5 SAMPLING

Introduction

5.1   The procedure used for price collection by a national statistical office in the production of a consumer price index (CPI) is a sample survey. In fact, in many countries, it might be better viewed as composed of many different surveys, each covering different subsets of the products covered by the index. We will therefore begin by outlining some of the general concepts of survey sampling thta need to be kept in mind when looking at a particular survey such as price collection for a CPI.

5.2   There is a *target quantity,* for example a CPI, which is defined with respect to:
− a *universe* consisting of a finite population of units (e.g. products);
− one or more *variables* that are defined for each unit in the universe (e.g. price and quantity);
− a formula that combines the values of one or more of these variables for all units in the universe into a single value called a *parameter* (e.g. the Laspeyres index).
The interest is in the value of this parameter.

5.3   The universe usually has three dimensions. There is a *product dimension* consisting of all purchased products and varieties of products. There is a *geographical and outlet dimension* consisting of all places and channels where a product is sold. Finally, there is a *time dimension* consisting of all sub-periods within an index period. The time dimension will be given less attention since price variation is usually smaller over a short time span and since temporal aspects may be dealt with in product and outlet specifications.

5.4   In this chapter, the first two dimensions will be regarded as being static over the time periods considered in the index. In other words, it will be assumed that the same products and outlets are in the universe in both periods, or that replacements between old and new products or outlets are one to one and without problems. For the complications arising from dynamic changes in the universe, please refer to Chapter 8, where replacement, resampling and quality adjustment are discussed.

5.5   Why take only a sample of units? Apart from the near physical impossibility and prohibitive cost of trying to cover all products in all outlets, the data are likely to be of better quality if there are fewer units to deal with because of the use of more specialized and better trained data collectors. Also, the time required to complete the exercise is shorter.

5.6   In *probability sampling*, the units are selected in such a way that each unit (an outlet or a product) has a known non-zero probability of selection. For example, outlets are selected at random from a business register in which each outlet has an equal chance of being selected. Traditionally, however, *non-probability sampling* methods have mainly been used in the compilation of a CPI for choosing outlets or products. The representative item method is particularly popular for selecting items. Other methods used are cut-off sampling and quota sampling (see below). There are also instances of a mixture of the two methods of sampling; for example, outlets are selected using probability sampling techniques, whilst products are selected using the representative item method.

5.7   Having decided to sample, there are two issues to be considered: how to select the sample; and how to use the sample values to estimate the parameter. The former reflects the

choice of sampling design, and the latter constitutes the estimation procedure. We first take a look at sampling design.

Probability sampling techniques

5.8   This section presents some general concepts and techniques of survey sampling that have important applications for price indices. This brief presentation covers those concepts of survey sampling that are of immediate interest in price index applications. For a full treatment of the subject, please refer to one of the many textbooks available, for example Särndal, Swensson and Wretman (1992) or Cochran (1977).

5.9   Survey sampling theory views the universe as composed of a finite number (N) of observational units denoted $j = 1,...,N$. Sampling then amounts to selecting n units out of N by attaching an inclusion probability, π_j, to each unit. For price indices there are two sampling designs that are of particular interest.

5.10   In *simple random sampling* and *systematic sampling* each unit is sampled with equal probability and we have $\pi_j = n/N$. In simple random sampling, all units are selected using a random mechanism. In systematic sampling, the sampling units are selected at equal distances from each other in the frame, with random selection of only the first unit. These techniques are usually recommended in situations where the units are relatively homogeneous.

5.11   In *probability proportional to size (pps) sampling* the inclusion probability is proportional to some auxiliary variable x_j. Units for which the probability $\pi_j = nx_j \Big/ \sum_{k=1}^{K} x_k$ is larger than one are selected with certainty, while the inclusion probabilities for the remainder of the units are calculated after excluding the large ones.

5.12   The universe may be divided into strata, denoted $h = 1,...,H$. In each stratum, there are then N_h units and we have $\sum_{j=1}^{N} N_h = N$. The purpose of stratification is usually to group units together that are either homogeneous in some sense or satisfy some administrative convenience such as being physically close together. Each stratum is a mini-universe with sampling taking place independently in each one. In a CPI, the practice is to use elementary aggregates as strata. In the remainder of this chapter, we look at sampling in a single stratum, corresponding to an elementary aggregate, and drop the subscript h.

Implementing probability sampling in consumer price indices

5.13   A *sampling frame* is a list of all (or most) of the N units in the universe. A frame may have overcoverage to the extent that it includes units that are not in the universe or includes duplicates of units. It may have undercoverage to the extent that some units in the universe are missing from the frame.

5.14   Sampling frames for the outlet dimension could be:
- Business registers. These should include locations of retail trade businesses with addresses and be updated regularly. If a size measure (turnover or number of employees) is included in the register, it is a useful device for

performing probability proportional to size (pps) sampling and this size measure would then be included in the universe parameter also.

- Telephone directories ("yellow pages"). These usually do not include size measures, so simple random sampling or systematic sampling would then be necessary. Sometimes informal knowledge of the importance of different outlets could be used to stratify the universe into two or more categories and then draw a relatively larger sample from the more important strata.
- Records of local administrations, organizations of enterprises, and so on, could be used for local markets and suchlike, which are especially important in developing countries.

5.15   Sampling frames for the product dimension could be:
- product lists provided by major wholesalers showing sales values for varieties in an earlier period. Sales values provide an obvious size measure for weights and pps sampling;
- outlet-specific lists of products. These lists could also be drawn up by the price collectors themselves by noting the products displayed on the shelf. Shelf space could then be used as a size measure for pps sampling.

Sampling techniques based on probability proportional to size
5.16   Several techniques exist for drawing pps samples. They fall into two main categories according to whether the size of the sample is fixed or random. A fixed, predetermined sample size is clearly desirable for CPIs, since the sample size in each stratum is often small and a random size would entail the risk of an empty sample. We therefore present two techniques here that provide fixed size pps samples.

