On the statistical objective of a Laspeyres' price index

By Jörgen Dalén[1]

Introduction

The present paper tries to give some kind of theoretical background to questions concerning different sampling, replacement and quality adjustment methods currently used in the European Union. It also draws some conclusions for the short term.

Two approaches will be tried. First, I will ask the question of the statistical objective of a Laspeyres' index as applied for example to the HICP. The statistical objective of any kind of survey could be seen as answering the following question.

> *Suppose we could observe the whole universe of measurement units and obtain exact information about all relevant variables for those units. What would then be our desired result?*

An answer to such a question has three components: 1) A *universe* consisting of a finite set of measurement units, 2) a number of exactly defined *variables* defined on these units (prices, quantities and qualities/characteristics of products in our price index case) and 3) a *statistical measure* (*formula*) combining the variable values for all the units into a single number.

It will unfortunately turn out that the result of this first approach is not decisive in the sense of leading to a single prescribed formulation of the objective. As a second approach, we will then turn to a so called g *factor* analysis for which single replacement methods are analysed. In the end some short term conclusions for harmonisation is drawn.

1. The statistical objective

Our starting point for discussing the objective is a two-level structuring of the universe of Products (Goods and Services) and Outlets which is considered in scope for the price index. We call the two levels:

i) The *aggregate* level. At this level we have a fixed structure of item groups (or perhaps a fixed cross-structure of item groups by regions and/or outlet types) within an index link. Genuinely new goods and services (GNGS) would be defined in terms of new groups at this level and moved into the index only in connection with a new index link (examples of such groups in recent years might be PCs, in-line skates or access to Internet).

---

[1] **Acknowledgement:** This paper was written within the author's contract for Eurostat, the Statistical Office of the European Communities. It has grown out of extended electronic discussions between the author and Don Sellwood.

ii) The *elementary* level. Within this level we (may) try to catch the properties of a changing universe in our index by comparing new and old products. In some way we must then define how the micro comparison from 0 to 1 is to be carried out when new products/outlets enter into or old products/outlets disappear from the market.

The common starting point for three alternatives at the elementary level to be presented here is the usual Laspeyres formulation of price change from period s to period t at the aggregate level:

$$\boldsymbol{I}^{st} = \frac{\sum_h \boldsymbol{Q}_h^s \boldsymbol{P}_h^t}{\sum_h \boldsymbol{Q}_h^s \boldsymbol{P}_h^s} = \sum_h \boldsymbol{W}_h^s \boldsymbol{I}_h^{st} \text{ , where } \boldsymbol{W}_h^s = \frac{\boldsymbol{Q}_h^s \boldsymbol{P}_h^s}{\sum_h \boldsymbol{Q}_h^s \boldsymbol{P}_h^s} \text{ and } \boldsymbol{I}_h^{st} = \frac{\boldsymbol{P}_h^t}{\boldsymbol{P}_h^s} \tag{1}$$

The alternatives now enter at the elementary levels, in the definition of I^{st}_h. As a further common starting point we define the set of product/outlets belonging to h in period *u (=s or t)* as Ω^u_h. We introduce the concept *observation point* and say that $\Omega^u_h=\{1,...,j,\ldots,N^u_h\}$. For each observation point $j \in \Omega^u_h$ there is a price p^u_j and a quantity sold q^u_j. In many item groups an observation point is a tightly specified item in a specific outlet.

1.1 The intersection universe

We define

$$\boldsymbol{I}_h^{st} = \frac{\sum_{j \in \Omega_h^s \cap \Omega_h^t} \boldsymbol{q}_j^s \boldsymbol{p}_j^t}{\sum_{j \in \Omega_h^s \cap \Omega_h^t} \boldsymbol{q}_j^s \boldsymbol{p}_j^s} \text{ ,} \tag{2}$$

i.e. the elementary index is defined in the usual Laspeyres' fashion but only over those observation points existing in both s and t. One may also call this index the **identical units index**. With this formulation the quality change problem at first sight disappears completely

In (2) the universe would decrease successively until it eventually becomes empty. The relative attraction of this definition for a certain item group now depends on how long this process is. If it is very long, such as for apartments (see more below), this alternative seems quite nice. Also, when working with short index links in annual chained indexes this is a viable option and some procedures for example in the Swedish CPI are best interpreted as stemming from an intersection universe idea. One might think that a chain index should generally make (2) applicable. But, as all index makers know, the rate of disappearance of some items (e.g. in women's clothing) is very fast. An even more fundamental difficulty with this formulation for some items (fashion clothing, new novels) is that time itself is a quality factor so that the value of exactly the same physical good for the consumer decreases over time. In these cases (2) would result in a serious understatement of inflation.

1.2 The double universe

The polar opposite to (2) is to estimate average prices separately in the two periods under consideration. One could then talk about a double universe situation; one universe in period s and another in period t. In an additional procedure we would then estimate the average quality change between the universes. In mathematical terms we obtain the following formulation:

$$I_h^{st} = \frac{\bar{P}_h^t}{\bar{P}_h^s \, g_h^{st}} \text{ , where } \bar{P}_h^u = \frac{\sum_{j \in \Omega_h^u} q_j^u p_j^u}{\sum_{j \in \Omega_h^u} q_j^u} \text{ for } u=s \text{ and } t. \qquad (3)$$

g_h^{st} is the average quality change in h (also interpretable as a *quality index*), which of course needs further definition. As this is not important for our main line of argument we defer it to an example in Annex 1.

