^t_{gi} = \prod_{\tau=1}^{t}\left(\frac{\Sigma_{b\in B_{gi}}p^\tau_{gb}x^0_{gb}}{\Sigma_{b\in B_{gi}}p^{\tau-1}_{gb}x^0_{gb}}\right) \ .$$

This can be estimated by

---

7) A specific case is when $p^0_{gb}x^0_{gb} = p^0_{gb'}x^0_{gb'}$ for all b, b' $\in B_{gi}$.

8) With respect to the class of estimators introduced by Valliant (1991) one must assume that when outlet b is in the period t sample the ratio p^t_{gb}/p^0_{gb} can be observed. This is an unrealistic assumption.

(19) $\Pi_{\tau=1}^{t} \underline{P}_{gi}^{\tau-1,\tau}$,

where $\underline{P}_{gi}^{\tau-1,\tau}$ is defined by expression (17c) after replacing 0 by τ-1 and t by τ.

The estimators (15) and (17a,b,c) can also be justified on other grounds. Suppose that the sample $\hat{B}_{gi}^{s}$ (s=0,t) is selected with probabilities proportional to base period quantities. That is, the selection probability of outlet $b \in B_{gi}$ at period s is $n(\hat{B}_{gi}^{s})x_{gb}^{0}/\Sigma_{b\in B_{gi}}x_{gb}^{0}$. Then the design-expectation of $\Sigma_{b\in\hat{B}_{gi}^{s}}p_{gb}^{s}/n(\hat{B}_{gi}^{s})$ is $\Sigma_{b\in B_{gi}}p_{gb}^{s}x_{gb}^{0}/\Sigma_{b\in B_{gi}}x_{gb}^{0}$. Thus under this sampling design expression (15) is an approximately design-unbiased estimator of P_{gi}^{t}.

Alternatively, suppose that the sample $\hat{B}_{gi}^{s}$ (s=0,t) is selected with probabilities proportional to base period sales. That is, the selection probability of outlet $b \in B_{gi}$ at period s is $n(\hat{B}_{gi}^{s})p_{gb}^{0}x_{gb}^{0}/\Sigma_{b\in B_{gi}}p_{gb}^{0}x_{gb}^{0}$. Then the design-expectation of (17a) and (17b) is equal to P_{gi}^{t}. If the matched sample $\hat{B}_{gi}^{0}\cap\hat{B}_{gi}^{t}$ can be considered as having been selected with probabilities proportional to base period sales, then (17c) is also a design-unbiased estimator of P_{gi}^{t}.

A third route is to assume that certain superpopulation models are true. Suppose that the prices p_{gb}^{s} ($b\in B_{gi}$) (s=0,t) are distributed with expectation $E_{M}(p_{gb}^{s}) = \mu_{gi}^{s}$. Suppose further that the population B_{gi} and the sample $\hat{B}_{gi}^{s}$ are both large so that the expectation of a ratio is approximately equal to the expectation of its numerator divided by the expectation of its denominator. Then the model-expectation of the difference between P_{gi}^{t} and expression (15) is approximately zero.

On the other hand, if the price relatives p_{gb}^{t}/p_{gb}^{0} ($b\in B_{gi}$) are distributed with expectation $E_{M}(p_{gb}^{t}/p_{gb}^{0}) = \mu_{gi}$, then the model-expectation of the difference between P_{gi}^{t} and expressions (17a, b, or c) is zero. The latter assumption is equivalent to assuming that B_{gi} is approximately an Hicksian aggregate.

It is difficult to judge the relative merits of the three approaches discussed. The validity of the first and the third - the covariance approach and the superpopulation approach respectively - depends critically on population characteristics that are subject to empirical checking.[9)] In order to implement the second approach - the sample design approach - one must have extensive knowledge of the population quantities or values. In all cases we need information that is hard to obtain in practice. Which of these approaches should be chosen cannot be prescribed a priori. The choice depends on the specific commodity considered and the information available. This information does not need to be precise. Also vague knowledge can help in determining which assumption is plausible. The appropriate estimator then follows in the way discussed above.

4. Accounting for changing universes of commodities and outlets

Until now we assumed that the commodity groups A_k and the sets of outlets B_g were fixed during time. As every statistician knows, this assumption is patently untenable. Commodities appear and disappear, and the same holds for outlets. We can model this by attaching a time variable to A_k and B_g.

It appears to be very difficult to incorporate the phenomenon of changing universes of commodities and outlets into the individual household price indices (2). We proceed therefore by directly modifying the group price index. The *object of estimation* then becomes instead of (4)-(6) the following set of expressions:

$$(20)\quad P^t = \sum_k w_k^0 P_k^t \ ,$$

where

$$(21)\quad P_k^t = \prod_{\tau=1}^{t} \left(\frac{\Sigma_{g\in Ak(\tau-1)} \Sigma_{b\in Bg(\tau-1)} p_{gb}^{\tau} x_{gb}^{\tau-1}}{\Sigma_{g\in Ak(\tau-1)} \Sigma_{b\in Bg(\tau-1)} p_{gb}^{\tau-1} x_{gb}^{\tau-1}} \right)$$

for k=1,...,K are the commodity group price indices and

$$(22)\quad w_k^0 \equiv \Sigma_{g\in Ak(0)} \Sigma_{b\in Bg(0)} p_{gb}^0 x_{gb}^0 / \Sigma_k \Sigma_{g\in Ak(0)} \Sigma_{b\in Bg(0)} p_{gb}^0 x_{gb}^0$$

for k=1,...,K are the aggregate base period expenditure shares.[10] Notice that (20) cannot be written as an average of household-specific price indices. This is due to the multiplicative structure of (21).