5.17   *Systematic pps sampling*. The procedure is best explained by an example. In Table 5.1 we show how a sample of 3 outlets can be drawn from 10. In this case we have the number of employees as our size measure. We look at the list, where we have included the cumulative sizes and the inclusion intervals. We take the total number of our size measure, which is 90 in this case, and divide it by the sample size, 3. This gives us a sampling interval of 30. We next choose a random number between 1 and 30 (random number functions are given in, for example, the Excel spreadsheet software). Say that we get 25. The sample will then consist of the outlets whose inclusion intervals cover the numbers 25, 25 + 30 and 25 + 2x30.

5.18   Systematic sampling is easy to execute. If, however, the frame has some overcoverage, the sample size will not be the predetermined one. Let us say that, at the first visit to the outlets, we discover that outlet 6 does not sell the products in the product sample. We would then be left with a sample of only two outlets. We would either be content with that, or somehow seek a replacement for the invalid outlet, which is not determined by the basic sampling procedure. Moreover, the selected sample depends on the order in which the outlets or products are listed. This might be important, especially if the listing order is correlated to the size measure.

Table 5.1. Systematic sample of 3 out of 10 outlets, based on probability proportional to size

Outlet	Number of employees = x	Cumulative x	Inclusion interval	Included when starting point is 25
1	13	13	1–13	

2	2	15	14–15	
3	5	20	16–20	
4	9	29	21–29	X
5	1	30	30	
6	25	55	31–55	X
7	10	65	56–65	
8	6	71	66–71	
9	11	82	72–82	
10	8	90	83–90	X

5.19 *Order pps sampling*. This is a relatively new method for drawing pps samples. Rosén (1997a, 1997b) gives its theory. In this case, a uniform random number U_i between *0* and *1* and a variable $z_i = n x_i / \Sigma_k \, x_k$, where x_i is a size variable, are associated with each sampling unit and a *ranking variable* is constructed as a function of these two variables. The units in the universe are then sorted in ascending order and the *n* units with the smallest values of the ranking variable are included in the sample. Two important examples of such ranking variables Q_i are:

- for sequential pps sampling: $Q_i = U_i / z_i$;
- for Pareto pps sampling: $Q_i = U_i (1 - z_i) / z_i (1 - U_i)$.

5.20 For the same universe as above and with Pareto pps as our example, we show in Table 5.2 how this works. We have now ordered the universe in ascending order with respect to the ranking variable. Our first sample turns out to consist of outlets 1, 6 and 9. Say that we now discover, however, that it is inappropriate to include outlet 1. We then turn to the fourth unit in order – outlet 4 – and include that one instead. Thus, order pps sampling is easy to combine with a fixed sample size and more flexible than systematic sampling.

Table 5.2: Pareto sample of 3 out of 10 outlets, based on probability proportional to size

Outlet	x_i	U_i	z_i	$F(U_i)$	$F(z_i)$	Q_i	Sample
1	13	0.198082	0.433333	0.112246	0.347498	0.323013	(X)
6	25	0.755509	0.833333	0.125918	0.203743	0.618026	X
9	11	0.277131	0.366667	0.175516	0.265052	0.662196	X
4	9	0.463730	0.300000	0.324611	0.160881	2.017709	X
5	1	0.067941	0.033333	0.065676	0.031069	2.113910	
7	10	0.709046	0.333333	0.472697	0.096985	4.873939	
3	5	0.500162	0.166667	0.416802	0.083306	5.003241	
2	2	0.297524	0.066667	0.277689	0.046832	5.929506	
10	8	0.834138	0.266667	0.611701	0.044230	13.830048	
8	6	0.915131	0.200000	0.732105	0.016974	43.131461	

5.21 Neither of the two order sampling procedures is, however, exactly pps, because the obtained inclusion probabilities vary somewhat from the desired ones. Rosén (1997b) shows, however, that for the purpose of estimating means and variances, they are approximately pps. In the case of the price index, this still holds when there is sample substitution resulting from overcoverage. Pareto pps is marginally better than sequential pps and should therefore be preferred.

5.22 Order pps sampling is at present used in many areas of the Swedish CPI, for example for sampling:

- outlets from the business register (the size measure is number of employees + 1);
- products from databases provided by major retail chains (the size measure is historic sales);
- car models from the central car register (the size measure is number of cars registered in the reference period).

5.23 Further details on the application of these procedures are given in Statistics Sweden (2001). Rosén (1997b) shows that Pareto pps and systematic pps are the two optimal pps sampling methods. Pareto pps permits an objective assessment of estimation precision. With regard to final precision, however, Pareto pps is best in some situations whereas systematic pss is best in other situations. The choice between them is therefore a matter of judgement and practicality in a particular sampling situation. The great flexibility of order pps sampling with regard to imperfections in the frame, an aspect of importance in CPI applications, leads us to make this procedure our first recommendation among pps procedures.

Sampling methods used by the US Bureau of Labor Statistics

5.24 The US Bureau of Labor Statistics (BLS) uses probability methods in all stages of sample selection. In the last stage, individual items in outlets are selected in a process designed to approximate pps sampling with respect to the sales of each such item. To this end, the BLS field representatives are allowed to use any of four procedures for determining the sales proportions (U.S. BLS, 1997):

- obtaining the proportions directly from a respondent;
- ranking the subgroups/items by importance of sales as indicated by the respondent and then obtaining the proportions directly or using pre-assigned proportions;
- using shelf space to estimate the proportions where applicable;
- using equal probability.

5.25 The advantages of this procedure, according to the BLS, are that it ensures an objective and efficient probability sampling procedure, where no other such procedure would be available. It allows broad definitions of the item strata so that the same tight specification need not be priced everywhere. The wide variety of specific items greatly reduces the within-item component of variance, reduces the correlation of price movement between areas, and allows a reduction of the sample size needed for a given variance.

