

Constraining unwarranted variations in sampling and quality adjustment practices in the construction of CPI elementary aggregates

Paper for the Ottawa Group Conference Washington April 1998
by Don Sellwood

Abstract

This paper argues that CPIs are not defined by broad concepts or principles but by the practices that have evolved to resolve the many measurement problems faced in their construction, in particular sampling and quality adjustment practices. Economic theories of inflation measurement have contributed little to the actual practical decision processes of index construction and practitioners have failed to develop more than rudimentary conceptual frameworks. As a result a variety of practices are followed without warrant and without any assessment of their effect. A closer association between theory and practice is advocated so that the theory is sufficient to choose between the available practical measures without making unnecessary assumptions about consumer behaviour. The impact on CPIs of currently available procedures should be examined and where differences are found the arguments for each choice should be examined by reference to concepts that bear upon that choice. An elaboration of the Laspeyres concept in relation to the universe of prices to be represented is advocated as a guide to the construction of elementary aggregate indices the formula for which embraces the essential procedural options for treating changes in quality.

Keywords: Price index, universe represented, quality adjustment, standard reference index.

Introduction

Work within the European Community to harmonize the construction of CPIs has shown that major differences arise not because of differences between the concepts followed but because of differences in the actual sampling and quality adjustment practices. All European CPIs aim to measure broadly the same thing. The central concept is the Laspeyres index, which implies a measure unaffected by changes in the quantities or qualities of goods bought or consumed. All CPIs hold quantities constant and aim to measure changes in price after allowing for any change in the quality of goods and services priced. The indices are all formed as a combination of a fixed set of item groups (within a link of one year or a number of years) and a dynamic universe within each of these groups. However, existing definitions of the measurement objective for the CPI are insufficient to determine unambiguously actual practice in any given context. Various procedures have therefore been developed in order to effect the actual construction of CPIs. It is these procedures, or rather the actual practices to which they lead, that determine what the indices in fact measure.

This paper therefore examines the existing practices for sampling and quality adjustment and considers what justification can be made for these. It seeks to develop a conceptual framework sufficient to provide consistency of practice both within and between CPIs. That is to elaborate the broad index concepts currently guiding CPI construction only as far as is necessary to provide a clear basis for preference among the available procedural options. If approved by Eurostat's Working Party on the harmonization of consumer price indices rules for sampling and quality adjustment will be followed not only by existing Member States of the Community but also by other European States that are seeking membership.

Sampling of Prices

Sampling in essence involves the same basic actions for all CPIs. There is an initial selection of a sample of items for price observation and recording. This is followed by either repeat observation of the prices of items matching in specification those of the initial selection, or selection of replacement items where matching fails. Periodically there is re-sampling, a selection of a new sample. As a result, for any given month-to-month period, the prices recorded depict three basic situations that must be handled in the index compilation:

- a) Matched items with prices for all, or nearly all, months and with a one-to-one correspondence of specification from month to month
- b) Non-matched items with one-to-one correspondence but with changed specifications between the initial sample items and their replacements. The observations may overlap for one or more months or there may be observations missing before the replacement.
- c) Re-sampled items with many-to-many correspondence, changed specifications and one month overlap between the initial sample and the new sample.

Combination of Prices

An elementary aggregate may be seen as comprising three strata depicting the three cases above. Schultz has shown that the way in which these situations are handled can all be expressed by a modified elementary aggregate formula (see note 1). Thus the essence of the procedures may be written as a function of the prices of the three strata with three quality adjustment factors that are in turn functions of the observed prices and/or direct valuations of quality changes. In the case of the “ratio of average prices” formula the process may be characterised as follows:

$$EAI = \frac{\frac{1}{m} \sum g_m P_t + \frac{1}{n} \sum g_n P_t + g_r \cdot \frac{1}{r} \sum P_t}{\frac{1}{m} \sum P_{t-1} + \frac{1}{n} \sum P_{t-1} + \frac{1}{r} \sum P_{t-1}} \quad (1)$$

Where there are m matched prices, n non-matched prices and r' re-sampled prices replacing r initial prices. The factors g are allowances for quality change. For brevity the individual items are not suffixed, the sums are over i=1 to m etc. The factor g_r covers additions and deletions for which there are no direct replacements and will in practice be complex. **Note** that the formula is an over-simplification and is not to be taken literally. In particular re-sampling, where there is linking, can only be described by reference to three months. Nevertheless, an effective quality adjustment can always be obtained from the ratio of the change over a given period in the actual index computed to the corresponding change in an index computed using the prices as if they were all matched, i.e. assuming no change in quality.

