Statistical targets for price indexes in dynamic universes[1]

Paper presented at the Sixth Meeting of the International Working Group on Price Indices
Canberra, Australia, 2-6 April 2001

Jörgen Dalén
jorgen.dalen@stockholm.mail.telia.com

Abstract: This paper argues that there is a need for defining the target of measurement for a consumer price index in a manner that is usual for statistical surveys. Such a target formulation serves as a bridge between the ideal economic target and the estimator used for practical measurement. This target needs to recognise explicitly the fact that the market of products and outlets is not fixed but constitutes a universe that is changing over time. A number of possible targets are presented and discussed. Three targets of special importance are the replacement index, the quality adjusted unit value index and monthly chaining and re-sampling. The latter method has recently been proposed by Aizcorbe et al (2000) as an alternative to hedonic regression for computers and other high-tech products. A recent proposal for minimum standards on representativity and sampling for the Harmonised Index for Consumer Prices in Europe, which recognises the dynamic nature of the universe, is also presented.

[1] The author is a consultant on price indexes for Eurostat, Statistics Sweden and other statistical agencies. Opinions expressed are those of the author only and not of any of these institutions. Don Sellwood has provided helpful comments on an earlier draft.

1. Introduction

This paper is concerned with specifying the statistical target of a consumer price index in a world of continual changes in the market of goods, services and outlets. For the concept of statistical target we adhere to the way of thinking that characterises survey sampling. For example, Särndal, Swensson and Wretman (1992, p. 4-5) draw up a skeleton outline of a survey target involving i) a finite set of elements called a finite population (we will instead use the word *universe*), ii) one or more *variables* of study associated with each element and iii) population characteristics or parameters which are functions of the study variable values (in the price index context we will refer to them as *aggregation formulas*).

A well-defined statistical target is a necessary prerequisite for a rational choice of a sampling design and for the application of sampling theory. In price indexes, sampling decisions involve not only initial sampling but also replacement (one-to-one) sampling and re-sampling (=the drawing of a completely new sample in a stratum). Furthermore, probability sampling is not the prevailing practice in price indexes, a fact that further complicates the issue.

When determining a statistical target according to the survey sampling paradigm, we will not be able to meet exactly an *ideal index* (such as a pure price index or a cost-of-living index) but we will want to approximate such a concept as closely as possible. A statistical target is defined only on observables. As we shall see there is no uniquely best way to formulate a target. Instead there are many possibilities, which all involve some compromises in relation to an ideal index and where the choice between them depends on the characteristics of the market concerned and on the available data. We will distinguish between

i) *primary targets*, involving also variables which, although clearly describing real entities, are not practical to observe within the time and budget constraints facing a statistical agency and
ii) *secondary targets,* involving only those variables that are actually observed by the statistical agency.

The advantage of defining a clear statistical target in between an ideal index and the estimator of it is that it facilitates the determination of an appropriate sampling strategy (design and estimator) and a focused discussion of errors (biases and variances). For example, it is essential to distinguish between *conceptual biases* that are caused by the imperfection of the statistical target in relation to an ideal index and *statistical biases* that are caused by the sampling strategy used for estimating the statistical target.

In our discussion, we will avoid the extra complications arising from seasonal products, within-period inflation and from certain types of specialised targets that are sometimes used in owner-occupied housing and some other service areas. In the discussion below the word "period" could either be thought of as a full year in a low-inflation economy or as a week or a month but then for an index only consisting of non-seasonal products.

The present discussion develops ideas that were put forward in Dalén (1992) and Dalén (1999). The paper is divided into three main sections. The main Section, 2, presents and discusses various formulations of the statistical target. In Section 3 we offer some remarks on the kind of sampling strategies that would be appropriate for estimating some of these targets. In Section 4 we discuss current HICP harmonisation issues against this background. An Annex presents a recent proposal on sampling by a Eurostat Task Force, where some of the principles discussed in this paper are given a regulatory formulation.

2. Defining the statistical target

The consumer market ultimately consists of an enormous (but finite!) number of transactions, where goods and services (products) are purchased by consumers. However, it is not feasible to compare transactions directly between periods. Like the physicists who divide matter successively into molecules, atoms and nucleons, we have to bring some structure into our market universe as a prerequisite for a measurement procedure.

We therefore distinguish two fundamental levels of our market universe. In the *upper level*, there is a universe of products and outlets existing in each time period. The basic unit in the universe is a *product-offer* (PO)– a specific product sold in a specific outlet[2]. For successive time periods there will be changes in the composition of the universe. New products and outlets will appear and old ones will disappear. This means that the universe of POs is *dynamic* in that there will be a continual flow of entries into and exits out of it. When the focus is on one period only one could speak of a *static universe* of POs. POs in periods s and t (s<t) are naturally divided into three subsets: the *intersection (matched) set*, consisting of ongoing POs, existing in both periods, the *entry set* of *new* POs (in period t but not s) and the *exit set* of *disappearing* POs (in period s but not t).

In the *lower level* there are sub-universes consisting of *transactions* for each PO in a certain time period. A transaction takes place when a consumer enters into an agreement to purchase a certain volume of a PO at a specified price. For each transaction there is a well-defined price and a certain volume (number of units) purchased at that price. However, there are no dynamics at the transaction level. Unlike POs, transactions have no life; they are one-off events that have no counterparts (matches) in another period.

There is no firm threshold between the upper and the lower aggregation level. An ideal economic definition would try to define the PO as a "good" but economic theory is presently unable to provide us with a clear recipe for this (Pollak, 1996). A minimum requirement is that within a PO substitutability (and price elasticity of demand) must be very high. Some geographic aggregation needs to be included into the PO (e.g. all outlets in a certain shopping centre or market area), similar product varieties (like vanilla and strawberry ice cream or different package sizes for which a price per kg or litre is used). With good quality adjustment methods (like hedonic regression) in place, one may choose a higher level PO (like all PC systems sold in a country) combined with a quality adjusted unit value index (see below), as the statistical target. In practice the setting of this borderline should be an integral part of the process of statistical target setting.

2.1 Upper level targets

Our upper level problem is thus to specify an aggregation principle over changing universes of POs for a range of successive periods *0,1,...,T* to obtain a series of index numbers $I_{01}, I_{02}, ... I_{0T}$. The key difficulty facing index compilers is that the aggregation formula must somehow be based on matched POs, despite the universe dynamics. The formulation of the statistical target involves two distinct sub-problems.

The chain-link specification. We decide on how to divide up the whole range of periods into *k* links so that we have

[2] The term product-offer, coined by Martin Ribe at Statistics Sweden, is central to most but not all of the specifications in common use in price index practice. Outlets could be municipalities for health-care and social protection. A product-offer in electricity or telecom services could be defined as a tariff component in a certain provider's price list. The traditional term used for the same concept has been "item" but this term is vaguer and has been given different meanings in different situations. Moreover, it is not easily translated to other European languages – a major consideration for the HICP.

Link 0: $I_{0,1}, I_{0,2},\ldots, I_{0,t_1}$

Link 1: $I_{t_1,t_1+1}, I_{t_1,t_1+2},\ldots, I_{t_1,t_2}$

…

Link k: $I_{t_k,t_k+1}, I_{t_k,t_k+2},\ldots, I_{t_k,T}$

Note that there is not any distinction between a fixed base and a chain index here. It is only a question of the number of indexes in each link. So called chain indexes with one-year links have twelve consecutive monthly index numbers in each link, whereas so called fixed base indexes would have at least 24.

PO aggregation within the link. Any single link involves the aggregation over POs, where somehow a one-to-one price matching principle has been imposed upon POs in the two non-identical universes, possibly also including imputations and deletions. This aggregation could be formulated as a static or a dynamic estimation problem and we will discuss both of these classes of target formulations below.

To add to the complexity, it is also possible to *intertwine* chain-linking and PO aggregation. For example, we could have the following steps in our upper level target specification:

1. Chain-linking with December reference periods and 12 monthly indexes in each link at the top level.
2. Laspeyres aggregation over, say 200 product groups within each link.
3. Within the product groups, we use monthly chaining in 20 of the product groups, thereby obtaining month-to-month links. In the other 180 groups the chain-link specification from the first step is retained.
4. Within the product 3 defined groups, further aggregation is made.
5. Etc.

This combination of chain-linking and aggregation takes us from single POs all the way up to the All Item Index between two periods.

We introduce the term *index segment* to describe the practice of diversified target setting used in practical index compilation. By this term we understand a particular area of the index where a particular statistical target is used. For example, the European HICP uses Laspeyres type aggregation over a segment consisting of a fixed set of some 90 coicop groups at the highest level. At the elementary aggregate level, another segment, a geometric mean may be used and the telecommunication sub-index would use another particular target of its own and constitute a segment of its own. A segment could cover one or several sub-indexes but it could also define some intermediate stage of aggregation.

The *variables* defined for the POs are of three kinds: i) an (average) price, ii) a purchase volume (or, equivalently, value=price*volume) and iii) a set of characteristics which theoretically defines the user value as well as the production cost of the PO.

The *aggregation formulas* that we will discuss below are variants of superlative and Laspeyres indexes in different chain-linking contexts, where we try to give explicit expressions for the dynamic aspects of the universe. We generally take the view that some variant of a superlative index must be the primary statistical target. The basis for this view could be either a cost-of living argument or a simple symmetry argument (such as noting that a combination of the two pure price indexes Laspereyes and Paasche is needed). Where the information needed to estimate a superlative index is not available, the Laspeyres index or an elementary aggregate formula enters into the picture as a secondary statistical target, whose appropriateness depends on its ability to approximate a superlative index.