We may think of this index as a **quality adjusted unit value index**.

Advantages of this formulation are:

1. It is a natural formulation of average price change, treating the time periods s and t symmetrically.

2. It effectively separates the representativity and the quality adjustment problems by isolating quality change to a separate factor in (3).

3. There is only one quality adjustment factor in each group h. This facilitates a separate, perhaps EU-centralised, estimation procedure of the quality adjustment factor *g*.

But there are also problems with this formulation.

1. The present sampling and estimation procedures based on matching items are quite far from being optimal for estimating (3). The kind of sampling designs that come to mind for estimating it are separate (although possibly co-ordinated) samples from different sampling frames in each time period.

2. This formulation is very sensitive to obtaining a correct estimate of *g*, the quality change. This problems would, however, be smaller in those cases where one is prepared to accept that g=1 on average.

3. It is hard to formulate an actual definition of what constitutes average quality change in the universe. In some cases hedonic regression could be of help but at the present state of art it will probably have to be left to democratic agreement or convention in many cases. This problem is not unique for this particular formulation, though.

In spite of these problems we believe that this is the most theoretically satisfactory formulation of the universe and objective of the Laspeyres index and in many cases it is also possible to apply in practice.

1.3 The replacement universe

Neither the intersection nor the double universe bear a close resemblance to most actual practices carried out when constructing consumer price indices. We therefore turn to a formulation which is closer to that practice.

For this purpose we define replacements in each case where an observation point disappears. That is, for each $j \in \Omega_h^s$ and $j \notin \Omega_h^t$ we define an $a_j \in \Omega_h^t$ whose price enters into p^t_j:s place in the formula. In addition to a replacement we may now also have a quality change from j to a_j which gives rise to a quality adjustment factor g_{ja_j} , interpreted for example as the number with which p^s_j must be multiplied for the average consumer to be indifferent between j and a_j in period t:

$$I_h^{st} = \frac{\sum_{j \in \Omega_h^s} q_j^s p_{a_j}^t}{\sum_{j \in \Omega_h^s} q_j^s p_j^s g_{ja_j}} \quad , \tag{4}$$

The advantage of this formulation is that it quite closely mimics index practice in most countries and tries to blow this up to a universal scale. For some item groups such as clothing, an enormous amount of quality adjustment factors g_{ja_j} need to be defined but these could be indirectly defined through an automatic procedure such as hedonic regression. But the formulation still has two quite serious problems:

1. A replacement for each disappearing item needs to be defined. It is hard to see a general, operational way of making this definition. “Most like”, “most sold” or “lasts longest” which are popular decision rules are very hard to define strictly.

2. Not all observation points in t enter into the definition since the average is taken over all replacements rather than over the whole set of items existing in the two time periods. This makes it hard to think of this index as the average price change from s to t.

In order to further analyse the suitability of procedures when working in a replacement-like universe we will, below, turn to a so called *g factor analysis*.

1.4 Examples

We will look here at five very different kinds of item groups and we will give some formulations of what we believe are the best formulations of the statistical objectives in each case.

Bananas
We start off with the simplest case that we can think of. Basically we believe that *bananas* is very close to a perfectly homogeneous item group. It is sold per kg or pound and in any particular country probably only one of these. The only quality differences that could be imagined are those between outlets. After dividing the outlet universe into more or less homogenous subgroups of outlet type and perhaps region, the double universe formulation with formula (3) and g=1 would be our preferred choice.

Toilet paper
Our next example is a product which may still be thought of as quite homogenous. But the sizes and number of rolls in a package may differ greatly as well as (to a lesser extent) the actual paper quality. What this means is that the *price* variable needs to be exactly defined (price per roll, per meter, per gram or per package?) and quantities defined accordingly. After this a subgrouping into package sizes (as well as outlet type and region) could be done and within this subgrouping the double universe formulation with (3) and g=1 would again be our first choice.

Women's dresses
This is the case of an extremely heterogeneous good. The double universe with (4) is still the best formulation of the object of measurement! A decision on a subgrouping of dresses according to some characteristics and to outlet type would then be a first step. As for g it could either be estimated by hedonic regression or set to 1.

Rents for apartments
This is a case of a very slowly changing population of observation units. It would be quite possible to have no subgrouping at all, that is to consider all apartments of the whole country as the item group h in (1). We could actually use any of the three universe definitions above. Although not common, there are cases of apartments torn down or changed into offices etc. so that the intersection universe according to (2) is not a bad choice. (4) is less natural to replace a disappearing apartment does not seem necessary since there are so few cases. Using (3) would mean that newly built apartments are to be included in the numerator. We would then probably want to make some further homogenising subgrouping into regions, sizes etc. and, since new apartments usually (?) have a higher standard with regard to equipment etc. we would probably also want to estimate g rather than setting it to 1 which makes the double universe a more complex and sensitive choice. (Sweden is presently moving from a definition close to (3) to one close to (2).)

Local transport
Here, we think of expenditure for public transport arrangements used for daily travels within a local area (typically a city with suburbs). There are many modes of payment: for a single ticket, for a pack of 10 (or some other number) tickets, for unlimited travel during a whole day, a whole week, a month or some other period. The amounts paid may also be different for different travel zones.