---

9) For example, both covariance assumptions can not be true at the same time. If $\text{covar}(p^s_{gb}, x^0_{gb}) = 0$ (s=0,t) then also $\text{covar}(p^t_{gb}/p^0_{gb}, x^0_{gb}) = 0$, but $\text{covar}(p^t_{gb}/p^0_{gb}, p^0_{gb}x^0_{gb}) \neq 0$ unless $\text{covar}(p^t_{gb}/p^0_{gb}, p^0_{gb}) = 0$. The latter condition is in general not satisfied.

In the running computation process the problem is the estimation of the τ=t factor of (21). The structure of this factor, however, is the same as that of expression (5). Thus we can retrace the steps from (7) with obvious changes of notation.

5. Can geometric averages be justified?

The approaches discussed in section 3 lead basically to only two admissible estimators for the outlet-stratum price indices P^t_{gi}, namely the ratio of average prices (15) and the matched sample average of price relatives (17c). Advocates of the use of geometric averages would estimate the outlet-stratum price indices by respectively

$$(23)\quad (\prod_{b\in\hat{B}^t_{gi}} p^t_{gb})^{1/n(\hat{B}^t_{gi})} / (\prod_{b\in\hat{B}^0_{gi}} p^0_{gb})^{1/n(\hat{B}^0_{gi})} \quad ,$$

or

$$(24)\quad \prod_{b\in\hat{B}^0_{gi}\cap\hat{B}^t_{gi}} (p^t_{gb}/p^0_{gb})^{1/n(\hat{B}^0_{gi}\cap\hat{B}^t_{gi})} \quad .$$

Within the present context however I fail to see which kind of assumption would justify the use of (23) or (24). In the remainder of this section I will scetch two different approaches which lead to geometric averages of price relatives. Both of these approaches require a modification of the object of estimation.

The first approach is in the spirit of the so-called axiomatic theory of price indices. Returning to (12), P^t_{gi} is replaced by a function of price relatives only, $m(\ldots,p^t_{gb}/p^0_{gb},\ldots)$ where $b\in B_{gi}$. It is assumed that the function $m:R^N_{++} \rightarrow R_{++}$ satisfies the following conditions:

(i) m is separable, that is $m(z_1,\ldots,z_N) = g(z_1)*g(z_2)*\ldots*g(z_N)$ where g is a continuous nonconstant function and * is a continuous, associative and cancellative operation;

(ii) $m(z,\ldots,z) = z$;

(iii) $m(rz_1,\ldots,rz_N) = rm(z_1,\ldots,z_N)$ $(r>0)$;

10) It is assumed that, at least over a reasonable time span, it is not necessary to introduce or delete commodity groups.

(iv) $m(1/z_1,\ldots,1/z_N) = 1/m(z_1,\ldots,z_N)$.
Then, using a theorem of Aczél (1987), it follows that

$$(25)\ m(\ldots,p^t_{gb}/p^0_{gb},\ldots) = \prod_{b\in B_{gi}}(p^t_{gb}/p^0_{gb})^{1/n(B_{gi})},$$

which can be estimated by (24). Similar results are stated by Diewert (1993), (1995) and Martini (1994).

The second approach is in the spirit of the so-called economic theory of price indices. At the lowest level of aggregation, the outlet-stratum level, substitution between outlets is allowed.[11] That is, the following behavioral hypothesis is assumed to hold:

$$(26)\ \Sigma_{b\in B_{gi}} p^0_{gb} x^0_{gb} =$$

$$\min\ \{\ \Sigma_{b\in B_{gi}} p^0_{gb} x_{gb}\ |\ U_{gi}(\ldots,x_{gb},\ldots) = u^0_{gi}\ \} \equiv$$

$$C_{gi}(\ldots,p^0_{gb},\ldots;u^0_{gi})$$

where U_{gi} is a utility function with arguments x_{gb} ($b\in B_{gi}$), C_{gi} is the dual expenditure (cost) function with arguments p_{gb} ($b\in B_{gi}$) and u_{gi}, and u^0_{gi} is the base period value of the utility function. Then in the object of estimation P^t_{gi} is replaced by the ratio of expenditure function values

$$(27)\ C_{gi}(\ldots,p^t_{gb},\ldots;u^0_{gi})/C_{gi}(\ldots,p^0_{gb},\ldots;u^0_{gi}).$$

