

12 ERRORS AND BIAS

Introduction

11.1 The CPI, like all other statistics, may be subject to error arising from a variety of sources. Errors may occur during any stage of the estimation process. This chapter discusses the general types of potential error to which all price indices are subject. The literature on consumer price indices (CPIs) discusses these errors from two perspectives, and this chapter presents the two perspectives in turn. First, the chapter describes the sources of sampling and non-sampling error that arise in estimating a population CPI from a sample of observed prices. Second, the chapter reviews the arguments made in numerous studies that attribute bias to CPIs as a result of insufficiently accurate treatment of quality change, consumer substitution and other factors. It should be emphasized that many of the underlying issues discussed here are dealt with in much greater detail elsewhere in the manual.

Types of error

11.2 One of the main objectives of a sample survey is to compute estimates of population characteristics. Such estimates will never be exactly equal to the population characteristics. There will always be some error and the precision and accuracy of the estimate is affected by both sampling and non-sampling error.. Table 11.1 gives a taxonomy of the different types of error. See also Balk and Kersten (1986) and Dalén (1995) for overviews of the various sources of stochastic and non-stochastic errors experienced in calculating a CPI. Two broad categories can be distinguished: sampling errors and non-sampling errors.

Table 11.1. A taxonomy of errors in a consumer price index

Total error:

Sampling error

Selection error

Estimation error

Non-sampling error

Observation error

Overcoverage

Response error

Processing error

Non-observation error

Undercoverage

Non-response

Sampling error

11.3 *Sampling errors* are due to the fact that an estimated CPI is based on samples and not on a complete enumeration of the populations involved. Sampling errors vanish if observations cover the complete population. As mentioned in previous chapters, statistical offices usually adopt a fixed weight price index as the object of estimation. A fixed weight index can be seen as a weighted average of partial indices of commodity groups, with weights being expenditure shares. The estimation procedures that most statistical offices apply to a CPI involve different kinds of samples. The most important kinds are:

- for each commodity group, a sample of commodities to calculate the partial price index of the commodity group;
- for each commodity, a sample of outlets to calculate the elementary price index of the commodity from individual price observations;
- for each commodity group, a sample of a day or a time span of the month when the data collection has to be carried out. Concerning this issue, the introduction of scanner data (that in general cover more than one week of a month) in CPI compilation could reduce the potential errors arising in the traditional data collection for this dimension of sampling.
- a sample of households needed for the estimation of the average expenditure shares of the commodity groups. (Some countries use data from national accounts instead of a household expenditure survey to obtain the expenditure shares.)

Sampling error can be introduced at any of the stages of the sample selection process. The potential for sampling error is greater in the selection of outlets and greatest for products because there is no comprehensive frame from which to select units for sampling.

11.4 The sampling error can be split into a selection error and an estimation error. A *selection error* occurs when the actual selection probabilities deviate from the selection probabilities as specified in the sample design. The *estimation error* denotes the effect caused by using a sample based on a random selection procedure. Every new selection of a sample will result in different elements, and thus in a possibly different value of the estimator.

Non-sampling error

11.5 CPI surveys involve many operations, all of which are potential sources of non-sampling error. The *non-sampling errors* arise from the survey process, regardless of whether the data are collected from the entire universe or from a sample of the population. . They can be subdivided into observation errors and non-observation errors. *Observation errors* are the errors made during the process of obtaining and recording the basic observations or responses. The most general categories of observation errors are overcoverage, response error, and processing error.

11.6 *Overcoverage* means that some elements are included in the survey which do not belong to the target population. For outlets, statistical offices usually have inadequate sampling frames. In some countries, for instance, a business register is used as the sampling frame for outlets. In such a register, outlets are classified according to major activity. The register thus usually exhibits extensive overcoverage, because it contains numerous outlets which are out of scope from the CPI perspective (e.g. firms that sell to businesses rather than to households). In addition, there is usually no detailed information on all the commodities sold by an outlet, so it is possible that a sampled outlet may turn out not to sell a particular commodity at all.

11.7 *Response error* results from the collection of incorrect, inconsistent, or incomplete data. Response error may arise because of the collection of data from inappropriate respondents, respondent memory or recall errors, deliberate distortion of responses, interviewer effects, misrecording of responses, pricing of wrong items, misunderstanding or misapplication of data collection procedures, or misunderstanding of the questions or survey needs and/or lack of cooperation from respondents. In household expenditure surveys, for example, households appear to systematically underreport expenditures on commodity groups such as tobacco and alcoholic beverages or games of chance. In price surveys, where the main price collection method is by persons who regularly visit outlets, they may return with prices of unwanted commodities.

11.8 *Processing error* occurs after the survey data are collected, during the processes that convert reported data to published estimates and consistent machine-readable information. Each processing steps, such as coding, data entry, transfer and editing (control and correction), can generate errors. For example, at the outlets the price collectors write down the prices on paper forms or using a dedicated software on a tablet. In the first case (paper-and-pencil data collection), after the collectors have returned home, a computer is used as the input and transmission medium for the price information. It is clear that this way of processing prices is susceptible to errors. The second case (computer assisted data collections) is less risky because it has built-in validation checks but it could be susceptible to errors for other reasons such as the lack of adequate controls during the registration of the prices. The occurrence of processing errors is strongly influenced by survey planning, and to some extent the survey's resources (e.g., staff and budget, devices) and constraints (e.g., elapsed time between data collection and publication).