This is actually the most complex case for a universe formulation! We need to decide i) what should be the pricing unit and ii) a subgrouping which is stable over time. There is more than one way to do this but one possibility would be the following.

The first subgrouping is into separate local transport systems. Within each such system there are, say, 1000 "stops" (underground stations, bus stops etc., a new stop would be a GNGS) in the base period s. Each pair of stops (there would in this case be 499500 such pairs) is a group h according to (1). If we choose price per single trip as our pricing unit we would have to count the number of journeys between these stops in period s, determine the price per trip that is paid each time (by, e.g., dividing the price of a monthly card by the actual number of trips in that month travelled by the person doing the journey). In this way we would eventually be able to apply the double universe formula (4) in each group. We would almost always be satisfied with setting g=1. (In practice we would not need to be as specific as this. All stop pairs within the same zone pair could be treated as one group as long as the division into zones is not changed but the most detailed formulation would be a fall-back position when the tariff system changes.)

Another possible formulation is to treat monthly cards etc. as separate groups. When the tariff system changes we would then have to resort to a replacement universe formulation according to (3) which gives more scope to differences in detailed procedures. Although simpler, it would be hard to make this formulation as unambiguous as the one above.

New cars
A first question to ask is whether there is a fashion element in new cars. If not, then one could divide the car market into fairly homogeneous groups based on size (small, medium, big) and other main characteristics (ignorance of car terms in English prevent me from further discussion). Within such groups a double universe definition seems a reasonable choice in theory.

In practice procedures chosen are much more like (4), where a successor to a given car model is usually (?) sought within the same brand. This is hard to justify theoretically, though. A newly introduced small Korean car model is not a genuinely new good but rather something that could well be compared with an already existing small car model of a Western make.

For quality adjustment there are various methods for direct adjustment available (see below).

Personal computers
The replacement universe, although probably similar to what is done in many countries, is an even less reasonable choice in the PC case than in the car case. The PC market is characterised by several models living their life in parallel with new high end models continually entering the market and low end models disappearing. In between, former top end models move successively lower on the quality ladder. This means that the disappearing product is logically replaced by something which is already in the index!

In this situation the only approach that makes theoretical sense is a collective price comparison with a proper direct quality adjustment according to the principles of the double universe with option cost or hedonic adjustments.

2. Analysis of g factors in a replacement universe

The g factor interpretation of quality adjustment has a long tradition. The pioneering work was done by Hofsten (1952). Recent papers in this tradition are, e.g., by Schultz (1995), Moulton and Moses (1997) and Boon and de Haan (1997). In order to apply this kind of analysis, however, we need to be in a consistent replacement universe with one-to-one matched comparisons.

In situations of item discontinuity in price measurement, a large number of procedures are used. In this section we take the replacement as given: A move from measuring the item A price in period t-1 ($\boldsymbol{p}_A^{t-1}$) to measuring the item B price ($\boldsymbol{p}_B^t$) in the immediately following period t. We try to list the possible procedures that are currently used in this situation, interpret the implicit quality adjustment factor inherent in these methods and, based on this interpretation, to discuss some criteria for their suitability.

First we introduce the concept *total price change*, defined as

$$\boldsymbol{T}^{t-1,t} = \frac{\boldsymbol{p}_B^t}{\boldsymbol{p}_A^{t-1}} \qquad (5)$$

Averaging total price changes in an item group leads to what has elsewhere been called *Standard Reference Indexes*. (A slightly revised paper on Standard Reference indexes, presented at a CPI Harmonisation Working Party Meeting in October 1997 is given in Annex 2.) Total price change is decomposed into *pure price change* ($\boldsymbol{I}^{t-1,t}$) which is our measurement objective and *quality change* ($\boldsymbol{g}^{t-1,t}$ or just $\boldsymbol{g}$ for short). $\boldsymbol{g}$ is also called the *QA* (quality adjustment) or *g factor*. By definition we thus have

$$\boldsymbol{T}^{t-1,t} = \boldsymbol{I}^{t-1,t} * \boldsymbol{g} \qquad (6a)$$

and thus $$\boldsymbol{g} = \frac{\boldsymbol{T}^{t-1,t}}{\boldsymbol{I}^{t-1,t}} = \frac{\boldsymbol{p}_B^t}{\boldsymbol{p}_A^{t-1}} * \frac{1}{\boldsymbol{I}^{t-1,t}} \qquad (6b)$$

Note that g=1 means that there is no quality change.

2.1 Some procedures used in cases of replacement

Unadjusted price comparison
(Also called direct price comparison.) This corresponds to a judgement that the quality difference between A and B is small. We have

$$\boldsymbol{I}^{t-1,t} = \frac{\boldsymbol{p}_B^t}{\boldsymbol{p}_A^{t-1}} \text{ and } \boldsymbol{g} = 1. \tag{7}$$

Comment on suitability: A bad condition for applying this method is in areas of rapid quality change (usually improvement) such as high-tech goods and especially PCs. On the other hand it may work quite well in areas where quality changes are more random in nature and tend to cancel on average. This is true for many kinds of low tech goods, such as clothing, furniture and household utensils. Where a "most like" criterion for replacement is applied, this simple procedure is further improved.

