Item & Outlet replacements and quality adjustment

by Ralph Turvey

Substitution bias in an index is generally understood to mean shifts in purchases between elementary aggregates towards those whose relative price has fallen, causing a current-weighted index to show a smaller increase than a base-weighted index. We know, in theory, how to deal with this; compute a Fisher index. This is not practicable, but at least the amount of bias can be measured ex post and minimised by frequent re-weighting.

Similar price-induced substitution by consumers can occur within elementary aggregates. The absence of weights makes even an ex post measurement of the resulting bias impossible, though the nature of this bias is the same.

This paper is about a different kind of substitution, namely that resulting when the statistician replaces some items or outlets by items or outlets which have hitherto not been in the sample used to compute the index. Computation of price change then implies an implicit or requires an explicit quality comparison of the new items with the items they replace.

Principles: the universe of prices

Before examining the replacement by the statistician of old by new items within the sample used to compute an elementary aggregate index, let us consider what it is that is ideally to be estimated when some items cease to be sold and new items start to be sold.

Consider two periods, 0 and 1, where period 0 is both the price reference-period and the weight-reference period. P's and Q's fall into three groups

c available in 0 and continue available in 1
n newly available in 1
d disappeared from the market after 0.

A Laspeyres index can be defined for c items. It would overstate the "true" increase because of the familiar substitution effect, namely:

1. substitution by consumers of those c items which have risen less for those c items which have risen more

But the "true" change will also depend upon the cost in 1 of:

2. substitution of some n items for c items
3. substitution of c items and n items to provide the reference population with the same utility as all the d items.

A “true cost of living index”, which would measure the cost of providing an unchanged level of utility to unchanged consumers, would allow for all three kinds of substitution.

What Laspeyres index would best approximate this? Such an index — comparing the period 1 cost of doing what consumers did in period 0 and would still be able to do in period 1 — would appear to be limited to c items. Its scope would shrink from month to month as what had been c items disappear.

This problem can be escaped if “consumption in period 0” is defined as the precise set of items bought in period 0 only for c items. For d items, period 0 consumption must instead be defined in broader commodity terms supplemented by more detailed definitions in characteristics terms. Thus if a specified model of a widget is available in both periods in a particular outlet, it can be priced in both periods. But if it is not identically available in period 1, the price of some other model of that kind of widget should be taken, corrected for differences in characteristics using implicit prices for those characteristics. If P_0 is the price of a d item, P_1 is the price of a selected c or n substitute, x are the price-relevant-characteristics and b_1 the implicit period 1 prices of such characteristics, a Laspeyres index can be formulated as:

$$\frac{\sum\left[P_1 - \sum b_1\left(x_1 - x_0\right)\right]Q_0}{\sum P_0 Q_0}$$

The x_1 equal the x_0 for all c items.[1]

According to this approach, bias arises if widespread and systematic differences in price-relevant characteristics are ignored or if incorrect implicit prices are used.

A similar argument applies mutatis mutandis to the concept of a Paasche index.

Practice: a sample of prices

Replacements made by the statistician seek to model the substitutions taking place in the universe for which the price index is to be estimated. These replacements are of two kinds which are treated differently in practice.

[1] This formula is not uniquely defined, since for any d item there might be more than one possible period 1 substitute whose quality-adjusted price could be used.

1) Optional replacements

The range of items covered by the index is usually revised when an index is reweighted. Even between reweightings, the range of items covered by the index may be revised, either routinely at regular intervals, or when the importance in the market of new items becomes apparent. Such revisions deliberately update

- the sample of representative items, or of varieties for given representative items, or
- the sample of outlets, so that varieties for the given representative items have to be selected in each new outlet.

The resulting replacements may be between elementary aggregates as well as within them.

