

Constraining macro effects of the use of different methods at the basic level in the project to harmonise consumer price indices in the European Union

by Don Sellwood - EUROSTAT

INTRODUCTION

1. The problem of the formula to be used in micro-level aggregation exemplifies a more general question of how to tackle biases in consumer price indices arising from what may be termed basic operational decisions. That is the choices that index compilers have made as to the concepts, methods and practices (hereinafter practices) to be followed in constructing actual national CPIs. This discussion paper explores the problem in the context of international comparisons of inflation and the project to harmonise consumer price indices within the European Union. It attempts to draw on lessons learned in the debate as to which formula should be used in micro-level aggregation to develop a strategy for removing other differences between practices in order to produce CPIs which are comparable one with another.

THE PROBLEM

2. The question of the formula for micro-level aggregation has for many years defied resolution. It is characteristic of the problem of bias which faces not only the compiler of a national Consumer Price Index (CPI) but also anyone addressing the task of producing any but the simplest of statistical measures. The debate on which formula to use has become increasingly complex as more empirical studies and theoretical comment have been added but the central problem can be illustrated by reference to the two most commonly used formulae.

A	The average of relative prices	$\frac{1}{N} \sum \frac{P_{it}}{P_{oi}}$
R	The ratio of mean prices	$\frac{\frac{1}{N} \sum P_{it}}{\frac{1}{N} \sum P_{oi}}$

The problem is how to reach an agreement on which formula to use given that there are arguments in favour of each and no overriding case can be made for either.

3. It is generally accepted that A will give a higher measured rate of price change than R. It is not, however, universally accepted that one formula is right and the other wrong. The choice between A and R is a choice that requires some higher level of concept or generalisation, beyond what appears appropriate or acceptable at the micro-level. An agreed point of reference from which to the alternatives. If one takes a sufficiently narrow view, the fact that R gives higher weight to changes in the higher prices seems to be a conclusive argument for A at least in the situation where prices in the aggregate are widely dispersed. Conventionally a wider view is obtained by reference to what is to be estimated by A or R. Then by applying some appropriate criteria or empirical evidence one can decide which gives the better estimate. However, this strategy has not proved successful.

4. It may, for example, be a Laspeyres CPI or, alternatively, a Fisher index that is required. Then deduction using Fisher's tests produces an argument against A in the latter case. However, it may be used as an argument in favour of A in the former case since the Laspeyres index also fails the test. Appeal to empirical evidence to decide which formula provides the better estimate requires an estimate of bias in each case. However it is not possible to compute the required index without first deciding the micro-level formula. The problem would therefore seem to be to clarify what it is that is to be measured. That is to agree on a precise index concept, e.g. a Laspeyres index to be interpreted in a particular way in specific situations. But what are these index concepts in practice and how are they formed?

The Index Concept

5. A number of index compilers would maintain that they follow the concept of a "pure price index". This might be interpreted in practice as the strict adherence to a particular Laspeyres type formula, either fixed or chained. Consistency of practice is built through time as particular problems are addressed and resolved by re-interpreting the requirements of the formula. Thus, for example, the meaning of a "fixed basket of goods and services" has to be interpreted in a situation, say, where outlets change. If the solution is "outlets should, in principle, be fixed" then practices are set

according to an appropriate procedure. Thus precedents are created in practical necessities and these serve to define inflation in particular contexts and become part of the "index concept". They cannot then be readily questioned and must be followed in subsequent decisions on the grounds of maintaining consistency. They thus become "givens" in later debates over issues such as whether to change the micro-level aggregation formula.

6. One concept which is supported by a great deal of theoretical work is the "cost-of-living" concept. This seems to have found favour among generations of economists and to have the advantage of providing a framework to ensure coherence and consistency in the resolution of CPI problems. Why then it is not accepted as the basis of reference for more than a few national CPIs? Much of the opposition to this approach, though not always well informed, is well entrenched particularly among those who justify other approaches. Whilst few may argue over the theory many question the assumptions about consumer behaviour on which it is based. Perhaps then a better theory of consumer behaviour is needed to persuade doubters that the cost-of-living concept has a better claim to the status of reference than alternatives.