Quantity augmenting.
This procedure is typically carried out when package sizes change. Usually the g factor is then taken to be exactly proportional to the change in package size. We have

$$\boldsymbol{g} = \frac{\boldsymbol{k}_B^t}{\boldsymbol{k}_A^{t-1}} \text{, where k stands for the quantity (package size)} \tag{8}$$

Comment on suitability: This method is recommended in cases where the package size change is limited – say between ½ and 2. It could also be seen as a particular case of direct quality adjustment (next method).

Direct quality adjustment
In this case an explicit decision is taken on the quality difference between A and B. Under this general heading we put the following "submethods":

Option cost adjustment. This method values an extra characteristic, now included in the price of a composite product, as its actual price as an optional extra in the previous period or as a fixed portion of that price. For example, ABS brakes on cars or extra hard disk space in a PC could be given a value this way.

Production cost adjustment. Here, the producer is asked how much the extra facility costs to produce, as a proportion of the total production cost of the item and this factor is used as $\boldsymbol{g}$.

Hedonic regression. In this method, the value of a characteristic is given by estimated coefficients in a multiple regression equation. If the regression is given a semi-logarithmic form and the characteristics are given as binary variables (existing or non-existing), the coefficients could be interpreted as percentages of total price.

Assessment, by experts, commodity specialists or interviewers. At present, these assessments could take on different forms. They could be expressed as a fixed amount of money to be added or subtracted to the reference price or (less common) to the comparison price or they could be expressed as percentages of the price difference between the two products (A and B).

It is possible to express all these types of adjustments as multiplicative corrections to the price ratio, resulting in a QA factor of the following form:

$$\boldsymbol{g} = \prod_{k} \gamma_{k}^{t-1,t} \quad , \tag{9}$$

where the γ are adjustment factors due to single factors k which are interpreted as the percentage of quality change that is due to this specific factor. Expressions (8) and (9) may be called *external* adjustments, since they bring in new information into the index data base which is not already there.

Comment on suitability: Of these methods option cost pricing or hedonics should be the first choices, since they provide some kind of objective basis to the adjustment. Assessments are also acceptable, if well documented and evaluated so as to reflect consumer preferences. Production cost is only to be used, where it can be demonstrated to be a good proxy for the consumer evaluation.

Overlap pricing

In this situation prices are available for both item A and item B at time t, so that price change up to t could be based on item A but after t on item B. In this situation we thus have

$$\boldsymbol{I}^{t-1,t} = \frac{\boldsymbol{p}_{A}^{t}}{\boldsymbol{p}_{A}^{t-1}} \text{ and } \boldsymbol{g} = \frac{\boldsymbol{p}_{B}^{t}}{\boldsymbol{p}_{A}^{t}} \tag{10}$$

Comment on suitability: This method stands and falls with the assumption that the price differential between A and B is a genuine reflection of the consumer's perception of quality difference. A good condition for this method to be used is where A and B are sold in parallel for a long time. This is often true in the PC market today, where different quality levels coexist. Bad conditions for this method are:

i) where there is just a final clearance sale of A at time t or

ii) where quality is in itself a function of time since introduction (examples: fashion clothing, new novels).

Resampling

This method could also be labelled *aggregate overlap pricing*. It means that in period t a new sample of items and/or outlets is priced. Thus, up to t the estimate of price change is based on the old sample, called A, and after t on the new sample, called B. In the same manner as for overlap pricing we could thus write:

$$I^{t-1,t} = \frac{\bar{p}_A^{\,t}}{\bar{p}_A^{t-1}} \text{ and } g = \frac{\bar{p}_B^{\,t}}{\bar{p}_A^{t}} \tag{11}$$

The expressions now involve average prices in the sample rather than simple prices.

A special type of resampling is done in some countries which use annual chain indexes, where new samples are drawn for use in December each year.

Comment on suitability: As for simple overlap pricing, resampling depends on the differences between average prices between A and B to be genuine reflections of quality differences. This fact is not always recognised. When resampling is only of outlets this assumption is generally acceptable, although far from obvious. But when resampling is of items, care must be taken that none of the bad conditions for overlap pricing is at hand:

i) There is a tendency towards final clearance sales of A items at time t or

ii) Quality is in itself a function of time since introduction so that the old sample will have a lower quality than the new sample.

Imputation

Here an average price change is imputed from some higher aggregate (AGG) of which a missing item is a part. We obtain

$$I^{t-1,t} = I_{AGG}^{t-1,t} \text{ and thus } g = \frac{p_B^t}{p_A^{t-1}} * \frac{1}{I_{AGG}^{t-1,t}} \tag{12}$$

A (worse) variant of this method is where imputation is done of the whole price change from the reference period b to t and the observed price change from *b* to *t-1* is thus dropped.

Comment on suitability: Here it is essential that price change pattern is unrelated to the disappearance pattern. A bad condition for this method is where sellers always change prices in connection with the introduction of new items. A variant of this method, used in the U.S., which may be considered in this latter situation is so called *class mean imputation* where the aggregate AGG in (8) is taken to be all replacements where direct adjustment methods have been used, excluding cases of matching items.

Automatic linking

This method is sometimes also labelled *link-to-show-no-change*. Items are simply called “non-comparable” and the price level is considered unchanged. We have

$$I^{t-1,t} = 1 \text{ and } g = \frac{p_B^t}{p_A^{t-1}} \tag{13}$$

Comment on suitability: This method generally gives a systematic underestimation of inflation and is therefore already banned by EU regulation. It is defensible only to the extent that it is a good approximation to some other appropriate method. For example, where there is a price of A in period t and $p_A^{t-1} = p_A^t$, it coincides with the overlap method.