2.2 Lower level targets

The lower level target covers aggregation from single transactions to an average PO price over a period of some length. *We will argue that this kind of aggregation is of a fundamentally different nature from that of the upper level aggregation of POs.* Since we do not any longer aggregate different goods but the same goods, neither traditional economic nor axiomatic index theory applies any longer. There are no natural pair-wise relations between these sub-universes in different periods and hence no question of matching, there is just the question of defining the mean price over all transactions in one period. The natural definition is the simple unit value – the volume weighted average price.

There is some controversy on this point. Our view tends to conform to Diewert (1995) and Hawkes and Piotrowski (2000) but not to Balk (1998) and Bradley (2000), who both speak of the bias of the unit value index in relation to a true price index. In our opinion there is no "true price index" defined at the very lowest level, as long as aggregation is over the same good. In each period there is a number (possibly zero) of transactions for a PO. Since transactions are unique in each time period there is no sensible matched comparison possible from a single transaction in one period to another one in the next period.

Of course, to the extent that transactions differ in quality and characteristics within a PO, the simple unit value index will not be unbiased. This fact calls for defining a concept for price change that is able to take quality variation at the transactions level into account – we will call this a *quality adjusted unit value index.*

2.3 Options for dynamic universes

What are the options for dealing with dynamic universes in this context? We will argue that there are three possible approaches that could be combined in various ways:

- To make index links so short that all non-matching POs can be omitted. From each link to the next, the universe is updated. In the link period, a collective overlap method is effectively applied. We have earlier (Dalén, 1999) called this approach an *intersection universe*. For each link we will then want to apply a superlative index formula as our primary statistical target.
- To define an index link so that it specifies one-to-one matching between POs in the two successive universes. These are then either *essentially equivalent* for the consumer so that their prices could be compared without adjustment *or* made *equivalent by* a *quality adjustment* procedure. We have earlier (Dalén, 1999) called this approach a *replacement universe*. Again we will want to apply a superlative index formula here.
- To use a quality adjusted unit value index, which somehow takes the implicit prices of the quality characteristics into account in the index formula. This way of addressing the problem involves a modification of the lowest level calculation in the index. In this context superlative index aggregation is irrelevant. We have earlier (Dalén, 1999) called this approach a *double universe*.

The first two targets are choices for the upper level of the index whereas the third one is applicable to the lower level.

2.4 The intersection universe target with short links

We consider here the combination of the following components:

i) A traditional superlative index formula applied only to those POs that existed in both periods.
ii) Short index links so that the non-intersecting parts of the universes are kept small. Two cases are of special interest: 1) A one-year link with twelve consecutive months sharing the same reference period. 2) A monthly chained index where each month is the base period for the next month.

This target concept, where the superlative index is approximated by a Laspeyres index is what is naturally in index practitioners' mind when specifying the index at higher levels. In each link we have a fixed set of product groups and there is never a question of replacements or additions of groups within a link.

Another application of this target is when POs that cease to exist are deleted instead of replaced and price change computed over the remaining, matching POs. This procedure is often used in connection with annually chained indexes, where the samples are renewed each year.

One particular condition could cause large under-estimating biases in an intersection index, compared to an ideal index. This is when a product group is characterised by *obsolescence* effects in the sense that the consumer value of a certain PO decreases over its lifetime. Examples of such POs are today's newspaper, a newly published novel, typical Christmas food or decorations (which can sometimes be found for sale at low prices in January) or fashion clothing whose market life often ends with sales[3]. For such product groups we must take care to define POs in a manner that eliminates obsolescence effects. One way of doing this is to include "time since market introduction" into the PO specification. Another way is to apply unit value aggregation higher up in the index hierarchy.

We believe that the intersection universe is the natural interpretation of many commonly used index practices, which will likely remain in use for a long time to come.

2.4.1 Monthly chaining and re-sampling

A particular example of the intersection universe target is ***monthly chaining and re-sampling (mcr)***. Here, a superlative index is applied to all POs with non-zero sales in both of two consecutive months. These month-to-month indexes then become links in a longer chain. This target has been proposed for PCs by Turvey (1999). Its attraction is due to the fact that, in a product group where the rate of turnover is high, all POs are included in the target except in their first month in the market. Furthermore, difficult and costly explicit quality adjustments are avoided by relying on the implicit overlap quality evaluation.

Recently, Aizcorbe, Corrado and Doms (2000) have formulated conditions for the mcr method (called the *matched model* method by them) to be a good estimate of an ideal economic index. Under the CES (constant elasticity of substitution) aggregator function, they find that the relationship between the exact price index and the matched-model Törnqvist index can be expressed as follows:

$$I_{t-1,t}^{EXACT} = I_{t-1,t}^{MATCHED} \left(\frac{1 - w_t^{ENTRY}}{1 - w_{t-1}^{EXIT}} \right)^{\frac{1}{\sigma - 1}} \qquad (1)$$

[3] Another way to look at obsolescence in womens' clothing is to note that the assortment is larger in the early season. Later the more popular items get sold out whereas the less popular ones remain to the late season, when they are sold at reduced prices. Also under this interpretation, the average quality of the product mix is lower in the late season, so we have obsolescence.

Here, w_t^{ENTRY} denotes the market share of the new POs in period t and w_{t-1}^{EXIT} the market share of the disappearing POs in period t-1. σ is the (constant) price elasticity of substitution in the product group. From (1), it can be seen that only one of three conditions is necessary for the mcr/matched model index to be a good approximation: i) the market shares of entry and exit products are small in the first and last periods, respectively, of their market lives, ii) the market shares of entry and exit product are about equal in the first and last periods, respectively, of their market lives *or* iii) the price elasticity of demand is large.

There is both a priori reasons and empirical evidence to suggest that within relatively homogenous product groups of durables, such as PC systems, TVs, refrigerators etc. price elasticity is high, thus satisfying the last condition. Furthermore, Aizcorbe et al give empirical evidence in the form of market share data for the PC market that condition i) is fairly well satisfied.

Comparing the mcr with hedonics, they draw the following interesting conclusion[4]:

> ”A comparison of the matched-model indexes compiled using a superlative index number formula with those generated using a hedonic regression technique suggests that the hedonic approach yields noisy and imprecise period-by-period measures of price change.”

The imprecision of the hedonic index would be due to the fact that relevant weighting information (purchasing values) is sometimes not included in the regressions when the hedonic coefficients are estimated.

A problem with (1) that needs more attention is that, as links become shorter the number of links in a chain grow larger. The errors in (1) would therefore grow linearly with the number of links, unless some offsetting factor is at work. This problem requires more study.

Usually all the information for the estimation of a superlative index is not available for applying the mcr method. A secondary statistical target would then be monthly chaining of a simple transitive elementary aggregate index like the ratio of average prices (Dutot formula) or the geometric mean (Jevons formula). A non-transitive index like the Laspeyres index would run the risk of having very large biases in this situation (see, e.g., Dalén, 1998). There are informal indications from small undocumented experiments, which indicate that a transitive unweighted mcr index does not lead to large biases. This hypothesis should obviously be subject to more research.

Turvey (1999) discusses the application of the mcr method (called multi-period overlap by him) to the PC market, where he considers it appropriate. He also believes it to be appropriate for other technological goods, mentioning TVs, digital cameras, hi-fi and other electronic goods.

2.5 Dynamic superlative indexes with adjusted prices

Economists propose the use of imputed reservation prices for products that are sold only in one of the two periods compared. Here we will examine what a superlative index with adjusted prices could look like, drawing heavily on two recent papers by de Haan (2001) and Silver and Heravi (2000).

When modifying a superlative index for a dynamic universe setting, we consider it to be crucial to retain the symmetry properties of the superlative formula. This means that we need to define procedures for dealing with period 0 and t POs that are in some sense mirror reflections of each other. Our approach here will be to provide a definition of a superlative index on the union of the two universes of POs, which will then be composed of three subsets: i) **M**=matched, ongoing POs, existing in both

---

[4] Apparently, this paper is what motivated the famous Alan Greenspan statement that “hedonics are by no means a panacea”.

periods, ii) **N**=new, entering POs, existing in the second period ***t*** only and iii) **D**=disappearing, exiting POs, existing in the first period ***0*** only.

We will look at algebraic form of this index under the Fisher formulation according to de Haan (2001). The notation is the obvious one where *P* stands for prices and *Q* for volumes. The superscripts 0 and t are time periods. In the period in which a PO does not exist the price $\hat{P}$ has to be estimated[5].

Dynamic Fisher: $$I_{0t}^{DF} = \sqrt{\frac{\sum_{j\in M} P_j^t Q_j^0 + \sum_{j\in D} \hat{P}_j^t Q_j^0}{\sum_{j\in M} P_j^0 Q_j^0 + \sum_{j\in D} P_j^0 Q_j^0} \; x \; \frac{\sum_{j\in M} P_j^t Q_j^t + \sum_{j\in N} P_j^t Q_j^t}{\sum_{j\in M} P_j^0 Q_j^t + \sum_{j\in N} \hat{P}_j^0 Q_j^t}} \qquad (2)$$

The Laspeyres variant of the index would be:

Dynamic Laspeyres: $$I_{0t}^{DL} = \frac{\sum_{j\in M} P_j^t Q_j^0 + \sum_{j\in D} \hat{P}_j^t Q_j^0}{\sum_{j\in M} P_j^0 Q_j^0 + \sum_{j\in D} P_j^0 Q_j^0} \qquad (3)$$

Note that whereas the dynamic Fisher index in (2) is able to take account of the new POs, the dynamic Laspeyres index as defined in (3) is not able to do so, since the set of new products is excluded. They would instead have to be brought into the index by chaining and re-sampling.