11.9 *Non-observation errors* are made when the intended measurements cannot be carried out. The most general categories of non-observation errors are undercoverage, and non-response error. *Undercoverage* occurs when elements in the target population do not appear in the sampling frame used for sample selection. The source of undercoverage error is the sampling frame itself. For instance, it is likely that there are delays in updating the outlet frame to include new in-scope units or to exclude mail order firms and non-food market stalls from their outlet sampling frame. Undercoverage means that some outlets where relevant commodities are purchased cannot be contacted

11.10 *Non-response* is another category of non-observation error. Non-response errors may arise from the failure to obtain the required information in a timely manner from all the units selected in the sample. A distinction can be drawn between total and partial (or item) non-response. Total non-response occurs when selected outlets cannot be contacted or refuse to participate in the price survey. Another instance of total non-response occurs when mail or electronic questionnaires and collection forms are returned by the respondent and the price collector, respectively, after the deadline for processing has passed. Partial (item) nonresponse occurs when a responding unit does not complete an item or items on the survey questionnaire, or the response(s) obtained are unusable. Mail or electronic questionnaires and collection forms that are only partially filled in, scanner data with missing information concerning specific outlets or GTINs (Global Trade Item Number) in the sample and web-scraped prices where some information are not downloaded from the web, , are examples of partial non-response. If the price changes of the non-responding outlets differ from those of the responding outlets, the quality of the price change estimates will be affected.

11.11 Total and partial non-response may also be encountered in a household expenditure survey used to derive the CPI basket weights. Total non-response occurs when households drawn in the sample refuse to cooperate. Partial non-response occurs, for instance, when certain households refuse to give information about their expenditure on certain commodity groups.

11.12 Other source of errors is the failure to measure the price actually paid. This failure may be caused by the use of list prices (e.g. for cars) and by the presence of discounts, coupons, bargaining etc., which are not accounted for. In many countries the “discounting” of prices becomes more common and the importance of “discounted” prices is increasing. Another source of error is due to the tendency of price collectors to choose an excessive proportion of regularly priced items in the price reference period, whereas the proportion of sales prices in the prices collected becomes proportional to their real share later in the year.

Measuring error and bias

Estimation of variance

11.12 The variance estimator depends on both the chosen estimator of a CPI and the sampling design. The 2012 ILO Survey of country practices gives an overview of the sampling methods that are applied in the compilation of CPIs by national statistical institutes. It appeared that only one out of three use some sort of probability techniques for locality selection, outlet selection, one in five for outlet selection, and only one in ten uses probability sampling for item selection. In the absence of probability techniques, so-called judgemental and cut-off selection methods are applied.

11.13 In view of the complexity of the sample designs in compiling a CPI (where the samples of localities, outlets and items are just partially connected), an integrated approach to variance estimation appears to be problematic. That is, it appears to be difficult to present a single formula for measuring the variance of a CPI, which captures all sources of sampling error. It is, however, feasible to develop partial (or conditional) measures, in which only the effect of a single source of variability is quantified. For instance, Balk and Kersten (1986) calculated the variance of a CPI resulting from the sampling variability of the household expenditure survey, conditional on the assumption that the partial price indices are known with certainty. Ideally, all the conditional sampling errors should be put together in a unifying framework in order to assess the relative importance of the various sources of error. Under rather restrictive assumptions, Balk (1989a) derived an integrated framework for the overall sampling error of a CPI.

11.14 There are various procedures for trying to estimate the sampling variance of a CPI. Design-based variance estimators (that is, variances of Horvitz-Thompson estimators) can be used, in combination with Taylor linearization procedures, for sampling errors arising from a probability sampling design. For instance, assuming a cross-classified sampling design in which samples of commodities and outlets are drawn independently from a two-dimensional population, with probabilities proportional to size (PPS) in both dimensions, a design-based variance formula can be derived. In this way Dalén and Ohlsson (1995) found that the sampling error for a 12-month change of the all-commodity Swedish CPI was of the order of 0.1–0.2 per cent.

11.15 The main problem with non-probability sampling is that there is no theoretically acceptable way of knowing whether the dispersion in the sample data accurately reflects the dispersion in the population. It is then necessary to fall back on approximation techniques for variance estimation. One such technique is quasi-randomization (see Särndal, Swensson and Wretman (1992, p. 574)), in which assumptions are made about the probabilities of sampling commodities and outlets. The problem with this method is that it is difficult to find a probability model that adequately approximates the method actually used for outlet and item selection. Another possibility is to use a replication method, such as the method of random groups, balanced half-samples, jackknife, or bootstrap. This is a completely non-parametric class of methods to estimate sampling distributions and standard errors. Each replication method works by drawing a large number of sub-samples from the given sample. From each sub-sample the parameter of interest can be estimated. Under rather weak conditions, it can be shown that the distribution of the resulting estimates approximates the sampling

distribution of the original estimator. For more details on the replication methods see Särndal, Swensson and Wretman (1992, pp. 418–445).

Qualitative descriptions of non-sampling errors

11.16 It is still more difficult to obtain quantitative measures of the non-sampling errors. Thus the use of qualitative indications is the only possibility. For instance, the coverage of the sampling frames as a proxy of the target populations can be addressed (including gaps, duplications and definitional problems). The percentage of the target outlet samples from which responses or usable price data were obtained (i.e. the response rates) can be provided. Any known difference in the prices of responding outlets and non-responding outlets can be described, as can an indication of the method of imputation or estimation used to compensate for non-response. Indicators such as IQIs (Implicit Quality Indices), that compare indices with and without prices adjusted for quality changes, are other examples of quantitative measures of non-sampling errors. Moreover, processing errors tend not to be well-reported or well-documented, and are seldom treated in the survey research literature. Occasional studies of these processing errors in the CPI have shown them to be extremely small. Several categories of non-sampling errors provide the bulk of the bias issues discussed below.

Procedures to minimize errors

11.17 The *estimation error* can be controlled and minimized by means of the sampling design. For example, by increasing the sample size, or by taking selection probabilities proportional to some well-chosen auxiliary variable, the error in the estimated CPI can be reduced. The choice of an adequate sampling design for the CPI is an extremely complex matter (see Dorfman et al. (2006)). The target population is the set of all goods and services that are acquired, used or paid for by households from outlets in a particular time period. A proper probability sampling procedure selects a sample by a random mechanism in which each good or service in the population has a known probability of selection. In combination with a Horvitz-Thompson estimator, such a probability sampling design will produce an index that is (approximately) unbiased and precise.