Aggregating g factors

If we work entirely in a replacement universe setting and apply one of the prescribed elementary aggregate formulae we could divide the sample into two parts: a matched part called M and a replaced part called R. The following formulae show how it works out in two cases:

The ratio of average prices: $$\boldsymbol{I}^{t-1,t} = \frac{\frac{1}{n}\left(\sum_{j\in M}\boldsymbol{P}_j^t + \sum_{j\in R}\boldsymbol{P}_j^t\right)}{\frac{1}{n}\left(\sum_{j\in M}\boldsymbol{P}_j^{t-1} + \sum_{j\in R}\boldsymbol{g}_j\boldsymbol{P}_j^{t-1}\right)} \qquad (14)$$

and

the geometric mean: $$\boldsymbol{I}^{t-1,t} = \frac{\left(\prod_{j\in M}\boldsymbol{P}_j^t \times \prod_{j\in R}\boldsymbol{P}_j^t\right)^{\frac{1}{n}}}{\left(\prod_{j\in M}\boldsymbol{P}_j^{t-1} \times \prod_{j\in R}\boldsymbol{g}_j\boldsymbol{P}_j^{t-1}\right)^{\frac{1}{n}}} \qquad (15)$$

In (15), but **not** in (14), it is possible to separate the g factors and obtain:

$$\boldsymbol{I}^{t-1,t} = \frac{\left(\prod_{j\in M\cup R}\boldsymbol{P}_j^t\right)^{\frac{1}{n}}}{\left(\prod_{j\in M\cup R}\boldsymbol{P}_j^{t-1}\right)^{\frac{1}{n}}} \bigg/ \prod_{j\in R}\boldsymbol{g}_j \quad , \qquad (15b)$$

where the first part can be interpreted as a geometric form of the Standard Reference Index and the second part as an Implicit Quality Index (IQI). The IQI could then also be decomposed into effects arising from different procedures.

The difficulty in applying (15b), however, lies in i) some microindices are ratios of averages which are not decomposable and ii) estimates from dynamic universes do not always take the form of matched replacements.

3. Conclusions for harmonisation

1. The above presentation gives an idea of the complexity involved in harmonising price index practices. Formulating the statistical objective (with the three components of universe, variables and formula) seems to be a first step. But within each possible objective there are still a large number of possible procedures. Full harmonisation, where we could feel 100% certain that comparability is achieved would require an agreement on all of these aspects.

2. Theoretically the double universe formulation according to expression (3) is the most attractive choice since it has a clear interpretation in terms of the average price change of the whole micro-universes which are in scope for the price index. But unfortunately it does not correspond to present index practice. One reason for present practice is the difficulty in estimating the quality change factors but is doubtful whether this excuse is enough in the long term.

3. At least in the short term, different item groups will need different statistical objectives and desired procedures. To the author's knowledge, no country today uses the same procedures all over its index.

4. Progress towards increased comparability therefore needs to be a stepwise process. One step forward would be to classify item groups according to criteria which make a particular formulation of the objective or particular replacement procedure more suitable than another. Such criteria are:
   - i) The pace of quality change in the item group. Is an assumption of average g to equal 1 acceptable or not?
   - ii) The rate of item attrition in the item group.
   - iii) The kind of data on prices, quantities and characteristics which are available in each case.

5. Study groups working on specific item groups (clothing, cars, durables) might be asked to provide g factor interpretations and estimates of methods which are used or recommended.

6. Resampling which is a popular method for keeping samples up-to-date needs careful rules. Resampling is best interpreted as aggregate overlap sampling with similar strengths and shortcomings as elementary overlapping. It is essential that the difference in average prices between successive samples for the same item group are approximately equal to differences in average quality.

7. One may relate the concept of total price change above to the Standard Reference Indices (SRI) previously discussed in the WP and TF. The "exact SRI" is in principle the unit value index without quality adjustment in (3) whereas g is the implicit quality index. This adds further weight to asking all Member States to actually compute such indices.

Note:

Scanner data do not always contain all the information needed for calculating all possible index definitions. For example the desired subgroupings are not easily done in the data sets that I have access to and not either is there information of package sizes, at least not easily accessible.

References:

Boon, M. and de Haan, J. (1997): A database of quality change estimates for harmonised indices of consumer prices. Internal Research Paper. Statistics Netherlands, Voorburg.

Hofsten, E. (1952): Price Indexes and Quality changes. Forum, Stockholm and George Allen & Unwin Ltd, London.

Moulton, B. and Moses, K.(1997): Addressing the Quality Change issue in the Consumer Price Index. Brookings Papers on Economic Activity, 1:1997.

Schultz, B.(1995): Treatment of Changes in Product Quality in Consumer Price Indices. Proceedings from the 2nd meeting of the International Working Group on Price Indices, Stockholm.

