ITEM SUBSTITUTION, SAMPLE SPACE AND NEW PRODUCTS 8

Introduction

8.1 As new items are introduced and old items no longer sold, the universe of items from which prices are sampled changes. Yet index number methodology may constrain the sampling to subsets of the universe. The samples selected from such subsets are referred to here as the "sample space" of the index. A focus of this chapter is the limitations of such sample spaces. In Chapter 7 the use of the matched models method was recognized as the accepted approach to ensuring that the measurement of price changes was untainted by changes in quality. It was noted, however, that the approach might fail in three respects: missing items, the limited sample space, and new goods and services (in the remainder of this chapter "goods" is taken to include services). In Chapter 7, several implicit and explicit methods of quality adjustment to prices, and the choice between them, are discussed as ways of dealing with missing items. In this chapter, attention is turned to the two other reasons why the matched models method may fail: sampling concerns (the limited sample space) and new products. The three sources of potential error are first briefly outlined below.

8.2 *Missing items*. A problem arises when an item is no longer produced. An implicit quality adjustment may be made using the overlap or imputation method, or the respondent may choose a replacement item of a comparable quality and its price may be directly compared with the missing item's price. If the replacement is of a non-comparable quality, an explicit price adjustment is required. This was the subject of Chapter 7, paragraphs 7.72 to 7.115. In paragraphs 7.125 to 7.158 a caveat was added. It was recognized that for items in industries where model replacements are rapid, continued long-run matching depletes the sample and quality adjustment becomes unfeasible on the scale required. Chained matching or hedonic indices are deemed preferable.

8.3 *Sampling concerns*. The matching of prices of identical items over time, by its very nature, is likely to lead to the monitoring of a sample of items that is increasingly unrepresentative of the population of transactions. Price collectors may keep following those selected items until they are no longer available. Thus, price collectors may continue to monitor old items, with unusual price changes and limited sales. With regard to item replacement, price collectors may select unpopular comparable items in order to avoid explicit quality adjustments. Thus obsolete items with unusual price changes may be replaced by near obsolete items, again with unusual price changes. That the replacement items are near obsolete will mean that their expenditure shares will be relatively small. This will compound the problem of unrepresentative samples. The substitution of an item with relatively high sales for an obsolete one has its own problems, since the difference in quality is likely to be substantial and substantive, beyond that which can be attributed to, say, the price difference in some overlap period. One item might be in the last stage of its life cycle and the other in the first stage of its life cycle. The problem has implications for sample rotation and item substitution.

8.4 *New products*. A third potential difficulty arises when something "new" is produced. There is a difficulty in distinguishing between new items and quality changes in old ones, and this is discussed below. When a quite new good is produced, there is a need for it to be included in the index as soon as possible, especially if the product is expected to be responsible for relatively high sales. New goods might have quite different price changes from those of existing ones, especially at the start of the life cycle. Furthermore, in the initial period of introduction there is often a welfare gain to the consumer. The new good is not a perfect substitute for the old good and this uniqueness gives economic value to the consumer which would have otherwise not been obtained, had the new good not been available (Trajtenberg, 1989). But by definition, there is no price for the new product in the period preceding its introduction. So even if prices of new products are obtained and included in the index from the initial introduction date, there is still something missing – the initial gain in welfare that consumers experience in the period of introduction. The difficulties in capturing such effects are discussed in paragraphs 8.59–8.60 and Appendix 8.2.

8.5 The problem of missing items was the subject of Chapter 7. This chapter considers sampling concerns arising out of the matched models approach and the problem of introducing new products into the index.

Matched samples

8.6 The matching procedure has at its roots a conundrum. Matching is designed to avoid price changes being contaminated by quality changes. Yet its adoption constrains the sampling to a static universe of items that exist in both the reference and base periods. Outside this matched sample, there are of course the items that exist in the reference period but not in the current period, and are therefore not matched, and similarly those new items existing in the current period but not in the reference period – the dynamic universe (Dalén, 1998a; Sellwood, 2001). The conundrum is that the items not in the matched universe – the new items appearing after the reference period and the old items that disappeared from the current period – may experience price changes that differ substantially from the price changes of existing matched items. This is because these products will embody different technologies and be subject to different (quality-adjusted) strategic price changes. The very device used to maintain a sample of constant quality, i.e., matching, may itself give rise to a sample that is biased away from technological developments. Furthermore, when this matched sample is used to impute the price changes of missing items (see Chapter 7, paragraphs 7.53 to 7.68), it will reflect the technology of a sample that is not representative of current technological changes.

8.7 A formal consideration of matching and the dynamic universe is provided in Appendix 8.1 to this chapter. Three universes are considered:

- an *intersection universe*, which includes only matched items;
- a dynamic *double universe*, which includes all items in the base comparison period and all in the current period, although they may be of different qualities;
- a *replacement universe*, which starts with the base period universe, but also includes one-to-one replacements when an item from the sample in the base period is missing in the current period.

8.8 It is, of course, difficult to ascertain the extent to which matching from the intersection universe constrains the penetration of the sample into the dynamic, double universe, since statistical agencies generally do not collect data for the latter. Its extent will in any event vary between products. Sellwood (2001) advocated simulations using the universe of scanner data. Silver and Heravi (2002) undertook such an experiment using scanner data on the consumer prices of washing machines in the United Kingdom in 1998. A matched Laspeyres index, based on price comparisons with matched models existing in both January and December, covered only 48 per cent of December expenditure on washing machines, as a result of new models introduced after January not being included in the matched index. Furthermore, the January to December matched comparison covered slightly more than 80 per cent of January expenditure, resulting from the exclusion of models available in January but not in December. A biannual sample rotation (rebasing) increased the December expenditure coverage to just over 70 per cent, while a monthly (chained) rotation increased that coverage to 98 per cent (see also Chapter 7, paragraphs 7.128 to 7.131 for further examples). Two implications arise from this. First, the selection of item substitutes (replacements) puts the coverage of the sample to some extent under the control of the price collectors. Guidelines on directed replacements in particular product areas have some merit. Second, chaining, hedonic indices (as considered in Chapter 7, paragraphs 7.125 to 7.158) and regular sample rotation have merit in some product areas as devices to refresh the sample. These are considered in turn.

Sample space and item replacement or substitution

8.9 When an item goes missing, one possibility is for the price collector to select a replacement item. The sample space of the index is thus the matched items initially selected and replacement items selected when matched items are missing. Price collectors are often best placed to select replacement items. The price collectors are often physically present in the same store as the missing item and thus any replacement price selected is likely to be unaffected by price differences which may be attributed to differences in the services (ease of location, parking, warranties, service) provided by different stores. It may also be the case that an obvious replacement is provided by a store which wishes to cater to the same market segment, and this will be conspicuous to the price collector. Sometimes the replacement may have a different code or model number which a desk officer may take to indicate a different item, but which the price collector can identify as simply being a difference of, say, a colour or packaging. Price collectors can also identify whether a new (replacement) model of an item has styling and other qualitative factors so different from the old model that in themselves they would account for substantial price differences. In such instances, a desk officer may only focus on the technical specifications and be unaware of these other differences. Against this, desk officers have additional information. This might include information from a similar store in a different location on the price of the missing item, which might be temporarily out of stock.

8.10 The price collector takes on the task of identifying whether an item is of comparable quality or not. If the price collector judges the item to be comparable when in fact it is not, the quality difference will be taken to be a price difference, resulting in bias where the unrecognized quality changes are in a consistent direction. Informed comparable substitution requires general guidelines on what makes a good substitute, as well as product-specific information on characteristics likely to determine price. It also requires timely substitution, to maximize the probability of an appropriate substitute being available.

8.11 Guidelines for selection of comparable items and monitoring of the nature of the selections are good practice. Liegey (1994) notes how useful the results from hedonic regressions are in the selection of items. The results provide an indication of the major quality factors that explain price variation in the product or service. Price collectors can thus be given guidelines as to which characteristics are important – in the sense that they are price determining – in the selection of the sample and replacement items.

