Improving Quality Adjustment in Practice

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

By Don Sellwood

Abstract
This paper seeks to establish what if anything can be done to improve quality adjustment practices in the short term. The capacity of existing theoretical frameworks to guide the treatment of quality change in index construction has not so far lead to widespread agreement on actual practice. The CPIs as actually constructed using simple quality adjustment procedures on sample data drawn from an unspecified universe, constitute the only substantive operational definitions of the diverse conceptual notions of "pure price change", "cost of living" or "inflation". Measured quality change is thus defined by the aggregate of actual quality adjustments. However, index compilers do not consider what it is they actually represent in this process nor how this might differ from what they seek to represent.

The operational definitions of CPIs differ only in minor respects, differences in the sampling and quality adjustment procedures used in their construction. There are in practice a limited range of procedures available and therefore only a limited range of possible definitions from which to choose. In the absence of an agreed operational definition for the CPI comparisons of practices between CPIs can be a route to improving existing indices. This is the approach taken in the European project of Harmonization.

"Overlap" is central to quality adjustment practice but its use is neither consistent nor coherent. The theoretical support for this notion provides no guidance on the limits of its applicability. Theoretical notions of utility or constant quality are of little consequence and may be positively misleading unless they determine the procedures to be applied to the observable entities that enter into the index computations.

We have the indices that we have because these are what we can actually produce. There are limits to the alternative practices that can be used and we must therefore focus our research efforts on discovering these and explaining what the indices as constructed imply for the generally held notions of quality change and quality adjustment. The way forward requires concepts firmly grounded in actual observation and an adherence to the tenets of sampling theory. With the availability of scanner data, index constructs for the universe of transactions, corresponding to acquisitions rather than consumption, may be defined and computed. A monthly chained index provides a measure that offers interesting perspectives on the notion of quality change.

Key words:- Consumer Price Indices, universe, sample, quality adjustment, overlap, product change, monthly chaining.

Introduction

1. We need to refashion our conception of the aims and limits of quality adjustment, to challenge claims to "self-evidence," "clarity," and "necessity" in suggestions for particular conceptual frameworks (note 1). This paper begins and ends with CPIs in practice. These are Laspeyres-type indices of acquisition prices constructed following a limited range of well-established procedures. The paper is a direct challenge to "Triplett's axiom". "The economic concept of consumption drives reasoning about consumer price index number issues" (notes 2 & 3.). It rejects the introduction of fictitious consumers and fictitious prices as serving mainly to confuse the practical issues. CPIs are necessarily about transactions, acquisition prices are what consumers see. Whilst certain aspects of the subsequent use of consumer products are in principle observable index compilers make no attempt to observe consumption. Nor do they have procedures for using such information if they could observe it. Readers are invited to set aside theories involving imaginary worlds and look at the facts of what index compilers actually do and the practical options for improving CPI construction.

2. The only theory that statisticians are bound to accept is sampling theory. Since CPIs are sample constructs, "sampling theory" must be taken as a minimum conceptual framework. A "population", or "universe", must be identified as the required representation and yet in the CPI there is a universal failure to state what samples seek to represent (note 4). Bias only has an operational definition in this sample /universe context. Bias, as in "the index is not what Boskin had in mind," provides no practical guidance until an alternative index construct is operationally defined. Earlier work (Dalén 1998) attempted to determine a universe consistent with the Laspeyres concept. The approach here is to relate the sample to an actual universe, regarding the Laspeyres concept as concerning the form but not the content of the representation.

3. Existing practice, to a large extent, sets the limits to what is possible at least in the short-term. What we can do is limited to what we can observe and the actual procedures we follow. The paper therefore seeks to identify those choices of practice that are critical to the measured price change, what representations particular procedures achieve and how choices between these might be made. Operational measures (population statistics) for the universe are explored in relation to the standard sample estimates. These are based on the Laspeyres model but this should not be taken as a rejection of alternatives (suggested by Diewert 1999 and others). Consideration of the population statistic serves to avoid the many confusions that arise in discussions that relate solely to sample practice. It also provides a guide to what can legitimately be inferred from the sample. Whilst recognising the practical constraints that face the index compiler it helps to locate the issues of quality as they arise in the universe and to say what must be observed to represent these in the sample.

4. The views expressed in this paper are the author's alone and take up themes developed in earlier Ottawa papers (Sellwood 1999). Nevertheless, the ideas arise out of work to formulate minimum standards for sampling and quality adjustment within the project to harmonize CPIs in the European Union. The subject of quality is infinitely complex and simplification can be dangerous but the author has attempted to distil the central points, the facts, as briefly as possible in the following paragraphs and in Figure 1. Some supporting argument is given in the notes and the index constructs are given in Annex I. Figure 2 illustrates the concept of overlap which is discussed in later paragraphs. The paper concludes with tentative proposals for empirical research to determine the actual effects of particular quality adjustment practices and to point to the choices for which standards might be set for sampling and quality adjustment practices. It is in the rationale for alternative operational choices that economic theory must seek its role.

The CPI in Practice

5. Any conceptual framework is only as good as the practices that implement it. No matter what they are said to measure, what a particular CPI actually measures is determined by its operational definition. That is by the actual practices followed, the sample selection, the observations made and the treatment of those observations. We therefore start by considering the few well-known procedures that form the basis of the limited range of operational definitions that currently exist.

6. The procedural options for making quality adjustment are strictly limited. These procedures are used to adjust an observed price difference between one sampled product and its sample replacement. The options include:

   a) Consult producer
   b) Consult panel of experts
   c) Train price collector to decide
   d) Hedonic regression
   e) Overlap, link or splice
   f) Treat as comparable value

7. The quality adjustments given by these procedures may be expressed as part of the elementary aggregate formula. (See Schultz 1995, the current paper follows this excellent description of the

treatment of product change in CPIs.) The adjustments, in effect, modify the elementary aggregate formula for combining current sample observations with a reference selection when a sample replacement indicates a quality change. No distinction between the sample and the universe is explicitly recognised in the descriptions of the procedures. The product pairs for which quality adjustments are considered and hence the number of quality adjustments made are determined by the sampling procedures. The choice and application of a particular adjustment procedure then determines the scale of the individual adjustments.