Annex 1

The quality index in the DU micro-index

In (3), g_h^{st} could be interpreted as an index of quality change in line with the interpretation:
Price index = Raw price index/quality index

How could g_h^{st} be computed? We introduce the following notation:
g_j^u is the quality of item $j \in h$ in period u (s or t). Quality is here defined relative to some standard item with g=1. We next relate to the hedonic index theory and say that quality is due to the values in characteristics space. We thus postulate that there is a fixed set of characteristics *(1,…,k, ….K)*, common to both periods in question. γ_{jk} is the component value of the characteristic (again relative to a standard). We could then define g_j^u in a multiplicative way (in line with a semi-logarithmic hedonic regression function) as

$$g_j^u = \prod_{k=1}^{K} \gamma_{jk} \tag{A1}$$

and we could then define a *unit quality index* as

$$g_h^{st} = \frac{\dfrac{\sum_{j \in \Omega_h^t} q_j^t g_j^t}{\sum_{j \in \Omega_h^t} q_j^t}}{\dfrac{\sum_{j \in \Omega_h^s} q_j^s g_j^s}{\sum_{j \in \Omega_h^s} q_j^s}} \tag{A2}$$

This formulation would be a natural extension of the kind of hedonic regression models used for clothing in the U.S. and Sweden, although defined in a double universe, where there is no matching and replacement.

The quality factor in the RU micro-index

In the Replacement Universe, we need to define the factor g_{ja_j}. According to (A1), it would be natural to put

$$g_{ja_j} = \frac{\prod_{k=1}^{K} \gamma_{a_j k}^t}{\prod_{k=1}^{K} \gamma_{jk}^s}, \tag{A3}$$

or in words as the ratio between the aggregate quality of the replacement item and the original item. (This formulation is very close to that of the Swedish hedonic clothing index.)

Comment

Although I mention hedonics above these definitions are not necessarily dependent on hedonic regression. The quality factors could be obtained through any kind of agreed procedure including quantity augmenting, option prices, expert guesses, democratic voting in the WP etc. Different procedures are likely to be adopted in different areas but the above formulation might be taken as point of departure.

Annex 2

Examples of standard reference indexes

The purpose of this note is to illustrate the standard reference index (SRI) idea, as described in the Eurostat document PTF2/97/87, with some real data and so make the discussion more concrete. Two examples of applications of the SRI idea will be given using actual CPI data for two countries – Luxembourg and Finland which have kindly provided their data for these analyses.

The SRIs are intended to be common yardsticks against which to measure existing practices. They are, so far, the only vehicle proposed that has any potential to serve this end. There are several ways to define an SRI.

1. The *exact method* requires the recalculation of each product index for a certain period without any adjustment to the observed prices. Prices should thus be directly compared in each case of replacement. The exact method is not defined in cases of new observations being added, old ones deleted or if an observation goes missing. In these cases I propose that the imputation method be appointed as the reference method. In its most ambitious version, it is defined as imputing the average price change for the other matching items within the same elementary aggregate for the shortest possible period needed. We call this method *Exact I*.

2. A version of the exact method which is easier to handle in practical computation differs only for new, deleted and missing observations. Here we impute price change for matching items for the whole time period under consideration. We call this method *exact II*.

3. A third version of the exact method is to retain the treatment of new, deleted and missing items as it is in the actual index. This method we call *exact III*.

4. The *weighted method* takes the ratio of the weighted average prices for the two time periods studied where averages are taken over all items in an elementary aggregate in each period and the weights are those actually used between the elementary aggregates.

5. The *unweighted* (previously called simple) *method* differs from the weighted method only in that it takes simple averages across all aggregates disregarding the weights.

The averages in 4 and 5 should be those actually used in the index, geometric or arithmetic; in this report we will use geometric averages. The version of the Exact method that will be used in this report is Exact II.

The ratio between an SRI and the actual adjusted index is the implied quality change in the sample between the two time periods. Since the sample should reflect the population this ratio can be interpreted as an *implicit quality index (IQI)*. For a short time period and a single item this interpretation is disturbed by haphazard replacements but for larger groups of items and/or longer time periods an IQI which is consistently well above or below 100 must be interpreted as saying something about the effect of the adjustment practices.

Finland

For the Finnish CPI (FCPI) we have had access to data for 10 representative items for the entire period 1990-94. The FCPI is a fixed base index with 1990 as the reference year. There are about 44000 observations each month divided into 401 representative items and 5 regions. For the 10 items in this study there were altogether 560 observations per month.

The Finnish CPI system uses quality codes, from 0 to 19 in cases of replacement which express the share of price change that is considered to be quality change. A quality code (Q) of 0 means no quality adjustment and a code of 10 leads to the whole price change being adjusted away.

Now the exact version (E) of the SRI in the Finnish case is interpreted as replacing the adjusted reference price (at *t=1*) with the initial reference price (at *t=0*). The weighted (W) and the unweighted (U) versions mean that (weighted) average prices are computed at both ends of the comparison.

Table 1 shows implicit quality indices (IQI 1-4) according to the exact method for successive 12-month intervals starting with December 90 and for the whole period (IQI T) which is a simple chaining of the four 12-month indexes. ACTIND is the actual index and SRI (E, W and U) the three alternative standard reference indexes.