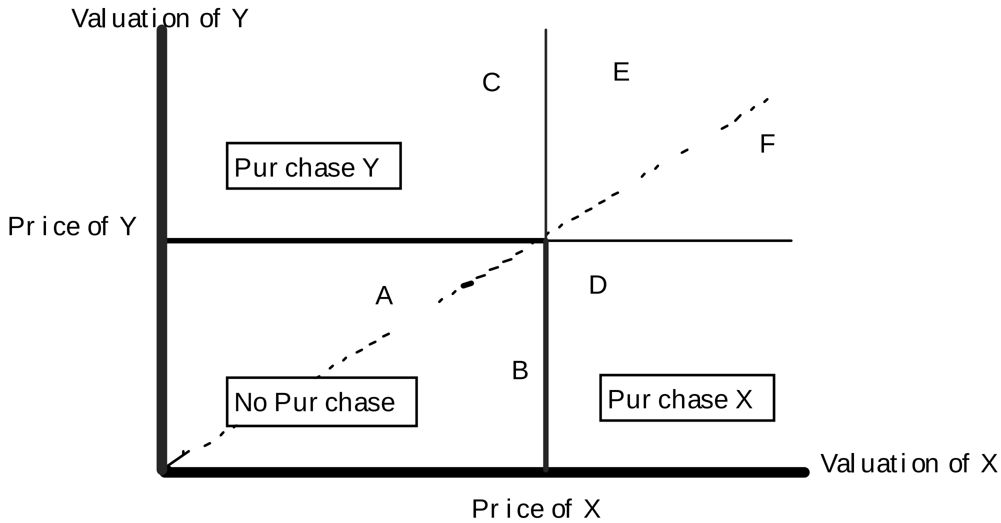
Overlap prices are normally obtained and the replacements brought into the index calculation by linking, so that they do not affect the index. Even if the replacements are a good sample of substitutions going on in the universe towards: new items (or outlets), this linking creates bias whenever the new items or the new outlets provide better or worse value for money than the items or outlets that they replace. It will create no bias only to the extent that the substitutions were made by consumers because of changes in their preferences — since the index attempts to measure changes in the cost of meeting base-period wants. Optional replacements to reflect only pure changes in consumer tastes should thus be avoided.

Linking in optional replacements in such a way as not to change the index is defended on the grounds that, for all prices that overlap, price relatives reflect quality relatives. The argument is that if a consumer chooses to buy both of two items during a month, or even a year, we can deduce that (s)he must prefer this combination of expenditure to any other way of spending the same amount of money, so that her/his marginal rate of substitution, and that of all other consumers buying both items, equals their price ratio.

Unfortunately, this logic does not apply to a large proportion of consumption:

- It unrealistically assumes that consumers and the market are in equilibrium. New products may be undergoing an increase in their market shares, old ones declining.

- There are many commodities on which most consumers, if they spend anything at all, buy only a single unit — one refrigerator, or one car, or one package holiday or one dining table. For such a purchase, consumer surplus (excess of maximum willingness to pay over price) must be larger than it would have been on any substitute, but price differences may poorly reflect quality differences.



In the diagram, consumers A and B value both X and Y at less than their prices, so they buy neither of them. C buys Y, but does not think much of X. D takes the opposite view. E and F would buy either X or Y if the other were not available, but E will obtain the bigger surplus on Y and F will obtain the bigger surplus on X.

Does the price ratio between X and Y, the slope of the dashed line, reflect the market evaluation of their quality difference? As drawn, the diagram suggests that on average, purchasers C, D, E and F value Y relative to X by more than the price ratio of Y to X. Is the right answer the mean slope of the lines from the origin to points A…F or the mean slope only for those consumers who buy?

Can we, less impractically, say that if Y is substituted for X in the index because it is believed that more consumers now buy Y than buy X, Y should be regarded as providing better value for money?

2) Forced replacements

These are made when it is decided to regard a missing observation as permanent. If a past price can be obtained, it should be, but overlap prices are not normally available in such cases. The choice forced upon the statistician is then:

- Carry on with a reduced number of items in the elementary aggregate until it is time to make the next set of optional replacements,
- Pick a replacement and use one of the methods that avoid any quality judgement

- Pick a replacement and make an implicit or explicit quality comparison (see Appendix 1), comparing an adjusted price of the new item with the actual price of the old item, or the actual price of the former with an adjusted price for the latter.

The statistician selecting a replacement for a disappeared representative item normally looks for one within the same elementary aggregate.

It is evident that minimising the number of forced replacements is one criterion of good item and outlet selection. Unfortunately, it may conflict with the criterion of representativity. Consider, for example, the choice between selecting the Bible or a popular novel as a representative book, or the choice between a plain skirt and a fashionable one!

Avoiding inconsistency

We have come across a paradox. If a forced replacement had been foreseen, the replacement item would have been selected in time to achieve an overlap in pricing. This would have allowed application of the customary overlap procedure where replacements are introduced in such a way that they neither raise nor lower the index. So the foreseen replacement would have automatically treated the whole of the price difference as reflecting a quality difference. Yet an explicit quality judgement, possibly yielding a different answer, might have been made had the same replacement been a forced one. Here is an inconsistency!

Such inconsistencies could be avoided.

- In principle, one way would be by making or avoiding an explicit quality adjustment in the case of optional replacement of an item whenever it would have been made or avoided in the case of its forced replacement. But quality adjustment in forced replacement always involves a quality comparison of just two items, the item replaced and the item replacing it. Optional replacement, however, unlike forced replacement, is often of one whole list of items for another list. This will preclude pairwise comparisons unless each new item is selected to replace a particular old item.

