Journal of Official Statistics, Vol. 11. No. 3, 1995, pp. 261–275

Quantifying Errors in the Swedish Consumer Price Index

Jörgen Dalén[1]

Approaches for evaluating, comparing and aggregating some of the most important errors in a Consumer Price Index are presented. Their application to Swedish CPI data for 1981–1992 is described and discussed.

Key words: Mean squared error; quasi-randomisation; bias estimation; nonsampling error.

1. Introduction

Attempts to measure and aggregate errors of a Consumer Price Index (CPI) are probably as old as the index itself. The first work of which we are aware is in the collected papers of Edgeworth (1925) which contains a paper first published in 1888. Edgeworth discusses three types of errors: (1) errors in weights, (2) errors in price relatives and (3) errors resulting from unrepresented product categories. He discusses these errors theoretically based on a mathematical model as well as empirically and gives advice that has since been passed on to subsequent generations of index practitioners:

"Take more care about the prices than the weights" (p. 320).

Since then a number of papers on this topic have emerged – see Biggeri and Giommi (1987) for a fuller reference list. Biggeri and Giommi themselves provide a detailed error analysis based on the mean squared error model. They also give an extensive list of all types of errors present in a CPI. However, when it comes to actual numerical estimation of the error components, only sampling errors are considered and low level sampling variation is disregarded.

Sweden also has a long tradition of dealing with the CPI precision problem. Early works are Ruist (1953) and Malmquist (1958), both in Swedish. Ruist looked at problems of regional, outlet and product sampling as well as errors due to quality change. Malmquist's main approach was to study the formula error of a Laspeyres index, but he also continued Ruist's approach with later data. Both works contain many numerical estimates. Andersson, Forsman, and Wretman (1987) studied the variance contribution from outlet sampling for some commodity groups in the Swedish Consumer

[1] Department of Economic Statistics, Statistics Sweden, S-115 81 Stockholm, Sweden.

Price Index (Konsumentprisindex or KPI). Dalén and Ohlsson (1995) give a more complete list of papers dealing with the CPI variance estimation problem.

A CPI is based on several levels of aggregation. From a technical statistical point of view, many CPIs are composed of a number of surveys covering different areas of private consumption. In the case of the KPI where no regional levels are involved, these surveys have independent designs.

In this report different error measures are presented along with a framework for their analysis. We aim to provide a unified structure for discussing, comparing and aggregating important errors in a CPI, considering that these errors originate from different sources and surveys. There is obviously a need for such a structure, given the interest among economists to assess the aggregate error or bias of a CPI. There are many examples of such analyses; the latest is Wynne and Sigalla (1994), who also cite earlier work of this kind.

Like Wynne and Sigalla and other economists, we take a broad view of what constitutes errors and biases in a CPI and try to combine the viewpoints of both economists and statisticians. Conceptual issues as well as formula choice play major roles when economists assess CPI accuracy. On the other hand economists pay less attention to sampling errors and the kind of nonsampling errors that survey statisticians are used to analysing.

This analysis is geared to the particular structure and procedures of the KPI and the details of our model may therefore not be generally applicable. But we believe that with suitable modifications our general approach may be applicable to the often similar circumstances that prevail in the CPIs of, e.g., many European countries. For the U.S. CPI with its use of full probability sampling in all dimensions some aspects of our approach may not be applicable.

There are various types of procedures for estimating KPI errors. Scientifically "hard" variance estimators can be used for errors arising from a probability sampling design. For the KPI, variance estimation is discussed in Dalén and Ohlsson (1995). But for other errors "softer" methods have to be used. In this article we use two such methods. One is *quasi-randomisation*, i.e., assuming that a given set of data has been generated by a random procedure although probability sampling was not strictly used. (The term quasi-randomisation is taken from Särndal, Swensson, and Wretman (1992, p. 574), where it is used for modelling the nonresponse distribution in surveys). The other method is *sensitivity analysis*, where the actual results from different methods are compared.

In Section 2 we present a structured list of types of errors that occur in a CPI and discuss general approaches to their measurement. In Section 3 the formal error structure is given in its mathematical form. In Section 4 several examples of numerical error estimation procedures are given. Although these examples are linked to the error model in Section 3, most of the quantitative estimates could also be applied within other model settings.

2. Types of CPI Errors and a General Approach to Their Measurement

When analysing errors in a statistical estimate it is necessary to start from the notion of a true value. It turns out, however, that this notion is difficult to apply to a CPI. The most reasonable starting point, as we see it, is the economic theory of the cost-of-living index (CLI). For a general review of this theory, see Pollak (1989); for a short summary relevant to the CPI error discussion, see Wynne and Sigalla (1994, p. 2). This theory is, however, in its pure form limited to one individual (or household). To come closer to the idea of a CPI, it is also necessary to aggregate the indices of individuals to the whole population in a country. Here we refer to the prevailing practice of plutocratic weighting, whereby the CLIs of individuals are weighted according to their total economic consumption.