8. Sample design and sample practices, particularly "matching", "replacement" and "re-sampling" determine the cases where an issue of quality change arises. For a given reference selection matching, the tracking of a "comparable" product from month to month, and replacement, when there is a failure to match, determine the current product selection. These procedures determine the current products that are covered and the dividing line between "matched" and "non-matched" replacements. Comparisons of numbers of replacement show large differences across the EU but neither the effects nor the implications of this are discernible from the sampling procedures. Though there is no intended representation stated these sampling practices in fact determine the universe that is actually represented and, along with other operational practices (such as quality adjustment), the adequacy of representation. To interpret matching and replacement we must consider what universe is represented by a particular sample. We offer as a statement of required representation take a universe that might reasonably be inferred from these practices.

9. The operational definition of a price index involves a stated coverage, Household Final Monetary Consumption Expenditure (HFMCE) in the case of the HICP. The expenditure coverage involves all prices and associated quantities without exception and there is no stated intention to exclude particular prices, e.g. those in certain outlets or in certain regions. The HICPs, as the CPIs from which they derive, are based on acquisition prices. Notwithstanding the fixed basket approach there is no intention to exclude certain current prices such as those for new models or new varieties of products. This coverage corresponds to the universe of all transactions. The prices normally observed are those quoted in outlets but for most CPIs the price concept is the "acquisition price" or the price incurred by consumers on acquisition of a product in a transaction, the transaction price. The advent of scanner data allows us to explore a transactions "universe" and to determine the actual effects of particular procedures as if they would have been followed in such a universe.

The Representations implied by sample practices (See Figure 1)

10. Figure 1 illustrates sampling from an elementary stratum of the transactions universe. An elementary stratum may be defined as the "lowest" level of an exhaustive subdivision of the transactions universe. This level is defined by and corresponds to that part of the universe represented by the elementary aggregate used in actual index practice. The elementary aggregate index (EAI) should be seen as the sample estimator for an elementary stratum index (ESI) of the universe. In practice the EAI is not a weighted index and is comprised of the prices of, so called, "representative" products selected from the elementary stratum. The EA is specified by a "reference sample selection" and by the subsequent procedures. The elementary stratum, so defined, is an appropriate representational requirement for the elementary aggregate in the absence of any stated intention to represent some alternative universe. From this we can draw some tentative conclusions about the representations that may be legitimately inferred from the existing sample practices against a requirement to represent the transactions universe.

11. **Reference sampling** or the initial sample selection determines the actual coverage achieved by the CPI in a given reference period. The "representation" of a purposive sampling can only be tested qualitatively but most index compilers would claim that their intention is to represent the totality of the index "coverage" by a selection of "representative product prices". Within an elementary stratum there would be at least some token effort to recognise that there are more transactions associated with some product prices than with others. Although there are practical constraints influencing this selection it is not intended, notwithstanding the lack of weights and purposive sampling, as a representation other than of the "transactions universe" in the price reference period. A probability sample of actual transactions would be universally preferred if costs were not of consequence.

12. **Matching** is a procedure that is taken as part and parcel of the Laspeyres concept but it serves to define a particular universe. In simple terms this is a "static sub-universe" of transactions (see note 5) in the current and reference periods for which there is a correspondence given by the matching criteria actually applied. That is where, for the purpose of the CPI there is deemed to be no product change and hence no quality change. A static transactions universe is assumed in much work on CPI formulas (see Diewert 2001). Matching of identical and "closely similar" sample products in pairs implies a universe that is only part of the complete universe of all transactions. It is those transactions in products available at both time t=1 (the reference period) and t=2 (the current period) but not transactions in those products only available at either t=1 or t=2 (but not both). [Note that matching in practice allows for similar or "comparable" "replacements" but to avoid confusion we do not use the term replacements for these.] The reference sample will include a selection from the remainder of products that were available at t=1 but are no longer available at t=2. The transactions universe will also include transactions in products that were not available at t=1 but are available at t=2. These transactions excluded from consideration by the "matching" are a "dynamic sub-universe" the complement of the static sub-universe.

13. The standard formula provides an operational definition for a Laspeyres index for a static universe of transactions (note 6). The numbers of transactions in the reference period provide the weights. The "matching" criteria actually applied (rather than those simply stated) define "constant quality" for the purpose of the CPI. Wherever a current product price is used without quality adjustment, notwithstanding "minor" product differences between the product and the reference product, quality is judged not to have changed. It is necessary to examine actual product matches to understand the representation achieved in this process. The quality change arises in practice when we incorporate the prices associated with the excluded transactions. We attempt to do this by replacement and re-sampling.

14. **Replacement** is a procedure designed to deal with product change when it is encountered in the sample. Where replacements are drawn from products in the transactions universe not represented in the "matching" reference selection they represent added or new products. Replacement then brings in a selection of those products available at t=2 but not at t=1 and in doing so it (normally) maps these one-to-one with products in the reference selection that are no longer available at t=2. Although there is nowhere a stated intention to limit coverage of CPIs, replacement rules with an emphasis on the reference selection may in fact exclude some of the products available at t=2 from the field for selection. The limitations of such replacement strategies are recognised by the need to re-sample.

15. The "Replacement" criteria actually applied together with the reference products where matching fails define "quality change" for the purpose of the CPI. These are the product pairs for which quality is deemed to have changed and for which a quality adjustment is therefore required. The extent to which the product change in the sample represents that in the universe depends on the replacement rules used.

16. **Re-sampling** does not in itself represent any change in the universe. It is a partial or a wholesale replenishment of the reference sample. In practice if an existing sample "seems" adequate apart from losses that have occurred through failure to make replacements then re-sampling may be a periodic maintenance of the sample to ensure adequacy of product coverage from the point of re-sampling. With annual chain indices this is conveniently done at the link month (usually December) (see annual re-sampling below). Otherwise re-sampling would normally be implemented when weights are updated. Re-sampling is further evidence of the intention to cover all available products.

17. Re-sampling to replenish the reference sample is designed to represent the products available in a new reference period and thus transactions in that period. It cannot in itself represent product change through time since it relates to a single time and is implemented at a single time without reference to earlier times.

18. **Annual re-sampling** is an annual replenishment of the sample. This is standard practice with some chain indices. It may or may not be comprehensive. In as much as it brings in new product varieties it reduces the likelihood of matching failures and hence of replacements and quality adjustments.

19. **Monthly re-sampling** uses re-sampling whenever a reference product cannot be matched rather than replacement with quality adjustment. A re-sampling can be made instead of replacement whenever there is a failure to observe a "matched" product. We call this "monthly re-sampling". In this situation there is no quality adjustment (or an implicit adjustment equal to the price difference) but clearly a different index results.