11.18 The following three probability sampling designs are used extensively in survey practice: simple random (SI) sampling, probability proportional to size (PPS) sampling, and stratified sampling with SI or PPS sampling per stratum. The advantage of SI sampling is its simplicity; it gives each population element the same probability of being included in the sample. PPS sampling has the advantage that the more important elements have a larger chance of being sampled than the less important ones. For instance, at Statistics Sweden the outlets are selected with probabilities proportional to some proxy for size, namely their number of employees. Unequal probability designs can lead to a substantial variance reduction in comparison with equal probability designs. In stratified sampling, the population is divided into non-overlapping sub-populations called strata. For instance, at the United Kingdom Office for National Statistics the population of outlets is split by outlet type (multiple, independent or specialist) to form different strata. In each stratum a sample is selected according to a certain design. One of the reasons why stratified sampling is so popular is that most of the potential gain in precision of PPS sampling can be captured through stratified selection with SI sampling within well-constructed strata. Stratified sampling is in several respects simpler than PPS sampling.

11.19 Because appropriate sampling frames are lacking, samples are frequently obtained by non-probability methods. Judgemental (or expert choice) sampling is one form of non-random selection. In this case an expert selects certain “typical” elements where data are to be collected. With skill on the part of the expert a fairly good sample might result, but there is no way to be sure. A more sophisticated non-probability method is quota sampling. In quota sampling the population is firstly divided into certain strata. For each stratum, the number (quota) of elements to be included in the sample is fixed. Next the interviewer in the field simply fills the quotas, which means in the case of outlet sampling that the selection of the outlets is ultimately based on the judgement of the price collectors. Another non-probability method is cut-off sampling, which means that a part of the target population is deliberately excluded from the sample selection process. In particular, this procedure is used when the distribution of the value of some auxiliary variable is highly skewed. For instance, a large part of the population may consist of small outlets whose contribution to total sales is modest. A decision may then be taken to exclude from the sampling frame the outlets with the lowest sales. Because the selection is non-random, non-probability methods usually lead to more or less biased estimates. Empirical results of research undertaken by Statistics Netherlands nevertheless show that non-probability selection methods do not necessarily perform worse, in terms of the mean square error, than probability sampling techniques (De Haan, Oppendoes and Schut, 1997).

11.20 Provided that the sampling design is given, the sampling variance of an estimated (all-commodities) CPI can in general be lowered by:

- enlarging the samples of households, commodities and outlets;
- the application of suitable stratifications to the various populations (e.g. grouping commodities with respect to similarity of price changes).

11.21 It is important to allocate optimally the available resources both between and within the different CPI samples, since badly allocated samples may lead to unnecessarily high sampling errors. The Swedish variance estimation results, presented in Dalén and Ohlsson (1995), show that the error resulting from commodity sampling is relatively high compared with the error resulting from outlet sampling. In this case, it is worthwhile increasing the sample size of commodities and reducing the sample size of outlets. Beisteiner (2008) stress the importance of allocating resources to those areas where the effect on the quality of the overall CPI is maximized, especially to goods and services with a high relative expenditure weight and to goods and services with high dispersion of prices. The paper presents a “ready-to-use” formula, the Neyman formula, for the allocation of the sample, which optimizes the precision of the CPI for given resources.

11.22 A systematic analysis of sampling errors offers possibilities for improving or reducing cost. The problem of optimum sample allocation is usually formulated as the determination of the sizes of the samples of commodities and outlets, and their distribution over the strata that minimizes the sampling error of an all-commodities CPI, subject to the available budget.

11.23 The sampling errors could also be reduced by making use of scanner data. For example, Bradley (1996) discusses the potential for scanner data to reduce the mean squared error of the CPI, so that calculations based on scanner data may have less sampling error than the corresponding official CPI component index. The use of scanner data also has a positive effect on the time dimension of sampling, covering time span longer than the one covered by the traditional data collection.

11.23 As already mentioned, a business register is usually not an adequate sampling frame for outlets, because it provides extensive *overcoverage*. It is recommended to set up an appropriate sampling frame by enumeration of the main outlets within each sampled municipality. Such enumeration yields a list of all outlets in a municipality together with the commodity groups that belong to their assortments. A less expensive way to organize an outlet sampling frame is to ask the price collectors – who may be assumed to know the local situation well – to make a list of outlets where purchases are made by households. It is important, that information about the quality of the sampling frame, in terms of overcoverage or undercoverage, and its completeness for the target population is known.

11.24 The populations of commodities (and varieties) and outlets are continually changing through time. The composition of most commodity groups is not constant over time, because commodities disappear from the market and new ones appear. The passage of time also plays a disturbing role with respect to the outlet population: outlets close, temporarily or permanently; new outlets emerge; the importance of some outlets diminishes or increases. The samples of commodities (and varieties) and outlets should be reviewed and updated periodically to maintain their representativity with respect to the current buying habits of the households.

11.25 *Response errors* caused by the underreporting of certain categories of household expenditure can be adjusted by using producer-based estimates from the national accounts (see Linder (1996) for an example). *Measurement errors* by price collectors can be reduced by providing them with hand-held computers for data entry. In this way the validation of observed prices can be executed at the point of price collection (i.e. in the outlet), by means of an automatic comparison of the currently observed price quote with the previously observed one (by setting a limit on the percentage price change) and with the price quotes obtained from other outlets (by setting suitable upper and lower limits). Details are provided by Haworth, Fenwick and Beaven (1997). It also believed that although using pricing forms that contain information on the previous period's price can reduce response variance it may as well cause response bias and lag.

11.26 It is useful to appoint data collection supervisors to conduct quality assurance checks on the data collectors. It is also a good idea to organize regular meetings where price collectors and statisticians from the head office can share their experiences. In this way, the statisticians will keep in touch with the conditions in the field, and may take the opportunity to provide more information about frequently made price collection errors and new representative goods.