5.26 A potential pitfall in this approach is that, if the sales value measure is taken during a very short period, it may coincide with a special campaign with temporarily reduced prices. It could then happen that an item with a temporarily reduced price is given a large inclusion probability. Since this price will tend to increase more than average, an overestimating bias may result. It is thus essential that the sampling of the item takes place at an earlier point in time than the first price collection or that sales values from an earlier period are used. Okamoto (1999) emphasizes this point for Japan, where price bouncing seems to be a very common phenomenon.

Non-probability sampling techniques

5.27 Modern statistical sampling theory focuses on probability sampling. Use of probability sampling is also strongly recommended and standard practice for all kinds of

statistical surveys, including economic surveys. But price index practice in most countries is still dominated by non-probability techniques. It may then be fruitful to speculate somewhat about the rational and irrational reasons for this situation. In the following section we discuss a number of such possible reasons, one by one. We then go on to consider some non-probability techniques.

Reasons for using non-probability sampling

5.28 *No sampling frame is available.* This is often true for the product dimension but less frequently so for the outlet dimension, for which business registers or telephone directories do provide frames, at least in some countries, notably in Western Europe, North America and Oceania. There is also the possibility of constructing tailor-made frames in a limited number of cities or locations, which are sampled as clusters in a first stage. For products, it may be noted that the product assortment exhibited in an outlet provides a natural sampling frame, once the outlet is sampled as a kind of cluster, as in the BLS sampling procedure presented above. So the absence of sampling frames is not a good enough excuse for not applying probability sampling.

5.29 *Bias resulting from non-probability sampling is negligible.* There is some empirical evidence to support this assertion for highly aggregated indexes. Dalén (1998b) and De Haan, Opperdoes and Schut (1999) both simulated cut-off sampling of products within item groups. Dalén looked at about 100 groups of items sold in supermarkets and noted large biases for the sub-indices of many item groups, which however almost cancelled out after aggregation. De Haan, Opperdoes and Schut used scanner data and looked at three categories (coffee, babies' napkins and toilet paper) and, although the bias for any one of these was large, the mean square error (defined as the variance plus the squared bias) was often smaller than that for pps sampling. Biases were in both directions and so could be interpreted to support Dalén's findings. The large biases for item groups could, however, still be disturbing. Both Dalén and De Haan, Opperdoes and Schut report biases for single-item groups of many index points.

5.30 *We need to ensure that samples can be monitored for some time.* If we are unlucky with our probability sample, we may end up with a product that disappears immediately after its inclusion in the sample. We are then faced with a replacement problem, with its own bias risks. Against this, it may happen that short-lived products have a different price movement from the price movement of long-lived ones and constitute a significant part of the market, so leaving them out will create bias.

5.31 *A probability sample with respect to the base period is not a proper probability sample with respect to the current period.* This argument anticipates some of the discussion in Chapter 8 below. It is certainly true that the bias protection offered by probability sampling is to a large extent destroyed by the need for non-probabilistic replacements later on.

5.32 *Price collection must take place where there are price collectors.* This argument applies to geographical sampling only. It is, of course, cheaper to collect prices near the homes of the price collectors, and it would be difficult and expensive to recruit and dismiss price collectors each time a new sample is drawn. This problem can be reduced by having good coverage of the country in terms of price collectors. One way to achieve this is to have a professional and geographically distributed interviewer organization within the national statistical agency, which works on many surveys at the same time. Another way of reducing the problem is to have a first-stage sample of regions or cities or locations

which changes only very slowly.

5.33   *The sample size is too small*. Stratification is sometimes made so fine that there is room for only a very small sample in the final stratum. A random selection of 1–5 units may sometimes result in a final sample that is felt to be skewed or otherwise to have poor representativity properties. Unless the index for this small stratum is to be publicly presented, however, the problem is also small. The skewness of small low-level samples will even out at higher levels. The argument that sample size is too small has a greater validity when it relates to first-stage clusters (geographical areas) that apply to most subsequent sampling levels simultaneously.

5.34   *Sampling decisions have to be taken at a low level in the organization*. Unless price collectors are well versed in statistics, it may be difficult for them to perform probability sampling on site. Such sampling would be necessary if the product specification that has been provided centrally covers more than one product (price) in an outlet. Nevertheless, in the United States (U.S. BLS, 1997) field representatives do exactly this. In Sweden, where central product sampling (for daily necessities) is carried to the point of specifying well-defined varieties and package sizes, no sampling in the outlets is needed. In countries where neither of these possibilities is at hand, full probability sampling for products would be more difficult.

5.35   In some situations, there are thus valid reasons for using non-probability techniques. We discuss two such techniques below.

Cut-off sampling

5.36   Cut-off sampling refers to the practice of choosing the *n* largest sampling units with certainty and giving the rest a zero chance of inclusion. In this context, the term "largeness" relates to some measure of size that is highly correlated with the target variable. The word "cut-off" refers to the borderline value between the included and the excluded units.

5.37   In general, sampling theory tells us that cut-off sampling does not produce unbiased estimators (see paragraphs 5.51 to 5.60 below for a discussion of bias and variance), since the small units may display price movements which systematically differ from those of the larger units. Stratification by size or pps sampling also has the advantage of including the largest units with certainty while still giving all units a non-zero probability of inclusion.

5.38   If the error criterion is not minimal bias but minimal *mean square error* (= variance + squared bias) then, since any estimator from cut-off sampling has zero variance, cut-off sampling might be a good choice where the variance reduction more than offsets the introduction of a small bias. De Haan, Opperdoes and Schut (1999) demonstrate that this may indeed be the case for some item groups.

5.39   Often, in a multi-stage sampling design there is room for only a very small number of units at a certain stage. Measurement difficulties that are sometimes associated with small units may then be a reason, in addition to large variances, for limiting price collection to the largest units.

5.40   Note that a hybrid design can also be applied in which there is a certainty stratum part, some probability sampling strata and a low cut-off point below which no sample at all is

drawn. In practice, this design is very often used where the "below cut-off section" of the universe is considered insignificant and perhaps difficult to measure.