7. Though there has been a good deal of research into consumer behaviour there is little agreement on explanatory theory. The classical model of the consumer maximising utility by applying a rational decision process to balance costs and benefits seems remote from common experience. One recent alternative, image theory*, suggests that consumer choices are made so as to promote and protect the goals and values of the individual as part of a continuously changing set of perceptions of self reformulated in the light of experience. This captures the apparent concurrent demand by consumers for both constancy and novelty and seems to imply a different CPI construction from any of those currently in use. (Is a situation where all prices remain constant but where consumers' values are changing to be regarded as inflationary?) It may have some way to go before it is demonstrably a better theory than that accepted by supporters of the cost-of-living approach. It is unlikely to encourage them to actively question their assumptions. Nor does it seem to offer specific answers to such problems as how to treat changes of outlets.

* See Lee Roy Beach. *Image theory - decision making in personal and organisational contexts*— John Wiley ISBN 0 471 923 77X which includes references to recent research.

8. To say that one is trying to estimate a Laspeyres, a Fisher or cost-of-living index is thus not sufficient to define an index concept. It does not determine what to do in practical circumstances. Index concepts have in practice been defined by such a choices of formula together with the sum total of micro-level decisions which go to make up the indices as they are actually produced. The concepts are reconstructed as new decisions are taken to tackle new consumer experience and in response to developments in concepts and theories in related areas. To seek agreement on the concept is to seek agreement on the practices to be followed in compiling a CPI. How then are agreements to be reached on the many differences which exist between CPIs as they are presently constructed?

9. Those responsible for CPIs are not readily persuaded that what others do is to be preferred. This is a prejudice which helps to maintain the integrity of national CPIs but at the same time constrains efforts to change them. Because of the political and public sensitivity of the CPI compilers have built up arguments to defend current practices. Alternative constructions are judged by these arguments which through time have gained the status of beliefs. The situation is further constrained in that for many CPIs change may be inhibited by law or by the vested interest of certain users. The consequences of admitting that "the CPI has been wrong " do not bear thinking about..

10. The question of the formula for micro-level aggregation is in some ways one of the more straightforward of the differences between CPIs. There has been some twenty years of work on the effects yet it is still unresolved. It is clear that some index compilers firmly believe that A is better than R while others believe the opposite. In other areas beliefs are stronger and the evidence weaker. Many compilers, for example, believe (act as if it were true) that "representative sampling of prices is without significant systematic bias" though none have any proof to that effect. Suggestions for probability sampling are therefore rejected as unnecessary and costly. In this respect index compilers are no more prejudiced than those who blindly advocate probability sampling or anyone else. Alternatives are always judged against accumulated current experience. That is both a strength and a weakness of human nature. The task, in the formula question, is to change the balance of argument in favour of one option in the absence of an overwhelming case for any particular formula.

11. The strongest argument against changing a particular CPI is that it serves the purposes for which it was designed. Thus while most compilers would accept the weakness of their defence of specific practices they see insufficient reason for change given the reaction this would provoke from users. The main argument for change, as exemplified by the formula question, is that different practices give different results and this cannot be justified. Added to this, for the European Union, is the growing need for better international comparisons.

The Eurostat Approach

12. The approach being followed in the project to harmonise consumer price indices in the European Union is firstly to demonstrate that different practices can have important numerical effects on the CPI (if this is indeed the case) and that alternative practices have at least an equal claim to validity. The acceptance of the need for comparable indices has provided the necessary impetus for the search for a consensus on the best or most appropriate of available concepts, methods and practices. These may include those followed outside the Union but will not involve major new solutions to long standing problems.
13. The opposition to change has in part been overcome by the agreement that the harmonised indices (HCPIs) should be distinct from existing indices. It is, nevertheless, clear that many of the harmonised practices, such the formula, will inevitably be followed in existing CPIs. Initial discussions have reached the tentative conclusion that the kind of measure required for international comparisons is different from that required for micro-economic purposes such as indexing wages to maintain living standards. The implications of this have not been worked out beyond arguing that imputed prices, such as rents for owner occupied dwellings and prices for own produce, should be excluded.

IDENTIFICATION OF DIFFERENCES - COMPARABLE

14. Discussions on the choice of the micro-level aggregation formula have provided a background for formulating a basis for determining what other differences in practice may be important. With 16 countries and international agencies represented in debates there are many views on what is most important. Eurostat has therefore set in train a number of studies to explore and quantify the effects of a number of practices (see

paper by Jörgen Dalén). It has also attempted to define an operational basis for deciding which practices should be harmonised in order to ensure comparability of HCPIs.