8.12 The matter of sample space requires consideration regarding the selection of replacement items/substitutes for missing items. The initial selection of items whose prices are matched may best be made at random, though such items are more often selected as those "typically" purchased. Similarly items "typically" purchased should be included as replacements. Not all price collectors should aim to sample the same "most typical" item. It is desirable to sample a distribution of items which broadly represents the distribution of purchases. For example, a particular brand – one that accounts for, say, 40 per cent of sales revenue – may be known to be the market leader. This common knowledge should not lead all price collectors to select that brand on rebasing. A representative sample is required.

8.13 Replacement items should intrude into the universe of transactions so that the sample is broadly representative of the dynamic universe. The inclusion of a popular replacement item to refresh the sample – one at the same point in its life cycle as the original popular one selected in the base period – allows for a useful and accurate price comparison, assuming that an appropriate quality adjustment is made. Substitute or replacement items should, where possible, not merely be comparable in quality, but should also be likely to account for a relatively substantial amount of sales value. It is of little merit to substitute a new item with limited sales for a missing item, again with limited sales, just because they have similar features, both being "old"; the index would become more unrepresentative. The replacement of an item only when the item is no longer available may be ineffectual with regard to the representativity of the index. In that case, items with relatively low sales would continue to be monitored until they died. And even replacement might not remedy the situation. If the replacement guidelines indicate that the price collector should select a similar item sold in the outlet, then the replacement selected will be almost as obsolete (Lane, 2001, p. 21).

8.14 Guidelines to select "similar" items are given to ease quality adjustment between the old and new items; at best the items are "comparable" and require no quality adjustment. The institutional mechanism devised to help in making quality adjustments to prices can lead to bias because of its adherence to a sample of items which do not enjoy the benefits of recent technological innovations and are unrepresentative of what is produced. Bear in mind that an index number methodology based on an initially selected matched sample and a sample of substitute replacement items, when items go missing, may not be representative of the universe of all items being consumed. In particular, if the index number methodology is biased to the selection of replacement items with relatively low sales, so that they are comparable with obsolete items, then the sampling from new items and the sample space of the index are biased. Quality adjustment and representativity are interrelated, since the former affects the sample space of the index.

8.15 The importance of, and care required in, the use of replacements to militate against sample depletion is worth reiterating. Consider the case where there is only one model of a product available in the market at the start of the price comparison in period t. A price collector includes it in the sample in period t and then monitors its price in subsequent periods. A new (replacement) model enters the market in, say, period $t+2$, but it is ignored since the original model continues to exist for several months. However, in, say, period $t+9$ the old item is no longer on the market and is replaced, with a quality adjustment, by the new item. The long-run price comparison between the new model's price in period $t+9$ and the old model's price in period t has no sampling bias. Both account for 100 per cent of the market in their respective periods, being the only items available. Both are near the start of their life cycles, so the price comparison is a fair one. If the new and old items have different price changes, sampling bias will occur between periods $t+2$ and $t+8$, when only one of the two items is being sampled, but sampling will be unbiased once the model is replaced in period $t+9$.

8.16 There is thus a case for managing the replacement strategy to minimize sample depletion. In that respect, the following points should be borne in mind:

- Replacements offer an opportunity to cut back on, and possibly remove, sample bias in the period of replacement, though not prior to it.
- The more frequent is the replacement, the less the sample bias.
- If there is more than one new (replacement) item in the market, there may still be bias as only the most popular one will be selected and it may well be at a different stage in its life cycle and thus be experiencing different price changes in comparison with other new (replacement) models.
- The analysis assumes that perfect quality adjustments are made on replacement. The less frequent the replacement, the more difficult this might be to achieve, as the very latest replacement item on the market may have more substantial differences in quality than earlier ones.
- If the best-selling replacement item is of comparable quality and at the same stage in its life cycle as the missing item, then its selection will minimize sample bias.
- If there is more than one replacement item and the most comparable one – having the old technology – is selected, it will have low market share and unusual price changes.
- Given advance information on market conditions, replacements that are included in the sample well before the old item becomes obsolescent are likely to increase the sample's share of the market, include items that are more representative of the market, and facilitate quality adjustment.

8.17 The problem of item substitution is analogous to the problems that arise when an outlet closes. It may be possible to find a comparable outlet not already in the sample, or a non-comparable one for which, in principle, an adjustment can be made for the better quality of service provided. It is not unusual for an outlet to close following the introduction of a new, more competitive outlet. Where the matching of prices between these outlets broadly follows the consumption patterns of the clients of the original outlet, there is an obvious replacement outlet. If, however, the new outlet has comparable prices but, say, a better range of items, parking and service, there is a gain to consumers from substituting one outlet for the other. Yet since such facilities have no direct price, it is difficult to provide estimates of their value in order to make an adjustment for the better quality of service of the new outlet. The index would thus have an upward bias, which would be lost on rebasing. In such cases, replacing the old outlet by a new one that provides a similar standard of service may be preferable to replacing it by one that has a different standard but serves the same catchment area. In their regression analyses for consumer durables, Liegey (2000), Shepler (2000) and Silver and Heravi (2001b) found "outlet type" to be a substantial and statistically significant explanatory variable for price variation, while for a particular outlet type – grocery outlets, for food and petrol prices in the United States – Reinsdorf (1993) found much smaller differences.

Sample rotation, chaining and hedonic indices

8.18 It is important to recognize the interrelationships among the methods for handling item rotation, item replacement and quality adjustment. When consumer price index (CPI) item samples are rotated, this is a form of item substitution, except that it is not "forced" by a missing item, but is undertaken for a general group of items to update the sample of items. It has the effect of making future forced replacements less likely. Yet the assumption implicit in its use is equivalent to that for the overlap adjustment technique: that price differences are an adequate proxy for the change in price per unit of quality between items disappearing from the sample and replacement items.

8.19 Consider the initiation of a new sample of items. This may be by probability or judgemental methods, or a combination of the two. Prices for the old and new samples are returned in the same month, and the new index is compiled on the basis of the new sample, the results being linked to the old. This is an implicit use of the overlap method, in which all price differences between the new and old items in that month are taken to be quality changes. Assume that the new sample is initiated, say, in January. Assume also that the prices of an old item in December and January are \$10 and \$11 respectively, a 10 per cent increase, while those for the new replacement item in January and February are \$16 and \$18 respectively, an increase of 12.5 per cent. The new item in January is of a better quality than the old, and this difference in quality may be worth $\$16-11=5$ to the consumer. That is, the price difference is assumed to be equal to the quality difference, which is the assumption implicit in the overlap method. Had the price of the old item in December been compared with the quality-adjusted price of the new item in January under this assumption, the price change would in this case be the same, 10 per cent (i.e. $(16-5)/10=1.10$). In practice, the need to simultaneously replace and update a large number of items requires the assumptions of the overlap method, the point being that this process should not be regarded as error free. In cases where the assumptions are considered likely to be particularly untenable (discussed in Chapter 7, paragraphs 7.44 to 7.52), explicit adjustments of the form discussed in paragraphs 7.72 to 7.115 should be used.

8.20 It was noted above that when samples are updated, any difference in the average quality of items between the samples is dealt with in a way that is equivalent to the overlap adjustment technique. Sample rotation to refresh the sample between rebasing is an expensive exercise. If rebasing is infrequent, however, and if there is a substantial loss of items in particular product areas, then sample rotation might be appropriate for those areas. A *metadata* system (described below) will aid such decision-making. More frequent sample rotation aids the process of quality adjustment in two ways. First, the new sample will include newer varieties. Comparable replacements with substantial sales will be more likely to be available and non-comparable ones will be of a similar quality, which facilitates good explicit adjustments. Second, because the sample has been rotated, there will be fewer missing items than otherwise and thus less need for quality adjustments.

8.21 A natural extension of more frequent sample rotation is to use a chained formulation in which the sample is reselected each period. In Chapter 7, paragraphs 7.153 to 7.158, the principles and methods were outlined in the context of sectors in which there was a rapid turnover of items. These principles are echoed here. Similarly, the use of hedonic indices (as outlined in paragraphs 7.132 to 7.152) or the use of short-run comparisons (discussed in paragraphs 7.159 to 7.173) might be useful in this context.