Formula 1 is intended to highlight and make explicit the effects of sampling and adjustment procedures that are not shown in the usual formulae, which take only matched prices. The procedures are not neutral with respect to the extent to which actual price and quality changes impact on the CPI. The factors g_m represent the change in quality of the matched items.

These are normally taken as unity assuming there to be no change in quality. The g_n 's are implicit or explicit adjustments made when there are replacements. The effective allowance for the average change in quality with re-sampling, which is usually ignored, is captured by the factor g_r . What a CPI measures is determined by the prices recorded and the parameters of formula 1. The parameters are determined by the rules actually followed in obtaining prices and in adjusting those prices. For a given universe of prices CPIs will differ only in so far as the parameters differ. The parameters may differ if there is a difference in measurement objective but for a given measurement objective CPIs may differ in parametric detail but should not differ in overall result. The same elementary aggregate index may be achieved by different procedural arrangements. Any difference in measurement objective should be discernible from the reasoning used to justify the choices of procedure. These are examined below.

Treatment of Matched Items

For each item there are two prices for $t-1$ and t . In some cases the specification will differ from one month to the next but the assumption here is that there is no change in quality. The balance between the number of matched and non-matched prices will depend on the actual decisions on what changes in specification can be discounted as not affecting quality. There is clearly scope for variation in practice and effect both within and between CPIs. It is not usual to test the extent to which g_m might depart from unity because quality changes have in effect been ignored.

The balance between matched and non-matched items in the sample will depend on the kind of description used for an item. Loosely specified items (e.g. most popular shirt) are easily matched from period to period and would not be likely to require replacing. Tightly specified items (e.g. brand X shirt made of material M) would be more likely to involve replacement in time. There will always be some rule, which determines the level of matching required, by which price collectors (or maybe the computer system) decide whether or not a currently available item is of “comparable” quality to a previously priced item. Until studies are undertaken to establish what effect the actual rules followed have on the balance of matched/non-matched judgements the choice of rule remains arbitrary.

Treatment of Non-matched Items

For each sample unit in this stratum there are prices for a pair of items serving a similar consumer purpose but with distinct specifications, there is one-to-one correspondence but no strict match. The prices may relate to the same period of time (overlap) or to two periods, one month, or occasionally more, apart. In all cases there is a difference between two observed prices that must be apportioned between what counts as inflation and what counts as quality change.

There are two main kinds of options for treating a replacement in the case of non-matched pairs:

- i) “External” quality adjustment applied, this covers explicit procedures such as quantity augmentation, option cost, hedonics, producer cost, expert guess, other guess. The quality adjustment factor g may be applied to either the price of the replacement or to the last price or the reference price of the replaced item. (g may be an overlap price not contained within the system).
- ii) “Internal” procedure followed, this covers any implicit adjustment procedure applied by program or manually using the price of the replacement and any previously recorded prices within the system. It includes imputation of a price change based on the price of the replaced item or of a reference price, overlap prices and link to show no change. These procedures can all be expressed in terms of an adjustment g that is a function of the difference between the last actual observed price of a particular item and the price of its replacement. They all thus assume that this observed price difference is an indication of quality difference.

Treatment of Re-sampled Items

In this stratum there are prices for two sets of items for the month where a planned re-sampling occurs. The two sets of items serve similar consumer purposes (that is, they are in the same category of good or service) but their detailed specifications will differ. It is not normal practice to consider the difference in either the average price or average quality between the replaced sample and the replacement sample. There is not, therefore, usually an explicit assumption that $g_{r,t} = 1$ nor is any warrant given for such an assumption and $g_{r,t}$ is invariably taken as unity in practice. Where re-sampling is on an annual or longer-term basis it may be argued that separate estimates of inflation before and after the re-sampling are justified. However, such a justification should be set within a broader framework of theory for all CPI sampling practices.