5.41   A particular CPI practice that is akin to cut-off sampling is for the price collector to select the most sold product in an outlet, within a centrally defined specification. In this case, the sample size is one (in each outlet) and the cut-off rule is judgemental rather than exact, since exact size measures are only rarely available. In all cases of size-dependent sampling in an outlet, it is crucial to take a long-term view of size, so that temporarily large sales during a short period of reduced prices are not taken as a size measure. Such products will tend to increase in price in the immediate future much more than the product group that they represent and thus create a serious overestimating bias.

Quota sampling
5.42   Many product groups, even rather small ones, are quite heterogeneous in nature, and the price varies according to a large number of subgroups or characteristics. There may well be different price movements going on within such a product group, and a procedure to represent the group by just one or a few tightly specified product types may then carry an unnecessarily great risk of bias.

5.43   The definition of quota sampling is that the selected sample shall have the same proportions of units as the universe with respect to a number of known characteristics, such as product subgroup, type of outlet, and location. The actual selection of sampling units is then done by judgemental procedures in such a manner that the composition of the final sample meets the quota criteria.

5.44   The following example illustrates the concept of quota sampling. A sample of 20 package holidays is desired. It is known that, in the universe, 60 per cent of the holidays are to Spain, 30 per cent to Greece, and 10 per cent to Portugal. Of the travel groups, 70 per cent comprise of 2 adults, 20 per cent comprise 2 adults +1 child, and 10 per cent comprise 2 adults +2 children. Of the sample, 20 per cent stay in 2-star hotels, 40 per cent in 3-star hotels, 30 per cent in 4-star hotels, and 10 per cent in 5-star hotels. With this information, it is possible to design the sample purposively so that all these proportions are retained in the sample, which then becomes self-weighted. Note that these proportions reflect volumes, not values, and may need to be adjusted depending on the elementary aggregate formula used.

5.45   Quota sampling requires central management of the whole sampling process, which may limit its usefulness in some situations. It is more difficult, but not impossible, to manage a quota sampling system where local price collection is used. One would then need to divide the price collectors into subgroups with somewhat different instructions for selecting products. A limitation of quota sampling, as in other non-probability sampling, is that the standard error of the estimate cannot be determined.

The representative item method
5.46   This is the traditional CPI method. The central office draws up a list of product types, with product type specifications. These specifications may be tight, in that they narrowly prescribe for the price collectors what products they are permitted to select, or they may be loose, giving the price collector freedom to choose locally popular varieties.

5.47   The method with tight specifications is in a sense diametrically opposite to the quota

sampling method discussed above. Unless the product groups are defined so as to include a very large number of product types, representativity will suffer in this procedure, since no products falling outside the specification will enter the index. Another disadvantage with the method is that it may lead to more missing products in the outlets and thus reduce the effective sample. Its main advantage is simplicity. It is easy to maintain a central control over the sample. If quality adjustments are needed, they can be decided in the central office, which may or may not be an advantage.

5.48 The method with loose specifications gives price collectors the chance to adjust the sample to local conditions and will normally lead to greater representativity of the sample as a whole. Where it is combined with the "most sold" criterion it will, however, systematically underrepresent the smaller brands and products that may be bought by important minorities.

Sampling in time
5.49 A CPI usually refers to a month, during which prices are not constant. The issue of sampling in time then arises. Often, this problem is ignored, for example by using the 15th day of the month, or the days surrounding the 15th, as the target date for price measurement. In some areas, there is a day-of-the-week effect on prices, for example in cinemas, theatres and restaurants, but this may be taken into account in the product specification rather than in sampling, for example by specifying a weekday evening price.

5.50 As far as is known, random sampling in time is not used anywhere. The method used in some countries is to spread price collection over several weeks according to some pattern, for example different weeks in different regions or for different product groups. In some cases, more frequent pricing than monthly is also used, for example for fresh produce. There is not yet any systematic knowledge about the pros and cons of such practices. Chapter 6 discusses the more practical aspects of distributing price collection over time.

Choice of sampling method
5.51 In this section, we discuss how choices in sampling method could depend on specific factors in a country. But first we consider the matter of sample size.

5.52 *Sample size*. The final precision of a sample estimate depends only on the size and allocation of the sample and not on the size of the country, so in this sense there is no need for a larger sample in a larger country. Larger samples are called for if regional differences in price change are of interest and if the amount of product disaggregation that is desired in presenting the indices is very high. Of course, the budget allocated to CPI work may be larger in large countries, allowing for larger samples.

5.53 Studies of bias (not the sampling bias described in paragraphs 5.61 to 5.64) and of sampling error show that bias in CPIs is generally a much greater problem than sampling error. This leads to the conclusion that, in many cases, smaller samples that are better monitored with respect to replacements, resampling and quality adjustment could give a higher quality index for the same budget. In some countries, local price collection is a fixed resource and it is therefore difficult to move resources from local price collection to central analytical work. Still, it is advisable to try to use local resources for higher quality price collection rather than just for many observations. The quality of price collection is further discussed in Chapter 6.

5.54   Monthly sample sizes in different countries seem to vary from several thousand to several hundred thousand. Often, the reasons for these differences lie more in tradition than in a rational analysis of the needs of precision. Countries with very large sample sizes would probably do well to look at ways of reallocating their total resources.

5.55   *Geographical distribution of price collectors*. Sampling is more expensive further away from the homes of the price collectors. If the organization for price collection is centralized in a few main cities, it will be difficult to sample outlets elsewhere. It should be borne in mind, however, that rural and urban inflation may well be different, so failure to collect prices in both rural and urban areas would be detrimental to efforts to achieve the best measure of average national inflation. It would be better to have at least a small sample in the rural areas so that this factor can be taken into account. The major part of the saving arising from allocating outlets close to price collectors can then still be realized.