Information requirements for a quality adjustment strategy

8.22 It should be apparent from the above that a strategy for quality adjustment must not only be linked to one relating to sample representativity, but must also require the building of a statistical metadata system. This is not an area where the approach for the index as a whole can be simply described, but one that requires the continual development of market information and the recording and evaluation of methods on a product-by-product basis.

Statistical metadata system

8.23 The methods used for estimating quality-adjusted prices should be well documented as part of a statistical metadata system. Metadata are systematic descriptive information about data content and organization that help those who operate the systems that produce statistics to remember what tasks they should perform and how they should perform them. A related purpose is to train new staff and introduce them to the production routines (Sundgren, 1993). Metadata systems also help to identify where current methods of quality adjustment require reconsideration, and prompt the use of alternative methods. They may also serve user needs, the oldest and most extensive form being footnotes.

8.24 The dramatic increase in volume of statistical data in machine-readable form, with a concomitant increase in metadata, argues for keeping the metadata in such a form. This is to enhance transparency in the methods used and help ensure that the methods are understood and continued, as staff leave the CPI team and new staff join it. Changes in quality adjustment methodology can in themselves lead to changes in the index. Indices produced using new procedures should be spliced onto existing indices. The metadata system should also be used as a tool to help with quality adjustment. Because so much of the rationale for the employment of different methods is specific to the features of the products concerned, data should be held on such features.

8.25 Statistical agencies should monitor the incidence of missing items against each Classification of Individual Consumption according to Purpose (COICOP) group. If that incidence is high, then the monitoring should be carried out by class within each group. Again, if that incidence is high, the monitoring should be done by elementary aggregate or selected representative items within each group, or at the most detailed level of the system. Where the incidence is high, the ratios of temporary missing prices, comparable replacements and non-comparable replacements to the overall number of prices, and the methods for dealing with each of these three circumstances, should also be monitored to provide the basis of a statistical metadata system. The advantage of a top-down approach is that resources are saved by monitoring only at the detailed level product areas which are problematic.

8.26 Product-specific information, such as the timing of the introduction of new models, pricing policies (especially with regard to months in which no changes were made) and the popularity of models and brands according to different data sources, should be included in the metadata as the system develops. An estimate, if available, of the weight of the product concerned should be given, so that a disproportionate effort is not given to relatively low-weighted items. All this will lead to increased transparency in the procedures used and allow effort to be directed where it is most needed.

8.27 For items for which replacement levels are high, the metadata system would benefit from contacts between statistical agencies and market research organizations, retailers, manufacturers and trade associations. Such links will allow staff to better judge the validity of the assumptions underlying implicit quality adjustments. Where possible, staff should be encouraged to be responsible for learning more about specific industries whose weights are relatively high and where item replacement is common.

8.28 Statistical staff should identify price-determining characteristics for product areas using hedonic regressions, information from market research, store managers, trade associations and other such bodies, and the experience of price collectors. This information should contribute to the statistical metadata system and be particularly useful in providing subsequent guidelines on item selection.

8.29 When hedonic regressions are used either for partial patching of missing prices or as indices in their own right, information on the specification, estimated parameters and diagnostic tests of the regression equations should be kept, along with the data and with notes as to why the final formulation was chosen and used. This will allow the methodology for subsequent updated equations to be benchmarked and tested against the previous versions.

8.30 The metadata system should help statistical staff to:

– identify product areas likely to be undergoing regular technological change;
– ascertain the pace at which models change and, possibly, the timing of changes;
– undertake an analysis of what have in the past been judged to be "comparable" replacements in terms of the factors that distinguish the replacement and old items;
– identify whether different price collectors are making similar judgements regarding comparable replacements, and whether such judgements are reasonable.

8.31 Price statisticians may have more faith in the use of some quality adjustment procedures than others. When such procedures are used extensively it might be useful to note, as part of the metadata system, the degree of faith the statistician has in the procedures. Following Shapiro and Wilcox (1997b), this may be envisaged as a traditional confidence interval: the statistician may believe at, say, a 90 per cent level of confidence that the quality-adjusted price change is 2 per cent (0.02) with an interval of plus or minus 0.5 per cent (0.005). There may be an indication as to whether the interval is symmetric, or positively or negatively one-sided. Alternatively, statisticians may use a simple subjective coding on a scale of, say, 1 to 5.

New products and how they differ from products with quality changes

8.32 The question arises of how to define new products (goods and services) and to distinguish them from existing products whose quality has changed. A new model of a good may provide more of a currently available set of service flows. For example, a new car model may differ from existing ones in that it may have a bigger engine. There is a continuation of a service and production flow, and this may be linked to the service flow and production technology of the existing models. A practical definition of a new good, as against quality changes in an updated existing model, is that, first, the new good cannot easily be linked to an existing item as a continuation of an existing resource base and service flow, because of the very nature of its "newness". For example, frozen foods, microwave ovens and mobile phones, while extensions of existing flows of services to the consumer, have a dimension of service that is quite new. Second, as discussed below, new goods can generate a welfare gain to consumers by their very introduction. The simple introduction of the new good into the index, once two successive price quotes are available, misses this gain.

8.33 Oi (1997) directs the problem of defining "new" goods to that of defining a monopoly. If there is no close substitute, the good is new. He argues that individual new books, new videos and new television serials may have quite small cross-price elasticities in some cases; their shared service is to provide entertainment and they are similar in this respect. Hausman (1997), however, found cross-elasticities for substitution to be quite substantial for new television serials (though see Bresnahan (1977)). There are many new forms of existing products, such as fashionable toys and clothes, which are not easily substitutable for similar items and for which consumers would be willing to pay a premium.

8.34 Bresnahan (1997, p. 237) notes that *Brandweek* counted over 22,000 new-product introductions for the United States for 1994 – the purpose of their introduction being, as differentiated products, not to be exact substitutes for existing ones, but to be distinct. Their distinctiveness is in many cases the rationale behind their launch. The extent of differentiated markets nevertheless makes the definition and treatment of such things as "new" impractical. Oi (1997, p. 110) sets out the pragmatic case: "Our theory and statistics would be unduly cluttered if separate product codes had to be set aside for Clear Coke and Special K." Furthermore, the techniques for including such products are not, as indicated below, readily applicable. The sound practical advice given by Oi (1997) to keep matters "uncluttered" is therefore not unreasonable.

8.35 The terminology adopted here is that used by Merkel (2000) for producer price index (PPI) measurement, but considered in a CPI context. It distinguishes between *evolutionary* goods and *revolutionary* goods. Evolutionary goods are replacement or supplementary models that continue to provide a similar service flow, but maybe in new ways or to different degrees. These are distinguished from revolutionary goods, which are entirely new goods not closely tied to a previously available product. Although revolutionary goods may satisfy a long-standing consumer need in a novel way, they do not fit into any established CPI item category (Armknecht et al., 1997). Problems are associated with incorporating distinctly new revolutionary goods. This is because a good, which by its nature is unique, is unlikely to be incorporated into the sample as a replacement for an existing item. It would neither be comparable nor be amenable to explicit adjustments to its price for quality differences with existing goods. Since a distinctly new item is not replacing an item, it does not have an existing weight; its introduction therefore implies a need to re-weight the index.

Incorporation of new products

8.36 There are three major concerns regarding the incorporation of new goods into the CPI. The first concern is the detection and identification of the new good; these are facilitated by close links with market research, and producer and trade associations. The second concern, which is related to the first, is the decision on the need and timing for their inclusion. This refers to both the weight and price changes of the new good. The third concern relates to the incorporation of the initial welfare to the consumer arising from the switch from the old technology.

8.37 Consider some examples on the timing of the introduction of new goods. The sales of mobile phones were at such a significant level in some countries that their early inclusion in the CPI became a matter of priority. They simply rose from nothing to relatively quickly account for quite a large proportion of sales in their product classification. Furthermore, their price changes were atypical of other goods in their product classification. Being new, they might be produced using inputs and technologies quite different from those used for existing telephones. Because of substantial marketing campaigns, many new goods command substantial sales and are the subject of distinct pricing strategies at launch. For radical innovations, however, there may be a delay in their incorporation into the index since they cannot be defined within existing classification systems.