11.27 It is important to check the collected price data for *processing errors* and, where possible, to correct these errors. This activity is called data editing. When editing is carried out on individual observations, it is called micro-editing. When the resources to spend on data editing must be minimized, while at the same time maintaining a high level of data quality, selective editing and macro-editing are possibilities. Selective editing is a form of traditional micro-editing, in which the number of edits is kept to a minimum. Only those edits which have an impact on the survey results are carried out. Macro-editing offers a top-down approach. The edits are carried out on aggregated data (for instance, the price index numbers of a commodity group) instead of individual records (for example, price observations). Micro-editing of individual records is then carried out only if macro-edits raise suspicion. In particular, attention should be paid to outliers among the observations.

11.28 *Non-response* reduces sample size, results in increased variance, and usually introduces selection bias. Nonresponse rates, in terms of missing observation, are often viewed as a proxy for the quality of a survey. While nonresponse rates are important, imputation rates alone provide no indication of nonresponse bias. There are three methods for the treatment of missing price observations. First, the corresponding price can be excluded from the data set of previous prices, so that the set of previous prices is “matched” with the set of current prices. Second, this matching can be achieved by using an imputed (or artificial) price for the missing one. The imputed price can be calculated by either carrying forward the previous price observation or by extrapolating the previous price observation using the change of other price observations for the same commodity. Third, there is the possibility to reweight the sample to minimize the effect of non-response error. The objective of reweighting is to inflate the weight given to the prices of the responding outlets. This compensates for those prices that are lost by non-response.

11.29 In a household expenditure survey, missing data are usually imputed with the help of information on the same household from a previous observation period or other households from the same observation period. To reduce bias in the average expenditure pattern arising from selective non-response, a household expenditure survey sample of households is generally post-stratified by a number of household characteristics, such as income, composition and size.

Types of bias

11.30 This section reviews several categories of error, either in pricing or in index construction, that potentially can lead to bias in the overall CPI. The emphasis here is on the categorization of these types of errors, along with some consideration of their likely size, but also on methods to reduce or eliminate these categories of errors. The question might arise of why such a discussion is necessary, since such issues as quality change, and the appropriate methods for handling them in the CPI, are dealt with at both a conceptual and operational level in other chapters.

11.31 The reason this chapter addresses the topic of CPI bias per se is the great surge in interest in price measurement problems during the mid-1990s. Especially in the United States, the view became widespread that the CPI was subject to systematic upward biases because of the failure to deal adequately with consumer substitution, product quality improvements, and the introduction of new items and services. Moreover, it was recognized, first, that the existence of such upward bias would have fundamental implications for the measurement of recent trends in output and productivity, and second, that the elimination of upward bias could substantially improve the government budget situation through reduced government expenditures and increased tax revenues (see, for example, Eldridge (1999) and Duggan and Gillingham (1999)). These discoveries led to a series of papers and reports on CPI measurement problems, often accompanied by point estimates of aggregate bias.

11.32 Prominent examples of these quantitative studies of bias are those by the Advisory Commission to Study the CPI (United States Senate, 1996), Congressional Budget Office (1994), Crawford (1998), Cunningham (1996), Dalen (1999a), Diewert (1996c), Lebow, Roberts and Stockton (1994), Lebow and Rudd (2003), Shapiro and Wilcox (1997b), Shiratsuka (1999), White (1999), and Wynne and Sigalla (1994). Responses and estimates by statistical agencies include those provided by Abraham et al. (1998), US Bureau of Labor

Statistics (1998), Ducharme (1997), Edwards (1997), Fenwick (1997), Johnson et al. (2006), Lequiller (1997), Moulton (1996b), and Moulton and Moses (1997). Among the many other discussions of the CPI bias issue are those reported by Baker (1998), Berndt (2006), Boskin et al. (1998), Deaton (1998), Diewert (1998a), Gordon (2006), Krueger and Siskind (1998), Nordhaus (1998), Obst (2000), OECD (1997), Pollak (1998), Popkin (1997), and Triplett (1997, 2006).

11.33 Two points are worth making at the outset with respect to measuring bias in CPIs. First, the issue has usually been addressed in the context of the cost of living index (COLI). That is, the CPI bias has been defined as the difference between the rate of increase in the CPI and the rate of increase in a true COLI. Many authors on bias have taken as given that the COLI should be the CPI's measurement objective. Somewhat different conclusions might be reached if the index objective were taken to be a pure price index. Notably, the gains in consumer welfare from a widening array of new goods, or the ability of consumers to substitute away from items with increasing relative prices, might be deemed irrelevant and an index that ignored those factors might not be judged biased on that account.

11.34 The second point is that CPI bias is not amenable to estimation with the same level of rigour as that used in CPI variance estimation. Since the COLI or other ideal target index is unobserved, analysts have been forced to rely in part on conjectures and on generalizations from fragmentary empirical evidence in order to quantify the extent of bias. The notable exceptions are with respect to substitution bias, when traditional Laspeyres indices and indices using superlative formulae can be computed using the same underlying price and expenditure data, and the differences construed as a measure of the upward bias from use of the Laspeyres formula.

11.35 Several different taxonomies of bias have appeared in the literature mentioned above. It is sufficient, however, to employ four categories roughly corresponding to those set forth in the best-known study, namely the *Final report of the Advisory Commission to Study the CPI* (the Boskin Commission), established by the United States Senate Finance Committee in 1995. These categories are: upper-level substitution bias; elementary aggregate bias; quality change and new goods bias; and new outlet bias.

11.36 These categories can be further broken down into two subgroups according to whether they refer to errors in individual price measurements or errors in computing index series. Quality change bias and new goods bias arise because of failures to measure adequately the value to consumers of individual goods and services that appear in (or disappear from) the marketplace. It should be recognized that discussions of "new goods" problems apply equally to all products, whether goods or services. At a conceptual level, it can be difficult to distinguish these two biases from each other. Operationally, however, quality change bias pertains to the procedures for comparing new products or models with the older products they replace in the CPI samples. In general, new goods bias can be thought of as applying to wholly new types of products, or products that would not enter samples routinely through forced replacement. New outlet bias, sometimes referred to as outlet substitution bias, is similar to new goods bias but is focused on the appearance of new types of stores or marketing methods that offer goods at lower prices or higher quality.