5.56   *Sophistication of price collectors*. If price collectors are well educated, they may be instructed to carry out more complex sampling schemes such as pps sampling in the outlets. Otherwise, simpler methods are called for.

5.57   *Access to sampling expertise in the central office*. Probability sampling requires access to methodological expertise in the central statistical office.

5.58   *Homogeneous versus heterogeneous product groups*. The representative item method is more suitable for homogeneous product groups. In heterogenous groups, it is more likely that important segments of the product universe, with different price movement, will be left out.

5.59   *Access to sampling frames and their quality*. Probability sampling requires sampling frames. But they do not necessarily have to be available at the national level. By applying geographical cluster sampling in the first stage (where the sampling frame is just a map), a list of relevant outlets can be constructed in each sampled cluster using telephone directories or local enumeration, as is done in the United Kingdom. This method is also used to select urban areas for the United States CPI (Dippo and Jacobs, 1983).

5.60   *Scanner data.* The discussion in this chapter is based on the traditional situation, where prices have to be collected locally and centrally, and entered individually into a central database. Where prices and possibly quantities are collected electronically, as is the case with sale point scanner data, sampling could be different. There is then no need for sampling products or varieties or points in time, since they are completely enumerated automatically. Nevertheless, not all outlets selling a product will be covered by scanner data in the foreseeable future. Since all kinds of outlets should be represented in the index, there will be continue to be a need to combine scanner data samples with traditional samples for non-scanner outlets.

Estimation procedures
5.61   There is a crucial distinction to be made between what is to be estimated, the *parameter*, which is defined for the whole universe, and the *estimator*, which is a formula to be calculated using the sample values as an estimate of the parameter. Now, in survey sampling in general we want to estimate a population total or a function of several such totals, for example a ratio of totals. So, if we have two variables y and z defined for each

sampling unit (for example, prices at two different periods), we may want to estimate the following parameters:

$$Y = \sum_{j=1}^{N} y_j \quad \text{and} \quad Z = \sum_{j=1}^{N} z_j \quad \text{or} \quad R = Y / Z$$

5.62   Several different estimators may be proposed for the same population parameter, in which case we need to decide which of these estimators to use. In assessing the quality of a sample estimator, i.e. how well it estimates the parameter, two measures are often considered in the probability sampling paradigm. The first measure is the bias of an estimator, which is the difference between the universe parameter and the average of the estimator over all possible samples that could be drawn under the specified sample design (referred to as the mean of the sampling distribution of the estimator). Note that this bias refers to something different from the index number bias discussed elsewhere in this manual. An estimator is unbiased if it has zero bias. The second measure is the variance of the estimator with respect to this sampling distribution. An estimator is considered good if both its bias and variance are small; that is, the estimator is on average very close to the parameter and does not vary much from its mean.

5.63   The good fortune of finding an estimator that minimizes both bias and variance at the same time does not often happen. An estimator with a small bias may have a large variance, and one with a small variance may have a large bias. So use is frequently made of a criterion called the mean square error, which is the sum of the bias squared and the variance. A "good" estimator is then one which minimizes this criterion.

5.64   Sampling theory tells us that the following estimators are unbiased, respectively, for the parameters Y and Z above: $\hat{Y} = \Sigma_{j \in S}\, y_j / \pi_j$, $\hat{Z} = \Sigma_{j \in S}\, z_j / \pi_j$, where S is the sample, and that $\hat{R} = \hat{Y}/\hat{Z}$ is approximately unbiased for R, subject to a (usually negligible) technical ratio estimator bias.

Implementing estimation procedures for consumer price indices
5.65   As stated earlier, sampling for CPIs is usually stratified, with elementary aggregates as strata. Let us assume that the universe parameter is I and that the parameter in stratum h is labelled I_h. Then we have:

$$I = \sum_{h} w_h I_h$$

where w_h is the weight of stratum h. The issue then is to estimate I_h for each stratum. In the following discussion, we therefore concentrate on estimating for a single stratum and drop the subscript h.

5.66   Depending on the content, degree of homogeneity, price elasticity and access to weighting information within the stratum, different parameters may be appropriate in different strata. The choice of parameter is an index number problem, to be solved by reference to the underlying economic concepts. As discussed in Chapter 20, it could be the unit value index, the Laspeyres index, the Lowe index, or the geometric Laspeyres index.

5.67   Suppose we have a sample of size n and that the units in the sample are labelled 1,2, ..., n. Very often, one of the three formulae below is used as an estimator of the stratum index:

The arithmetic mean of price relatives (Carli index):

$$r = \frac{1}{n}\sum_{j\in S}\frac{p_j^1}{p_j^0}$$

(5.1)

The ratio of mean prices (Dutot index):

$$a = \frac{\dfrac{1}{n}\sum_{j\in S}p_j^1}{\dfrac{1}{n}\sum_{j\in S}p_j^0}$$

(5.2)

The geometric mean (Jevons index):

$$g = \prod_{j\in S}\left(\frac{p_j^1}{p_j^0}\right)^{\frac{1}{n}}$$

(5.3)

For discussion below, we also need to introduce the ratio of harmonic mean prices:

$$h = \frac{\dfrac{1}{n}\sum_{j\in S}1/p_j^0}{\dfrac{1}{n}\sum_{j\in S}1/p_j^1}$$

(5.4)

5.68   When comparing the above estimators with the functional form of the parameters in Chapter 20, we realize that very special conditions are needed to make them unbiased estimators of those parameters. For one thing, unlike the parameters in Chapter 20, there are no quantities involved in the sample estimators.

5.69   We state, without proof, some results concerning the statistical properties of the above estimators (see Balk (2002) for details). Suppose we have N products in the universe labelled 1,2, ..., N. Let p_j^t, q_j^t be respectively the price and quantity for product j in period t ($t = 0$ for base period and 1 for current period), and let

$$w_j^0 = \frac{q_j^0 p_j^o}{\sum_{k=1}^{K}q_k^0 p_k^o}$$ $(k==1,...,K)$

be the base period expenditure share of product j. Then:

- Under simple random sampling, none of r, a or g estimates any of the population parameters without bias. Instead, weights need to be used in the estimators also.