8.38 Armknecht et al. (1997) cite the example of the incorporation of video cassette recorders (VCRs) into the United States CPI. VCRs were launched in 1978 with a sales value of US$299 million and estimated average retail price of US$1,240. Because the CPI was rebased every ten years, VCRs were introduced into the CPI only in 1987 when their sales value was US$3,442 million and average price had fallen to US$486. All the extraordinary price movements between 1978 and 1987 were thus missed by the index.

8.39 Dulberger (1993) provides some estimates for United States PPIs for dynamic random access memory (DRAM) computer memory chips. She calculated price indices for the period 1982 to 1988 with varying amounts of delay in introducing new chips into the index. The indices were chained so that new chips could be introduced, or not, as soon as they had been available for two successive years. Using a Laspeyres chained index, there was a fall of 27 per cent if there was no delay in introducing the new goods, as compared with falls of 26.2 per cent, 24.7 per cent, 19.9 per cent, 7.1 per cent and 1.8 per cent if the introductions were delayed by, respectively, 1, 2, 3, 4 or 5 years. In all cases the index was biased downwards because of the delay. Berndt et al. (1997) provide a detailed study of a new anti-ulcer drug, Tagamet. They found that the effects of pre-introduction marketing of the drug on its price and market share at introduction were substantial. Not unexpectedly, there were price falls for the generic form of a pharmaceutical on the expiry of the patent, but there were increases for the branded form; loyal customers were found to be willing to pay a premium over the price prior to the patent expiry (Berndt et al., 2003).

8.40 Waiting for a new good to be established or waiting for the rebasing of an index before incorporating new products may lead to errors in the measurement of price changes if the unusual price movements at critical stages in the product life cycles are ignored. Strategies are required for the early identification of new products, and mechanisms are needed for their incorporation either at launch (if preceded by major marketing strategies) or soon after (if there is evidence of market acceptance).

These strategies and mechanisms should form part of the metadata system. Waiting for new products to achieve market maturity may result in an implicit policy of ignoring the quite disparate price movements that accompany their introduction (Tellis, 1988, and Parker, 1992). This is not to say that new goods will always have different price changes. Merkel (2000) gives the example of "light" varieties of foods and beverages, which are similar to the original varieties but with fewer calories. The prices of "light" products are very close to the prices of the original products. The introduction of "light" varieties simply serves to expand the market. While there is a need to capture such expansion when the weights are revised, the price changes for the existing items can be used to capture those of the "light" items.

8.41 The second measurement concern with respect to new products is the incorporation of the effect of those products at launch. The preceding discussion is concerned with the incorporation of price changes into the index once two successive quotations are available. Yet there is a gain to the consumer when comparing the price in the first of these periods with the price in the period that preceded the introduction of the product. This latter price is a hypothetical price. It is the price which would make the demand from the community for the product equal to zero; that is, it is the reservation price which, when inserted into the demand function, sets demand to zero. If a demand system can be estimated, so too can the reservation price. The virtual reservation price is compared with the actual price in the period of introduction, and this is used to estimate the surplus from the introduction of the good. If the reservation price is relatively high, then the introduction of the new good is clearly of some benefit to the consumer. To ignore this benefit, and the change from the shadow price to the actual price in its period of launch, is to ignore something of the price movements that give rise to improvements in the standard of living. Of course, if a "new" good is a close substitute – at the price it is brought into the index – for goods already in existence, then no additional consumer surplus is generated.

8.42 It should be noted that a consumer may be in a geographical area in which a new good or service, say, cable television, a video rental outlet or health facility, is not present. The benefits of the new good on its introduction to different geographical areas will therefore develop over time as the new good becomes more generally available. The benefits will emerge again and again for each sector of the population that benefits from access to the new product. In practice, such items gain increasing weight as the index is rebased or the sample rotated.

8.43 The methods outlined below for the inclusion of substitute and new goods include both normal CPI procedures and exceptional treatments. In regard to the former, consideration is given in paragraphs 8.44 to 8.58 to the rebasing of the index, rotating of items, introduction of new goods as replacements for discontinued ones on rotating, and a strategy for dealing with new item bias. In regard to the latter, techniques that require different sets of data are outlined. The use of chained matched models and hedonic indices was discussed in Chapter 7 in the context of products experiencing rapid turnover in models. Analytic frameworks that consider new goods bias by way of reservation prices and substitution effects are considered in paragraphs 8.59 and 8.60 and Appendix 8.2. The data requirements and econometric expertise are much more demanding for these approaches.

Sample rebasing and rotation

8.44 A new good may be readily incorporated in the index at the time of rebasing the index, or when the whole or the pertinent part of the sample is rotated. If the new good has, or is likely to have, substantial sales, and is not a replacement for a pre-existing one, or is likely to command a much higher or lower market share than the pre-existing one it is replacing, then new weights are necessary to reflect this. New weights are only fully available on rebasing, not on sample rotation. There will thus be a delay in the new item's inclusion in the index. The extent of the delay will depend on how close the introduction of the item is to the next rebasing and, more generally, the frequency with which the index is rebased. This discussion of rebasing is effectively concerned with the use of new weights for the index. Even if the index is rebased annually and chained, there will be a delay until the annual rebasing before weights can be assigned, and there may even be a further six-month delay for the sampling and collating of the survey results for the weights. Such frequent rebasing allows for the early introduction of a new good and is to be advised when the weights are not keeping pace with product innovations.

8.45 At the elementary level of aggregation, an implicit weight equal to the expenditure share is given by the Jevons index, for example, to each price relative. The Dutot index gives each price change the weight of its price relative to the sum of the prices in the initial base period of the comparison (see Chapter 7). If a product area is expected to be subject to dynamic innovations, then the sample may be increased on rotation, without any changes to the weight for the group. There would simply be more items selected to form the arithmetic or geometric average price change. As new varieties become available, they could be substituted for some of the existing ones, there being a wider range from which to draw a comparable item, or less effort involved in the quality adjustment procedure for a non-comparable one.

8.46 Some statistical agencies rotate (resample) items within product groups. Opportunities exist to introduce new items within a weighted group under such circumstances. The resource practicalities of such schemes require items to be rotated on a staggered basis for different product groups. Product groups experiencing rapid change should be rotated more frequently. The incorporation of new goods using sample rotation allows some of the existing weight of the product group to be reallocated to the new good. Yet it implicitly uses the overlap method for the introduction of the new good of a different quality. The difference in prices in the overlap period of the new and obsolete items is assumed to be equal to their quality difference. The assumptions implicit in such procedures have been outlined above and their likely veracity needs to be considered. Since evolutionary items are defined as continuations of the service flow of existing (and possibly exiting) ones, the hedonic framework may be more suitable in some cases to the use of the overlap method. These and other methods and the choice between methods are discussed in Chapter 7.

8.47 In many countries, rebasing is infrequent and sample rotation is not undertaken, despite their advantages. The rotating of samples on a frequent basis should not, however, be considered a panacea. Sample rotation is an arduous task, especially when performed over a range of product groups experiencing rapid change. Even frequent rotation, say, every four years, may miss many new goods. Yet it is not necessary for statistical agencies to wait until an item is obsolete before a new one is introduced. It is quite feasible for statistical agencies to pre-empt the obsolescence of an old item and decide on its early substitution by a new one. In some product areas, the arrival of a new good is well advertised in advance of the launch. In other areas, it is feasible for a statistical agency to have general procedures for substitutions, as outlined below. Without such a strategy, and where rotation or rebasing is infrequent, a country would be open to serious new product bias.

8.48 In summary:

- The treatment of a new good as a replacement for an existing one can be undertaken if the old item's weight suitably reflects the new good's sales and if a suitable quality adjustment can be made to its price to link it to the existing, old, price series.
- If the new good does not fit into the pre-existing weighting structure, it can be included on rebasing, though this may be infrequent in some countries.
- Regular sample rotation provides a means by which the inclusion of such items can be formally reconsidered, though since this is undertaken on a staggered basis, only the weights within the product group are reallocated, not those between the groups.
- Directed sample substitution, as opposed to waiting for sample rotation, may be used to pre-empt the arrival of new goods.
- Revolutionary items will not fit into existing weighting structures and alternative means are required.
- The modified short-run or chained framework outlined in Chapter 7, paragraphs 7.153 to 7.173, may be more appropriate for product areas with high turnover of items.