20. **Imputation** covers a range of procedures for dealing with replacements without making explicit quality adjustments. There is room for ambiguity above to the extent that sample replacement for a selected reference product may involve selection from the static sub-universe. That is, the selection of a violin as the replacement for a piano, a change in the representative product. The inference to be drawn from such a replacement is not a straightforward matter of quality change. It may be justified as maintaining the matched sample in which case imputation is appropriate. They may be taken as procedures for dealing with the "matched" sample where the chosen replacement is taken from the matched or static universe. Where they are used for replacing a disappearing product with a matched product they cross the static/dynamic boundary and this is not a phenomenon observed in the universe. A reference product has disappeared from the universe but the procedure, in effect, removes this from the sample and product disappearances from the universe are not then represented. This is not a stated intention of the procedure.

21. **Replacement/re-sampling strategies** differ between CPIs. The difference between replacement and re-sampling is that the latter does not involve quality adjustment (or one might say it implies that any price difference between the average price between the old and new samples at the link is due to quality). It is reasonable to conclude that the balance between replacement and re-sampling will markedly affect the resultant index figures.

Index Forms

22. Apart from sampling practices the CPI is determined by the index form used. By this we include not only the formula but also the actual procedures by which the price observations are combined. Discussion is confined to the Laspeyres-type indices but the arguments are believed to apply to other forms. Annex I explores the basic forms of index construction for an elementary stratum or the "lowest" subdivision of the transactions universe as depicted in **Figure 1**. The Laspeyres index can be defined and computed for a static universe and hence an ESI for the static part of the elementary stratum is readily obtained. (The notion of constant quality implied by this construct is explored below.)

23. The problem of constructing an index for a dynamic sub-stratum of transaction universe looks, in **Annex I**, somewhat different from the elementary aggregate index (with explicit quality adjustments applied to reference replacement pairs) used as its estimator. It incorporates an overall explicit adjustment factor to allow for the fact that different products are available in the current period from those available in the price reference period. This factor simply locates product change in the universe. It is not intended to have a specific meaning but is to be interpreted by inference from the sample procedure. However, it must be accepted that this index form does impose some constraint on possible quality concepts.

24. In the dynamic stratum there is no necessary correspondence between the "added" products that exist in the current period but not in the reference period and those "deleted" products that existed in the reference period but not in the current period. It might be said that the suppliers or outlets have replaced a particular product with one or more other products but any notion of matching is necessarily vague. The universe includes all products whether or not they match. The price index for the dynamic part of the stratum must therefore be specified for transactions in different products. It is here that the problem of quality adjustment arises.

25. Annual or monthly chaining change the matching/replacement balance. The scale of the quality adjustment problem depends on the relative sizes of the static and dynamic components of the elementary stratum. This is in turn determined by the time between the reference and current periods, the rate of actual product change and the judgements as to what counts as change. If the time is short enough then the problem "apparently" vanishes. This can be exploited by using a chain index where product changes all occur at the links. In practice some countries use annual links but a monthly chained Laspeyres index can be defined to cover virtually the whole transactions universe.

Monthly Chained Laspeyres Index for The Elementary Stratum

26. As the time period from t=1 to t=2 of the Elementary Stratum Index in Annex I and Figure 1 is reduced then the numbers of transactions in the static sub-universe increases and the number in dynamic decreases. We can then conceive of a chain index of matched products.

$$I^{ch} = [\sum_{m}^{M1} P_{2m}N_{1m} / \sum_{m}^{M1} P_{1m}N_{1m}][\sum_{m}^{M2} P_{3m}N_{2m} / \sum_{m}^{M2} P_{2m}N_{2m}][\sum_{m}^{M3} P_{4m}N_{3m} / \sum_{m}^{M3} P_{3m}N_{3m}].....etc$$

The product sets M1, M2, etc gradually change over time. They each include all the available products except those products that disappear over the subsequent month. It is natural to see this as being estimated by an elementary aggregate index where the are no replacements but a re-sampling is made whenever a product disappears. However, such an index might not adequately represent additions to the product range unless re-sampling is targeted at additions.

27. This result is not new. The point is to show chaining as a limit to the normal matching and replacement constructs. There are no explicit quality adjustments because there are no product changes within a link. Clearly such an index would be a different measure of price change from one where product change is treated by specific quality adjustment of prices in the dynamic component with annual or, say, five yearly links. It is not known whether any EAIs in actual CPIs come close to monthly re-sampling though re-sampling whenever matching fails may be a widespread practice that approximates this.

Product Change v Quality Change

28. The analysis above suggests that the quality adjustment problem can be seen as the problem of allowing for product change in the dynamic strata. Using the terminology "product change" rather than "quality change" may avoid some of the ambiguity surrounding "quality" (notes 5 to 11 seek to elaborate the reasoning behind the approach here). The dynamic consists in changes in the range of products available to consumers. The reality here is of transactions in two sets of similar but essentially different products involved in two sets of transactions with, for the most part, different consumers but there may be a relatively static set of outlets. The product sets will be PCs or TVs or other goods or services. Actual samples provide a selection of reference/replacement product pairs from selected outlets. Whether these constitute representative samples of product changes may be examined empirically.

29. The central issue is how one compares prices for sets of different products. The quality adjustment procedures attempt to divide the reference/replacement price difference into a price and a quality component. The choice facing consumers is between existing products. The reference products have no necessary relevance. If a consumer is replacing a durable the replaced product may be several years old. Sample replacement is not intended to represent this. It is necessary to consider whether each sample replacement presents a peculiar or particular adjustment problem or whether there are general categories of quality differences for which different rules may be appropriate. A close examination of quality adjustments in our samples may suggest what we might legitimately claim these to represent. We can consider what arguments are used and whether these constitute a theoretical basis for choosing between the procedures or whether there are alternative procedures that might be theoretically justified.

"Overlap" and Product Mix

30. The products in the static strata remain unchanged within the limits of the matching criteria. However, the Laspeyres construct for the static sub-universe in Annex I not only allows for changes in the total numbers of transactions but also for changes in their mix. The static embraces changes in the mix of transactions in a fixed range of products. The allowance for changes in mix implies a particular view of quality, namely that quality difference may be treated as equal to price difference in the reference period. This corresponds to the notion supporting "overlap" quality adjustments (Armknecht and Wehback 1989). If this view of quality is extended to the dynamic components then price change for this component can be computed from price changes in the static component adjusted for changes in the mix of transaction across the elementary stratum as a whole, both static and dynamic components. Whether this is an acceptable procedure is a matter for debate. That it has implications for the representation of quality change and "overlap" is not.