Table 1: Standard reference indexes for the Finnish CPI, 9012-9412

ITEM CODE	IQI 1	IQI 2	IQI 3	IQI 4	IQI T	ACTIND	SRI E	SRI W	SRI U
2000	100.8	99.5	107.6	101.2	109.3	113.8	124.4	124.9	118.4
2001	100.4	96.3	99.1	100.6	96.3	103.5	99.7	99.7	112.8
2002	98.6	112.6	120	100.5	133.9	107.9	144.4	145.2	144.1
2003	123.7	109	91.7	100.1	123.7	115.7	143.1	144	124.1
2004	116.7	87.4	97.7	101.6	101.2	114.6	116	117.7	113.4
7000	102.5	98.5	102.2	98.1	101.2	119.2	120.6	120.2	120.9
7001	94.2	97.5	108.5	100.9	100.4	120.2	120.7	120.9	115.6
7002	98.8	97	104.7	99	99.3	116.6	115.8	115.9	116.9
7010	99.2	107.8	114.9	103.7	127.4	120.9	154	154.4	155.1
7011	100	100	112.1	105.8	118.5	116.9	138.5	138.5	145.8

Item codes 2000 – 2004 are respectively men's summer overcoat, men's leather jacket, raincoat, men's tracksuit and men's jacket for outdoor recreation wear. Item codes 7000- 7011 are respectively stereo system, video recorder, television, camera and video camera.

The pattern is quite irregular but two of the clothing items display large implicit quality increases on aggregate as do two of the home electronic items. SRI W is quite close to SRI E which implies that the weighted version works fairly well as an alternative reference index. This is not true for the SRI U, however, so this alternative could not be recommended.

Luxembourg

STATEC in Luxembourg has succeeded to construct a CPI system which allows an easily accessible and useful analytical database. There are 258 item subgroups immediately below the HICP COICOP level with no further stratification so the subgroups are also elementary aggregates with no weights inside. The total number of observations is about 6900 per month. Geometric means are used for aggregation.

A very nice part of the Luxembourg system are the "notes" which are applied when month-to-month comparisons between observations are in some way disturbed. Table 2 shows the number and distribution of notes up to and including July 97. (The LCPI has 1996 as its price reference year and in this year no replacements took place.)

Table 2:

# occurrences	*2*	*3*	*4*	*6*	*7*	*8*	*25*	*Sum*
9601-9707	65	135	16	424	161	31	1	833
9707 only	1	32	9	49	0	29	1	121

The meaning of these notes are given in Table 3.

Table 3: Author's interpretation of notes in the LCPI

Note	*Interpretation*
2	Monthly collection, observed price, quality change
3	1st month of missing price, old price is carried forward.
4	2nd month of missing price, old price is carried forward.
6	Monthly collection, observed price, substitution. Overlap pricing used.
7	Addition of new observation from a certain month and onwards. Before that month prices for the item were zeroed. A reference price is imputed, based on the movement of other items in the same COICOP subgroup.
8	Corrected price, usually involving a change in the reference price.
25	Seasonal price, quality unchanged

Tables 4 and 5 report on SRIs for 2-digit and 6-digit COICOP groups. The abbreviations are as in table 1. The last column - S_IQI - reports on the standard deviation between the IQIs at the 10-digit-level within the respective 2/6-digit level and is thus a measure of the variability of the IQIs. These tables give a clearly different picture compared with Table 1. IQIs are generally close to 100 and in those few cases where they are a little bit away the standard deviations are generally large so that no strong conclusions could be drawn. In general the LCPI for the studied period seems to be consistent with an assumption of unchanged overall quality.

In the LCPI SRI U and W coincide since there is no aggregation level below that of the item (COICOP 10). Usually, but not always, the W/U SRI version is quite close to the exact version.

Discussion

What can be inferred from these two examples as to the usefulness and practicability of working with Standard Reference Indexes in a larger scale? I suggest the following:

1. SRIs are quite easy to compute given a suitable database of microdata which now seem to be at hand in many Member States.

2. They are also easy to reproduce for many different items and time periods.

3. The “exact” method is not one single method in the case of a dynamic data structure. Further specifications have to be done. The “exact II” method used in this report, although not entirely optimal, should in general be quite easy to work with in a large scale. The weighted version seems generally also to be useful. The unweighted (simple) version is, however, not to be recommended.

4. SRIs and IQIs stimulate further interesting analyses of a kind that is relevant in a regulation and compliance context.

5. A remaining, more complex, task is to work out a standard partitioning of an IQI according to different adjustment methods, i.e. to divide up the difference between the actual and the reference index into parts corresponding to different methods of replacement and quality adjustment. This would require a standardised terminology as a first step. It would also require a more detailed and exact understanding of the actual measures taken in a Member State in cases of replacement. These are not entirely clear in the database setups of the kind reported here.

Table 4: Standard reference indexes for the Luxembourg CPI, July 96 – July 97, COICOP 2-digit level.

	2-digit group	**Weight**	**ACTIND**	**SRI E**	**SRI W/U**	**IQI E**	**S_IQI**
01	Food	162	102	101.4	101.1	99.4	6
02	Alcohol and tobacco	29.1	99	98.2	97.9	99.2	3
03	Clothing and shoes	117.3	101	101.1	101.1	100.1	6.2
04	Housing	132.7	102.9	103	103	100.2	1.4
05	Household equipment	120.3	100.9	100.9	100.9	100	7.7
06	Health	2.8	101.6	101.2	101.2	99.6	0.5
07	Transport	160.9	100.8	102.4	102.4	101.6	7.2
08	Communications	17	103	103.5	103.5	100.5	0.7
09	Recreation and culture	137.6	101.1	98.9	102.2	97.9	20.8
10	Education	3.4	100	100	100	100	.
11	Hotels and restaurants	63.6	102.2	102.1	102.1	99.9	0.8
12	Miscellaneous	53.3	101.6	101.6	101.9	100	2

Table 5: Standard reference indexes for the Luxembourg CPI, July 96 – July 97, COICOP 6-digit level.