It is well known that if U_{gi} has the Cobb-Douglas form (unitary elasticities of substitution) then (27) can be written as a weighted geometric mean of price relatives

$$(28)\ \exp\ \{\ \Sigma_{b\in B_{gi}} p^0_{gb} x^0_{gb}\, \ell n(p^t_{gb}/p^0_{gb})/\Sigma_{b\in B_{gi}} p^0_{gb} x^0_{gb}\ \}.$$

The weights are the base period within-outlet-stratum expenditure shares. Now in order to justify (24) as an estimator of (28) we need a further assumption, namely either

- that the finite population covariance between $p^0_{gb}x^0_{gb}$ and $\ell n(p^t_{gb}/p^0_{gb})$

11) This approach is related to Moulton (1993). See also Diewert (1995).

is zero[12]; or
- that the matched sample is drawn with probabilities proportional to $p^0_{gb}x^0_{gb}$; or
- that the logarithmic price relatives $\ell n(p^t_{gb}/p^0_{gb})$ follow a certain distribution.

The foregoing scetches make clear that the route to (24) is paved with difficulties. In both approaches the object of estimation has to be modified, thereby distorting the interpretation of the CPI as a genuine group price index. Moreover, quite a number of additional assumptions appears to be necessary.

6. Conclusion

We considered the estimation of a Laspeyres-type CPI for a group of households. The discussion was directed to the index formula that should be used at the first step of the calculation process. The outcome is that there is no universally applicable formula, but that in every specific case a choice must be made which takes into account the specific circumstances. This position corresponds with the conclusion of Woolford (1994).

As a by-product we conclude that there seems to be no straightforward justification for using a (ratio of) geometric average(s) of prices as a first step.

References

Aczél, J., 1987, A Short Course on Functional Equations (Reidel, Dordrecht etc.).

Balk, B. M., 1995, On the Use of Unit Value Indices as Consumer Price Subindices, Report (Department of Statistical Methods, Statistics Netherlands, Voorburg).

Carruthers, A. G., D. J. Sellwood and P. W. Ward, 1980, 'Recent Developments in the Retail Prices Index', The Statistician 29, 1-32.

Dalén, J., 1991, 'Computing Elementary Aggregates in the Swedish Consumer

12) Cf. footnote 7.

Price Index', Journal of Official Statistics 8, 129-147.

Dalén, J., 1994, Sensitivity Analyses for Harmonising European Consumer Price Indices, Paper presented at the International Conference on Price Indices, 31 October - 2 November, Ottawa.

Diewert, W. E., 1993, 'Symmetric Means and Choice under Uncertainty', Chapter 14 of Essays in Index Number Theory Volume 1, edited by W. E. Diewert and A. O. Nakamura (North-Holland, Amsterdam etc.).

Diewert, W. E., 1995, Axiomatic and Economic Approaches to Elementary Price Indexes, Mimeo.

Martini, M., 1994, Elementary Indices, for the Bilateral and Multilateral Comparison, in an Axiomatic Approach, Mimeo (First Draft).

Morgan, D. D. V., 1981, 'Estimation of Retail Price Changes', The Statistician 30, 89-96.

Moulton, B. R., 1993, 'Basic Components of the CPI: Estimation of Price Changes', Monthly Labor Review, December, 13-24.

Sabag, M. and Y. Finkel, 1994, 'The Israeli Consumer Price Index. Alternative Computing Methods for an "Item Price Index"', Statistical Journal of the United Nations ECE 11, 95-118.

Schultz, B. J., 1994, Choice of Price Index Formula at the Micro-Aggregation Level: The Canadian Empirical Evidence, Paper presented at the International Conference on Price Indices, 31 October - 2 November, Ottawa.

Sellwood, D. J., 1994, Constraining Macro Effects of the Use of Different Methods at the Basic Level in the Project to Harmonize Consumer Price Indices in the European Union, Paper presented at the International Conference on Price Indices, 31 October - 2 November, Ottawa.

Silver, M., 1994, Elementary Aggregates, Micro-Indices and Scanner Data: Some Issues in the Compilation of Consumer Price Indices, Report (Cardiff Business School, University of Wales College of Cardiff).

Szulc, B. J., 1987, 'Price Indices below the Basic Aggregation Level', ILO Bulletin of Labour Statistics. Reprinted as Appendix 7 in Turvey (1989).

Triplett, J. E., 1990, 'The Theory of Industrial and Occupational Classifications and Related Phenomena', Proceedings of the 1990 Annual Research Conference (Bureau of the Census, Washington DC).

Turvey, R., 1989, Consumer Price Indices, an ILO Manual (International Labour Office, Geneva).

Valliant, R., 1991, 'Variance Estimation for Price Indexes from a Two-Stage

Sample with Rotating Panels', Journal of Business and Economic Statistics 9, 409-422.

Winkler, O. W., 1993, 'A Foundation for the Realistic Measurement of Price Levels', Contributed Paper 49th Session of the International Statistical Institute, Firenze.

Woolford, K, 1994, A Pragmatic Approach to the Selection of Appropriate Index Formulae, Paper presented at the International Conference on Price Indices, 31 October - 2 November, Ottawa.