15. The preliminary definition of "comparable" which has been decided for use in the project is as follows :

"HCPIs are comparable if differences between them in the measured rate of price change reflect only the different impact of price changes on consumers. Indices which differ only on account of differences in the national consumption preferences are comparable but those which differ as a result of differences in the concepts, methods or practices used in their definition and compilation will not be comparable. The Commission shall set down concepts, methods or procedures to be followed to achieve comparable HCPIs where either:

- a) the difference between the effect on the annual change in the "all items" HCPI of a particular concept, method or procedure and the effect of its alternatives is more than one tenth of one percentage point.

or,

- b) it is agreed that a particular concept, method or practice is to be preferred to its alternatives."

16. It is proposed to use this as a guide to decisions on what practices should be harmonised by laying down guidelines such as stating which micro-level aggregation formula to use. The definition is not necessarily adequate for its purpose but it is hoped that it will be refined in the course of its application so that an appropriate definition will emerge. The decisions will be taken collectively by the Member States. It is thought important to have a minimum effect for decision purposes as this places limits on the harmonisation process as well as reducing the scope for argument. The difference of 0.1% would for example require the harmonisation of the micro-level aggregation formula where a difference of 0.5% would not.

17. The first step is to determine differences in the concepts, methods or practices which have an effect of 0.1 % or more on the estimate of annual rates of change from the CPIs. Some work has been set in train on this

using Swedish and British data sets and it is expected that other countries will contribute. It will then be necessary to draw up guidelines for the harmonisation of the divergent practices. Other practices where quantification is not possible may be harmonised by general agreement of the Member States.

18. The definition of comparable raises a number of issues which relate to well known problems of the detection and elimination of bias in CPIs. However, it is hoped that the context of international comparisons will throw new light on them. The issue of removing differences between CPIs is not one of reducing bias. As such it avoids the question of what should be measured. The point of reference for any one HCPI becomes the other HCPIs. However, the potential scale of some biases may be inferred from estimates of differences between CPIs arising from different practices.

19. The strategy leaves many questions to be answered but central to these are four :
 - a) how large a difference in annual rates of change between HCPIs on account of their method of construction is acceptable to users ?
 - b) how can specific differences be quantified ?
 - c) how can the effects of differences be aggregated and will the application of the 0.1% rule for comparability yield an appropriate level of residual difference?
 - d) how should systematic differences be viewed in decision processes?

They are amplified but not answered in the following paragraphs.

Categorisation of Differences

20. Annex A gives a typology of the sources of differences between CPIs. This is intended as an aid to the determination of ways to eliminate such differences (other categories might generate better ideas for solutions). It distinguishes differences in concept, from differences in methods and from

differences in practices. Conceptual differences such as purpose should be partly removed by the agreement on the aim of the CPI. Which for the HCPI is international comparisons of consumer price inflation in a macro-economic context and not a cost-of-living index. The scope may also be defined by agreement e.g. to include or to exclude tourist expenditure and to include or to exclude specific payments such as betting or life insurance. Such differences in definition may be quantified where there are countries already having the wider scope. Differences between the contexts in which measurements of price change have to be made such as in health and education pose particular problems.

21. Methodological differences in computation may likewise be removed by agreement and quantified by countries with sufficiently flexible systems. Differences in survey methods are less easily removed because of the costs of making changes in this area. They are also difficult to quantify. The alternatives are not readily simulated as their effects are not to change the way the data are used but to change the data itself. The possible extent of some effects may be obtainable by sub-sampling within existing data sets.

HOW LARGE IN AGGREGATE

22. CPIs and annual rates of inflation based on these are customarily published to one decimal place. Sampling errors are not generally available but estimates that have been produced indicate errors well in excess of 0.1 percent. For the purpose of argument 0.5 percent (two standard errors) may be a reasonable assumption. Many users will be indifferent to errors at this level. Most, even the more sophisticated of users, will not think about the question at all. Indeed there is little evidence that many CPI compilers have had to give the question much thought.
23. In a national context it is possible to make some conjecture on the consequences for certain users or uses of an error in the CPI. Often the aggregates of income, taxes and investment which are directly linked to the CPI can be computed. In the UK 0.1 per cent on the index equates to around £100 Million per annum in pensions and related payments. Changes in the CPI are also a background or starting point for wage and contractual agreements. Thus directly or indirectly small errors in the CPI can have large aggregate impact on payments by governments and employers. From time to time it is suggested that the failure to treat quality

change in a national CPI leads to an overstatement of inflation. Whether this is true or not it is clear that those making the observation regard effects of 0.1 percent as serious. In recent debates on the possibility and consequences of zero inflation, analysts are increasingly looking at the economic effects of relatively small differences in rates of inflation. It is therefore time for index compilers to take this question seriously.