11.37 The other categories of bias refer to the procedures for constructing index values from component series. As noted throughout this manual, CPI construction can be thought of as taking place in two steps, or at two levels. At the lower level, individual price quotations are

combined; at the upper level, these basic indices are aggregated together. Corresponding to these two levels are two forms of potential bias. Elementary aggregate bias involves the averaging formulae used to combine price quotations into basic indices. Upper-level substitution bias applies to the formulae used to combine those elementary aggregates into higher-level indices. These components of potential bias, and the means used to measure them, are discussed in more detail below.

Components of bias

Upper-level substitution bias

11.38 Upper-level substitution bias is perhaps the most widely accepted source of CPI bias, and the kind with which economists are most familiar from textbook expositions of price index theory and practice. Simply stated, it arises when CPIs employ the Laspeyres formula (see **Chapter 17**), which is well known to provide an upper bound on a cost of living index under certain assumptions about consumer behaviour. The assumption in the definition of the Laspeyres-type price index is that substitution among goods is zero, which is against one of the cornerstones in the theory of demand. As noted in paragraph 11.34 above, quantitative measures of upper-level substitution bias can be generated by comparing Laspeyres price indices to Fisher ideal, Tornqvist or other superlative indices. Under certain assumptions about, for example, constant preferences, these will stand as relatively precise bias estimates.

11.39 Genereux (1983) and Aizcorbe and Jackman (1993) provide such index comparisons and estimates of upper-level substitution bias using actual CPI index series for Canada and the United States, respectively. Other early studies by Braithwait (1980) and Manser and McDonald (1988) estimate the substitution bias in United States national account indices. In lieu of superlative indices, the Braithwait study uses estimated exact cost of living indices based on demand system estimation. A similar estimate for the Netherlands is provided by Balk (1990). In these studies and in the more recent analyses of U.S. CPI data by Shapiro and Wilcox (1997a) and Cage et. al. (2003), the existence of an upward bias from the Laspeyres formula is demonstrated consistently. The biases in the annual index changes in individual years are relatively small, typically 0.3 percentage points or less, and depend empirically on such factors as the distance from the Laspeyres base period, the level of index detail at which the alternative formulae are applied, and whether the superlative index is of the fixed base or chained variety.

11.40 The major differences between Laspeyres and superlative indices derive from the variation in relative prices over the period being compared, and from the shift in quantities consumed towards those index categories that have fallen in relative price. This leads to several conclusions:

- If index movements are characterized by continuing, uniform drift in relative prices over time, with accompanying drifts in consumption, the size of the annual Laspeyres bias will tend to increase with the distance from the base period. (Greenlees (1997)) notes, however, that there is little evidence for this phenomenon in the United States; see also Szulc (1983).)
- Under the same circumstances, reducing the expenditure weight chaining interval will work to reduce the upper-level substitution bias in the Laspeyres CPI. The more frequent chaining will increase the weight given to indices that are falling in relative price, thereby reducing the rate of CPI growth. Conversely, if there is “bouncing” in

relative index movements, frequent chaining can lead to an upward “chain drift” in a Laspeyres index.

- Upper-level substitution bias will tend to be larger during periods of higher inflation, if these periods also have greater relative price variation. Little empirical evidence exists on this point, however.

11.41 The concept of upper-level substitution bias has been derived and discussed in the context of cost of living index theory, but an equivalent bias may be defined from the perspective of the pure price index. If the Fisher ideal or other superlative index is judged preferable on the basis of its symmetric treatment of base period and current period expenditure patterns, then the difference between that index and a Laspeyres could be interpreted as a measure of representativity bias. A similar argument could be applied with respect to lower-level substitution bias within elementary index cells.

11.42 Recently, Lebow and Rudd (2003) have defined and estimated another category of bias related to upper-level aggregation. They concluded that the consumer expenditure survey weights used in the United States CPI were subject to error because of, for example, underreporting of alcohol and tobacco expenditures. This will lead to a weighting bias if the errors in relative weight are correlated with component index changes. (Sources for, and problems in, expenditure weight estimation are discussed in detail in [Chapter 4](#).)

11.43 Haan and Heymerik (2009) have identified another problem that affect the bilateral superlative price indexes when scanner data are being used. In order to handle the problem of high attrition rate of items, the use of high-frequency chaining, can create drift in the index series when prices and quantities oscillate or bounce arising from sales. Therefore new methods have been developed for price measurement based on scanner data. For example, the approach proposed in Ivancic, Diewert and Fox (2009) that provides drift free, superlative-type indexes through adapting multilateral index number theory seems to provide a solution to the problem. The methods proposed, however, pose some practical challenges and require some more evidence before getting it widely accepted. For overview of methods for price measurement using scanner data see Chessa Verburg and Willenborg (2017).

Elementary aggregate bias

11.43 Elementary aggregate bias can be divided into two components: formula bias and lower-level substitution bias. An elementary index in the CPI is biased if its expectation differs from its measurement objective. The term formula bias (or functional form bias) is used here to denote a situation in which the elementary index formula has an upward bias relative to the pure price index. When the measurement objective is a cost of living index, the elementary index formula suffers from lower-level substitution bias (or within-stratum substitution bias) if it does not reflect consumer substitution among the items contained in that index cell. Thus, given any elementary index formula, the two forms of bias can be distinguished according to the objective of the elementary index.

11.44 [Chapters 9 and 20](#) of this manual discuss the characteristics of alternative elementary index formulae. A key result is that the Carli formula for the arithmetic average of ratios has an upward bias relative to the trend in average item prices. Consequently, Eurostat has prohibited use of this formula in computations for the Harmonized Indices of Consumer Prices (HICPs). The weighted formulae used in basic indices of the United States CPI had

some characteristics of the Carli formula prior to procedural and computational changes made in 1995 and 1996. The problems and the methods chosen to address them are discussed, for example, by Reinsdorf (1998), Reinsdorf and Moulton (1997) and Moulton (1996b).