With an approach by Feenstra (1995), used by Silver and Heravi (2000) for a scanner data set on TVs, the adjusted price is instead brought into the matched part itself.

Feenstra-Fisher: $$I_{0t}^{FF} = \sqrt{\frac{\sum_{j\in M} P_j^t Q_j^0}{\sum_{j\in M} \hat{P}_j^0 Q_j^0} \; x \; \frac{\sum_{j\in M} \hat{P}_j^t Q_j^t}{\sum_{j\in M} P_j^0 Q_j^t}} \qquad (4)$$

Feenstra-Laspeyres: $$I_{0t}^{FL} = \frac{\sum_{j\in M} P_j^t Q_j^0}{\sum_{j\in M} \hat{P}_j^0 Q_j^0} \qquad (5)$$

In this formulation the POs are interpreted as heterogeneous product groups and none of them are directly comparable to POs in another period – all comparisons have to be done through an adjustment which is described below as a quality adjusted unit value index (QUVI) and within which new products can be brought into the index, *if they conform the definition of the product group j.* For example, for TVs Silver and Heravi defines the product group j as a particular make (Philips, Sony etc.) in a particular outlet type (independents, catalogue etc.). Within this product group new models are introduced through hedonic regression which defines $\hat{P}$ as an adjustment to the unit value. For example, for the reference price a linear adjustment would be as follows, where the $\bar{p}_j^0$, $\bar{x}_{cj}^t$ and $\bar{x}_{cj}^0$ are volume weighted average prices and average units of characteristics, respectively and b_{cj}^0 coefficients for the quality characteristics:

[5] This price can also be interpreted as a Hicksian reservation price, if its definition in the many-consumer case is worked out.

$$\hat{P}_j^0 = \bar{p}_j^0 + \sum_{c \in C_0} b_{cj}^0 \left(\bar{x}_{cj}^t - \bar{x}_{cj}^0 \right) \qquad (6)$$

A new make or outlet type, on the other hand, would, in Silver-Heravis case be excluded from a single index link and have to be brought into the index by chaining (re-sampling with overlap).

The advantage of the formulation (4)-(6) is thus that a matched formulation can be kept. The price for this advantage is the exclusion of some new POs from within a link. It could be seen as a formulation, intermediate between the matched models and a fully dynamic index like (2).

2.6 Replacement POs

In this section we attempt to formulate a target concept that "interprets" the one-to-one replacement method, commonly used in price indexes when POs disappear. Still, the basic idea is to have a superlative index with a symmetric treatment of the two periods in the comparison. Therefore, we have to look for a "backward replacement" from the comparison period t for POs that were not sold in the reference period 0 as well as the usual "forward replacement" from the reference period for POs that are not sold in the comparison period t.

Depending on the replacement method used ("most like", "most sold" etc.), a product-offer r_j that replaces j may enter the index on zero, one or many occasions (in addition to it being represented in "its own" period). r_j may further belong to the matched or the unmatched part of the universe. It is therefore difficult, if not impossible, to make a definition which is such that each PO is included according to its true purchasing quantities.

The forward replacement is what would represent a disappearing period 0 PO in period t and the backward replacement is a period 0 PO that would represent a new period t PO. The purpose of a replacement is to make a comparison of prices between the initial PO and its replacement possible, so there are two possibilities. Either we define an *essentially equivalent* PO or one that can be made *equivalent by quality adjustment*. In the first case the observed prices can be related without adjustment but in the second case we have to use a quality adjusted price. In addition we postulate that for the matched part of the universe we shall make price comparisons for *identical* POs. There are only these three possibilities. Unless there is a valid procedure for making the prices of the two POs comparable no price comparison is possible. This means that the definition of the *replacement index* depends on the access to valid quality adjustment procedures. If no such procedures are available, replacements must be limited to essentially equivalent ones and for cases where no such replacements can be found the PO has to be deleted, which takes us back to the intersection universe. So it is also possible to define *hybrid targets* combining equivalent replacement with deletion of non-comparable POs[6].

Ribe (2000) discusses metrics in terms of a dissimilarity function between POs and replacement mappings for defining these one-to-one relations in the universe. He distinguishes between characteristics-preserving replacements and consumer-fit-preserving replacements and draws a number of practical conclusions regarding the choice of replacement strategy.

As for the algebraic representation of replacement based indexes, we could still use equations (2) and (3) from above. However the adjusted prices would now equal the price of the replacement PO denoted P_{rj}, multiplied with an adjustment factor g_j, that in principle adjusts the price of r_j so that a

[6] One such practice, used in Sweden for food and daily necessities, is to allow only package size replacements, with a threshold as to the change in package size (<50%).

consumer becomes indifferent between buying r_j at price $P_{rj}g_j$ and buying j at price P_j. Thus we have $\hat{P}_j^t = P_{rj}^t g_j$ and $\hat{P}_j^0 = P_{rj}^0 g_j$ and obtain:

Replacement Fisher: $$I_{0t}^{RF} = \sqrt{\frac{\sum_{j\in M} P_j^t Q_j^0 + \sum_{j\in D} P_{rj}^t g_j Q_j^0}{\sum_{j\in M} P_j^0 Q_j^0 + \sum_{j\in D} P_j^0 Q_j^0} x \frac{\sum_{j\in M} P_j^t Q_j^t + \sum_{j\in N} P_j^t Q_j^t}{\sum_{j\in M} P_j^0 Q_j^t + \sum_{j\in N} P_{rj}^0 g_j Q_j^t}} \quad (7)$$

and

Replacement Laspeyres: $$I_{0t}^{RL} = \frac{\sum_{j\in M} P_j^t Q_j^0 + \sum_{j\in D} P_{rj}^t g_j Q_j^0}{\sum_{j\in M} P_j^0 Q_j^0 + \sum_{j\in D} P_j^0 Q_j^0} \quad (8)$$

A problem with these indexes is that each PO does not necessarily enter the index according to its own economic importance. This depends on the replacement metric used. In case a similarity metric ("take the most like PO as replacement") is applied, more dissimilar offers will be underrepresented. These may include major improvements (or the opposite) and bias the index, although they are still included at least in "its own" period.

In the replacement Laspeyres index the new subset N is by definition excluded and, especially in an expanding product universe, new POs have no chance to be covered adequately. This is because they are only included as substitutes for disappearing ones from period 0. If a "most like" replacement criterion is used, the ability of this index to represent new POs is very limited indeed.

If a product group is characterised by "natural succession" between disappearing and new POs then this concept may be more attractive. New cars, where new vintages replace old ones are to some extent of this kind but there are also completely new models that are not easily related to old ones. However, this case is more the exception than the rule and this concept does not help much if there are more new than disappearing POs or if there is no natural successor to a disappearing one.

We therefore express our scepticism concerning the possibility to provide the widespread replacement practices in price indexes with a strict interpretation. At least, there appears to be much work remaining in order to underpin these practices with a clear statistical estimation target.

2.7 Unit values

A completely different aggregation principle is given by the unit value index (UVI). The unit value index is not a price index in a theoretical sense and cannot therefore be used for aggregating POs. However for *aggregation over transactions* within what, in an economic sense, could count as "the same product", the UVI is a clearly valid procedure.

The basic definition of the UVI, for a PO, is

$$I_{0t}^{UVI} = \frac{\sum_{k\in h^t} q_k^t p_k^t \Big/ \sum_{k\in h^t} q_k^t}{\sum_{i\in h^0} q_i^0 p_i^0 \Big/ \sum_{i\in h^0} q_i^0} \quad (9)$$

(We use small cap p and q here to denote transaction rather than PO prices and quantities, respectively.)

In this notation we emphasize the fact that the UVI automatically takes care of dynamic changes in the universe, by summing over $\boldsymbol{h^0}$, the set of transactions in period 0, in the denominator and $\boldsymbol{h^t}$, the set of transactions in period t, in the numerator. Since it is not a price index, there is no question of one-to-one matching in the UVI and the $\boldsymbol{h^0}$ and $\boldsymbol{h^t}$ may well contain a different number of items.

The UVI fails and is biased to the extent that the user value ("quality") of the products transacted in $\boldsymbol{h^0}$ and $\boldsymbol{h^t}$ differs. Conversely, within a PO, which is defined so that user value variation within it is small, the UVI is often the best choice for a statistical target.

2.8 Quality adjusted unit values

It is possible to add quality adjustment factors to a UVI to obtain a *quality adjusted unit value index* (*QUVI*). By doing this, we could apply it to more heterogeneous POs than the simple UVI and thus move the threshold between upper and lower level aggregation upwards. We will speak of the QUVI as *aggregating over transactions with different user value*. We believe that a QUVI has to be the statistical target for products, where hedonic regression or another well-defined quality adjustment procedure is used.