31. Since price differences (overlaps) are central to quality adjustment practices it is important to understand the role they play. Theoretical justification must deal with the relevance of the context in which each sample "reference/replacement price difference" is observed for the particular quality adjustment made. If it is argued that the price difference between two models of a product is a market valuation of the quality difference then it is necessary to say in what contexts this is so and what must be observed as evidence that the contextual requirements are met. It would also be necessary to say why the price difference between the same model in two outlets is not likewise a market valuation of quality difference.

32. **Figure 2** (see notes) suggests that the Laspeyres (or other fixed basket index) can be seen as exploiting the concept of overlap in a strict way. The index for transactions in a fixed set of products adjusts for changes in the pattern of transactions over time and changes in the aggregate number of transactions. The figure relates to an elementary stratum but the notion of overlap applies between categories. The suggestion here could alternatively be seen as treating price/quality differences within elementary stratum in the same way as between strata and between category differences are treated in the Laspeyres construct. There is no a priori reason for any given categorisation of products in the CPI and a particular categorisation should not affect the treatment of product prices unless some justification is established. In practice overlap is used consistently between categories, between outlets or outlet types but not consistently between products within categories. Overlap is only used on occasion at this level as part of quality adjustment on replacement or arguably with re-sampling.

33. "Overlap" is essentially the idea that price differences at a given point of time reflect only quality differences. Accepting that there will be noise in any given situation let us take a strict interpretation of overlap. Then it follows that for all "i and j" at time t the ratio of any pair of product prices P_{it} / P_{jt} is a measure of their relative quality. Thus the ratio $R_{it} = P_{it} / \{\sum_i n_{it} P_{it} / \sum_i n_{it}\}$ is the quality ratio for the product i relative to the average quality at time t. Similarly the ratios of any averages of prices are measures of quality relatives.

34. Under the "economic" approach, theory says under certain assumptions prices are determined by the market. Price differences are therefore market valuations of quality differences and price differences are a measure of quality difference. Then provided that the assumptions supporting the theory are met any price difference is a measure of quality difference. The decisions as to when overlap applies seem either arbitrary or subjective. It is not clear how one might establish when the "market assumptions" are met and therefore no guidance from the theory on the applicability of overlap. Arguments that Laspeyres indices do not take account of "outlet substitution" suggest that price differences between outlets are not market valuations of quality differences. Likewise price differences between brands are sometimes taken to represent quality and other times not.

35. **Hedonic indices** take overlap prices as measures of quality difference associated with product characteristics. Following economic theory the regression coefficients are measures of quality.

However, this is the case if, and only if, the assumptions of the theory are met for the specified characteristics. This seems most unlikely and is anyway impossible to prove. In practice the regression shows only the association between price and characteristics - it cannot prove a causal link. Under overlap as used in the Laspeyres construct the regression procedure can be equated to matching (see Silver 1999). As such added and deleted characteristics present the same problem as added and deleted products. They may only be treated by using "overlap" between the added characteristics and the matched characteristics at t=2 and between the deleted and matched characteristics at t=1 as depicted in Figure 2. If this is acceptable for added/deleted characteristics then one might ask "Why not for added/deleted products?"

36. A key issue in relation to overlap prices as a measure of quality difference is whether in practice producers set prices and consumer simply accept them. Model change is an opportunity to change prices and there is evidence that some prices remain stable at their initial level (see Hoffmann 1999). There are no obvious mechanisms whereby producers can establish consumer valuations prior to the launch of a new product. Their confusion over quality is likely to be as great as ours. If new products involve price increases then overlap, re-sampling and chaining will not represent this. The same problem arises with new characteristics

Empirical Issues

37. A key issue for empirical research is to establish the effects of replacement/re-sampling practices in the alternative index forms. This might be done using scanner data to test the sensitivity at the level of the universe. An analysis of current quality adjustments should seek to establish the decisions taken and whether there can be specific rules to guide these. A deeper analysis along the lines herein is called for. What do the actual matching criteria amount to? What implications do outlets and brands have for the concept of quality?

38. Scanner data also provide the opportunity to explore index forms for a transactions universe. They may also facilitate the exploration of various sampling strategies. Sampling schemes for matching, replacement and re-sampling might be simulated. Besides these it is necessary to examine different matching, replacement and re-sampling regimes to establish how these can affect the number of quality adjustment made. An in depth study of explicit quality adjustments, what is actually done rather than what procedures attempt to achieve, should then enable us to understand the concepts of quality implicit in our CPIs. The various applications of the use of "overlap" should be catalogued and standards for its consistent use should be developed within some coherent framework.

39. In the European Union the strategy followed has been one of agreeing a framework of minimum standards leading to comparable indices. "Comparable" is defined by the criterion that annual rates of change shall not differ by more than 0.1% points on account of differences in the concepts, methods or practices used in their definition and compilation. This criterion has served both to focus effort on the most important issues and as a spur to change in practice. In the absence of an operationally defined reference, potential for bias can be explored by reference to other CPIs. The extent to which differences in sampling and quality adjustment practices actually affect HICPs remains the most important unresolved issue.

Summary

40. Any conceptual framework is only as good as the practices that implement it. CPIs are based on samples and, like any sample, they can only have meaning in terms of the populations from which they are drawn - this is their representation and it must necessarily relate to some real universe. That universe may be inferred from the actual sampling practices followed. There is an onus on index compilers to say what universe they represent and this cannot be an imaginary or fictitious universe.

41. Since all CPIs make allowance for quality change, they embody some concept of quality in their operational definition. To understand the quality concept of a particular CPI we must understand what they seek to represent. There is no separating quality adjustment from sampling.

42. Sample “matching” and “replacement” rules have clear implications for what it is in the universe that counts as “constant quality” and as “quality change” respectively. Where these procedures are implemented, largely or exclusively, by price collectors then, notwithstanding any theoretical claims, it is the price collectors who decide the meaning of quality. To change that concept we must change what the price collectors actually do.

43. The classical economic approach to CPI design based on the concept of utility has, so far, failed to provide specific operational guidance for the quality adjustment practices actually used in CPI construction.

44. The pure price approach involves the counterfactual notion, “what would be the current cost of a reference period basket if it still existed?” This “what if” question leaves the concept of quality to be defined by actual practice.

45. The universe of transactions comprising the expenditure coverage of the CPI is the obvious representational requirement.