11.45 The ratio of arithmetic averages (Dutot) and geometric mean (Jevons) formulae eliminate formula bias as defined here, and both are permitted by Eurostat. Their expectations differ, however, when item prices do not change at a uniform rate. The differences provide one way of evaluating the potential importance of lower-level substitution bias. The geometric mean formula is exact for a cost of living index if consumers follow the Cobb-Douglas behavioural model, whereas the formula based on the ratio of arithmetic averages corresponds to zero-substitution behaviour. Thus, if the goal is to approximate a cost of living index, the geometric mean formula is likely to be judged preferable.

11.46

Scanner data provide new opportunities for measuring and addressing elementary aggregate bias. The availability of both prices and quantity information in scanner data remove the need for unweighted index formula, at least for those items where unit values are available, and it allows calculation of elementary CPI indices by employing superlative formulae. By using scanner data Gabor and Vermeulen (2015) compute product category level elementary price indexes using nine different index formulas (Carli, Dutot, Jevons, Laspeyres, Paasche, Fisher Ideal, Lowe, Geometric Lowe and expenditure weighted Jevons) and compare the resulting indices with the Fisher index. The main findings are that across product categories, mean levels of annual elementary index bias vary between -0.53 percentage points and 0.55 percentage points depending on the index.

11.47 The method used by the statistical agency for sampling items within a category will determine the effectiveness of formula choice in dealing with lower-level substitution bias. For example, if only a single representative item is chosen to represent the category, the index formula will fail to reflect the consumer response to any relative price change in the universe of items. Therefore a larger sample of representative items should yield a smaller sampling variance for a given elementary index. More generally, the geometric mean formula index suffers from an upward bias in very small samples (less than 5 observations), so lower-level substitution bias may be underestimated in empirical comparisons of the geometric mean to other index formulae. White (1999) discusses the relationship between sampling error and bias estimates. McClelland and Reinsdorf (1999) also study the impact of small sample sizes on the index and conclude that it has the effect of rising the expected values of an index based on non-linear formulas, especially the geometric mean formulae. . To mitigate the impact of small sample the best alternative is to increase sample size. More extensive use of scanner data may mitigate the problem all together given that sample sizes in a typical scanner data set are large.

11.48 The impact of formula choice can be estimated with some degree of precision over a given historical period. Any corresponding bias, however, can be estimated only by assuming that the geometric mean or other functional form successfully approximates the index's measurement objective.

11.49 As implied by the above discussion, the importance of elementary aggregate bias will vary by country, depending on the particular index formulae used, the degree of heterogeneity within index strata, and the sampling methods employed. Also, as with upper-

level substitution bias, elementary aggregate bias will vary with the overall level of inflation in the economy if absolute and relative price changes are correlated.

11.50 The performance of any formula for elementary aggregate calculation will also be affected by the methods used by the statistical agency to handle special situations, such as seasonal goods and other products that are temporarily unavailable. Armknecht and Maitland-Smith (1999) discuss how the failure to impute missing prices can lead to bias in the modified Laspeyres and other index formulae.

Quality change and new products bias

11.51 Discussion of potential CPI biases arising from inadequate quality adjustment has a long history. For example, the Stigler Committee report on United States price statistics (Price Statistics Review Committee, 1961) indicated that “if a poll were taken of professional economists and statisticians, in all probability they would designate (and by a wide majority) the failure of the price indices to take full account of quality changes as the most important defect of these indices”. In most studies of bias, unmeasured or mismeasured quality change is also the largest contributor to the total estimated bias. Just as quality adjustment is widely recognized as an extremely difficult process, however, it is correspondingly difficult to measure any quality change bias.

11.52 Unlike substitution bias, which can be estimated by comparison of alternative formulae, quality change bias must be analysed on a product-by-product basis. Products and their associated index components will experience widely varying rates of quality change over time. Moreover, the methods used for quality adjustment will also vary. Whereas the linking method may dominate in terms of frequency of use, important index components may employ production cost, hedonic adjustment, or the other methods described in **Chapters 7 and 21**.

11.53 A crucial point to recognize is that the direction of overall quality change does not imply the direction of any quality change bias. Non-experts sometimes assume that the CPI does little or no quality adjustment, and that it therefore must overestimate price change in view of the many demonstrable improvements over time in the quality of goods and services. Rather, for any component index, the issue is whether the direct or indirect method chosen for quality adjustment overestimates or underestimates the relative quality of replacement items in the CPI sample. The resulting bias can be either positive or negative.

11.54 Empirical evidence on quality change bias has been based largely on extrapolation from individual studies of particular products. These individual studies may involve, for example, comparisons of hedonic regression indices to the corresponding CPI series or estimates of the value of some product improvement that is ignored in CPI calculations. Although the majority of such studies have suggested upward rather than downward bias, the reliance on fragmentary evidence has led to criticism by observers who point to evidence of quality declines that have not been subjected to systematic analysis.

11.55 Especially for services, overall quality trends can also be a matter of subjective valuation. New technology has led to unambiguous improvements in the quality of many consumer durables and other goods. By contrast, in service sectors such as mail delivery, public transport and medical care, it can be difficult to evaluate changes in quality. Airline

travel, for example, has become safer and faster but perhaps less comfortable and reliable in recent decades, and the lack of cross-sectional variation in these characteristics makes the use of hedonic quality adjustment problematic.

11.56 New product bias, like elementary aggregate bias, can be divided conceptually into two components. The first concerns the failure to bring new products into the CPI sample with sufficient speed. This can lead to upward bias if those new products later experience large price reductions that are not reflected in the index. The second component is the welfare gain that consumers experience when a new product appears. This may not be viewed as a bias, however, when the cost of living index is not accepted as the CPI's measurement objective.

11.57 As discussed in **Chapter 8**, “new goods” can be: products that replace predecessor items, for example cloud storage areas replacing physical storage devices; product varieties that widen the range of consumer choice, such as no alcoholic beers and ethnic restaurants; or products that represent wholly new categories of consumption, such as multi task robots for cooking or smartphones.