- Alternatively, consistency could be secured by always treating the whole of the price differences in the case of forced replacements as reflecting quality differences. This would have the advantage of making any resulting bias of the index the same for forced and optional item replacements — and (as noted below) for forced and optional outlet replacements.

Which is more important: consistency, or reducing bias by making and using quality adjustments in those cases where they are feasible?

Jörgen Dalén has recalculated annual changes in the Swedish Consumer Price Index without quality adjustments using two alternative methods:

1) The price of the replacement is directly compared with that of the replaced item — equivalent to assuming equal quality.

2) The price of the replacement is linked in without showing any price change — equivalent to assuming that the price difference measures the quality difference.

The difference is small but in the direction that might be expected in four of the five years, since when prices were rising/falling one might expect replacements to be mostly more/less expensive than the substituted items.

Year	Alternative 1	Alternative 2
1989	104.02	103.89
1990	105.71	105.20
1991	103.09	103.06
1992	97.49	97.89
1993	102.17	102.18

Quality adjustment methods

This term is a little misleading. It means adjusting a price in the light of a quality difference to make the two prices comparable. The market values of characteristics possessed by the one and not by the other are used to increment or decrement the price of a replaced item or that of its replacement, applying:

- information about differences in production costs, or
- option prices, or
- explicit coefficients derived from hedonic regression, or
- statisticians' or expert advisors' judgement.

Outlet replacement

In principle, the characteristics of a consumption item include not only those characteristics intrinsic to it, but also those of the outlet where it is acquired. However, no feasible procedure for evaluating outlet quality differences has yet been developed. For the great majority of elementary aggregates, linking is therefore unavoidable in practice. This implicitly assumes that the price differences between outlets reflect quality differences so that the impact of outlet replacement on the index is zero. The use of hedonic regressions which included dummy variables for different outlet types would have the same effect.

Developments in retailing appears to have been towards a growth in the number of lower price outlets and, possibly, the closure of higher price outlets. Even if the cheaper outlets provided an inferior level of service, poorer accessibility and so on, the fact that consumers have shifted their purchases towards them indicates that they have gained, so that replacements of such outlets in the sample, reflecting these changes, should have lowered the index. US data studied by Reinsdorf suggest that the upward bias in the index for food could have been between 0.25% and 2% per annum during the eighties and between 0.25% and 0.9% for unleaded petrol.

A similar contrast between the movement of mean prices and the index is observable for some food prices in Europe:

Item	Index	Mean price ratio
Austria 1987—1993		
Frozen chicken	100.2	92.63
Cooking oil	118.32	97.04
Hazelnut biscuits	111.23	98.62
Teabags	109.34	72.59
Leberkäse	126.14	127
Goulash soup	125.82	126.39
Beefsteak	121.26	129.5
Germany 1985—1993		
Pork cutlets	111.0	108.6
Cooked ham	116.5	116.7
Fresh milk, tetrapack, 3.5% fat	112.7	109.0
Apples	92.5	88.8
Edam or Gouda, 45% fat	112.2	107.9
Dark bread	125.9	127.0
Ground coffee	67.3	62.1
Sweden Jan. 1988—Nov.1994		
Loaf of bread	115.6	115.1
Fancy biscuits	113.1	100.0
Potatoes	158.3	129.5
Apples	91.9	94.5

A French study[2] based on the annual food expenditure surveys has estimated that, during the nineteen-eighties, shifts by French consumers in their pattern of purchasing by outlet caused the average food prices of 27 families of products to rise by 0.1% — 0.7% per annum less than their price indexes; probably by 0.4% for food products in general.

[2] Dominique Dubeaux & Alain Saglio, "Modification des circuits de distribution et évolution des prix alimantaires" *Economie et Sttaistique* , 1995-5/6

A Consumer Price Index cannot reflect changes in the cost of shopping due to shifting retail market shares. Keeping the sample up to date and fully representative through optional replacements, though desirable, does not remedy this bias.

How are quality differences expressed

There are three possibilities:
1. As a % of price difference,
2. As a % of the price of the replaced item
3. As the absolute value of the difference.

The third must be treated as continuously variable, whereas 1 and 2 could either be treated as continuously variable or could be coded for 10% or 20% ranges.

1 is awkward when there is a quality difference but there is no price difference. Surely the person making the judgement will start by estimating 2 or 3 and then convert into 1, which makes this form of expression unnecessarily complicated.