46. Since quality change necessarily derives from product change, samples must represent both static product sets and the dynamic disappearances and appearances of products. Sampled product pairs on which quality change is to be assessed must represent product change in the universe.

47. “Overlap”, the price difference between products at a given point in time, is central to the Laspeyres (and Fisher) concept, to the Hedonic approach and to the economic notion of market determined prices. It can be exploited to address quality change but there needs to be a clear rationale and consistency in application. There is an interesting equivalence between added characteristics and added products.

Conclusions

48. The terminology “static” and “dynamic” universe are consistent with the standard practices of sampling, matching, replacement, re-sampling and help to avoid confusion on what is or can be represented by CPI sample estimators. The static embraces changes in the mix of transactions in a fixed range of products. It is represented in the sample by matched products and an index form that holds the mix constant. It is the changes in the range of products available to consumers that constitutes the dynamic universe covering transactions in products not included in the static universe. It is represented by sample replacement and explicit quality adjustments to reference replacement pairs. Initial sampling and re-sampling do not represent product change but simply ensure a representative reference.

49. This terminology is purely descriptive and should be non-controversial. However, it implies a particular view of “quality” as operationally determined in the CPI practices of matching and replacement by price collectors and others. The CPI is critically dependent on both the actual sample representation of product change and the aggregate sample allowance for the change in product range, quality adjustment. The representation, individually and collectively, of elementary aggregates should be explored by reference to the elementary strata of the transactions universe to which they correspond.

50. The Eurostat “minimum standards” approach is an appropriate way forward. Common standards should be developed and agreed for the progressive improvement by selection and refinement of existing sampling practices, where divergent practice leads to divergent CPIs, within the framework of sampling theory. The practice not only embraces matching, replacement and annual or monthly re-sampling, but also the design of the reference sample and the choice and numbers of elementary aggregates. Replacement coverage of new product varieties should be appropriate for an agreed representation so that explicit quality adjustments is confined to sample observations that are representative of product change in a target universe.

51. If it proves impossible to develop sensible standards for explicit quality adjustments we should consider agreeing to follow automatic procedures, such as the monthly replacement formula, that will guarantee CPIs that are comparable in the sense of not differing on account of differences in sampling and quality adjustment practices. The behaviour of such constructs should be examined using scanner data.

52. A consistent and coherent use of "overlap" should be elaborated. This should include a rationalisation of category specific (either product kind COICOP, product variety, brand or outlet) applications where these differ. Small model differences should not be ignored, implying overlap is all price, where overlap is used to justify treating large model differences as all quality. Evidence on the price setting practices of producers and sellers for new product varieties should be sought from the extensive literature on the subject.

Caveat

53. The emphasis placed on the Laspeyres index in this paper is simply an acknowledgement that most CPIs are of this form and that arguments as to choice of index form can be addressed separately (see Diewert 1999). However, the choice of the Laspeyres has been a choice on the grounds of practice and not of principle. This choice does not apply to the elementary aggregate where no weights are used. The elementary stratum index defined herein need not therefore be the measurement objective on practical grounds. The Fisher form might be preferred on the grounds of symmetry and, at the micro level on the basis of empirical evidence from scanner data, against the Laspeyres form. The arguments herein should hold for a Fisher index. The issue would then be to what sample design and estimator form result in bias. A Fisher monthly chained formula for scanner transactions data should therefore be investigated.

Notes

1. The **approach** taken here follows from a central theme in Wittgenstein philosophy (Floyd 2000). This was to overturn the notion that there is a necessary correspondence between "words" and "the world" and this has had considerable impact on the science of the social. He emphasised the need to look at how words are actually used. He had much to say "on certainty" and on the place of mathematics in philosophy which challenge the way philosophers and others think about society. The author believes his ideas apply particularly to the way we attack problems like quality adjustment and has attempted to use his approach by seeking to focus on what index compilers actually do in contrast to what they or others say they should do.

2. **Triplett** (1999) comments in his content versus rhetoric argument that " The substantive content of the COL (Cost-of-living) index can be stated in a simple axiom: *The economic concept of consumption drives reasoning about consumer price index number issues.*" This is simply a rhetorical ploy. The concept certainly drives certain reasoning but it does not justify that reasoning and it does not drive practical reasoning. The axiom is not simple and, as Khun would say is highly theory laden, inviting the reader to accept the quite unrealistic assumptions of an un-proven theory and the notion of systematic forces controlling the consumer. Triplett maintains that the role of economic theory is to provide a conceptual framework for discussion of practice. In as much as theory guides practice it helps to ensure consistency but despite the eloquence of its proponents, Coli theory has failed to guide quality adjustment practice in CPIs. CPI statisticians might wonder whether the main intent in this case is to find an application for a theory that has so far failed either to meet its critics or to explain any actual consumer's behaviour.

3. One of last year's Nobel Prize winners in economics, Daniel L. McFadden of the University of California at Berkeley, who has developed ways to analyse how people make everyday decisions such as where to live, when to get a job and how to travel, makes the following observation on the Chicago-man model:- "*It is fals*e. Almost all human behaviour has a substantial rational component, at least in the broad sense of rationality. However, there is overwhelming behavioural evidence against a literal interpretation of Chicago-man as a universal model of choice behaviour." His further view perhaps offers hope for a more constructive contribution to CPIs in future:-"*It is unnecessarily strong*. Many of the core objectives of economic analysis are attainable with weaker forms of rationality that relax perception-rationality, and

permit some important deviations from preference rationality (e.g., mutable preferences) and process-rationality (e.g., bounded rationality). Both users and critics of the model sometimes interpret it in unnecessarily restrictive ways. For example, immutability of preferences does not imply that consumers are unaffected by history or incapable of learning, but only that preferences develop consistently following a "rational" template." But the reservations in the notes below and elsewhere should be addressed (see Kahneman and Scharz 1999, Hodgeson 1988, and Sellwood 1999).