There are many possibilities for defining a QUVI and the preferred concept will depend on the type of data and kind of quality adjustment procedure used. Here we will provide two alternative definitions. The first one is from de Haan (2001), who, for a group of two varieties, postulates a quality adjustment factor $g^t_{2/1}$ that serves to 'change the quantity bought of good 2 in period *t* into a quantity of good 1'. He assumes that the representative consumer attains the same level of satisfaction from the consumption of one unit of good 2 as from the consumption of $g^t_{2/1}$ units of good 1. Hence, the consumer is indifferent between consuming q^t_2 units of good 2 and consuming $g^t_{2/1}q^t_2$ units of good 1. An average price in period *t* of goods 1 and 2 can now simply be computed as

$$p^t_{2/1} = \frac{q^t_1 p^t_1 + g^t_{2/1} q^t_2 \tilde{p}^t_2}{q^t_1 + g^t_{2/1} q^t_2} = \frac{q^t_1 p^t_1 + q^t_2 p^t_2}{q^t_1 + g^t_{2/1} q^t_2}, \qquad (10)$$

where $\tilde{p}^t_2 = p^t_2 / g^t_{2/1}$ might be called the quality-adjusted price of good 2; hence, $p^t_{2/1}$ might be referred to as a *quality-adjusted unit value*. At the next higher aggregation level, for a Fisher index from t-1 to t, the prices p^{t-1}_1 and $p^t_{2/1}$ would be used together with the quantities q^{t-1}_1 and $q^t_1 + g^t_{2/1}q^t_2$. For the index going from *t* to $t+1$ the prices $p^t_{2/1}$ and p^{t+1}_2, and the quantities $q^t_1 + g^t_{2/1}q^t_2$ and q^{t+1}_2 would be used.

A generalisation of de Haan's definition to many varieties might be the following: We want to find a quality adjustment (QA) factor g_k that serves to make the quantity purchased in a certain transaction equivalent to another quantity of some numeraire transaction with $\boldsymbol{g=1}$. In microeconomic terminology one might say that the representative consumer attains the same level of satisfaction from the consumption of one unit of the quality adjusted PO as from $\boldsymbol{1/g_k}$ units of the numeraire. We then define:

$$I^{QUVI1}_{0t} = \frac{\sum_{k\in h^t} q^t_k p^t_k \Big/ \sum_{k\in h^t} g_k q^t_k}{\sum_{i\in h^0} q^0_i p^0_i \Big/ \sum_{i\in h^0} g_i q^0_i} \qquad (11)$$

QUVI1 leaves the exact definition of the g_k open. If hedonic regression is used it would be a function of the difference in characteristics between PO k and the numeraire. The QA factors are applied individually to each transaction, both in the numerator and in the denominator. This definition does not, however, come close to any method known to be used in practice. For scanner data it would possibly be applicable to changes in package size.

Another possibility would be to bring coefficients of characteristics from a hedonic regression (or other sources such as option prices) collectively into the index expression instead of into each PO as in (11) . The formula would then depend, among other things, on the functional form of the regression. In the case of a linear model estimated for period 0, we could give our second definition, for QUVI2, as:

$$I_{0t}^{QUVI2} = \frac{\sum_{k \in h^t} q_k^t p_k^t \Big/ \sum_{k \in h^t} q_k^t}{\sum_{i \in h^0} q_i^0 p_i^0 \Big/ \sum_{i \in h^0} q_i^0 + \sum_{c \in C_0} b_c^0 \left(\bar{x}_c^1 - \bar{x}_c^0 \right)} \tag{12}$$

Here b_c^0 is the ("true") regression coefficient for characteristic c, and $\bar{x}_c^1$ ($\bar{x}_c^0$) is the (quantity weighted) average of characteristics per unit in period 1 (0). If a logarithmic regression model were used the correction would be multiplicative rather than additive. This definition applies the quality adjustment collectively to all transactions in period 0 to make them "quality equivalent" to those in period t. This method is used by Silver and Heravi (2000) in a hedonic study on televisions and is also used in actual price index production[7].

Any specification of a QUVI will need a great number of variables and it would be an additional specification problem how many and which these variables should be. Also the functional form of the index is open to many possibilities. Add to this the cost and difficulties with observing all these variables and it becomes clear, why these concepts have not acquired widespread use in practical index work. As described above, Aizcorbe et al (2000) provides theoretical arguments as to why a well constructed mcr (matched model) index can be a good alternative to a hedonic index.

2.9 Other recent developments with implications on the choice of target

In the Canberra meeting of the Ottawa group several papers were presented that in different ways implied new approaches to the setting of statistical targets for low level subindexes. These papers can be found in this volume. Here we give a short description of each one of them.

Sellwood's two-stratum index

Sellwood (2001) proposes the division of each elementary stratum into a static and a dynamic sub-stratum. The static substratum is identical to our matched stratum ***M*** above and for this sub-stratum a Laspeyres subindex is defined. The dynamic sub-stratum consists of both the new (***N***) and the disappearing (***D***) POs (he calls them products) and for this stratum the desired index is a unit value index with new POs in the numerator and disappearing products in the denominator, with an aggregate quality adjustment factor ***G*** to allow for quality change between the disappearing and the new products. He does not define ***G*** further however. Sellwood also gives expressions for the weights to be used for aggregating the two substrata to the elementary stratum index.

---

[7] An index formulation of this kind is used for the Swedish House Construction Index. It is described (in Swedish) in Byggnadsindexkommittén (1971). A similar idea was also used in the Swedish CPI subindex for rents up to the late 1990's. In the housing context this formulation has practical advantages, since q, the purchased quantities, are by definition all equal to one.

One problem with Sellwood's definition is that the dynamic sub-stratum is undefined in cases, where there are either no new or no disappearing POs. Another problem when looking at his definition as a candidate for a target index is that the weight for the dynamic sub-stratum only contains the disappearing POs and not the new ones, which would thus not be taken into account according to their economic importance.

Koskimäki-Vartia's hedonic price functions

Koskimäki and Vartia (2001) give a general formulation of a hedonic model, that brings in all data on prices and characteristics from both periods under comparison into a regression equation but which leaves the exact form of the price function open. The price index is estimated directly from this model by estimating the coefficients of the model and depends on the *standard quality point*, where it is read. Two natural candidates are then the *reference period quality point* defined as an average of the characteristics in the reference period and the *current period quality point* defined as an average of the characteristics in the current period.

The KV index could thus be seen as a competitor to or perhaps as a generalisation of the quality adjusted unit value indexes defined above. A seeming problem with this formulation is that the purchased quantities do not explicitly enter the index. They could be taken into account either through a weighted regression or by including all purchases into the regression directly in their correct proportions. The KV definition leaves the functional form and the set of characteristics to include as an estimation problem rather than a specification problem.

Diewert's analysis of hedonic functional forms

Diewert (2001) makes a number of simplifying assumptions to a traditional economic model that serves to give a microeconomic interpretation to hedonics. The objective of his analysis is to obtain exact hedonic indexes under a specified model. Diewert's final expressions can be seen as variants of equations (2) and (3) above, where explicit meaning has been given to the imputed prices under the hedonic model.

Diewert further compares different hedonic specifications and concludes on theoretical grounds that linear regressions should be avoided, whereas formulations with a logged price would remain in scope. However, after going through a number of specifications of the hedonic equations he concludes that each specification has its advantages and shortcomings and so none of them completely dominates the other.

He further recommends to bring in quantity weights into the hedonic regression. Diewert also notes that traditional superlative indexes based on matched models can give more or less the same answer as a hedonic approach if the market share of the matched models is relatively large.

De Haan's generalised and adjusted Fisher indexes

De Haan (2001) defines what he calls a *generalised Fisher index*, which is identical to our dynamic Fisher index in (2) above. Based on an idea of Balk (2000) he further defines an *adjusted Fisher index* as the Fisher index based on the matched part of the universe, multiplied by the factor $\left(\frac{1-w_t^{ENTRY}}{1-w_{t-1}^{EXIT}}\right)^{\frac{1}{\sigma-1}}$ (the same factor as the one appearing in Aizcorbe et al result in Eq. (1) above).

Under certain assumption this adjustment can provide an estimate of an exact (cost of living) price index.

There are a number of problems with this formulation, though. The elasticity σ is difficult to estimate, the underlying assumption of equal elasticity within the product group is rather strong and the exit and entry weights are not always easy to estimate (although in scanner data they are available).

De Haan also experimented with the adjusted as well as a matched model Fisher index on scanner data for 10 product groups. A further problem that was discovered in this experiment was that the new and disappearing products were difficult to distinguish from seasonal products that need a different treatment.

2.10 Summary points on the statistical target

The analysis presented here could be summarised as follows. A dynamic universe of products and outlets can be represented in a price index in three different ways. We would argue that any serious index design has to involve all of the following three components.

1. Matching of identical product and outlet groups in both time periods in a single index link.
2. Bringing in new products and outlets into a new index link and multiplying consecutive index links into a chain index. Overlaps are then created which serve as implicit adjustments for the quality change between the old and the new universe.
3. Applying a unit value index for aggregating transactions for homogeneous POs or a quality adjusted unit value index for aggregating heterogeneous POs and transactions.

Index design could be seen as the segmentation of the index according to circumstances in different markets and to data availability. In each segment we have to formulate a particular statistical target.

In general, there are no (conceptually) unbiased statistical targets. Under different circumstances, and in relation to an underlying ideal index concept, the bias of a certain statistical target can be analysed. The objective of target setting for a certain segment of the index has to be based on considerations of minimising the conceptual bias as well as on practical considerations of data availability, cost etc.