11.58 Like quality change bias, new product bias has sometimes been estimated primarily by generalization from individual product evidence. A frequent approach has been to measure the price change for a product or category during a period prior to its entry into the CPI sample. Studies by Hausman (1997, 1999) of breakfast cereals and cellular telephones provided quantitative measures of the consumer surplus gain from the new products, but this complex econometric approach has not been applied widely. Some of the Boskin Commission's estimates of new product bias, notably those for food, were necessarily based on conjecture.

11.59 Also, like quality change bias, new product bias could be negative if the range of products decreases, if valuable consumer goods disappear from the market, or if the index fails to capture phases of rapid price increase for items. Most observers, however, seem to agree on the direction of bias as upward, and that the uncertainty concerns the magnitude. The extent of the new product bias depends on the importance of the new products in terms of the proportion of consumer expenditure spent on new products not yet introduced in the CPI basket, and the extent of the price decline from the initial price. **11.60** One of the risks of downward bias in CPI is associated with producers that reduce the package size of household goods keeping the price stable (“shrinkflation”) or repackaging the old product. This phenomenon is strictly linked with the minor changes in product packaging or product characteristics (so called “product relaunches”). These should be properly handled, in particular in scanner data, because the product relaunched in most of the cases presents a new GTIN but it is directly comparable with the product before the relaunch. When scanner data are used in CPI and big amount of data are processed on weekly basis, it is not possible to observe and report all changes incurred in the size or in the characteristics of the product and assessing the comparability of the previous and the replacing product. Therefore automatic procedures have to be carefully implemented to link different GTINs in contiguous months and to correctly estimate the price change, avoiding the risk of bias that usually is downward.

New outlet bias

11.60 Conceptually, new outlet bias is identical to new product bias. It arises because of the failure to reflect either price changes in new outlets not yet sampled, or the welfare gain to consumers when the new outlets appear. The explanation for its existence as a separate bias

category is twofold. The first reason is historical: new outlet bias was identified by Reinsdorf (1993) as a potentially major explanation for anomalous movements in the United States CPI. Second, the methods used to sample and compare outlets differ from those used with products, and the problems in controlling new outlet bias are somewhat different.

11.61 A failure to maintain a current outlet sample can introduce bias because the new outlets are distinctive in their pricing or service policy. Reinsdorf (1993), and more recently Hausman and Leibtag (2004, 2005), focus on the growth of discount stores. It should be noted, however, that the problem could also be geographical in nature; it is important to employ outlet sampling frames that reflect new as well as traditional shopping locations, although the wide spread and increasing weight of online shops have changed the dimension of this issue.

11.62 One way that new products enter the CPI sample is through forced replacement, when exiting or less successful products disappear from shelves. Outlet disappearance is less frequent, and agency procedures may not provide for automatic replacement. Moreover, when a new outlet enters the sample there are no standard procedures for comparing data at the new and old outlets. Thus, the index will not incorporate any effects of, for example, lower price or inferior service quality at the new outlet.

11.63 Reinsdorf (1993) estimated the degree of new outlet bias by comparing average prices at outlets entering and disappearing from United States CPI samples. There has been little or no empirical work, however, on the measurement or consumer valuation of outlet quality. As a consequence, there is little evidence on which to evaluate the accuracy of new outlet bias estimates.

11.64 Greenlees and McClelland (2011) confirms the potential importance of new outlets bias in the CPI. They provided new and detailed evidence on the impact of the appearance and growth of new types of product outlets on the CPI. Using BLS-collected price data for 2002-2007, they observed a continuous increase in the market share of discount department stores and warehouse club stores, and significantly lower prices at discount department stores and at warehouse club stores than at large grocery stores. The increasing shares of lower-priced store categories reduced the average prices collected by the BLS. Changes in the distribution of outlets within categories also led to a substantial decline in average prices. Hausman (2004) also documents the growing role of discount outlets and provides a specific example of outlet bias.

11.65 Like with new product bias, most observers, seem to agree on the direction of bias as upward. The extent of the new outlet bias depends on (i) the components of the CPI basket that are likely to be affected, (ii) the change in market share of new outlets for these items, and (iii) the percentage difference in quality-adjusted prices between new outlets and old outlets. Estimates of the size of outlet-substitution bias must take into account the fact that the market price of an item depends on both the quality of the commodity and the quality of the retail outlet where it is purchased, based on such factors as the level of service and the convenience of the location.

Summary of bias estimates

11.64 The 1996 Boskin Commission report gave a range of estimates for the total upward United States CPI bias of 0.8 to 1.6 percentage points, with the point estimate being 1.1 percentage point. This total reflects the straightforward summation of the component bias estimates. As reported by the United States in United States General Accounting Office (2000), however, changes in CPI methods subsequent to 1996 led the Boskin Commission members to reduce their estimates of total bias. Lacking evidence to the contrary, additivity of biases has been assumed in most such studies. Shapiro and Wilcox (1997b) provide probability distributions and correlations of their component bias estimates, yielding an overall confidence interval for the total bias. Most detailed studies of bias also conclude that the CPI bias is in an upward direction, although there have been numerous criticisms of that conclusion.

11.65 It is apparent that statistical agencies cannot compute or publish CPI bias estimates on a regular basis. Many of the same obstacles that prevent the elimination of bias also stand in the way of estimating bias. These include the lack of complete data on product-level consumer preferences and spending behaviour, and the inability to observe and value all differences in quality among items in the marketplace. Without such information it is impossible to calculate a true cost of living index, and similarly impossible to measure the divergence between its rate of growth and the growth rate of the CPI.

11.66 Statistical agencies have been reluctant to provide their own estimates of CPI bias. In some cases, they have accepted the existence of substitution bias, recognizing that the use of a Laspeyres formula implies that the CPI usually will overstate price change relative to a cost of living index. Statistical agencies have, however, been reluctant to draw even qualitative conclusions from the fragmentary and speculative evidence on quality change, new products and new outlet bias.

11.67 The CPI bias may appear to a different extent in different countries. Hanousek-File (2001) show that bias was especially high in countries in transition, and during the period of high inflation. They argue that the substitution bias increases with the increase of the variance of relative price changes. They argue that bias arising from substitution increases in line with the level of inflation, because as the rate of inflation increases, so does the variance of relative price changes.