4. The **CPI sample** is simply an operational device to avoid taking the whole universe. It does not exist for any purpose other than to represent the universe. The required representation should be known a priori. It should not be left to be invented in the sampling procedures. The position taken here is that although most CPIs appear to be defined by reference only to sample data we must step back in order establish or choose an appropriate measure of price change for an agreed universe. **"Bias"** in statistical theory has a well-defined meaning. It is the difference between the expected value of the sample estimate and the defined statistic for the universe. This bias cannot be assessed for CPIs where no universe measure is defined. The concept of bias used by the many critics of existing CPIs seems to mean a departure from a more general but as yet undefined objective. "Consistency" in the statistical sense requires that sample estimates converge on the universe measure as samples approach the size of the universe. The advent of scanner data has facilitated the exploration of the alternative universe measures and the sample/universe relationship. (See Jan de Haan 1999)

5. **The Static/dynamic distinction** is a straightforward reductionist approach to the problem. The static and dynamic universes are sub-universes in both space and time. The point here is to see what can be learned by treating them as separate universes for the purpose understanding the representations that may be inferred from the sample. The static/dynamic distinction can not be readily made without full knowledge of the universe but it can be inferred from the sample practices and could, with some difficulty, be achieved by post stratification. However this is not the intention. The aim is simply to elaborate the requirement to represent the universe as it is. We can proceed to ask what index we would calculate if we observed all that there is to observe about transactions.

6. The static universe is a universe in which a basket of transactions can be regarded, for practical purposes, as fixed. This universe is supposed as the starting point of most CPIs and has been used in much empirical work (see Diewert) but it has not been distinguished in CPI design. It will not normally be the case that the same consumers are involved in buying the fixed basket, particularly where major durables are involved, but the basket is available for purchase.

7. "Quality" is a word used over and over again in everyday life in all kind of contexts. It seems self-evident that its meaning is not constant but is given by its use in any particular context. "Constant quality" is used in CPI compilation when comparing the descriptions of two products. If their specifications are the same then there is no "quality difference." But what if the two products actually relate to two different times? They may be "identical" but by the later time the product may have become obsolete because of the introduction of new products. So "constant quality" is also context dependent. Since context always changes quality never means the same thing between any two uses. What then can we make of "quality change" or "quality difference?" Index compilers have sought to solve the problem by imagining some undefined "thing" that is "quality" and which can be held constant. Maybe if we use "product change" instead of "quality change" we will be less inclined to make this mistake. Can we make any rule to guide our treatment of "product change" (or difference) in sample observations? What do we ask price collectors to imagine when they assess a replacement against the recorded specification of its reference product?

8. "Utility" is another word that we use either to make the idea of "quality" more specific or to capture what consumers can be said to maximise by their choices. That consumers make choices in what they consider to be in their best interest is a truism. To suggest that "their best interest" is a thing "utility" known to the consumer and preserved from choice to choice is a nonsense. In what contexts does "utility" differ from "quality?" As "usefulness" or function in fashion perhaps, but does not this year's fashion and even this year's colour serve to make the wearer look "with it"? "Constant utility" is, according to Triplett,

a concept that is better understood than “a constant basket.” He says " ‘standard of living’ is a popular non-technical concept that has an intuitive meaning for most people”. People certainly talk of “standard of living” and a sense of wellbeing but what is this “intuitive meaning” that is understood? Not any “thing” that can be observed in a “product difference.” In what circumstance would Triplett say that the “utility” of the mathematical function corresponds to what a single consumer has in mind.

9. **“Consumers”** are real observable entities and we can also observe the transactions they make. We can say that a particular purchase involves a choice but not in what general sense does this reveal some “thing” that is a “preference” for one product over another. Choice is context dependent and contexts are never preserved and so each preference is unique. An individual consumer may choose between a TV and a fridge. The “preference” thus “revealed” could be dominated by the discomfort of a fuzzy TV picture against a too small freezer box with no reference to the latest technical wizardry incorporated in alternative TVs.

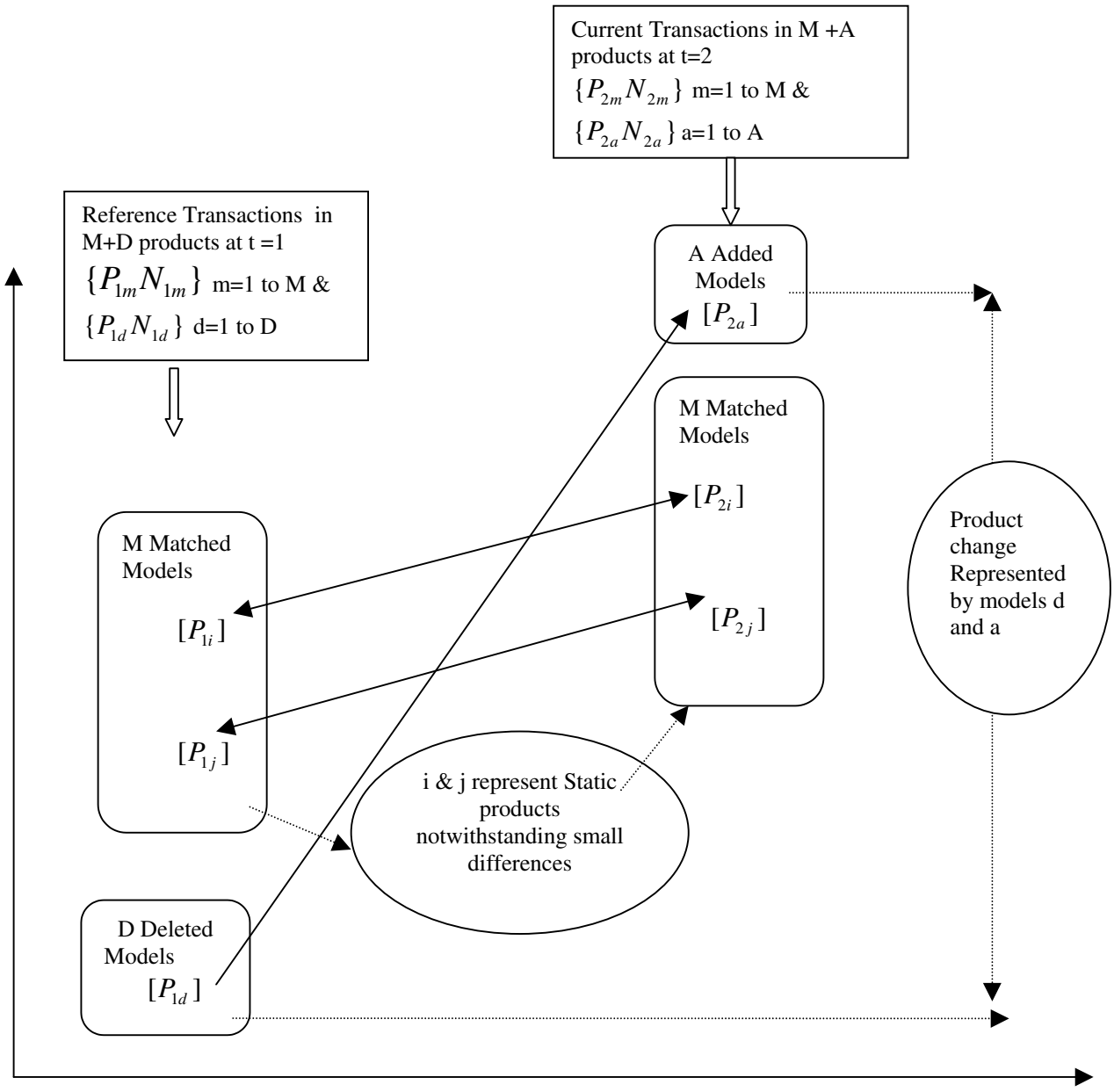
10. “**Rational consumers**,” it may be safely said, make the best choice on the information available to them and which they see as relevant to the choice. We may wish to describe this as maximising behaviour but it does not follow that they thus maximise some “thing” that is “utility” and which is preserved from context to context. Much research (Kahneman and Scharz 1999) suggests that sloppy (non-economic/rational) thinking is characteristic of human thought. The author makes no claim to be free of this indeed he admits to having not only held but also propagated the “sample” view of the CPI for more years than most. However, he maintains that rejection of the Coli theory is the rational choice for a fully informed statistician.