- Under pps, if $\pi_j \propto w_j^0$ for all j then r, the average of relatives, is unbiased for the Laspeyres index (the symbol "$\propto$" means "proportional to").

- Under pps, if $\pi_j \propto q_j^0$ for all j, then a, the ratio of averages, is approximately unbiased for the Laspeyres index.

- Under pps, if $\pi_j \propto w_j^0$ for all j then g is approximately unbiased for the geometric Laspeyres index. In this case $\log g$ is unbiased for the logarithm of the geometric Laspeyres index. The remaining bias tends to be of a similar order to that of a.

5.70   All these results are somewhat theoretical in nature since neither w_j^0 nor q_j^0 are known

at the time when the sample could be drawn. This is a reason for introducing the Lowe index:

- Under pps, if $\pi_j \propto q_j^b$ (where b is some period before 0) for all j then a is approximately unbiased for the Lowe index.

5.71 There is no simple way to relate any of the estimators to the unit value index. In fact, estimating that index requires separate samples in the two time periods, since its numerator and denominator refer to different universes.

- Under two separate sample designs, one for period 0 and one for period 1, which are both pps and where $\pi_j^0 \propto q_j^0$ and $\pi_j^1 \propto q_j^1$ then a is approximately unbiased for the unit value index. In this case, however, the interpretation of the a formula will be different, since the samples in the numerator and the denominator are different.
- Under two separate sample designs, one for period 0 and one for period 1, which are both pps and where $\pi_j^0 \propto v_j^0 = p_j^0 q_j^0$ and $\pi_j^1 \propto v_j^1 = p_j^1 q_j^1$ then h, the ratio of harmonic mean prices, is approximately unbiased for the unit value index. The following algebraic reformulation of the unit value index helps to clarify that fact:

$$UV = \frac{\sum_{j \in S} v_j^1 \Big/ \sum_{j \in S} v_j^1 \big/ p_j^1}{\sum_{j \in S} v_j^0 \Big/ \sum_{j \in S} v_j^0 \big/ p_j^0}$$

As for a, however, the interpretation of the h formula will be different, since the samples in the numerator and the denominator are different.

5.72 The phrase "approximately unbiased" needs some explanation. It refers to the fact that the estimator is not exactly unbiased but that the bias is small and decreases towards zero as the sample size and the size of the universe simultaneously go to infinity in a certain, mathematically well-defined manner. In the ratio estimator case applicable to a, the sign of this bias is in-determinate and its size after aggregation is probably negligible. In the case of the geometric mean, however, the bias is always positive, i.e. the sample geometric mean tends to overestimate the universe geometric mean on average over many sample drawings. In the case of simple random sampling and an unweighted geometric mean in both the universe and the sample, the bias expression is: $b \approx \sigma^2 / 2n$, where σ^2 is the variance of the price ratios. For small universes, a finite population correction needs to be multiplied to this expression. This result is easily derived from expression (4.1.4) in Dalén (1999b). The bias may be significant for small sample sizes, so that a caution against very small samples in a stratum may be warranted when the geometric mean is applied.

Variance estimation
5.73 A CPI is a complex statistic, usually with a complex design. It is thus not a routine task to estimate the variance of a CPI. To the extent that samples are not probability based, variance estimates need to make use of some kind of model in which random sampling is assumed. In the absence of systematic and generally accepted knowledge, the approaches to variance estimation used in four countries will be briefly described.

Variances of elementary index formulae
5.74 As a preliminary, some variance estimators for elementary aggregate formulae will be provided. In order not to overburden the text with formulae, the variance estimators, not the exact variance, will be given. The variance estimators are approximately unbiased under

simple random sampling, where the corresponding universe parameter is unweighted. They are also applicable to the case of pps sampling for a weighted universe parameter, where the size measure is the same as the parameter weight. For definitions of the formulae, see equations (5.1)–(5.3).

$$V(r) = \frac{\sigma_r^2}{n} \text{, where } \sigma_r^2 = \frac{1}{n-1}\sum_{j \in S}(r_j - r)^2 \text{ and } r_j = \frac{p_j^1}{p_j^0} ; \tag{5.5}$$

$$V(a) = \frac{1}{n(\bar{p}^0)^2}\left(\sigma_1^2 + r^2\sigma_0^2 - 2r\sigma_{01}\right), \tag{5.6}$$

where $\sigma_1^2 = \frac{1}{n-1}\sum_{j \in S}(p_j^1 - \bar{p}^1)^2$, $\sigma_0^2 = \frac{1}{n-1}\sum_{j \in S}(p_j^0 - \bar{p}^0)^2$, $\sigma_{01} = \frac{1}{n-1}\sum_{j \in S}(p_j^1 - \bar{p}^1)(p_j^0 - \bar{p}^0)$,

$\bar{p}^1 = \frac{1}{n}\sum_{j \in S}p_j^1$ and $\bar{p}^0 = \frac{1}{n}\sum_{j \in S}p_j^0$.

This estimate follows from the fact that a, unlike r, is a ratio of stochastic variables. See, for example, Cochran (1977) for a derivation of this formula.

5.75 The geometric mean is more complex, since it is not a linear estimator. However, Dalén (1999b) derived the following easily applied variance expression, which holds with good approximation if price ratios do not have too extreme variation ($\sigma_r/r < 0.2$, say):

$$V(g) = \frac{\sigma_r^2}{n}\left(1 - \frac{\sigma_r^2}{r^2}\right) \tag{5.7}$$

The United States approach

5.76 The United States CPI uses sampling and estimation procedures which are in many ways unique in comparison to those of other countries. The exact design obviously varies somewhat over time. The following description is based on U.S. BLS (1997) and Leaver and Valliant (1995).