24. In the context of international comparisons Eurostat is seeking to establish a requisite degree of comparability. One of the central uses of the CPIs which are to be harmonised will be the assessment of the economic convergence of economies in the run up to Monetary Union. The Treaty requiring this is not precise on how convergence should be assessed but, broadly, individual rates of consumer price inflation will be compared with that of the three best performing countries. A difference of no more than 1.5 percent will be the criterion to be satisfied for entry into the Monetary Union.
25. The required reliability of any statistic depends on the purposes for which it is used. Given the wide range of purposes for which the CPI is used it is difficult to imagine that a high degree of accuracy is not required. Nevertheless, it is difficult to argue conclusively that a target error around 0.1 percent (per annum) is justified over a target of around 0.5 percent.

QUANTIFICATION OF SPECIFIC DIFFERENCES

26. Quantification of the aggregate effects of specific differences in the formula used for micro-level aggregations has been carried out successfully by many countries. Alternative formulae can be readily applied to a single set of prices. Quantification of some other differences such as annual versus less frequent updating of weights and different methods of revaluing weights can be computed by some countries. Those who use probability sampling can simulate the difference between probability and purposive sampling. Other practices such as different validation procedures and different treatments of missing prices can also be simulated at least to some extent.
27. An alternative approach to differences arising from operational failures or short cuts is to test the assumptions on which they are based. Thus purposive sampling of outlets assumes that there is no systematic

difference between the movement of prices from the outlets chosen (according to some rule) and the population of outlets as a whole.

AGGREGATION OF DIFFERENCES

28. If the HCPIs in the EU are to be comparable it will be necessary to reassure users that the aggregate difference are within acceptable bounds. Some estimate of what this might amount to before harmonisation and how much it is improved by harmonisation would be desirable.
29. Ideally, for the European Monetary Institute (EMI), a statement such as the following is what is required:

"The difference between the annual rate of inflation in Member State M and that of the three best performing States is less than or equal to 1.5 percent (or alternatively is greater than 1.5 percent)".

However, taking account of the remaining differences in practice and the sampling errors requires a more elaborate statement.

30. One approach for which there is at least some theoretical framework is the idea of the power of a statistical test. Thus one might attempt to ensure that tests of convergence based on the HCPIs would neither wrongly exclude a Member State which actually met the convergence criterion nor wrongly include one which had not in fact met the criterion. The statement might be reasonably rephrased as testing the hypothesis that

"The difference is not greater than 1.5 percent" and requiring there to be a risk of no more than 1 in 20 of non-rejection if the actual difference exceeds, say, 2.0 percent, and similarly of rejection if the actual difference is less than 1.4 percent.

31. The actual difference will be of the form :

$$\Delta = \hat{X} \pm f_n(s, d_1, d_2 \dots d_n)$$

where \hat{X} is the measured difference, s the sampling error and d_1 to d_n are differences arising from various sources. Some of these differences will be similar to the micro-level aggregation formula difference. That is of known direction (\pm) but uncertain in scale. Other differences will be no more than

possibilities of uncertain direction and scale. The question to be answered is whether it is or will be possible to make any useful statement about Δ given that it should be possible to discover something about the differences d_i etc.

32. The differences d arise from differences in practices which may be identified. The magnitude and direction of d will depend on the particular combination of price changes over the period of 12 months concerned. The d are not random variables but are the results of specific decisions on the treatment of a particular data set. They arise from the choice of formula or of procedure and are likely to persist at least for a period of months. They may be independent of each other but may, on the other hand, interact. There is, for example, some evidence to suggest that the effects of different validation procedures may depend on the formula for micro aggregation. The function f is not therefore likely to be a simple addition of the differences.
33. The programme of work set in train by EUROSTAT might be expected to provide the following kind of information about the differences:-