A crucial choice is where to set the threshold between the aggregation of transactions into POs on one hand and the aggregation of POs into sub-indexes and the All Products Index on the other hand. In practice, this choice again depends on data availability and the availability of advanced methodology for quality adjustment.

Within the upper level of POs, we have to define groups and subgroups in a number of stages. For a highest level, it is always necessary to apply an intersection target within a link by using a fixed product group classification.

For the upper level, where POs are aggregated into a price index, there are only two possibilities for a target.

We could apply an intersection (matched) concept and take only those product groups or POs into account that have sales in both periods. In this case we are within the world of traditional price index theory with a fixed vector of products. We then have to use chain-link definitions that serve to minimise the loss of new POs within the links. When moving from one link to the next, a collective overlap method could be applied. This target has, under different circumstances two opposite risks for bias. For products with obsolescence effects (e.g. ending with sales prices) there is a risk for an under-estimating bias. For high-tech products, there is a risk for an over-estimating bias if new, superior models rapidly gain market shares from old ones.

Or we could apply a replacement universe concept where one-to-one relations are defined between a PO sold in only one period and a “twin” in the other period. The problem with this target, however, is that we cannot accomplish an accurate representation of new (or disappearing) products according to their economic importance. Also, it is a very complex target to specify in detail. We therefore tend to think of replacement practices more in the light of maintaining the size of the old sample. Another and possibly more fruitful way to evaluate replacement practices is to look at them as estimators of a quality adjusted unit value index for a product group. This approach needs more research, though.

Dynamic Laspeyres indexes are by their definition unable to give an adequate representation to new POs, since these were not sold in the reference period and thus would have a volume weight of zero. One-to-one replacements, starting from POs in the reference period, can neither represent an expanding nor a contracting universe of POs adequately.

At the low level we aggregate transactions into (average prices for) POs. In our view traditional index number theory does not apply here. Since there is no natural matching of transactions between the two periods, the only possible option is to use unit values. If POs are not sufficiently homogeneous a quality adjusted unit value index needs to be defined.

The threshold between the upper and lower level of the index is not a priori fixed, neither in the product nor in the geographic/outlet dimension. A quality adjusted unit value index makes it possible to raise this threshold so that a PO is more heterogeneous, for example to the point where a hedonic model could be used (a generic product, POs that share a common set of characteristics).

Considering the cost and complexity of hedonic regression in a month-to-month production environment, a promising target formulation is the combination of an intersection universe with short index links and overlaps at the upper level and unit values at the lower level. A special case of this is the monthly chaining and re-sampling (***mcr***) method.

3 Some observations on classification and sampling strategies

We have already identified the segmentation process as an integral part of the target setting process. Segmentation is closely connected to *classification*, which is the process of organising POs into (cross-classified) groups of products on one hand and geographical areas and outlets on the other. This classification process continues in a number of stages into subgroups, subsubgroups etc. into what we could call final groups. In a sampling context each final group is often a *stratum*, but there is also the possibility of using *cluster* sampling by taking a sample of subgroups instead of all. In practice, some small groups are left out of the index, resulting in under-coverage of the whole universe. The word stratum is therefore inappropriate to use in the stage of target specification and should be reserved for discussions of sampling design.

A *sampling strategy* defines the process of *selecting final groups and POs* for measurement as well as an *estimator*, a formula that aggregates the information obtained in the sample, of the target price index. Selection could be purposive or according to a probability design. It is important to distinguish the index estimator from the index target - they do not necessarily look the same[8]. We want to look at the estimator in a dynamic context, i.e. what does it imply with regard to the dynamic target

[8] For example, if the target is a weighted index but the sample is drawn with probability proportional to the weights, then the appropriate index estimator would be unweighted.

universe. The estimator should of course have a minimal error with respect to the statistical target, under the budget allocated to the work.

The sampling strategy should also define a process for selecting transactions for a PO within the period. This part of the strategy is only important, where there is price variation but there often is. For many product groups, the easy way out has often been to take the price according to a potential transaction taking place at a certain date, for example when a price collector visits the outlet, but this is not necessarily the best solution.

The uppermost product groups are always treated as strata; i.e. there is sampling in all of them. In most countries there are at least some 200 product groups that are treated this way. Within product groups there exist various possibilities for geographical and outlet sampling, including multi-stage sampling etc.

The *elementary aggregate* (ea) is a somewhat strange bird in this context. In today's CPI practice, the domain of an ea is often over upper level POs and lower level transactions at the same time. Take as an example an elementary aggregate defined as "*breakfast cereals sold in independent shops in Scotland"*. Within this aggregate there is a sub-universe of outlets, another sub-universe of product varieties and there are 28-31 days in each month of consideration. The primary statistical target as discussed above would be a Fisher index over product varieties but unit values over days and probably also outlets. But in the sample perhaps only one price per outlet is observed, and the product varieties will differ between outlets.

This fact adds to the complication of establishing appropriate sampling strategies for eas. A principled sampling strategy could be to take a first stage sample of POs (certain brands and package sizes in certain regions) and then a second stage sample of transactions and observe their prices and volumes to estimate the average price of the PO. When the two sampling stages are mixed together the estimation problem becomes ill defined. The best interpretation of many current sampling practices is probably to regard the ea samples as upper level samples of POs whose average price per month (unit value) is estimated by the single price observed at the moment of price collection. Access to scanner data for the whole transactions universe changes this situation. However, it should be realised that in one respect scanner data is also static. This is because new outlets are usually not immediately included in a scanner data set.

The important difference between sampling for a UVI and for a traditional price index is that the UVI sample needs to represent directly the universe in each time period. The UVI sample for each period therefore needs to be drawn directly from a sampling frame representing the universe in the relevant period. Timeliness considerations have traditionally made such a sampling strategy impossible but this situation has changed with the appearance of scanner data.

Estimating a quality adjusted unit value index is more complex since it also involves estimating the adjustment factor or the hedonic coefficients.

3.1 A sampling strategy for the mcr target

We start by quoting the conclusion of Aizcorbe et al:

> "The logical conclusion of our findings is that high frequency data on both prices and quantities of high technology goods should be collected in a single survey instrument. Unfortunately, such survey instruments are rare in official statistics, especially at a high frequency and at the level of detail required … Under both the conventional and welfare-based approaches to price measurement, however, such data are required for the accurate measurement of prices indexes

for high technology goods such as computers and semiconductors, and, by implication, the productivity performance of the aggregate economy."

The distinguishing feature of the mcr method is the necessity to renew the sample every month by entering new models in their very first month in the market. A valid sampling and estimation procedure would therefore have to be defined with respect to all models sold in the market each month. The issue of how to design efficient sampling strategies for this purpose is a subject that has to be given more thought. Here we just want to point out that, since the large part of the cost for local price collection is associated with sending a price collector to an outlet, it is probably cost efficient to take a relatively small sample of outlets but to enumerate all or nearly all models marketed in that outlet each month. For this procedure speaks also the fact that market competition tends to produce similar price movements between outlets for the same product variety.

Although both the product and the outlet dimension are dynamic, the turnover of outlets is of a smaller order and the need for re-sampling of outlets could therefore be made less frequently.

4. Special issues concerning the HICP

Much of the research generating the ideas in this paper has been generated by the need for rules and guidelines for sampling and quality adjustment for the EU Harmonised Index for Consumer Prices (HICP). We would therefore like to present some recent proposals concerning rules for sampling both in its static and dynamic contexts.

But first I would like to make a few remarks on the general aggregation principles in the HICP in the light of the above discussions on targets.

4.1 The statistical target of the HICP

The HICP is by declaration a Laspeyres type index and does not attempt to estimate anything else and therefore the assumption of a superlative index as the primary statistical target would seem irrelevant for its purpose. Yet, it has also required a transitive index (ratio of averages or the geometric mean) to be used at the elementary aggregate level. Since especially the geometric mean is incompatible with the Laspeyres index, we interpret this as a recognition that the Laspeyres index is not the primary statistical target at the lower levels of the index.

Numerous empirical studies demonstrate that at the uppermost level of the index (of 100-200 groups or so) the Laspeyres index approximates a superlative index fairly well (the difference is usually between 0.1 and 0.2), whereas this approximation is much worse at low index levels (differences are on average around 0.5). The economic interpretation of these facts is that the price elasticity of demand is much higher at low index levels. (People are more prone to substitute coffee in another store in the same town for coffee in one's normal store than they are to substitute clothing for food.) For this reason, the geometric mean, which is consistent with a price elasticity of demand that is equal to one, is usually a better estimator of a superlative index than a Laspeyres index, at lower index levels. (In elementary aggregates, where substitution is small, the ratio of averages is more appropriate than the geometric mean.)

A possible interpretation of the present, at first sight contradictory, HICP aggregation rules at upper and lower index levels is that the primary statistical target at both levels is that of a superlative index. Under the existing practical constraints, the prescribed aggregation methods could then be demonstrated to be the most appropriate ones. The Laspeyres index at the higher levels and the geometric mean or ratio of averages at the lower level could then be taken as secondary statistical targets rather than primary ones. This interpretation would have the advantage of adding some form of consistency to the HICP aggregation rules and to act as a guide, when deciding on aggregation rules at intermediate levels, where no explicit rules yet exist.

4.2 Static and dynamic sampling

A Task Force has recently proposed a regulation on minimum standards for the HICP concerning sampling, attempting to address both the static and the dynamic aspects of the problem[9]. The proposal is attached to this paper as Annex 1. The objective of the proposal could be seen as achieving representative samples, for the sake of comparability and unbiasedness of the index. We summarise the key proposals here but refer to the main text of the proposal for the technical details.