5.77 The United States CPI is composed of building blocks consisting of geographical areas crossed with product strata to a total of 8,487 "basic CPI strata" corresponding to elementary aggregates. The 88 geographical areas were selected by pps in a controlled selection procedure and 29 of them were included with certainty (self-representing). Within each basic CPI stratum an estimation procedure is applied in which indices for a particular time period are based on the overlapping sample units (outlets and items) between this time period and the immediately preceding period. The period-to-period indices are then multiplied to obtain an index from the base period to the current period. Sampling within the basic CPI strata is approximately pps according to the description above.

5.78 Variance estimation for this design proves to be too complex for the use of a direct design-based variance estimator. Instead a random group replication method, using the so-called VPLX software, is applied. Other methods have also been tried.

5.79 Leaver and Swanson (1992) provide a detailed account of the variance estimation methods used up to then. They also present the following numerical estimates of (median) standard errors for CPI changes for various intervals over the 1987–91 period: 1 month –

standard error 0.074; 2 months – standard error 0.103; 6 months – standard error 0.130; and 12 months – standard error 0.143.

The Swedish approach

5.80 The following outline summarizes the description given by Dalén and Ohlsson (1995). The Swedish CPI uses a primary stratification into product groups, which are measured in separate and independent price surveys. The first step in the Swedish approach is therefore to note that the variance of the all items price index is a weighted sum of the variances of the separate surveys:

$$V(I) = \sum_h w_h^2 V(I_h)$$

(5.8)

5.81 The reason that all these surveys can reasonably be assumed to be independent is that there is no common regional sampling scheme used in them. Altogether, there are about 60 different surveys. Some of them cover many product groups and have a complex design and there is stochastic dependence between them. Other surveys cover only one group and have simple designs. Some cover their universes, without any sampling, so they have zero variance.

5.82 In many simple product groups it is fairly reasonable to assume that the price ratios obtained are effectively random samples. In some cases this may lead to some overestimation of variance since there is in fact some substratification or quota sampling within the group. In those product groups, stratum variances could then be estimated according to formulae (5.5)–(5.7). When a price survey is stratified, formula (5.8) can be applied at lower levels above the elementary aggregate.

5.83 Some price surveys are more complex, however. This is especially the case for that large part of the index where outlets and products are simultaneously sampled. In the Swedish case these surveys are called the local price survey and the daily necessities survey. In both these cases, outlets are sampled by probability (pps) from the central business register. Products are sampled by pps in the daily necessities survey but by the representative item method in the local price survey. In the Swedish variance estimation model, the final sample is in these cases considered as drawn from a two-dimensional universe of products and outlets. The final sampling units are thus sampled products sold in sampled outlets – a cross-classified sample.

5.84 In a cross-classified sample, the total variance can be decomposed into three parts:
–   variance between products (in the same outlet);
–   variance between outlets (for the same product);
–   outlet and product interaction variance.
Dalén and Ohlsson (1995) provide the exact formulae used.

5.85 In the daily necessities survey, the cross-classified model comes fairly close to the actual sampling design. In the local price survey, it is more of a model, since the products are in fact purposively drawn. It has nevertheless been considered a useful model for the purpose of getting a first idea of the sampling error and for analysing allocation problems.

5.86 The total variance of the Swedish CPI, according to this model, was estimated to be 0.04, corresponding to a 95 per cent confidence interval of ±0.4. This estimate appeared to be

fairly stable over the period 1991–95 for which the model was tried.

The French approach

5.87   In France, variance calculation at present only takes into consideration items accounting for 65 per cent of the total weight of the index.

5.88   The smallest element of the calculation is a product-type in an urban area. For these elements one of two formulae are applied, depending on whether the product is homogeneous (ratio of arithmetic means) or heterogeneous (geometric means). A two-stage random sample is assumed, first of urban areas and then of a particular item (variety) in an outlet. The variance obtained is thus the sum of a "between urban areas" and a "within urban areas" component. Linearization based on second-degree expansions is done, because of the non-linear nature of the estimators. Higher-level variances are obtained by weighting the elementary level variances.

5.89   After an optimization exercise that took place in 1997, the standard deviation of the all-products index (for 65 per cent of the total weight of the index) was computed as 0.03. This value is close to that estimated in 1993, although the number of observations was reduced. The precision of a number of sub-indices was also improved.

5.90   Covariance terms are ignored. In fact, this makes a very small difference in the "between urban areas" component. In the "within urban areas" component it has undoubtedly a greater influence. The effect is, however, seen as limited because of a rule which limits the number of products observed in the same outlet.

5.91   For the 35 per cent of the weight that is at present excluded from the variance calculation (called the "tariffs"), such calculations will be introduced for insurance. The necessary elements for variance calculation are also present for physicians' and dentists' services. Variances will soon be calculated for these products, as well as for new cars. For a certain number of sub-indices (tobacco and pharmaceuticals) the sample is in effect a total count. Their variances are thus zero.

5.92   A 95 per cent confidence interval for a 12-month comparison can be expressed as the estimated index ± 0.06 for the ordinary, non-tariff items. If zero variance is assumed for the remaining 35 per cent of the index, the confidence interval for the all-products index would become ± 0.04. This assumption is clearly too optimistic, but from the work on variance estimation done so far, it can be concluded that the confidence interval is certainly smaller than 0.1.

5.93   More details on the French computations can be found in Ardilly and Guglielmetti (1993).

The Luxembourg approach

5.94   The Luxembourg CPI can be described as a stratified purposive sample with 258 product strata. There are slightly fewer than 7,000 observations each month, giving an average of 27 observations per stratum. In each stratum, observations are taken from several different outlets; but the same outlet is represented in many product strata. The outlet is here used as the identifier for the price-setting organization (for rents it is a landlord, for insurance it is the companies, and so on). In each stratum, there are observations from several outlets.