11. If **Coli theory** does give a good approximation to reality and the market determines the fully informed consumer value of quality, then the rational consumer should take price as the best (least search cost) indicator of quality. The observation that consumers tend to choose lower priced goods, substitute away from higher prices (Diewert in conversation) suggests that they are choosing lower quality goods. Such observation needs to be qualified by some reference to quality differences between products. It may be the case that economists and index compilers substitute between products. It is less evident that individual consumers do.

Acknowledgements
I should like to thank Alexandre Makaronidis for his ideas, constructive comment, and his continual support. Leendert Hoven and Thierry Lacroix for their insistence on the primacy of practice over theory. Jörgen Dalén for many fruitful discussions. Martin Ribe for reminding me that a sample has no sense other than the population from which it is drawn. John Astin for giving me the reason and opportunity to pursue this project and for his encouragement.

Figure 1

Sampling an Elementary Aggregate from an Elementary Stratum of a Transactions Universe



The Elementary Aggregate Index (EAI) is $\hat{I} = \dfrac{P_{2i} + P_{2j} + g_{d/a}P_{2a}}{P_{1i} + P_{1j} + P_{1d}}$ where $g_{d/a}$ is the quality adjustment factor for the replacement of the deleted reference sample representative product d by a current product a.

This EAI **represents** both the price change in the transactions in "matched" products and the price change after allowing for "quality" in transactions of deleted and added products. The product d in the reference period is part of the representative selection at t=1. The selection of its replacement depends on the replacement rules. To be representative of all added products the selection rules should not exclude any added products. Any restriction on this selection defines a different universe from the transactions universe and this should be recognized and justified.

Annex I

A Laspeyres-type index for an elementary stratum of the transactions universe (See Figure 1)

1. Figure 1 shows the representation implied by the elementary aggregate index (EAI). Each elementary aggregate represents some stratum of a universe but this is not usually defined. We therefore suppose it to be an Elementary Stratum of the Transactions Universe.

2. The Elementary stratum index may be divided into two sub-indices one for the static or matched sub-stratum and the second for the unmatched or dynamic sub-stratum. The index involves all transactions in products at t=1 (the reference period) and t=2 (the current period). There are M matched products available at both t=1 and t=2, D deleted products available at t=1 but not at t=2 and A added products available at t=2 but not t=1.

We can define an Elementary Stratum index (ESI) for the static and dynamic combined as

$$I = W^c I^c + W^y I^y \tag{1}$$

Where the Static sub-stratum index

$$I^c = \sum_m^M P_{2m} N_{1m} / \sum_m^M P_{1m} N_{1m} \tag{2}$$

is a straightforward Laspeyres index

The Dynamic sub-stratum index

$$I^y = [\sum_a^A P_{2a} N_{2a} / \sum_a^A N_{2a}] / [\sum_d^D P_{1d} N_{1d} / \sum_d^D N_{1d}][G_{2a/1d}] \tag{3}$$

might be a unit value index adjusted for the change in the average quality between the deletions and additions. The quality adjustment term $G_{2a/1d}$ is the appropriate allowance for the quality change for the product change between additions and deletions whatever it is decided this should be.