3 is obviously appropriate whenever the replaced item or items like it had options which are present or absent in the replacement item and whose prices can be ascertained. Any remaining quality differences could then be expressed in any of the three ways. 2 is the same as the coefficients of a semi-logarithmic hedonic function.

Which differences are important?

Item groups where
1. there are many replacements of dissimilar items,
2. which have a large weight
3. where quality improvement or deterioration is thought to be considerable.

Clothes, shoes, domestic appliances, TV's and cars seem to be the major categories in question. Don Sellwood suggests that there may have been significant deterioration in the quality of some public transport services.

Selecting replacements

The instructions to the price collectors may be
- to apply the same rule as applied originally (e.g. "most sold");
- or, in order to minimise the difficulty of adjusting price for quality difference, to choose the "most similar" item currently available;
- or, to minimise the risk of having to make adjustments in the future, to choose the item most likely to remain on sale.

Thus there is a choice between maximising representativity and minimising the need for quality adjustments. The appropriate choice for any elementary aggregate will depend upon whether or not a method has been established for quality adjusting its prices and how much confidence is felt in its results.

The special case of fashion items.

New models of fashion items — clothes and shoes — are introduced at the beginning of a season; towards the end of the season they are reduced in price and are then superseded by the new season's goods at higher prices. There are therefore two interacting problems:

- In the spring, winter clothes disappear and the new fashion that become available are summer clothes. Then, in the autumn, these disappear and new fashion winter clothes come into the shops.

- During each season, as the fashion becomes less new, prices generally decline. The result is that, for fashion items, the items in the old sample will differ from those in the new sample not only because they are in the new fashion, but also because they are winter instead of summer clothes, or vice versa.

- In the last month of the old season, the proportion of fashion clothes in the sample whose prices are Sale prices will be much larger than in the new sample introduced in the following month for the new season.

The first problem might be dealt with by treating winter and summer clothes as different commodities and applying a variant of the Rothwell procedure outlined in my paper **Three kinds of monthly CPI**.[3] Alternatively, hedonic regression or manufacturers' cost data might allow prices to be quality adjusted in a way covering the major differences in characteristics between winter and summer clothes.

As regards the Sale price problem, it is believed that there is a downward bias in micro indexes for fashion clothing and shoes. (The clothing index in Sweden is estimated to have required upward correction by 1.3%-2.8% to offset it in recent years.) This bias arises because, as the season progresses, more and more fashion items which become in effect less fashionable have their prices reduced. In consequence, the micro index falls. The true price decline is overstated because no correction can be made for the (unmeasurable) change in the characteristic of fashionability.

---

[3] The procedure requires the calculation of weighted average prices for the price-reference year, which is not possible when there were replacements during the season unless (i) approximate monthly sales data are available and (ii) quality adjusted prices were computed on the occasion of each successive replacement.

This would not matter if it were compensated by too large a rise in the micro index when the new sample set of fashion items replaces the old one. But, even if the prices observed for the new fashion items and the old ones do overlap for one month, pairwise quality comparisons are precluded and the prices of the new sample of items can only be linked into the index. Hence the micro index for fashion items does not rise when new fashions are introduced to offset its excessive decline when the previous fashions were gradually ceasing to be new.

The same problem is avoided in the case of forced replacements made during a season. There is then a one to one correspondence between the item removed from the index and its replacement, so pairwise comparison is possible and a quality adjustment for differences in observable characteristics can be made. If the new item, being more in fashion, turns out to be more expensive after adjustment, a price increase will be recorded. There is then symmetry of treatment between the failure to treat the gradual decline in fashionability of the old item as a reduction in quality, and the failure to treat the greater fashionability of the new item as higher quality.

A remedy for the downward bias involved in the seasonal change in the sample is to try to remove it at the end of the sales life of the old fashions by replacing their actual (Sales) prices in their final month in the index with:

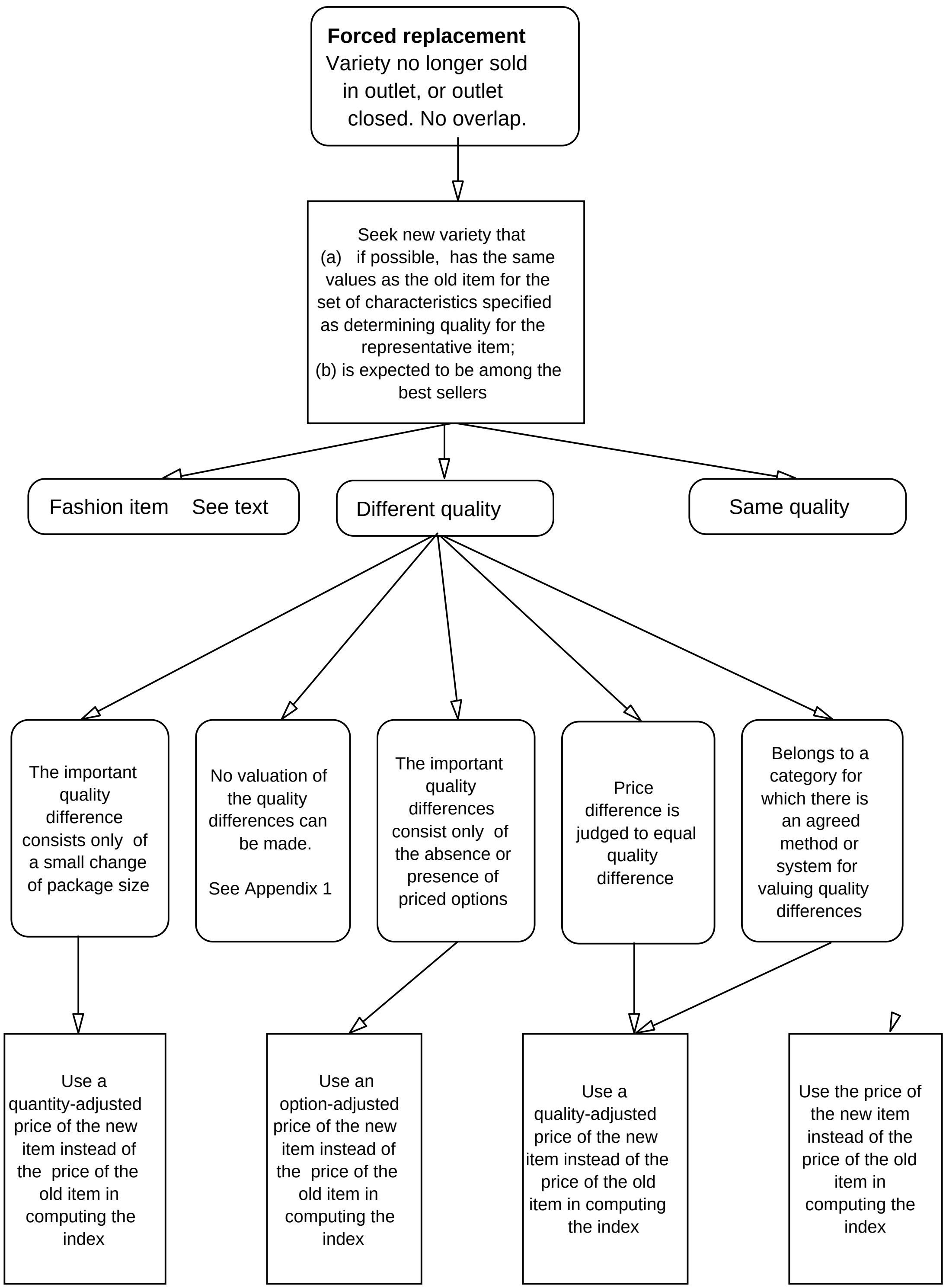
> either, the Regular prices recorded for them some months back when they were still fashionable,
>
> or, especially when inflation is rapid, a forward extrapolation of those earlier prices, using the price movements of similar but non-fashion items,
>
> thus raising the micro index in that final month to which the micro index for the new sample is linked.

The Swedish procedure, applied annually when a new sample of clothes is selected, is to calculate "Sales-price-effects" both (i) for the old sample and (ii) for the overlapping new sample as the mean ratio (or ratio of means) between Sale prices and Regular prices. (The new sample too may include some Sale prices.) The new index is then multiplied by (ii) ÷ (i).

These procedures can fail to capture the true price increase when, as apparently often happens, sellers seize the opportunity of introducing the new fashion items to raise prices.

24.8.95

Appendix 1


Forced replacement
Variety no longer sold in outlet, or outlet closed. No overlap.
Seek new variety that
(a) if possible, has the same values as the old item for the set of characteristics specified as determining quality for the representative item;
(b) is expected to be among the best sellers
Fashion item See text
Different quality
Same quality
The important quality difference consists only of a small change of package size
No valuation of the quality differences can be made.
See Appendix 1
The important quality differences consist only of the absence or presence of priced options
Price difference is judged to equal quality difference
Belongs to a category for which there is an agreed method or system for valuing quality differences
Use a quantity-adjusted price of the new item instead of the price of the old item in computing the index
Use an option-adjusted price of the new item instead of the price of the old item in computing the index
Use a quality-adjusted price of the new item instead of the price of the old item in computing the index
Use the price of the new item instead of the price of the old item in computing the index