- Article 3.1 deals with static representativity against the background of the predominantly purposive sampling practices now used in European countries. It sets thresholds in terms of the maximum part of the universe that is allowed to be excluded from the "sampling frames", that is deliberately not taken into account in the sampling designs. It is recognised that such thresholds need to be set both in the product and in the outlet/geography dimensions, although thresholds for the product dimension need to be sharper due to its greater variability.
- Article 3.2 and 3.3 deals with the two major tools for upholding dynamic representativity. These are one-to-one replacements of disappearing POs and re-sampling of entire product groups. These two methods are put on a par with regard to their appropriateness.
- For replacements (3.2), it is required that the replacing and the replaced POs be equivalent so that prices can be validly compared. Equivalence can either be direct (essentially equivalent) or be obtained by an adjustment procedure (equivalent by quality adjustment).
- Where such replacements are not possible the POs have to be deleted so that price change is effectively computed over the remaining POs. However, there is a threshold to the total rate of such deletions set at 20% of a coicop group. Re-sampling frequencies must be set so that this rate is not normally exceeded.

For re-sampling (3.3) a fairly general requirement is set. However, in no case is it allowed to keep old samples and internal weights for more than five years without updating them to represent current universes. Re-sampling necessarily implies the use of a "collective overlap" method for dealing with quality changes.

5 References

Aizcorbe, A., Corrado, C. and Doms, M. (2000): *Constructing Price and Quantity Indexes for High Technology Goods.* Paper presented at the National Bureau of Economic Research Summer Institute 2000 session on Price Output, and Productivity Measurement. Cambridge, MA, July 31, 2000.

Balk, B. (1998): *On the Use of Unit Value Indices as Consumer Price Subindices.* In Proceedings of the Fourth Meeting of the International Working Group on Price Indices. Bureau of Labor Statistics. Washington D.C.

Balk, B. (2000): *On Curing the CPI's Substitution and New Goods Bias.* Paper presented at the 5th meeting of the Ottawa Group. Reykjavik.

[9] The proposal has been discussed at two different Working Party meetings in Luxembourg, in October 2000 and in February 2001. No decision is yet taken on its implementation and, although there is a widespread concensus around its general principles, its final adoption can be expected to be some time away due to the desirability to integrate it with rules on quality adjustment metods.

Bradley, R. (2000): *Price Index Estimation Using Price Imputation for Unsold Items.* Unpublished manuscript, Bureau of Labor Statistics, December 2000,

Dalén, J. (1992): *Computing Elementary Aggregates in the Swedish Consumer Price Index*. Journal of Official Statistics, Vol. 8, No.2, 129-147.

Dalén, J. (1998): *Experiments with Swedish Scanner Data.* In Proceedings of the Third Meeting of the International Working Group on Price Indices. Statistics Netherlands, Department of Statistical Methods. Voorburg, Netherlands.

Dalén, J. (1999): *On the Statistical Objective of a Laspeyres' Price Index.* In Proceedings of the Fourth Meeting of the International Working Group on Price Indices. Bureau of Labor Statistics. Washington D.C.

Diewert, W. E. (1995): *Axiomatic and Economic Approaches to Elementary Price Indexes.* Dept. of Economics, University of British Columbia, Working Paper Series, 95-01.

Diewert, W. E. (2001): *Comment on The Measurement of Quality-Adjusted Price Changes.* Paper presented at the 6th meeting of the Ottawa Group. Canberra.

Feenstra, R. C. (1995): *Exact Hedonic Price Indexes.* Review of Economics and Statistics, LXXVII, 634-54.

de Haan, J. (2001): *Generalized Fisher Price Indexes and the Use of Scanner Data in the CPI.* Paper presented at the 6th meeting of the Ottawa Group. Canberra.

Hawkes, W.J. and Piotrowski, F.W. (2000): *Using Scanner Data to Improve the Quality of Measurement and the Measurement of Quality in the Consumer Price Index.* ACNielsen Company. Unpublished manuscript, September 16, 2000.

Koskimäki, T. and Vartia, Y. (2001): *Beyond Matched Pairs and Griliches-type Hedonic Methods for Controlling Quality Changes in CPI Sub-indices*. Paper presented at the 6th meeting of the Ottawa Group. Canberra.

Pollak, R. (1996): *The Consumer Price Index: A Research Agenda and Two Proposals.* Washington University in St. Louis, Dept. Of Economics.

Ribe, M. (2000): *Price index, statistical inference and harmonisation.* Paper presented at the Methodological Issues in Official Statistics conference, Stockholm, October 2000.

Sellwood, D. (2001): *Improving Quality Adjustment in Practice*. Paper presented at the 6th meeting of the Ottawa Group. Canberra.

Silver, M. and Heravi, S. (2000): *Scanner Data and the Measurement of Inflation*. Cardiff Business School, Cardiff University.

Silver, M. and Heravi, S. (2001): *Quality Adjustment, Sample Rotation and CPI Practice: An Experiment.* Paper presented at the 6th meeting of the Ottawa Group. Canberra.

Särndal, C.E., Swensson, B. and Wretman, J. (1992): *Model Assisted Survey Sampling.* Springer.

Turvey, R. (1999): *Incorporating New Models into a CPI: PCs as an Example.* In Proceedings of the Measurement of Inflation Conference. Cardiff University, 1999.

Annex 1

Proposal by Task Force V for a Regulation

Concerning the representativity of the HICP sampling

Whereas Council Regulation (EC) No 2494/95 (Article 3 Scope) requires that the HICP shall be based on the prices of goods and services available for purchase in the economic territory of the Member State for the purpose of directly satisfying consumer needs. And whereas Council Regulations (EC) 1687/98 and 1688/98] define the coverage of the HICPs as "Household final monetary consumption expenditure" That is that part of final consumption expenditure that is incurred:

- By households irrespective of nationality or residence status, and
- In monetary transactions, and
- On the economic territory of the Member State, and
- On goods and services that are used for the direct satisfaction of individual needs or wants, and
- In one or both the time periods being compared.

Whereas Article 8.1 of Regulation (EC) No 2494/95 requires that the HICP shall be compiled each month.

Whereas Article 9 of Regulation (EC) No 2494/95 requires that the HICP be a Laspeyres-type index

Whereas Council Regulation (EC) No 2494/95 (Article 3) requires the HICP to be based on the prices of goods and services available for purchase in the economic territory of the Member State for the purpose of directly satisfying consumer needs;

Whereas Article 5(3) of Regulation (EC) No 2494/95 requires further implementing measures for ensuring comparability of HICPs and for maintaining the reliability and relevance of the HICPs; and

Whereas Article 13 of Regulation (EC) No 2494/95 requires that such implementing measures as are necessary should be adopted taking the greatest account of cost-effectiveness;

Whereas Commission Regulation 1749/96 (Art 8 Minimum standards for sampling) requires that HICP target samples shall «have sufficient elementary aggregates to represent the diversity of items within the category' and sufficient prices within each elementary aggregate to represent the variation of price movements in the population».

Whereas Commission Regulation 1749/96 (Art 2) defines

> "target sample"(as the set of prices of goods and services which the Member State plans to obtain for the production of the HICP from January 1997 or at some subsequent date in order to meet the Member States' own or any European standard for reliability and comparability.)
>
> "replacement price" (as an observed price for a good or service which is taken as a direct substitute for a good or service the price of which was in the target sample.)
>
> "sampling" (as any procedure in the construction of the HICP where a subset of the population of prices faced by consumers is used to estimate the price change for some category of the goods and services covered by the HICP) and
>
> "reliability" (as to be assessed according to 'precision' which refers to the scale of sampling errors and 'representativity' which refers to lack of bias).

Article 1
Aims

The aim of this Regulation is to set minimum standards for sampling sufficient to ensure the representativity, reliability and comparability of HICPs

Article 2
Definitions

For the purpose of this Regulation

1. The 'universe', or specifically the ‘transactions universe’, is defined as the set of all transactions comprising household final monetary consumption expenditure and taking place in a given period of time.

2. A **stratum** is any sub-set of transactions in the 'universe'.

3. A **‘product-offer’** is an instance of a specific kind of good or service offered for sale in a specific outlet at a specified time.

4. The **'weight reference universe'** is the 'transactions universe' for the weight reference period of the HICP. The **weight reference period** is a period of one or more years.

5. **The 'sampling frame'** is the confined part of the 'universe' that the Member State represents by the HICP. In addition to actual transactions it may also include potential transactions, that is non-observed transactions for product-offers the prices of which can be observed. The **'sampling frames'** are these three sets:-

i). The **‘product frame’** is the set of products that are available for selection using the sampling procedure defined by the Member State.

ii) The **‘outlet frame’** is the set of outlets that are available for selection using the sampling procedure defined by the Member State.

iii) The **'time frame'** is the set of days or intervals during a month for which prices are available for observation using the sampling procedure defined by the Member State.

6. A **'representative sample'** is a sample of product-offers that reflects the expenditure pattern of transactions in the weight reference universe and from which HICPs for the transactions universe can be estimated with controlled errors [comment: in terms of bias and variance].

7. The 'reference selection procedures' are the procedures for selection of the reference product-offers for price collection in the price reference period. The **price reference period** is a period of one or more months.