Selection strategies

There is currently only a limited amount of probability sampling for CPIs. Purposive sample selection either centrally determined or purposive selection by price collector are most commonly used. [It has yet to be established whether these produce different results in practice] Probability sampling normally requires a definition of the universe whereas purposive sampling seems to have imposed no such constraints on actual practice. It is possible, therefore that some CPI samples may, by their design, exclude certain products, e.g. fashion items, the prices of which may be important for appropriate representativity. However, if the population or universe from which prices are to be drawn for the construction of the indices is not defined, then it cannot be said what the CPI sample is required to represent. This is not standard statistical practice although it seems to be common practice.

The Requirements of Theory to Support Actual Practice

The foregoing assertion is that CPIs differ on account of unwarranted differences in the detail of actual practices of sampling and quality adjustment followed in their construction. Studies in the EU indicate considerable differences in the rates of failure to match prices and in the procedures followed when failures occur. Sample selection and re-selection strategies also vary. Any effect will be systematic across sub-indices and it can reasonably be concluded that aggregate differences well in excess of 0.1% on annual rates of inflation between the resulting CPIs will exist on these counts. **There do not appear to be any fundamental differences in the measurement objectives nor in the rationalisations for existing strategies that can justify such differences in results.** Nor does it seem likely that differences between the products and institutions of consumer markets would justify such differences between CPIs. Whilst in theory consumers may make different evaluations of quality both within and between CPIs the methods for treating quality show no systematic recognition of such effects.

All CPIs assume that in certain situations observed price difference is an indication of quality difference but there appears to be no clear rules as to when the assumption is justified and hence little consistency in practice. While it must be accepted that differences in quality between two products will frequently account for a difference between their prices this is not, per se, a justification for assuming that an observed difference in price is wholly or partly a measure of the difference in quality. There is clearly a need for a systematic justification as to

when all or part of an observed price difference can be taken as a measure of quality difference.

The requirement is, therefore, for a measurement objective that will provide a warrant for procedural choice and ensure consistency of application. Agreement on a common measurement objective would limit the diversity of practice and thus reduce the non-comparability of results. For the EU the aim is to find a common objective that is compatible with the majority of current practices and sufficiently differentiating to indicate preferred practices where current differences lead to non-comparability. The requirement is not for a completely new objective but for one that is sufficient to determine what CPI samples should represent and how that representation should be achieved. The objective should lead to standard procedures for quality adjustment and for sampling. The objective involves a statement of the universe of goods and services the prices of which are to be represented and the mode of their representation. This includes a statement of what should and should not count as quality change in the universe and how it is to be represented and treated in the sample.

Possible Measurement Objectives – Universes to be Represented.

None of the common index concepts taken for CPIs, Laspeyres, “fixed basket” “pure price index”, “true price index” or “cost of living”, define a sampling universe. Loosely the Laspeyres index might be said to take “the universe of all prices paid by consumers in a given reference period”. In practice there is sometimes the implication that it comprises only prices matched between the base and current periods but seldom is such a continuously reducing universe used for long periods. Indeed, the notion of “matched prices” is not a measurement objective in itself but is only a device for capturing the fixed “quality” aspect of the Laspeyres concept. This is not to say that a matched sample cannot yield the best, or an adequate estimator for a CPI covering a non-matched universe.