$$-0.1 < d_1 < 0.1$$

$$0 < d_2 < 0.2$$

$$0.1 < d_3, d_4 < 0.1$$

in the circumstances where it has been possible to test the effects of differences in practices 1 to 4. That is to say that practise 1, maybe "representative sampling using rule R rather than rule S", affects the measured annual rate of inflation by amounts in the range $\pm 0.1\%$. Practice 2, say "the choice of micro-level formula", adds to the rate by up to 0.2%. Practices 3 & 4 acting together have an effect in the range $\pm 0.1\%$. From such results, while it not strictly logical to reach any general conclusions, it is necessary to modify the proposition that practices 1 to 4 do not affect the measured rate of inflation. One might say that "taken together practices 1 to 4 add up to X% to the measured rate". If $X=0.01\%$ then one would be fairly sure that the statement was true. Whereas if $X=0.4\%$ one would be less confident in its truth but might nevertheless be willing to accept it. That is to act as if it were true and, for example, to change practices 1 to 4 or to allow for an error of 0.4% in the measured difference between rates of inflation.

34. The forgoing reasoning is akin to that used under the heading of "fuzzy logic" where a considerable amount of theory has been developed to handle possibilities as opposed to probabilities. Unfortunately the writer has been unable take the argument to a useful conclusion. It may nevertheless be helpful to speculate on what an acceptable result might look like. It would be a decision rule that could be applied to actual differences between measured rates of inflation, known differences in the practices used to construct the CPIs and the above kind of information on the effects of such practices. Such a rule might be:-

"Given a measured difference of 1.7% and differences in practices a) and b) there is a strong possibility that the actual difference exceeds 1.9%, and a weak possibility that it is less than 1.35 or exceeds 2.1%. Therefore reject the hypotheses that it is no greater than 1.5%."

35. If the only difference in practice were in the micro level formula, AvR, there would be a very strong possibility that a measured difference of 1.7% would be an actual difference of 1.8%.

Conclusion

36. This paper has considered the reasons why it has proved so difficult to resolve the problem of which micro-level aggregation formula to use in compiling a consumer price index. It suggests that there is a lack of an agreed point of reference on what is to be measured. This partly reflects the absence of a generally accepted definition of consumer price inflation and partly the multiplicity of operationally measures which have in effect become the definition of inflation in different countries. Further, it is suggested that inflation will only be defined from the bottom up as index compilers solve the practical problems of measurement. The process should be guided by the requirements of consistency and coherence and the needs of users. It should not ignore the search for general principles but aim for acceptance that inflation be defined in terms of the measure which is most widely accepted as being the best that can be achieved in the current circumstances.
37. The choice between different formula and between different practices can be informed by a process of international discussion and debate. Bias in CPIs may be assessed in terms of differences between actual operational measures. Specific differences may be quantified by applying different

methods to single data sets. However a new framework or methodology is necessary in order to be able to say how in aggregate biases affect comparisons between CPIs. Decision makers need a rationale for handling the possibility that measured differences reflect differences in concepts, methods and practices. It will be necessary to look outside statistical theory for this since statistical theory has notably failed to provide a framework within which practitioners can tackle the long standing problems of bias. This paper is, however, intended only as a contribution to the discussion of the problem as it relates to CPIs.

Typology of Differences between Consumer Price Indices

Conceptual

Differences in what should be measured. Differences in principle in what should be the aim in constructing a CPI. These can arise from differences in :

Purpose

Pure price index
Cost of living index
Laspeyres index

Definition

Scope

All/particular - consumption
All/particular - households/resident/tourist

Price

Gross/net
Acquisition/use/payment

Quality

Producer/user/hedonic

Context

Institutional

Public/private provision of goods and services

Information

Administrative/survey data sources

Methodological

Differences in how the conceptual requirements of the CPI are to be met. The method of approach to the resolution of operational problems. Difference in design of operations. These may be subdivided into :

<u>Computation</u>	<u>Aggregation</u> formulae micro/macro
	<u>Processing</u> algorithms correction/validation
<u>Survey</u>	<u>Design</u> data specification/processing
	stratification/selection
	<u>Procedures</u> methods of collection/post/telephone
	clerical/automated

Operational

Differences in the way the intended methods are interpreted, in the practices that are actually followed. These are threefold :

<u>Price collectors</u>	Expertise/training
<u>Price collection</u>	Achieved scale/items/outlets
	Descriptive detail/quality of recording - error rates
<u>Control</u>	Audit/validation/feedback
	Management statistics, independent checks.