8. The 'current selection procedures' are the procedures for the selection of product-offers for price collection for any month after the price reference period.

i) A '**reference product-offer'** is a product-offer the price of which is collected in the price reference period, to be taken into the denominator of the EA index.

ii) An ‘identical’ product-offer is a current product-offer that meets the same product-specification as a given 'reference product-offer'.

iii) An 'essentially equivalent' product-offer is a current offer of a product serving the same purpose without significantly different functionality from that of a given 'reference product-offer''

iv) An **'equivalent by quality adjustment'** product-offer is a current offer the price of which has been adjusted by a quality adjustment procedure to reflect changes in functionality so that it may be regarded as serving the same purpose as a given 'reference product-offer', whereby the remaining difference in functionality between them is equivalent to some difference in price.

9. **'Re-sampling'** is the selecting of a new representative sample for a stratum or a set of strata to represent the universe for a new price reference period.

Article 3
Selection of Sample Prices

1) **Minimum standard for reference selection procedures.** The 'reference selection procedures' shall assure that the sample is representative with respect to the pattern of all actual transactions of 'household final monetary consumption expenditure' in a recent weight reference period. It is required that any distinguishable set of products that accounts for more than 0.1 per cent of this expenditure shall be covered by the HICP through inclusion in the product frame. It is required that any distinguishable set of outlets that accounts for more than 1.0 per cent of the total ‘household final monetary consumption expenditure’ shall be covered by the HICP through inclusion in the outlet frame. Product-offers of a unique character that make a repeated price collection unfeasible are exempt from the requirements.

2) **Minimum standard for current selection procedures.** The ‘current selection procedures’ shall, on a one-to-one basis, select product-offers that are identical, essentially equivalent or equivalent by quality adjustment to the reference product-offers. If this is not possible, the product-offer shall be deleted from the current selection. However, the current selection and re-sampling procedures must be planned so that the number of deletions does not normally exceed 20% of the original sample in a coicop group.

   Member States shall apply explicit consistent rules for the current selection procedures. The rules for current selection procedures used by Member States shall assure that the sample is representative with respect to the consumer purposes of products in the price reference period, also where other products fulfil the same purposes in the current period. Therefore the rules for the current selection procedures shall explicitly make it possible to select product-offers that are similar in use or physically similar, and newly introduced product-offers that are essentially equivalent or equivalent by quality adjustment.

3) **Minimum standard for re-sampling.** Re-sampling shall be made frequently enough to assure that a currently representative sample is used as reference at all times. The minimum standards for 'reference selection' shall apply to re-sampling. In no case shall there be more than 5 years between re-sampling periods for strata within a particular COICOP/HICP class.

Article 4
Comparability

HICPs following the rules of Article 3 of this Regulation, or other rules resulting in an HICP not differing systematically by more than one tenth of one percentage point on the average over one year against the previous year from an index so constructed, shall be deemed comparable.

Article 5
Quality control

Member States shall provide the Commission (Eurostat) at its request with relevant information on the universe represented in sampling practices and of the operational arrangements for achieving and validating representation of that universe by the sample observation and index processes.

Specifically, Member States shall provide the Commission (Eurostat) with information needed for evaluating the precision of the index [subject to an agreed comparable approach and subject to funding in accordance with Article 13 of Council Regulation 2495/96].

Explanatory memorandum

Introduction

1. The Regulation is based on the following premises

   A. Representativity cannot be defined by reference to any given sample but must relate to a specified universe. The sample is the means not the end of representation. Sample procedures, such as 'replacement', must be determined by the representation that they are required to achieve.

   B. The HICP regulatory framework has defined the universe to be represented, all transactions in available products, and the form of representation, a monthly Laspeyres-type index.

   C. Statistical sampling theory provides the conceptual framework for minimum standards of good practice.

   D. Each transaction refers to a specific product in a specific outlet at a specific time. The question of representativity must be translated from the universe of all transactions to the space of all such instances of a product in an outlet at a time. These instances are termed 'product-offers'. Price comparisons refer to corresponding 'product-offers' in two time periods or at two points in time.

   E. The set of 'product-offers' in a certain period may be viewed as the cross-classification or Cartesian product of the set of products and the set of outlets.

   F. The operational procedures for achieving appropriate representation are a matter of subsidiarity, for MSs to determine. However, there must be minimum standards that must be met by such operational practices in order to ensure a representative and reliable index.

2. The Regulatory framework is given in the 'whereas' clauses. It has been rightly argued that the pure price index following the Laspeyres concept is not completely well defined. This Regulation is meant to meet that criticism.

3. The regulation seeks to establish minimum standards for the design of a 'sampling frame' according to product, outlet and time and for the adequacy of current and reference sample selection from that frame. Where current samples cease to represent the current universe of transactions, which includes new models or varieties, then re-sampling is required. Probably the product dimension is most critical for representativity, rather than the outlet or time dimensions, although outlet dynamics should not be forgotten. As regards the time of month at which prices are observed this is a matter for MSs.

4. Further guidance on sampling may be obtained from the Task Force report to the Working Party.

Article 1

i) The Regulation specifies the minimum representation of the universe of transactions required of HICP samples. It does not specify how they should be designed. This is a matter of subsidiarity. The notes on Art 3.1 below therefore give some suggestions on how representativity might be achieved.

ii) Current terminology derives from discussions of CPIs on the basis of sample practices without reference to the universe to be represented. There is therefore some need for new terms. The Regulation seeks to define the options for current selection and rules for ensuring an appropriate representation. The Member State decides the degree of detail of the product-specification and what counts as equivalent to this. Rules relating to quality adjustment and thus constraining the use of the last three definitions have yet to be established.

Article 2

1. The **'universe'** is all monetary transactions that make up the expenditure covered by the HICP. The index is required to provide a description of changes in the prices of these transactions. Failure to properly represent all transactions can be a potential for bias. The form of description, the measure of inflation, shall be a Laspeyres-type index. The following terms elaborate distinctions necessary to define this form of index.

2. A **stratum** is defined as any sub-set of transactions in the 'universe'. The set of all non-overlapping strata make up the full transaction universe. For future use it may be useful to define an Elementary Stratum (ES) as the lowest level of stratification for which explicit expenditure weights are established for the purpose of constructing the HICP. The universe may be 'unambiguously' and exhaustively stratified by COICOP categories (down to sub-index level and below) and by outlet-types (or individual outlets) except to the extent to which these do not include new (future) products or outlet-types but also by regions, sizes of cities, etc. A **stratum** refers to any cell or aggregation of cells in this cross-classification (or perhaps a finer classification which remains exhaustive). The purpose is to recognise that the categories used in practice form the basis of representation of the 'universe' through an aggregation of representation of its parts.

3. **"Product-offer"** is an observable entity that is distinguished from an actual transaction price since actual transactions may not be observed in practice. The term 'product-offer' corresponds to a set of actual or potential transactions, having the identifying dimensions product, outlet, and time. Where there is a reference 'product-offer' and there is a current 'product-offer' differing from the reference only in time then there is an observable 'static' correspondence in the universe.

4. The **'weight reference universe'** is the transactions that took place over any period of time used as the weight reference period, which defines the Laspeyres- type index for the Member State. This period is one year (or more) and thus also covers transactions that occur at certain seasons only (such as winter holidays) whereas the price reference and current periods normally relate to one month. A 'weight reference frame' might be defined as the sampling or observation frame (e.g. the frame used for the household budget survey, which may exclude the transactions of certain individuals, vagrants etc.) used to estimate the HICP weights. However, this is beyond the scope of this Regulation.

5. The **"sampling frames"** distinguishes the practical representational aim of sampling set by the MS, the operationally observable part of the universe from the unknown or difficult to observe. The frame itself is static but should be updated to meet the dynamics of the universe. It may exclude new goods not yet classified, certain times (nights or Sundays) and certain outlets or products of minor importance or difficult to sample (see Art 3.1). **Available for selection** is a technical term that is intended to generalise the notion of a non-zero inclusion probability in random sampling to various forms of multi-stage non-probability sampling, where a price collector makes the final selection. It means, for example, that where "representative items" are specified by the central office, any part of a product group that is not covered by these specifications is by definition not available for selection. On the other hand, any product-offer that falls within the central specification is available for selection by the price collector. The possible need for rules specifying the use of the terms **essentially equivalent** and **quality equivalent** is left for future consideration elsewhere. **The 'Sampling frame'** is that part of the 'universe ' that the Member State intends to repre-

sent directly by the HICP. A sampling frame can best be seen in terms of excluded transactions, e.g internet or mail-order transactions, perhaps also one-off rock concert transactions, and Sunday transactions. The Regulations then imposes a limit on the value of excluded transactions. The 'sampling frame' consists in the universe less certain transaction excluded a priori for practical (or other) reasons. **The 'product frame'** is defined by a priori product exclusions, **the 'outlet frame'** by a priori outlet exclusions and **the 'time frame'** by a priori` observation time exclusions. A priori exclusions are those for which there is no intention to directly represent the excluded transactions. This corresponds in probability sampling to a zero probability of inclusion in the sample. The Regulation seeks to ensure that the sampling frames used adequately represent the universe and that important parts of the universe are not excluded a priori.

6. A **'representative sample'**. A necessary condition for a sample to be representative in a static sense is that the corresponding frames cover those products and outlets which existed in the weight reference period. A first necessary, but not sufficient, condition for representativity in a dynamic sense is that the gradually increasing under-coverage over time is kept below a specified threshold level. To be representative in a full sense, the sample should be representative in both a static sense and a dynamic sense.