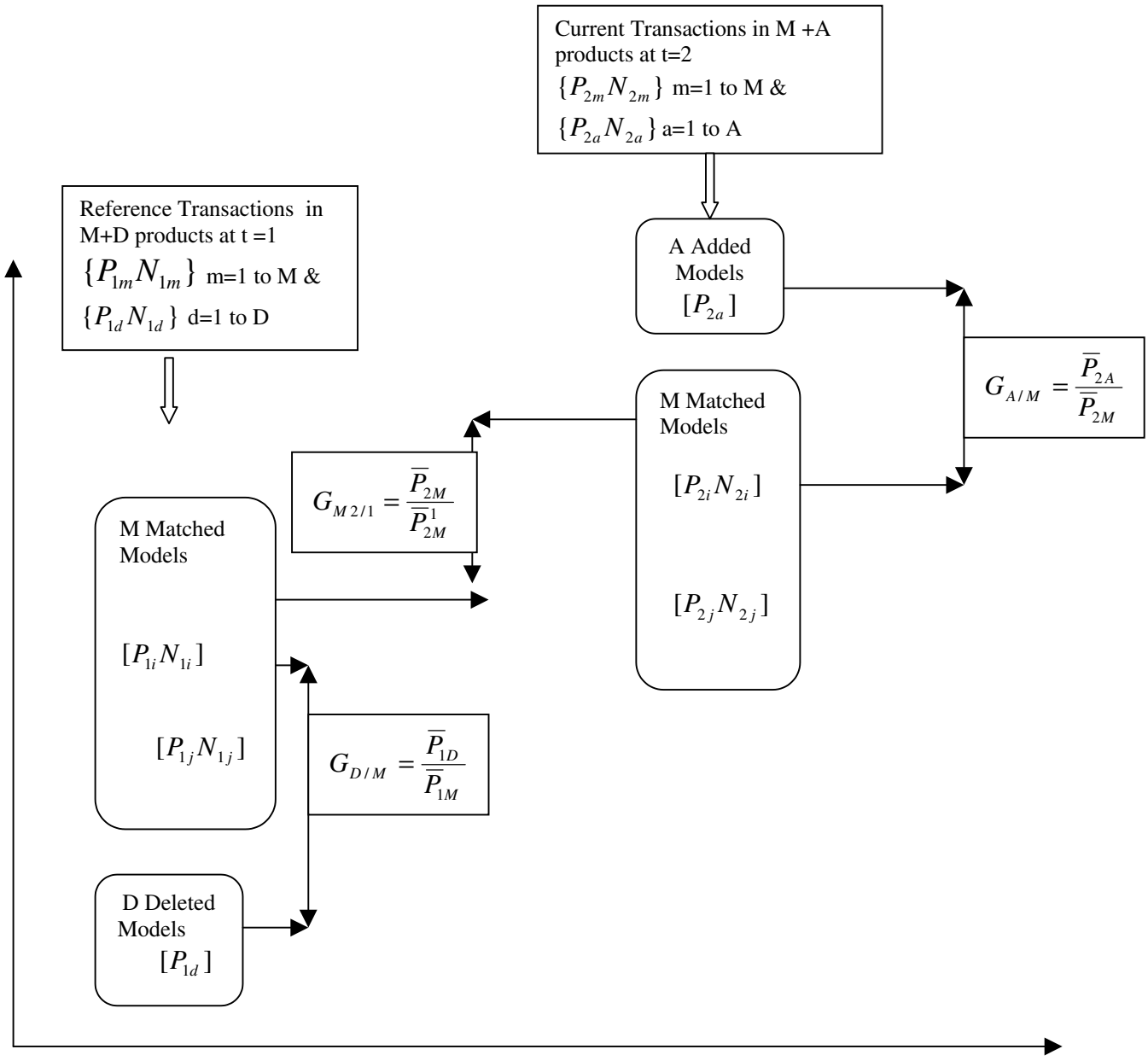
The weights are $W^c = \dfrac{\sum_m^M P_{1m} N_{1m}}{\sum_m^M P_{1m} N_{1m} + \sum_d^D P_{1d} N_{1d}}$ and $W^y = \dfrac{\sum_d^D P_{1d} N_{1d}}{\sum_m^M P_{1m} N_{1m} + \sum_d^D P_{1d} N_{1d}}$

2. The sample estimator for the ESI is an Elementary Aggregate Index that is based on a sample of matched products and of deleted and added products. In figure 2 for simplicity, two matched products are sampled and one deleted product is replaced by one added product with an explicit quality adjustment. A larger sample might involve more replacements and adjustments. The EAI is

$\hat{I} = \dfrac{P_{2i} + P_{2j} + g_{d/a} P_{2a}}{P_{1i} + P_{1j} + P_{1d}}$ where $g_{d/a}$ is the quality adjustment factor for the replacement of the deleted reference sample representative product d by the current replacement product a.

Figure 2

Quality changes "G" implied by "Overlap" in the Laspeyres Index



Overlap simply equate all price differences at a given point in time to quality differences. The Laspeyres index exploits this idea in adjusting the mean current price to allow for the change in the mix of transactions among a fixed set of products. The Laspeyres index for the static or matched stratum is

$$L = \frac{\sum_{m}^{M} P_{2m}N_{1m}}{\sum_{m}^{M} P_{1m}N_{1m}} = \frac{\sum_{m}^{M} P_{2m}N_{2m} / \sum_{m}^{M} N_{2m}}{\sum_{m}^{M} P_{1m}N_{1m} / \sum_{m}^{M} N_{1m}} * \frac{\sum_{m}^{M} P_{2m}N_{1m} / \sum_{m}^{M} N_{1m}}{\sum_{m}^{M} P_{2m}N_{2m} / \sum_{m}^{M} N_{2m}} = \frac{\overline{P}_{2M}}{\overline{P}_{1M}} * \frac{\overline{P}^{1}_{2M}}{\overline{P}_{2M}}$$

The first term is the change in the arithmetic mean price (the unit value index) and the second term is the allowance for quality change due to the change in product mix. It the reciprocal of the Paasche volume index adjusted for the change in the aggregate number of transactions.

References

Armknecht, Paul A and Donald Wehback 1989, “Adjustments for Quality Change in the U.S. Consumer Price Index.”, *Journal of Official Statistics*, Vol.5, No. 2.

Dalén, J 1998, “On the statistical objective of a Laspeyres price index” paper for the Fourth Meeting of the International Working Group on Price Indices .

Diewert, W. Erwin 1999, “The Consumer Price Index and Index Number Purpose” paper presented at the Fifth Meeting of the International Working Group on Price Indices, Reykjavik, Iceland, August 25-27, 1999; Revised: December, 1999.

Diewert, W. Erwin 2001, forthcoming draft for updated ILO manual.

Floyd, Juliet 2000, “Wittgenstein, Mathematics and Philosophy” in “the New Wittgenstein” edited by Alice Crary and Rupert Read, Routledge.

Hoffmann, Johannes 1999,“The Treatment of Quality Changes in the German Consumer Price Index”, Proceedings of the International Conference on Price Indices Iceland.

Hodgeson, Geoffrey M. 1988, *Economics and Institutions*, Polity Press Cambridge.

Haan, Jan de 1999, "Dealing with New and Disappearing Goods in the Fisher Price Index draft". Paper prepared for?for Eurostat’s Task Force on sampling December 1999. (Suggests how the many product changes found in scanner data might be handled)

Kahneman, D. and N. Schwarz 1999, *Well-Being-the foundations of Hedonic Psychology*, Russell Sage New York.

Schultz, B 1995, “Treatment of Changes in Product Quality in Consumer Price Indices”, Proceedings of the International Conference on Price Indices Stockholm.

Silver, Mick and Saeed Heravi 1999, “The Measurement of Quality-Adjusted Price Changes”, in *Proceedings of the Measurement of Inflation Conference* ., M. Silver and D. Fenwick (eds), Cardiff University, August 3- September 1, London: Office for national Statistics (Argue the equivalence between hedonic indices and the matched models approach.)

Sellwood, D. 1999, “Quality Change a Statistical Approach”, Paper for the Fifth Meeting of the International Working Group on Price Indices

Triplett Jack E. 1999, “Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?” in *Proceedings of the Measurement of Inflation Conference*, M. Silver and D. Fenwick (eds), Cardiff University, August 3- September 1, London: Office for National Statistics .