CODE2	Weight	ACTIND	SRI E	SRI W/U	IQI E	S_IQI
10101	25.4	102.7	104	102.7	101.3	5.7
10102	49.1	102.4	98.7	98.7	96.4	8.1
10103	7.7	98.9	98.3	98.3	99.4	1.5
10104	19.9	99.7	100.6	101	100.9	11.6
10105	5.3	101.1	102.6	102.6	101.5	3.3
10106	11.7	107.6	107.3	106.9	99.8	1.2
10107	13.9	100.1	100.3	100.1	100.2	0.9
10108	9.6	100.4	100.4	100.7	99.9	1.5
10109	3.4	100	100	100	100	0
10201	6.7	108.5	113.9	113.1	105	9.6
10202	9.3	99.2	98.5	98.2	99.3	1.5
20101	1.8	100.3	99.8	100.3	99.5	.
20102	11.7	102.8	101.2	101.2	98.5	4.6
20103	4.7	102.4	102.2	100.1	99.8	.
20201	10.9	93.4	93	93	99.6	0.3
30101	0.3	102.7	102.7	102.7	100	.
30102	90.4	101.1	101.4	101.4	100.3	6.8
30103	1.3	102.2	100.9	100.9	98.7	.
30104	2.7	97.9	97.9	97.9	100	.
30201	22	100.6	100.1	100.1	99.4	1.7
30202	0.6	104.1	104.1	104.1	100	.
40101	56.2	103.3	103.6	103.6	100.4	3.1
40301	12.4	100	100	100	100	.
40302	9	105.6	105.6	105.6	100	.
40401	4.9	100.3	100.3	100.3	100	.
40402	1.2	104	104	104	100	.
40403	5	100.9	100.9	100.9	100	.
40501	20.2	100.9	100.9	100.9	100	.
40502	10.1	105.4	105.4	105.4	100	0
40503	12.8	104.8	104.8	104.8	100	.
40504	0.9	101.2	101.2	101.2	100	.
50101	51.6	101.5	102.2	102.2	100.8	16.2
50102	4.3	98.3	98.3	98.3	100	.
50201	10.4	100.6	100.4	100.4	99.9	3
50301	12.4	98.7	97	97	98.3	2.3
50302	0.8	98.1	98.8	98.8	100.7	.
50303	0.3	101.1	101.1	101.1	100	.
50401	4.5	100.2	98.2	98.2	98.1	5
50501	3.2	99	90.8	90.8	91.7	3
50502	4	100.8	98.5	98.5	97.7	4.1
50601	13.1	99.8	102.6	102.6	102.8	1.8
50602	15.7	103.3	102.5	102.5	99.2	.
60101	1.2	101.3	101.3	101.3	100	.

60102	1.6	101.7	101.1	101.1	99.4	.
70101	77.4	99.6	102.5	102.5	102.9	12.1
70102	3.1	98.9	104.7	104.7	105.8	.
70103	2	100.2	100.8	100.8	100.6	.
70201	5.7	98.9	101.7	101.7	102.9	7.5
70202	32.2	102.2	102.2	102.2	100	0
70203	28.5	102.4	102.4	102.4	100	.
70204	4.4	103	103	103	100	0
70301	0.9	102.9	102.9	102.9	100	.
70302	1.3	100	100	100	100	0
70303	1.9	105.1	105.1	105.1	100	.
70305	0.5	100.8	100.8	100.8	100	.
70306	3	100	100	100	100	.
80101	1.5	100	100	100	100	.
80109	15.5	103.2	103.8	103.8	100.6	.
90101	8.6	96	93.6	93.6	97.4	9.6
90102	4.8	94.1	97.7	97.7	103.8	.
90103	3.8	94.2	106.4	98.6	113	.
90104	3.5	104.1	103.9	103.9	99.9	1.3
90105	5.8	99.9	98.8	98.8	99	1.3
90106	3.8	99.6	103	103	103.5	.
90107	9.4	98.8	101.1	156.6	102.3	.
90108	5	100.7	102.2	100.5	101.5	.
90109	1.2	100.3	100.3	100.3	100	.
90201	9.4	109	109.6	109.6	100.5	0.3
90202	12.2	101	103.3	100.6	102.3	19.5
90301	8.1	101.7	103.2	103.2	101.5	.
90302	9.4	101.5	101.5	101.5	100	.
90303	0.7	100	100	100	100	.
90304	1.9	98.8	99.6	99.6	100.8	.
90401	50	102.3	93.6	93.6	91.6	41.1
100104	3.4	100	100	100	100	.
110101	56.3	102.3	102.2	102.2	99.9	1
110102	3.4	100	100	100	100	.
110201	3.9	102.7	102.7	102.7	100	0
120101	13.2	104.4	104.4	104.4	100	.
120102	16.4	99.8	100.3	101.1	100.5	2
120201	7.9	101.1	101.3	101.3	100.2	0.7
120202	3	100.9	98.3	98.3	97.5	3.3
120402	0.5	100.1	100.1	100.1	100	.
120404	5	100	100	100	100	.
120501	0.3	112	112	112	100	.
120601	7	102	102	102	100	.