7. The **'reference selection procedures'** are those procedures used by the Member State in setting the sampling frame and the actual price observations and is the main determinant of the representation achieved. The selection governs the prices taken as reference and used in the denominator of the Elementary Aggregate formula used to compute the HICP. These prices are also used to revalue weights from the 'weight' to the 'price' reference period.

8. The **'current selection procedures'** are those procedures followed in month to month price observation following the price reference period.

9. **'Re-sampling'** is the process whereby the reference selection is comprehensively updated to represent the product-offers available in the new reference period. The product-offer prices in the re-sampled strata then become new reference samples and follow the rules relating to reference samples. In the sequel the new sample entirely replaces the old sample. There will have to be an **'overlap'** period in which the old sample is used for backward comparisons and the new sample for forward comparisons.

Article 3, general

Article 3 as a whole sets minimum standards for representativity. Point 1 deals with the static aspects in representing the reference period universe of products and outlets. Points 2 and 3 treat the two major approaches to dynamic representativity. These are one-to-one replacement of product-offers (point 2) and re-sampling of coicop groups or parts thereof (point 3).

There is at present no scientific basis for declaring either a replacement-oriented strategy or a re-sampling oriented strategy as the uniformly best choice. Replacements have the advantage of enforcing a deliberate evaluation of the quality change between two successive product-offers. Re-sampling, on the other hand, is the only way to take all changes into account that have taken place in the universe since the previous reference period, but it has the limitation of relying on the overlap method for assessing the quality change between the samples. Member States should be able to provide an explicit account of their rationale for choosing a particular strategy for combining replacement and re-sampling.

Article 3 point 1

i) This rule for 'reference selection' relates to the adequacy of the sampling frame and the sample selection. The reference sample aims at representing the universe in the weight reference period.

This requirement is important to keep in mind, especially when defining representative products. Products purchased in the full weight reference period (such as fresh strawberries or summer package holidays) that are not available in December (the price reference period) are not to be excluded from the sample other than under the 0.1% rule. Note that it is not necessary or even desirable to have the same weight reference period for all subindices and aggregation levels. The most recent reference period possible should be chosen in all cases, especially at low levels, in order to represent recent patterns of consumption. As for the selection of product-offers in an outlet it is usually practical to consider the universe at the point in time that it is first visited for the selection. This practice could be interpreted as having a short recent period (in some cases coinciding with the price reference period) as the weight reference period.

This standard also applies to appropriate representation of products that follow cycles ending with sales, where the time elapsed since the introduction of a new model has to be correctly represented in the initial sample and/or the estimation procedure. Especially in clothing, biases resulting from a failure in this regard can be very large.

The universe includes transactions throughout the month. Although no specific minimum standard is set for the 'time frame', Member States are advised to ensure that distinguishable times of the month that could bias the HICP are represented.

ii) Methods that can be used for achieving representativity are, e.g., probability sampling according to textbook procedures, cut-off sampling or quota sampling. Probability samples should be stratified by subgroups for which expenditure weights are available. Weights for individual products or outlets could be used for PPS sampling. Cluster sampling of geographical areas in which outlets are sampled according to sampling frames, which are tailor-made for CPI purposes is another possibility. Cut-off sampling should only be used where the relevant universe is markedly skewed with respect to weights. A quota sample is defined as a purposively selected sample in which the proportions, with respect to relevant price-determining characteristics, are kept at approximately the same level as in the universe. Quota sampling is one reasonable option, where sampling frames are hard to come by.

iii) Elementary strata relate to distinguishable sets of product-offers. A 'distinguishable set' is a term that can only be given a definition by example: Small cars (from cars), instant coffee (from coffee), or student homes (from all apartments) are good examples in the product dimension and mail-order outlets (from all outlets) and e-commerce outlets (from all outlets) are examples from the outlet dimension. Note that it is permitted to exclude several unrelated sets in separate product groups, which are each below the 0.1% threshold of total household expenditure. It is also permitted to exclude several unrelated outlet categories, which are each below the 1.0% threshold of total household expenditure. Note also that for a particular product group there is no precise requirement on outlet coverage. In spite of that some sub-sets of the transactions universe are excluded from the sampling frame, the sample should nevertheless be representative with respect to the product-offers available in the price reference period and the pattern of actual transactions in the weight reference universe.

iv) The exception provision covers cases like rock concerts and other unique performances, which cannot be followed over time.

v) The higher threshold for outlets is motivated by two factors: a) smaller price and price change variation for the same product between outlets than between products in the same outlet b) higher cost of covering more outlets than more products.

vi) The Regulation does not require probability sampling but it does challenge any complacency on the quality of non-probability sampling. MSs should take specific action to ensure that purposive sampling is representative. It is reasonable to expect those using purposive sampling to run some checks and to provide some evidence that their selection has effective safeguards against bias.

Article 3, point 2

i) By explicit consistent rules are understood rules that are determined in advance and have some generality over a defined set of products, outlets and/or situations.

ii) When an old product-offer can no longer be observed, one-to-one replacements should be planned so that it is possible to obtain a reliable measure of price change between the old and the new product-offer. This means that it must be possible to obtain a numerical estimate of the value of the quality change. One of the available options, primarily for product groups where quality changes slowly, is to make an adjustment according to the criterion «most like» in such a way that it is reasonable to take the old and the new product-offers as being essentially equivalent. It is expected that future regulations will go further in setting minimum standards for quality adjustment, including for such procedures that fall outside the scope of one-to-one replacements.

iii) Deletions should only be seen as a last resort, when nothing else is possible. When there are more than 20% deletions in a coicop group and this situation can be expected to remain, a Member State is expected to take some action, within a reasonable time scale, for bringing this percentage down.

iv) An advisable deletion procedure may be explained by the following example in which the ratio of arithmetic average prices formula is used. The imputed price of product-offer C in period 2 is the product of its price in period 1 with the average price change for the ongoing product-offers A and B from period 1 to 2. Note that with this procedure, the price change of the deleted product-offer C from period 0 to period 1 is retained:

Product-offer	Period 0 price	Period 1 price	Period 2 price	Period 3 price (if a deleted product-offer returns)
A	10	11	12	12
B	11	10	11	11
C	9	13	Deleted, imputed value = 13*(12+11)/(11+10)=14.24	14
Average	10	11.33	12.41	12.33

Article 3, point 3

i) Failure to sample a corresponding product-offer in the current sample will lead to some erosion of the reference sample through time. This rule for 're-sampling' requires MSs to refresh the reference sample where it, and hence the current sample, ceases to be representative of the universe. The aim should be to re-sample for EAs where erosion is most rapid. The need for re-sampling obviously varies between dynamic groups such as electronic products and telecom services on one hand and relatively static areas like municipal services on the other hand.

ii) Specific guidelines on re-sampling frequencies by product group should be set later, if the WP agrees that they are needed. Meanwhile MSs should ensure that the current sample does represent the current ratio of new to old models in the universe. Resampling should be considered whenever new product-offers, which are not in the current sample, become significant. This may be tested occasionally, for example by price collectors observing the full range of models in selected sampled outlets. The practices developed would provide a basis for future guidelines if any.

iii) Although not known to be currently used, it is possible to make some kind of "collective quality adjustments" in connection with resampling. For example, the dummy variable method in hedonic regression could be interpreted in this way. There is nothing in this regulation preventing the use of such explicit quality adjustment methods.

Article 4

The intention here is to take the 0.1% comparability requirement as a rule of thumb for investigating potential sampling bias. It may, for example. apply where a range of products is so clearly very stable over the years that re-sampling less frequently than every five years obviously cannot disturb the result by more than the prescribed amount.

Article 5

This regulation requires MSs to ensure that HICP samples are representative. Compliance will necessarily involve evidence to this effect. The form of evidence is a matter of subsidiarity. There are a large number of quality control variables and other items that could meet the rquirements of compliance monitoring and act as indicators of the effectiveness of sampling practices. For example:

Quantative data

- A list of excluded subsets according to art. 3.1
  This list could be a table containing information on outlet type, product type, share of HFMCE, which other product-offers may represent the same information
- A count of identical, essentially equivalent, equivalent by quality adjustment and deleted product-offers, for selected product groups. One could choose periods of 6 or 12 months after the reference period.
- Re-sampling frequencies in different groups.
- Implicit Quality indices
- Precision

Qualitative data.

- Report of market analyses for specific products
- A rolling programme of sample appraisals
- Systematic comparison of the product-offers included by other MSs

Precision: Although sampling errors have been estimated in some Member States, there is not yet an agreed procedure for their computation. Also, current evidence suggests that they are generally smaller than possible bias due to misrepresentation. But their size is by no means negligible and their estimation provides great opportunities for improving allocations of the sample in addition to their properties of being control statistics for reliability and comparability. The computation of sampling errors is therefore required.

The following rule could be seen as a «soft guideline» in this respect:

> "Tentative idea for a minimum standard of precision. The precision of a 12-month change of the All Items Index, expressed as a standard deviation, shall be kept below 0.2 percentage points. The precision shall be estimated according to a documented procedure, that takes full account of stratification and other features of the sampling design, and that also takes account of non-random selection uncertainty as the uncertainty of a corresponding random procedure."

A «rule» of this kind would however be premature for harmonisation use, since the theory and methodology required for its implementation is not sufficiently developed. Further research activities by Member States are required in order to establish an adequate conceptual framework for producing comparable and adequate estimates of precision.