Procedures to minimize bias

11.68 Although it is almost impossible to eliminate sources of bias, some measures can be taken to minimize them. These include:

(i) Use appropriate formulas in compiling elementary aggregate indexes, in particular use of the GM formula where appropriate or the RAP formula.

(ii) Use a superlative index formula rather than the Laspeyres, if current-period weighting data can be obtained on time. Where Laspeyres-type formula is used, the upper level substitution bias can be reduced by more frequent updating of expenditure weights, implementing them with minimal time lag. Given that significant part of the total measurement bias in the CPI may be caused by the fixed nature of the CPI basket, the commodity-substitution bias and some of the new-goods bias could be reduced by increasing the frequency at which weights are updated. Other options might be to use formulas that

allow substitution or assumptions about substitution between commodity groupings to be entered.

(iii) Closely monitor and update price samples to reflect changes in the outlets from which households purchase. For example, there is clearly a need to plan for the inclusion in consumer price indexes of purchases from outlets operating exclusively online but also from discount outlets, factory outlets, etc. whose importance has been increasing.

(iv) Include new goods into the CPI as soon as possible. For a fixed-weighted index such as Laspeyres, there would also be a need to update the fixed weights to allow for the inclusion of the new goods if they are substituting for all goods in general, or to adjust the weights within a commodity grouping if the new good is substituting for specific items. For example, one could argue that CDs were a new good, but as they were substituting for records and tapes they could be introduced into the commodity grouping for records and tapes, and weights between these items adjusted accordingly

(v) Make sure the most appropriate quality adjustment methods are applied. Whenever suitable and where price determining characteristics are quantifiable, make greater use of the hedonic technique to adjust for quality change .

(vi) Make greater use of the scanner data to deal with quality change, substitution and new products. They contain detailed and timely information on the prices and quantities of all consumer transactions. The role for scanner data cannot be underscored, given its ability to track market trends and detect the emergence of new products on the market, helping to reduce the lag in introducing new goods into the CPI basket. The use of scanner data also makes it possible to compile superlative price indexes at detailed aggregation levels since prices and quantities are available.

Conclusion

11.69 The CPI is subject to various types of errors that affected precision and accuracy of the CPI estimates. Several potential sources of errors and bias have been identified in the CPI and addressed, though there continues to be debate over to what extent and in what direction bias may still exist and the ways in which its accuracy can continue to be increased.

11.67 In order to ensure public confidence in a CPI, a detailed and up-to-date description of the methods and data sources should be published. The document should include, among other things, the objectives and scope of the index, details of the weights, and last but not least, a discussion of the accuracy of the index. A description of the sources and magnitude of the sampling and non-sampling errors (coverage, non-response rates, etc.) in a CPI provides users with valuable information on the limitations that might apply to their uses of the index. Examples of handbook of CPI methods are that published by the United States Bureau of Labor Statistics (2015), and Australian Bureau of Statistics (2009), which devote a section to the varieties and sources of possible error in the index.

11.68 Improving precision and accuracy of the CPI will take both time and resources. Resources should be allocated to those areas where the effect on the quality of the overall CPI is maximized, especially to goods and services with a high relative expenditure weight and to goods and services with high dispersion of prices. Further uses of scanner data can help statistical agencies deal with the quality changes, outlet substitution and new goods bias

problems. An investigation into factors affecting consumer choice and an expanded consumer expenditure survey would help identify consumer preferences for different outlet types and to improve accurate price measurement in areas where quality change is rapid. Opportunities presented by technology, such as the use of computer-assisted techniques, scanning and web scraping techniques, can minimize processing errors.

References to be removed:

Boon (1998)

References to be added:

Valentina Stoevska and Esther Namukasa, Methodologies of compiling consumer price indices 2012, ILO Survey of country practices, available at <http://beaver.ilo.ch:8888/ilostat>.

Alexandra Beisteiner (2008), Optimal allocation of prices in practice - using the Neumann formula as tool., <http://www.unece.org/index.php?id=14519&type=111>

Filer, Randall K., and Hanousek, Jan—Survey-Based Estimates of Biases in Consumer Price Indices during Transition: Evidence from Romania, *Journal of Comparative Economics*, [Volume 30, Issue 3](#), September 2002, Pages 476-487

Greenlees, J. S. and R. McClelland. 2011. “New Evidence on Outlet Substitution Effects in Consumer Price Index Data.” *Review of Economics and Statistics* 93 (2): 632–46.

Bradley, R. (1995). The use of scanner data as a means to reduce the mean squared error in the CPI. Working paper, U.S. Department of Labor, Bureau of Labor Statistics

Jan de Haan and Heymerik van der Grient, Statistics Netherlands 2009. Eliminating Chain Drift in Price Indexes Based on Scanner Data
<https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2010/zip.8.e.pdf>

Eniko Gabor and Philip Vermeulen (Bundesbank and European Central Bank) (2015): New evidence on elementary index bias
<http://www.stat.go.jp/english/info/meetings/og2015/01agenda.htm>

Antonio G. Chessa¹, Johan Verburg² and Leon Willenborg, 2017 A Comparison of Price Index Methods for Scanner Data,
https://www.bundesbank.de/Redaktion/EN/Downloads/Bundesbank/Research_Centre/Conferences/2017/2017_05_10_ottawa_group_07_1_paper.html?__blob=publicationFile

Ivancic, L., E.W. Diewert and K.J. Fox, 2009, Scanner data, time aggregation and the construction of price indexes, Discussion paper 09-09, Department of Economics, University of British Columbia, Vancouver Canada, V6T 1Z1

US BLS, 2015, Handbook of Methods, Chapter 17, Consumer Price Indexes
<https://www.bls.gov/opub/hom/pdf/homch17.pdf>

ABS, 2009, 6461.0 - Consumer Price Index: Concepts, Sources and Methods, 2009
<http://www.abs.gov.au/ausstats/abs@.nsf/mf/6461.0>