Since there is good reason to believe that each outlet has its own price-setting behaviour, prices and price changes in the same outlet tend to be correlated, resulting in positive covariances in the general variance expression:

$$V(I) = \sum_k w_k^2 V(I_k) + \sum_k \sum_l w_k w_l Cov(I_k, I_l)$$

(5.9)

5.95 In the sampling model, each separate outlet sample within a product stratum is regarded as a simple random sample. Further, a two-stage model was assumed such that, in the first stage, a simple random sample of outlets was assumed to have been drawn from a (fictitious) sampling frame of all outlets in Luxembourg. Then, in each sampled outlet, a second-stage sample of observations was assumed to be drawn in product stratum h so that the combined product–outlet stratum became the lowest computational level in the index. All second-stage samples were assumed to be mutually independent and sampling fractions to be small. This model resulted in three components of total variance:

- variance within outlets;
- variance between outlets;
- covariance between outlets.

Covariances are difficult to calculate, even with a computer. Luckily, however, it was possible algebraically to combine the last two components into one, with the number of summation levels reduced.

5.96 Numerical estimates were made with this model for 22 consecutive 12-month changes starting from the period January 1996 to January 1997 and ending with the period October 1997 to October 1998. The average variance estimate was 0.02 (corresponding to a standard error of 0.14), which is surprisingly small given the small sample size. The reason for this small value was not explored in detail but may lie in a combination of the special circumstances in the markets in Luxembourg and in procedures used in the index estimation system.

5.97 The full variance estimation model for the Luxembourg CPI and the results from it are presented in Dalén and Muelteel (1998).

Other approaches

5.98 A number of experimental models have been tried out and calculations done for the United Kingdom. None of them has so far been acknowledged as an official method or estimate. Kenny (1995 and earlier reports) experimented with the Swedish approach on United Kingdom data. He found a standard error of the United Kingdom Retail Price Index as a whole of around 0.1, which was reasonably constant over several years, although the detailed composition of the variance varied quite a lot. Sitter and Balshaw (1998) used a pseudo-population approach but did not present any overall variance estimates.

5.99 For Finland, Jacobsen (1997) provided partial calculations according to a similar design as in the Swedish approach. His analysis was used to suggest changes in the allocation of the sample.

Optimal allocation

5.100 Producing a consumer price index is a major operation in any country and a great deal of resources are spent on price collection. It is therefore worthwhile to devote some effort to allocating these resources in the most efficient way.

5.101   The general approach to sample allocation was established by Neyman and is described in any sampling textbook. It uses a mathematical expression for the variance of the estimate and another expression for the cost. Both variance and cost are functions of sample size. Optimal allocation then amounts to minimizing variance for a given cost or minimizing cost for a given variance.

5.102   Variance estimation was discussed above. As for cost, it is important to note that not all price observations are equally costly. It is less expensive to collect an extra price in an outlet that is already in the sample than to add a price in an outlet that is new to the sample. For example, in the Swedish CPI, the following cost function was used:

$$C = C_0 + \sum_h n_h \left\{ a_h + b_h \sum_g m_g r_{gh} \right\}$$

(5.10)

where C refers to total cost and C_0 to the fixed part of the cost that is independent of sample size,

n_h is the number of outlets in outlet stratum h,

m_g is the number of product varieties in product stratum g,

a_h is the unit cost per outlet and reflects travelling time to the outlet,

b_h is the unit cost per product, which reflects the additional cost for observing a product, when the price collector is already in the outlet,

r_{gh} is the average relative frequency of products in stratum g sold in outlets of stratum h.

5.103   In formula (5.10), a_h is usually much larger than b_h. This fact calls for an allocation with relatively more products than outlets, i.e. of several products per outlet. This allocation is further reinforced to the extent that variances between products in the same outlet and product stratum are usually larger than variances between outlets for the same product. This is the case, at least according to experience in Sweden.

5.104   With a specified variance function and a specified cost function, it is possible, using the mathematical technique of Lagrange multipliers, to derive optimal sample sizes in each stratum. It is usually not possible to obtain explicit expressions, however, since we run into a non-linear optimization problem for which it is not possible to find an explicit solution.

5.105   In a CPI, the all-products index is usually the most important statistic. Therefore, the allocation of the sample should be directed towards the minimization of its error. It is also important that other published sub-indices are of high quality, but the sub-index quality can often be taken as the criterion for publication, rather than the other way round.

Summary

5.106   The above discussion can be summarized in the form of a small number of specific recommendations.

5.107   *Clarity* – sampling rules should be well defined. In many CPIs, there is a wide range of sampling and other solutions for different product groups. A fairly well-defined method is often used for the field collection of prices, but the exact methods used for the central price collection of many products are commonly in the hands of one or a few responsible persons and are sometimes poorly documented. It is essential for the basic credibility of the CPI that rules for sampling and estimation (e.g. the treatment of outliers) are well defined and described.

5.108   *Probability sampling should be seriously considered*. The use of probability sampling

designs should be increased. In many areas, useful sampling frames do exist or could be constructed without excessive difficulties. Stratified, order pps sampling is an important type of design that ought to be considered in many situations. Size measures used for sampling must have a long-term interpretation, so that they are uncorrelated with price movements.

5.109 *Representativity – no large part of the universe should be left out*. When sampling designs are planned, the full universe of items and outlets belonging to the item group in question should be taken into account. All significant parts of that universe should be appropriately represented, unless there are excessive costs or estimation problems involved in doing so.

5.110 *Variance or mean square error should be as low as possible*. Samples should be reasonably optimized, based on at least a rudimentary analysis of sampling variance. As a first-order approximation, sample sizes could be set approximately proportional to the weights of the commodity groups. A better approximation is obtained by multiplying each weight by a measure of price change dispersion in the group. Variance and cost considerations together call for allocations where relatively many products are measured per outlet and relatively few outlets are contained in the sample. Since biases are generally a greater problem than sampling errors, smaller but better samples, allowing for more frequent renewal and careful monitoring of replacements and quality adjustments, generally make good sense.