The point of taking samples is to make inferences about a particular population or universe. Therefore the measurement objectives of national CPIs, if not explicitly stated, might be expected to be implied by the sampling schemas used. The universe to be represented in any CPI sample must be all or a sub-set of the universe consisting of all base period prices and all consumer prices for all subsequent periods. The main universes that might be inferred from the present European CPI sampling strategies [with the sampling practice given in brackets] are along the following lines:

- Base and current period prices that are matched i.e. excluding base and current period prices for which there are no matches, [In practice the base period samples are often reduced when no match is found, re-sampling may be ad hoc or only every few years]
- All base period prices and current matched prices plus the prices of the nearest substitutes when there are no matched prices in the current period, i.e. excluding current models where models more similar to those of the base period exist. [In practice the target sample is maintained by replacement with similar item when match not found, re-sampling only after a number of years]

- All base period prices and matched current prices together with prices of all similar goods and services in the current period, i.e. excluding only genuinely new goods. [In practice the target sample is maintained by replacement and/or re-sampling, which is purposive sampling of all possible replacements not just those most comparable]
- All base prices and current matched prices and all other prices including those for genuinely new goods and services. [In practice the target sample is maintained by replacing when no match is found and/or by comprehensive re-sampling to keep track of new models and genuinely new goods and services]
(See note 2)

Choosing between Objectives (Universes)

The choice of measurement objective or universe to be represented only matters (makes a difference to the result) where it makes a difference to actual practice through the choice between alternative procedures. However, it is instructive to consider whether there are reasons for preferring any particular universe. The EU has reached some agreements concerning the reliability, relevance and comparability of CPIs (HICPs). New goods and services should be brought into the index and weights should be updated where this is critical to the reliability of an index. Thus the universe for European CPIs must embrace all currently available varieties of consumer goods and services. This points to the latter wider universes above. Against this there is no wish to increase the costs of producing CPIs unless this can be justified and many existing CPIs correspond more closely to the earlier narrow universes. Thus while HICPs should provide a measure of inflation in currently available goods it is accepted that this should not necessarily require that the prices of all such goods be included in the HICP samples.

The choice of universe might also depend on context, in particular, according to the category of goods or services covered by an elementary aggregate. Categories where there are continually changing varieties or rapidly changing technologies may require different procedural strategies from categories where there are relatively static specifications.

Elementary Aggregates and the Choice of Universe

The central questions to be answered concern the practices that determine the parameters of formula 1 for the construction of elementary aggregate indices. What limits if any should be placed on:

- a) the sample balance of matched, non-matched, and re-sampled items?
- b) the adjustments for replacements where price difference can be taken as equal to quality difference?
- c) the procedures where part of a price difference is discounted as quality difference?
- d) direct quality adjustments based on information external to the regular sample data?

- e) re -sampling where the change in average price not fully accounted explained by the change in average quality.

The proposition of this paper is that these questions should be answerable, at least in part, by reference to a price index defined on a stated universe. Given that CPIs in the EU are Laspeyres type indices the point of reference proposed is a Universe Laspeyres Index or ULI (for want of a better title). Clearly there are other possibilities based on the Fisher and other formulae but these raise different issues. It is the considerable disparity between CPIs on account of sampling and quality adjustment procedures that is of immediate concern. Where the questions are not answered by reference to the ULI then further conceptual elaboration will be necessary.

The ULI is still to be agreed but whatever the chosen universe this index should provide a guide as to what must be represented in the CPI sample. It is likely that the chosen universe will be a changing universe and the ULI will therefore be an index of changes in the average (with fixed quantity weights) level of prices in that universe after allowing for changes in the average quality of the universe. The CPI is required to measure what is happening in the universe not what is happening in the sample. The concept of the ULI must not be remote from the operational decisions involved in CPI construction. It should provide clear guidelines on what it is in the universe that sampling and quality adjustment procedures are required to represent. The sample must track the changes in that universe and the procedures must make appropriate allowance for changing quality. Or, failing that, clearly inappropriate allowances should be avoided.

Full elaboration of the ULI might be justified on the grounds of correcting a serious oversight in CPI methodology. However, before devoting much time and resource to such a project it is important to consider whether it is likely to provide the sought after benefits. If, for simplicity, the problems of allowing for changing quantities are ignored the ULI may be written as a ratio of the universe average price index and the universe average quality index. That is $ULI = UAPI/UAQI$. This might be obtained by estimating the UAPI and the UAQI separately or by estimating their ratio directly. CPIs normally follow the latter course but the former course provides a different perspective that may be helpful.