Directed replacements for evolutionary items and directed augmentation of the sample for revolutionary goods are considered below.

Directed replacements and sample augmentation

8.49 For evolutionary goods in product areas where there is a rapid replacement and introduction of such goods, a policy of directed replacements might be adopted. Judgement, experience, discussions with store managers, market research companies and a statistical metadata system should help identify such products. The selection of replacements is directed to evolutionary items in order to ensure that the index maintains its representativity. If the new version of a product is designed as a replacement for an existing one, then substitution might be automatic. Once a substitute has been made, the prices require an adjustment for the quality difference using, perhaps, the overlap method, imputation, or an explicit estimate based on production or option costs, or a hedonic regression as discussed in Chapter 7.

8.50 The management of the directed substitution can take a number of forms. It can comprise instructions to price collectors who are informed of defined configurations of a product, such as "high end", "mainstream", "economy", "entry level" and "other" (Lane, 2001). Directions might also be given as to the proportions expected of items at these levels, say, 20 per cent of the market should be "high end". Such information should be based on actual data or judgement of specialists. The configurations are revised, say, every six months. What was "high end" at the start of the period may now be "entry level" and the price collectors will have new configurations indicating what the desired replacements should look like. They are directed to particular replacements. Alternatively, the price collector might be responsible for the selection of replacements, either after discussion with store managers or, if an indication of the market share of the popular makes is given, with probability proportionate to size. There are, of course, other variants. In such markets, the desired end effect is that replacement items likely to be representative of substantial sales are selected and that this selection is made earlier rather than later. The point is not to miss the birth of such items and to facilitate quality adjustment.

8.51 It is important to emphasize that, on the introduction of new versions of these evolutionary goods, a particularly high price may be charged to take advantage of segments of the market willing to pay a premium for the "newness" of the item. Alternatively, a particularly low price may be charged to introduce the good to the market in order to help gain acceptance. After a while, prices may be changed as the novelty of the item wears off or as it gains acceptance, or as competitors bring out improved products. Directed substitution is important in ensuring that the CPI captures the unusual price increases at the launch. It is also necessary in ensuring that the coverage of items becomes more representative. Although directed substitution allows for both, a caveat applies. If the overlap method is used, the item is introduced on the assumption that the price difference between the old and new items equates to their quality difference. For example, if a new type of detergent is introduced with a new, biological cleaning action, it may be that the typical consumer is willing to pay a price of 10 against the existing standard detergent's price of 8. With no explicit estimate of the additional usefulness or utility to be gained from the biological action, the overlap method implicitly assumes it is worth 2. Yet there may be an introductory launch price of 8 and the price may later increase to 10. At the time of the overlap, the two prices would be the same, there being no adjudged quality difference. In fact, the quality-adjusted price would be falling; there is a quality difference of 2, but this cannot be deduced by the statistical office. In general, therefore, when there is evidence of items being launched at unusual prices and the overlap method is used, it is better to make the replacement later when the market has settled.

8.52 For revolutionary goods, substitution may not be appropriate. First, they may not be able to be defined within the existing classification systems. Second, a major part of their uniqueness may be the manner in which they are sold, which will require extending the sample to such new sales channels. Third, there will be no previous items to match such goods against in order to make a quality adjustment to prices since, by definition, they are substantially different from pre-existing goods. Finally, there is no weight to attach to the new outlets or items.

8.53 The first need is to identify new goods. The suggested contacts with market research companies, outlet managers and manufacturers, mentioned above in relation to producing a supportive metadata system, are also pertinent here. Once the new goods are identified, sample augmentation is appropriate for the introduction of revolutionary goods. It is necessary to bring the new revolutionary goods into the sample, in addition to what already exists in the sample. This may involve extending the classification, the sample of outlets, and the item list within new or existing outlets. The choice of means by which the new goods are introduced is more problematic.

8.54 Once two price quotes are available, it should be possible to splice the new good onto an existing or obsolete one. This of course misses the impact of the new item in its initial period. As discussed below, however, including such effects is not a trivial exercise. Consider the linking of a good that is likely to be replaced in the market by the new good. For example, a relatively new electrical kitchen appliance might follow the price index for existing kitchen appliances up to the period of the link, and then the price changes for the new good in subsequent periods. This would create a separate and additional price series for a new good that augments the sample, as illustrated in Table 8.1. Item C is new in period 2 and has no base period weight. Its price change between periods 1 and 2, had it existed, is assumed to follow the overall index for products A and B. For period 3 onwards, a new linked price series is formed for C, which for period 3 is $101.40 \times 0.985 = 99.88$ and for period 4 is $101.40 \times 0.98 = 99.37$. New revised weights in period 2 show C's weight to be 20 per cent of all the items. The new index for period 3 is:

$$101.40 \times [0.8 \times (101.9/101.4) + 0.2 \times (99.88/101.4)]$$
$$= 0.8 \times 101.9 + 0.2 \times 99.88 = 101.50$$

and for period 4:

$$101.40 \times [0.8 \times (102.7/101.4) + 0.2 \times (99.37/101.4)]$$
$$= 0.8 \times 102.7 + 0.2 \times 99.37 = 102.05$$

8.55 If C were an evolutionary good replacing B, then there would be no need to introduce new weights and no need to augment the sample. The revolutionary good C has no weight in the base period; the splicing thus requires a revision of the weights at the same time. Both the selection of the series onto which the new item is spliced, and the product groups selected for the weight revision, require some judgement. Items whose market share is likely to be affected by the introduction of the new good should be selected. If the new good is likely to be responsible for a significant share of expenditure, such that it will affect the weights of a broad class of product groups, then there may be a case for a realignment of the overall weighting procedure. Such seismic shifts can of course occur, especially in the communications industries, and for a wider range of markets when trade barriers are relaxed in less-developed economies or when regulations are removed. The change in weights may also be required for disappearing goods no longer sold in an economy. In that case, the weights of these goods need to be reassigned. As noted in Chapter 7, paragraphs 7.132 to 7.158, chaining and hedonic indices may well be appropriate when there is a rapid turnover in such new and obsolete goods. Chaining is an extension of the above procedure and can be used to introduce a new good as soon as it is available for two successive periods.

Table 8.1 Example of sample augmentation

Products	Base weight	Revised weight	Period 1	Period 2	Period 3	Period 4
A	0.6	0.5	100.00	101.00	101.50	102.50
B	0.4	0.3	100.00	102.00	102.50	103.00
All items		*0.8*	*100.00*	*101.40*	*101.90*	*102.70*
C				100.00	98.50	98.00
Spliced C		0.2	100.00	101.40	99.88	99.37
All items (revised)			*100.00*	*101.40*	*101.50*	*102.05*

8.56 Item augmentation may also be used for evolutionary goods that are likely to be responsible for a substantial share of the market, while not displacing the existing goods. Say, for example, that a country has a local brewery and that a licensing agreement with a foreign brewery has led to the joint production of two beers, under different brand names. Say the market share for beer from the brewery remains the same, but one segment of the market now drinks foreign as opposed to domestic beer. Price collectors may be directed to a forced substitution of some of the sample of domestic beers for foreign ones, the weight remaining the same. This would be similar to a quality adjustment using a non-comparable replacement, as discussed in Chapter 7, paragraphs 7.72 to 7.115. Alternatively, the sample may be augmented since there is concern that a smaller sample of domestic beers may now not be sufficiently representative. The augmentation process may be similar to that outlined in Table 8.1, with the new foreign beer C accounting for 20 per cent of the market. If the advent of foreign beers displaced some of the alcoholic spirits market, say, then the revision of weights would extend into that product group. As noted in Chapter 7, paragraphs 7.125 to 7.158, chaining and hedonic indices may well be appropriate when there is a rapid turnover in such new and obsolete goods. With chaining, the good needs to be available for only two successive periods to allow for its introduction.

8.57 In some instances a directed replacement is required for evolutionary and revolutionary outlets. Forced augmentation of the sample of outlets may be implemented so that new goods available only in specific outlets are included. This is especially likely in the service sector, where a new service is particular to specific outlets, for example cyber cafés or online retailers. The procedures are similar to those described for items. For example, in the above example, instead of products A, B and C, consider C as a new outlet in addition to outlets A and B. Some estimate would be required of its expected sales share to form the revised weights.

8.58 The effect of a new outlet on the index depends on how it is included, as well as the nature of the market and its reaction to the new outlet. First, if a new outlet offers some innovation which induces some consumers to shop there, there is an increase in usefulness or utility. Because of imperfect knowledge about the new outlet or different preferences of different segments of the market, the old outlet may not close down. There is no natural prompt for the new outlet to be introduced into the CPI, as there is with the closure of an old outlet. The start-up of the new outlet may have been apparent to the statistical office. If the new outlet is expected to have substantial sales, it may augment the sample. It may be spliced onto the index in the manner of item C above. Such a methodology would not include the welfare gain to consumers arising from the uniqueness of the outlet (Trajtenberg, 1989), since price comparisons are only being undertaken once it has been introduced. The initial welfare effect is between the period prior to its existence and the period of its introduction. Second, all other outlets might lower their quality-adjusted prices to match those from the new outlet. The fall in price and gain in usefulness or utility arising from the new outlet's technology would then be captured by the CPI. Finally, outlets may appear that offer a wider range of options in terms of goods and service, which is valued by consumers and is therefore an improvement in the standard of living via the gain in utility. There is nothing in current CPI methodology that allows for the valuation of such gains (Shapiro and Wilcox, 1997a).

Reservation prices

8.59 Shapiro and Wilcox (1997a, p. 144) expressed concerns over:

> ...the rare new item that delivers services radically different from anything previously available. For example, even the earliest generation of personal computers allowed consumers to undertake tasks that previously would have been prohibitively expensive. This problem can be solved only by estimating the consumer surplus created by the introduction of each new item. Hausman (1994) [republished as Hausman (1997)] argues that this must involve explicit modeling of the demand for each new item. Although explicit modeling of demand may be of dubious practicality for widespread implementation in the CPI, strategic application in a few selected cases might be worthwhile.

8.60 The technical means for such estimates is recognized as being beyond the practical capabilities of a statistical agency. More disturbing is that the argument for the inclusion of such effects extends from revolutionary new goods to the clutter of evolutionary items such as new breakfast cereals. Appendix 8.2 provides some details of a generalized Laspeyres approach which takes account of substitution between new and old models. Given the complexity of the estimation systems involved, however, this manual envisages a pragmatic approach which would initially exclude such effects.

Summary

8.61 The need to consider the sample space of the items selected by the index number methodology and new goods arises out of a very real concern with the dynamic nature of modern markets. New goods and quality changes are far from being a new phenomenon. As Triplett (1999) has argued, it has not been demonstrated that the rate of new product development and introduction is much higher now than in the past. It is certainly accepted, however, that the number of new products and varieties is substantially greater than before. Computer technology provides cost-effective means for collecting and analysing very large sets of data. Chapter 6 considers the use of hand-held computers for data capture, and the availability of bar-code scanner data. The proper handling of such data requires consideration of aspects beyond those normally taken into account in regard to the static intersection universe which underscores matched samples. Appendix 8.1 to this chapter provides an outline of these sampling issues.

8.62 The following important points should be borne in mind:

- Where nothing much in the quality and range of goods available changes, use of the matched models method presents many advantages. The matched models method compares like with like, from like outlets.
- Statistical metadata systems are needed to help identify the product areas in which matching provides few problems, and to focus attention on those areas that are problematic. They show how to collect and provide the information that will facilitate quality adjustment. They also allow for transparency in methods and they facilitate retraining.
- Where there is a very rapid turnover in items such that serious sample depletion takes place quickly, replacements cannot be relied upon to make up the sample. Alternative mechanisms, which sample from or use the double universe of items in each period, are required. These include chained formulations and hedonic indices, as discussed in Chapter 7, paragraphs 7.125 to 7.158.
- Some new goods can be treated as evolutionary and incorporated using non-comparable replacements with associated quality adjustments. The timing of the replacement is critical for both the efficacy of the quality adjustment and the representativity of the index.
- Instructions to price collectors on the selection of replacements are important, for they too have a bearing

on the representativity of the index. The replacement of obsolete items with newly introduced items, in turn, leads to difficulties in undertaking quality adjustments, while their replacement with similar items leads to problems with representativity.

- Sample rotation is an extreme form of the use of replacements, and is one mechanism for refreshing the sample and thus increasing its representativity. Against this, however, is the possibility of bias arising from the implicit assumptions underlying the overlap procedure for quality adjustment not being met.
- Revolutionary goods may require the augmentation of the sample to make room for new price series and new weighting procedures. The classification of new goods into evolutionary goods and revolutionary goods has a bearing on the strategy for their introduction, directed replacement (substitution) and sample augmentation.
- The initial gain in consumer welfare arising from new items and loss in welfare because items disappear are not captured by either of these procedures. Econometric estimates of reservation prices provide an approach that is theoretically appropriate, although problematic in practice.

Appendix 8.1 Appearance or disappearance of products or outlets

1. In previous chapters, it was generally assumed that the target quantity for estimation could be defined in terms of a fixed set of products. Here we consider the complications arising from the fact that the products and outlets are continually changing. The rate of change is rapid in many industries. Sampling to estimate price changes is thus a dynamic rather than a static problem. Somehow, the prices of new products and the prices in new outlets have to be compared to old ones. Whatever methods and procedures are used in a price index to handle these dynamic changes, the effects of these procedures will always amount to an explicit or implicit estimation approach for this dynamic universe.

The representation of change in a price index

2. From a sample selection perspective, there are three ways of handling dynamic changes in an elementary aggregate universe (Dalén, 1998a), where varieties and outlets move in and out:

– by resampling the whole elementary aggregate at certain points in time;
– by a one-to-one replacement of one variety or outlet for another;
– by adding and deleting single observation points (items in outlets) within an index link.

Sample rotation

3. By resampling it is meant that the old sample is reconsidered as a whole so as to make it representative of the universe in a later period. This does not necessarily mean that all or even most sampling units have to be changed, only that a fresh look is taken at the representativity of the whole sample and changes are undertaken, as appropriate. The methods used for resampling could be any of those used for the initial sampling. In the case of probability sampling, every unit belonging to the universe in the later period needs to have a non-zero probability, equal to its relative market share, of being included in the sample.

4. Resampling (or sample rotation) is traditionally combined with the overlap method outlined in Chapter 7, paragraphs 7.45 to 7.52. It is a similar procedure to that used when combining two links in chain indices. The first period for which the new sample is used is also the last period for which the old sample is used. Thereby, price change estimation is always based on one sample only – the old sample up to the overlap period and the new sample from the overlap period onwards (see below). Resampling is the only method that is fully able to maintain the representativity of the sample. Resources permitting, resampling should be undertaken frequently. The appropriate frequency, of course, depends on the rate of change in a particular product group. It also relies on the assumption that the price differences between the old and new items are appropriate estimates of quality differences. At its extreme, resampling amounts to drawing a new sample in each period and comparing the average price between the samples, instead of the usual procedure of averaging price changes for matched samples. Although logical from the point of view of representativity, resampling in each period would aggravate the problem of quality adjustment by its implicit procedure of quality adjustment, and is thus not recommended.

Replacements

5. A replacement can be defined as an individual successor to a sampled product that has either disappeared completely from the market or lost market share in either the market as a whole or a specific outlet. Criteria for selecting replacements may differ considerably. First, there is the question of when to make the replacement. The usual practice is to do it either when an item disappears completely or when its share of the sales is reduced significantly. Another possible, but less-used rule, would be to replace an item when another variety within the same group, or representative item definition, has become larger with regard to sales, even if the old variety is still sold in significant quantities.