The concept of a “Standard Reference Index” (SRI) (see note 3) has been introduced in EU discussions as an indication of the effect of procedural choices at either the sub-index or overall index levels. At the elementary aggregate level the SRI is given by setting the adjustments (g 's) in formula 1 equal to one. That is the SRI is the simple ratio of average prices. The ratio of the SRI to the index for the elementary aggregate is the average quality adjustment for the aggregate or the Implicit Quality Index (IQI). That is $EAI = SRI/IQI$. In terms of the universe, if the SRI of the sample were an unbiased estimate of the change in the average prices in the universe (UAPI). Then the sample IQI would have to be an unbiased estimate of the change in the average quality in the universe (UAQI) in order for the EAI to be an unbiased estimate of the change in price in the universe. Alternatively any bias in the IQI would need to be balanced by an equal and opposite bias in the SRI.

The SRI and the IQI are useful constructs for exploring the adequacy of the CPI. Whatever the process whereby the changes in quality are estimated the results should stand up to scrutiny. Where an IQI is computed changes in quality can be compared between categories

and between countries. Critics of the CPI will not be able to say that no allowances have been made for quality and the credibility of adjustments can be tested. (See note 4)

Conclusions

Nothing in what index compilers say they are trying to measure justifies the apparent diversity of sampling and quality adjustment practice. Little of what economic theory has to say about what CPIs should measure has a constructive bearing on the procedural choices to be made. There is, therefore, a need for a theory that will inform procedural choice. This need may be met by following the requirements of statistical theory to identify a “population” or “universe” objective for measurement.

The proposition to be put to Eurostat’s Working Party runs as follows:

- In order to determine what the CPI sample must represent it is necessary to define a Laspeyres-type universe.
- The appropriate universe is a universe covering both the reference period and later periods.
- The required CPI is then an estimator of a Laspeyres-type index of the change in the average prices for the universe after allowing for the change in the average quality of goods and services in that universe.
- The CPI can be estimated from an initial sample that is replicated from month to month and updated to represent the changing universe by either re-sampling or replacement.
- Re-sampling, a planned re-drawing of part, or all, of the target sample, must be done in such a way that the difference in average prices between successive samples reflect differences in average quality. If this is not the case any imbalance between price difference and quality difference must be taken into account.
- Replacement always involves a specific adjustment for quality either by taking an external estimate of quality change or by following one of the agreed procedures based on available prices.
- Standard Reference Procedures (SRPs) or procedures giving comparable results must be followed for replacements where external estimates of quality change are not available. SRPs should be set according to context, in particular according to category:
 - Where quality is changing rapidly
 - Where quality is changing slowly
- External or direct quality adjustments should be separately recorded and aggregated included in a central database for use by other Member States.

- Standard Reference Indices (SRIs) and IQIs should be computed in the short-term to assess how and where procedural choices affect results and, in the longer term when those choices are limited, as a check on the uses of recommended procedures.

Many sampling issues, such as sampling of outlets, and quality issues, such as increasing variety, have been ignored in this paper in an attempt to focus on a central problem that is believed to be solvable, at least in part. The aim is for short-term improvement in CPIs but the hope is that the approach will provide a useful perspective for tackling those other issues.

Acknowledgements

The ideas in this paper owe much to discussions with Jörgen Dalén and to discussions both within Eurostat and with delegates of Member States engaged in the programme to harmonize CPIs.

Notes

1. In his paper for the Stockholm meeting Bohdan Schultz gave a comprehensive algebraic description of the common procedural adjustment, Jörgen Dalén gives a brief summary. See also CBS paper no 9719 by Martin Boon and Jan de Han.
2. Jörgen Dalén develops these universes in his paper
3. Jörgen Dalén has carried out a number of empirical studies. Work by Robin Lowe for the Stockholm meeting provided a stimulus for this concept.
4. Various studies have shown the value of calculating the quality changes that are implied by CPIs. Moulton and Moses (Brookings 1997), Lowe (Stockholm), Hoven (CBS Paper for Eurostat) Dalén (forthcoming).