6. The second question is how to select the replacement item. If the rule for initial selection was "most sold" or with probability proportionate to (sales) size, then the replacement rule could follow the same selection rule. Alternatively, the replacement could be that item which is "most like" the old one. The advantage of the former rule is that it produces better representativity. The advantage of the "most like" rule is that, at least superficially, it might reduce the quality adjustment problem.

7. It is important to realize that, under current conditions, replacements cannot adequately represent new items that are coming onto the market. This is because what triggers a replacement is not the appearance of something new, but the disappearance or reduced importance of something old. For example, if the range of varieties in a certain group is increasing, sampling can only represent this increase directly from the set of new varieties, say by sample rotation.

Adding and deleting

8. It is possible to add a new observation point into an elementary aggregate within an index link. For example, if a new brand or model of a durable is introduced without replacing any particular old model, it is desirable to add it to the sample, starting from the time of its introduction. In order to accommodate this new observation in the index system, a reference price needs to be imputed. A practical way to do this is to use the ratio of the price of the new item in the month of its introduction to the average of all other items in the elementary aggregate from the reference period to the month of introduction. In this way, the effect of the new item on the index for months up to the introduction month will be neutral.

9. Similarly, an item that disappears could just be deleted from the sample without replacement. Price change can then be computed over the remaining items. If no further action is taken, this means that the price change for the deleted item, which was measured up to the month prior to deletion, will be disregarded from the month of deletion. This may or may not be desirable, depending on the circumstances in the particular product group.

Formulating an operational target in a dynamic universe

10. A rigorous approach to statistical estimation requires an index estimation strategy, including both the operational target of measurement and the sampling strategy (design and estimator) needed for estimating this target. This strategy would have to consist of the following components:

– a definition of the universe of transactions or observation points (usually a product variety in an outlet) in each of the two time periods between which we want to estimate price change;
– a list of all variables defined for these units. These variables should include prices and quantities (number of units sold at each price), but also all relevant price-determining characteristics of the products (and possibly also of the outlets). This forms the price basis;

- the target algorithm (index formula) that combines the values of the defined variables for the observation points in the defined universe into a single value;
- procedures used for initial sampling of items and outlets from the defined universe;
- procedures within the time span for replacing, sample rotation, adding or deleting observations;
- the estimation algorithm (index formula) applied to the sample with the purpose of minimizing the expected error of the sample estimate compared with the target algorithm. In principle, the estimation needs to consider all the procedures undertaken in replacement and sample rotation situations, including procedures for quality adjustment.

11. Because of its complexity, the rigorous strategy outlined above is generally not used in practical index construction, although the associated information (statistical metadata) system is discussed in paragraphs 8.23 to 8.31 above. A few comments on such possible strategies are made below.

A two-level aggregation system

12. A starting point for discussing an objective of estimating a price index from a sample drawn from a dynamic universe is a two-level structuring of the universe of items and outlets that are considered in the scope of a price index. These levels are:

- the *aggregate* level: at this level there is a fixed structure of item groups $h = 1, \ldots, H$ (or perhaps a fixed cross-structure of item groups by regions and outlet types) within an index link. In terms of updating the universe of items, new goods and services would be defined as new groups at the aggregate level and moved into the index only in connection with a new index link;
- the *elementary* level: at this level the aim is to capture the properties of a changing universe in the index by comparing new and old items. The micro-comparison from period s to period t must be defined so that new products or outlets enter the market and old products or outlets disappear from the market.

13. The common starting point for the three alternative approaches at the elementary level presented here is a basket index from period s to period t at the aggregate level:

$$I_{st} = \frac{\sum_h Q_h P_h^t}{\sum_h Q_h P_h^s} = \sum_h W_h^s I_h^{st},$$

$$\text{where} \quad W_h^s = \frac{Q_h P_h^s}{\sum_h Q_h P_h^s} \quad \text{and} \quad I_h^{st} = \frac{P_h^t}{P_h^s}. \qquad \text{(A8.1)}$$

The quantities, Q_h, are for $h = 1 \ldots H$ item groups from any period, or functions of quantities from several periods, for example, a symmetric average of the base and current periods s and t. Special cases of such a basket index are the Laspeyres ($Q_h = Q_h^s$), Paasche ($Q_h = Q_h^t$), Edgeworth ($Q_h = Q_h^s + Q_h^t$) and Walsh ($Q_h = [Q_h^s Q_h^t]^{1/2}$) price indices outlined in Chapters 15 to 17. Alternative formulations for an elementary-level estimation strategy now enter in the definition of I_h^{st}. As a further common starting point, the set of items or outlets belonging to h in period u ($= s$ or t) is defined as Ω_h^u. The concept of an *observation point* is introduced, usually a tightly specified item in a specific outlet. For each observation point $j \in \Omega_h^u$, there is a price p_j^u and a quantity sold q_j^u. There are now three possibilities for defining the operational target.

The intersection universe

14. The elementary index is defined over the intersection universe, that is, only over observation points existing in both s and t. This index may also be called the *identical units index*. It is equivalent to starting out with the observation points existing in s and then dropping (deleting) missing or disappearing points. An example of such an index is:

$$I_h^{st} = \frac{\sum_{j \in \Omega_h^s \cap \Omega_h^t} q_j p_j^t}{\sum_{j \in \Omega_h^s \cap \Omega_h^t} q_j p_j^s} \qquad \text{(A8.2)}$$

The intersection universe decreases successively over time, as fewer matches are found for each long-run comparison between s and t, s and $t+1$, s and $t+2$ etc., until it eventually becomes empty. An attraction of the intersection universe is that there are, by definition, no replacements involved, and thus, normally, no quality adjustments. If the identical units index is combined with a short index link, followed by *resampling* from the universe in a later period, sampling from the intersection universe is a perfectly reasonable strategy, as long as the assumption implicit in the overlap procedure, that the price differences at that point in time reflect the quality differences, is valid.

The double universe

15. The polar opposite approach to the intersection universe is to consider P_h^s and P_h^t as average prices defined over two separately defined universes in the two periods. A double universe could then be considered as the operational target of measurement: one universe in period s and another in period t. This seems to be a natural way of defining the target, since both time periods should be of equal status and all products existing in either of them should be taken into account. The difficulty with this approach is that the two universes are rarely comparable in terms of quality. Some kind of adjustment for average quality change would need to be brought into the index. The natural definition of the average prices involved in this approach is based on unit values. This would lead to the following definition of a *quality-adjusted unit value index*:

$$I_h^{st} = \frac{\bar{P}_h^t}{\bar{P}_h^s g_h^{st}},$$

$$\text{where} \quad \bar{P}_h^t = \frac{\sum_{j \in \Omega_h^t} q_j^t p_j^t}{\sum_{j \in \Omega_h^t} q_j^t} \quad \text{and} \quad \bar{P}_h^s = \frac{\sum_{j \in \Omega_h^s} q_j^s p_j^s}{\sum_{j \in \Omega_h^s} q_j^s} \qquad \text{(A8.3)}$$

where g_h^{st} is the average quality change in h (also interpretable as a *quality index*), which of course needs further definition. For example, g_h^{st} could be thought of as a hedonic adjustment procedure, where characteristics are held constant. Equation (A8.3) was discussed in Chapter 7, paragraphs 7.142–7.149, as part of Laspeyres, Paasche, Fisher and Törnqvist indices (as opposed to unit value ones), in a form which includes explicit hedonic quality adjustments, g_h^{st}. This operational target is attractive for products where the rate of turnover of varieties is very fast, but where average quality changes only slowly or where reliable estimates of quality changes can be made. The commonly used representative-item method is not really compatible with a double universe target. It implicitly focuses on pre-selected primary sampling units that are used for both period s and t.

The replacement universe

16. Neither sampling from the intersection universe nor from the double universe bears a close resemblance to usual practices for constructing price indices. The most common sampling method used in practice – the representative-item method combined with one-to-one replacements – needs a rationalization in terms of operational targets which differs from these alternatives. Such a rationalization of sampling from a *replacement universe* is considered below.

17. For each $j \in \Omega_h^s$ and $j \notin \Omega_h^t$ we define replacement items $a_j \in \Omega_h^t$ whose price replaces that of j in the formula. Obviously, for $j \in \Omega_h^s$ and $j \in \Omega_h^t, a_j = j$. In addition to a replacement, a quality change from j to a_j is included. This gives rise to a quality adjustment factor g_j, interpreted as the factor with which p_j^s must be multiplied for the consumer to be indifferent between consuming items j and a_j at prices p_j^s and p_{aj}^t.

$$I_h^{st} = \frac{\sum_{j \in \Omega_h^s} q_j p_{aj}^t}{\sum_{j \in \Omega_h^s} q_j p_j^s g_j}. \qquad \text{(A8.4)}$$

18. This step towards an operational use of the formula requires, first, a definition of g_j, which is possible using a hedonic regression as described in Chapter 7, paragraphs 7.132 to 7.152. Second, there is a need to define a_j. A natural procedure is to use a *dissimilarity function* from j to a_j. The notation $d(j, a_j)$ is introduced for this function. The common procedure of choosing the most similar item in cases of replacement now corresponds to minimizing the dissimilarity function. Some further specifications nevertheless need to be made. When is the replacement defined to take place? In practice, this ought to be done when the first chosen variety is no longer representative. Mathematically, this could be defined as follows: observation point j should be replaced in the first period in which $q_j^t < cq_j^s$, where c is a suitably chosen constant between 0 and 1 (a modification being required for seasonal items). The choice of replacement point would then be governed by a rule such as: a_j should be chosen so that $d(j, a_j)$ is minimized for j. Since some priority should be given to observation points that are important in terms of quantities or values, that definition can be modified to become, for example: a_j should be chosen so that $d(j, a_j)/q_{aj}^t$ is minimized for j. Other rules for the choice of replacement point or function to be minimized can of course be chosen.

19. The dissimilarity function needs to be specified; it may depend on the item group h. In general, it must be some kind of metric defined on the set of characteristics of the product and outlet in question. For example, priority could be given to dissimilarity either to "same outlet" or "same product", concepts which could easily be worked into such a metric. A more troublesome concern is the inclusion of as many new points in Ω_h^t as possible in the index definition, in order to ensure that the sample is representative. As the above definitions now stand, the same new point could replace many predecessors, whereas there might be many new points which will not be sampled unless there is a need for a replacement. This shortcoming of the replacement universe is an inherent trait in the replacement method as such. The replacement method is designed to maintain only the representativity of the old sample, not that of the new sample.

Appendix 8.2 New goods and substitution

1. An alternative approach to estimating the effect of introducing new goods is to see new goods as a special case of substitution. In each period a consumer, faced with a set of prices, decides what to consume. The relative sales of the different items sold may change over time. Consumers may decide to consume less of one existing item and more of another existing one, or substitute consumption of an existing old item by a new one not previously available, or discontinue consumption of an existing item and substitute it by consumption of an existing or new one. Such changes are generally prompted by changes in relative prices. In many cases the ''decision'' of the consumer is tied to that of the producer or retailer, as items are no longer produced or sold so as to make way for new ones. Such substitutions between items apply as much to radically new goods as to new models of existing goods. In economic theory, the *elasticity of substitution*, denoted as σ, is a measure of the change in the quantity of, say, item i relative to item j, that would arise from a unit change in the price of item i relative to item j. A value of zero would imply that a change in price would lead to no substitution between the consumption of items and $\sigma > 1$ implies that the change in expenditure arising as a result of substituting items is positive: it is worth switching.

2. There is an intuition here that, if σ is known, and the extent to which substitutions occur in terms of their expenditure shares is also known, then estimates of the underlying price change that prompted the substitution can be derived. This applies as much to substitution between existing items as to substitution between existing, discontinued and new ones. The framework for operationalizing this institution for CPI use is proposed by Shapiro and Wilcox (1997b) – see also Lloyd (1975) and Moulton (1996a) – whereby the usual Laspeyres formulation is generalized to include the (demand) elasticity of substitution:

$$\left[\sum_{n \in 0,t} w_0 \left(\frac{p_{it}}{p_{i0}}\right)^{1-\sigma}\right]^{1/(1-\sigma)} \qquad \text{(A8.5)}$$

where w_0 are expenditure shares in the base period and the summation is over matched items available in both periods. The correction, using σ, incorporates a substitution effect into the basic Laspeyres formula. If $\sigma = 0$, the formula is the traditional Laspeyres one. As $\sigma \rightarrow 1$, the formula tends towards a base-period weighted geometric mean. To use this formulation to generalize across the items in the summation, the restriction must apply that for any pair of items, the elasticity of substitution must be the same. The elasticity of substitution must also be the same over time. Such forms are referred to as constant elasticity of substitution (CES) functional relationships.

3. Feenstra (1994), Feenstra and Shiells (1997) and Balk (2000b) have extended the substitution to discontinued and new items. The advantage of equation (A8.5) is that, given an estimate of σ, a cost of living index which includes an estimate of substitution effects can be measured in real time. The incorporation of the effects of new and discontinued items follows directly from this. Alternative frameworks for including substitution effects (discussed in Chapter 17) require expenditure data for the base and current periods.

4. To extend the framework to new items, it is necessary to know how expenditures shift between new, existing and discontinued items. Let λ^t be the expenditure share of matched existing items out of the total in period t. The total includes existing and new items, so $1 - \lambda^t$ is the share of new items in period t. Similarly, $1 - \lambda^0$ is the expenditure share of old, discontinued items in period 0. The generalized Laspeyres index, which includes substitution between existing and old and new items, is given by:

$$\left[\frac{\lambda^t}{\lambda^0}\right]^{1/(\sigma-1)} \left[\sum_{n \in 0,t} w_0 \left(\frac{p_{it}}{p_{i0}}\right)^{1-\sigma}\right]^{1/(1-\sigma)} \qquad \text{(A8.6)}$$

Like the usual Laspeyres index, it requires only the price relatives, the base period weights, the ratio of expenditure shares and an estimate of the elasticity of substitution. It can be derived in a number of alternative forms, including generalized, Paasche, Fisher or Sato–Vartia indices.

5. While there is an intuition behind the above formula, its formal correspondence to an index of consumer prices defined in economic theory is given by Balk (2000b). De Haan (2001) shows how the Fisher equivalent could be derived from a decomposition of a Fisher index when there are new and disappearing goods. The derivations show how the framework requires that $\sigma > 1$, a factor prompting Balk (2000b) to argue for its use for lower-level index aggregation, where this is more likely. The remaining problems are the estimation of σ, the availability of data on current expenditure shares, and the validity of the implied constant σ. There are also some conceptual issues. Increases in utility are regarded as having resulted from increases in the desirability of the items included in the above aggregation. If such items improve, then utility increases. Yet there are other goods outside the aggregation or system of demand equations. Deterioration in such goods will lead to increases in the desirability of the included items and decreases in utility. For example, if a consumer switches to private transport as a result of a deterioration in public transport, this should not be measured as a welfare gain resulting from better private transport, even though the expenditure flows in equation (A8.6) shift that way (Nevo, 2001).

6. The direct estimation of σ requires considerable econometric expertise. This puts it outside the routine construction of index numbers (see Hausman, 1997). Balk (2000b) shows how an alternative numerical routine might work. De Haan (2001) used scanner data to apply the methodology to a generalized Fisher index. He applied Balk's routine to nine product groups, using data from the Netherlands CPI, and found values of σ that exceeded unity. He advised the use of chained indices to maximize the matching of ongoing items, a principle discussed in Chapter 7, paragraphs 7.153 to 7.158. De Haan (2001) found major discrepancies between a generalized and ordinary Fisher index for at least six of the products, arguing for the need to incorporate the effects of new goods (see also Opperdoes, 2001). He also demonstrates how sensitive the procedure is to the selection of σ: for a share in current expenditure for new items of 4.8 per cent, and $\sigma = 1.2$, a Paasche-type index which includes new goods would be 93 per cent below the Paasche price change for ongoing goods only. For $\sigma = 5.0$ and the same expenditure share, the discrepancy falls to 34.1 per cent. For very large values, say $\sigma > 100$, the two indices would be relatively close. In such cases, the goods are almost identical, being near-perfectly substitutable; a switch to a new good would have little effect, the new and existing goods having similar prices.