As a first approximation to this theory the abstract CLI is typically replaced with some kind of fixed basket index. Usually, but not in Sweden, a Laspeyres or "pseudo-Laspeyres" index (in which base quantities represent an earlier time period than base prices) is used for this purpose. Better index formulae using current quantity weights are, e.g., the Fisher, Törnqvist, or Edgeworth formulae. The reason that these are not generally used in current index production is the difficulty of obtaining up-to-date weights in time. Note, though, that the KPI uses a formula similar to the Edgeworth index (Dalén 1992) for its long-term index.

The error resulting from using a Laspeyres' index formula, often referred to as (high level) substitution bias, is best estimated by comparison with a better formula that is computable at a later stage. See, e.g., Aizcorbe and Jackman (1993) for studies of the substitution bias of the U.S. CPI, a pseudo-Laspeyres index. A related type of error is the new goods bias resulting from the rapid price decline usually experienced for new goods, not yet represented in the CPI. For an economic analysis of this kind of error, see Hausman (1994).

In this article we will take a fixed weight index formula as our starting point. With this perspective the types of errors discussed below are likely to be the most important ones:

2.1. *High level weight errors*

Weights for aggregate product groups are usually estimated from one of two sources. Most countries use a household budget survey but some (such as Sweden) use the National Accounts, which in turn may use budget surveys combined with various other data, like retail sales data. Budget surveys are based on sampling and are thus affected by sampling as well as nonsampling errors – see, e.g., Biggeri and Giommi (1987) or Balk and Kersten (1986) for approaches where these errors are treated as variance components. National Accounts data are usually estimated in a less formal manner. One way, which is used here, of assessing their size is to compare early versions of these data, which are the ones actually used as CPI weights, with later and more reliable versions (revisions). This method regards the errors as biases.

2.2. *Nonrepresented consumption*

Certain products, often services, are not included in a CPI although they are part of the private consumption target. In the KPI these were mainly (1992) financial services, public child care and care of the elderly and certain international transport services. How should we regard this type of error? There are basically two alternatives: One is to regard it as a bias, since there is no random mechanism involved in the selection. For a given time period, it may be possible to use expert judgement to assess the possible error but generally speaking bias measures are rarely available.

Another, more promising alternative is to base the error estimate on the observed dispersion of the represented product group indices. This procedure is based on the assumption that the mean and the dispersion of price change of all product groups are the same as those of the included product groups. This assumption, although less likely for a short time period, is not too unreasonable in a long term perspective. One of the things that we are aiming at is a quantitative measure of how important it is to develop new measurement systems for the excluded groups and to answer that question we need a measure that will remain relevant for several years. In addition to its computational feasibility, this is the reason we use this method here. The method could be seen as a quasi-randomisation model, in which the observed product groups constitute a simple random sample from the population of all product groups. The approach is virtually identical to that of Edgeworth (1925, p. 311) who, in a footnote, speaks of the observed groups as a sample from the universe.

2.3. *Elementary aggregate bias*

An elementary aggregate refers to the most detailed level at which price data are used to form an index. Below a certain aggregation level there is often insufficient information for computing quantity or value weighted indices. Instead simplified formulae, only based on observed prices and perhaps some crude size measures of outlets, etc., are applied. This problem is the main theme of Dalén (1992). The errors resulting from using a crude formula are considered a bias; the size of the bias is assessed by comparing the result based on the crude formula with that of a better one, i.e., a form of sensitivity analysis.

2.4. *Low level substitution bias*

A consumer can take advantage of changing relative prices between outlets and product varieties in order to minimise his or her expenditure for the same standard of living ("get more value for money"). Treating the CPI as an approximation to a CLI tends to give any fixed basket index an upward bias, at least when the product and outlet variation is increasing. This type of error is, in practice, often confounded with elementary aggregate bias.

2.5. *Quality adjustment errors*

These errors result from the implicit or explicit adjustments done in cases of item substitution compared with some ideal procedure based on the concept of the CLI. They are considered biases and can be assessed by contrasting different approaches, again a kind of sensitivity analysis.

2.6. *Errors in the recorded price*

There are various reasons for errors and ambiguities in the actual recorded price. These reasons include direct mistakes, the use of list instead of transaction prices, the presence of discounts, coupons, bonus systems, and other kinds of price differentiation which result in the recorded price being more or less irrelevant to a large number of consumers. These errors are also best interpreted as biases and their sizes could be assessed in various ways based on known facts. An example of this kind of sensitivity analysis is discussed in Section 4.4.

2.7. *Selection bias, undercoverage and nonresponse*

If purposive sampling is used a selection bias results which is equal to the difference in price change between the representative product/outlet and the whole product/outlet group. If probability sampling is used we have traditional forms of non-sampling error such as nonresponse and undercoverage in the price surveys. However, nonresponse for outlets is usually very small, since price measurement is done by field persons employed or contracted by the statistical agency. Undercoverage of products/outlets in random sampling is a close relative to selection bias due to purposive sampling and should normally be a much smaller problem as long as the rate of undercoverage is not extremely large. In this article, no error measures of this category are considered.

2.8. *Conceptual ambiguities in certain product groups*

The most familiar example is owner-occupied housing, where a large number of basic approaches exist in different countries. Its large weight makes it a great problem when comparing price change between different countries. Examples of other areas with conceptual difficulties are: insurance and other financial services, gambling and lotteries and fees for certain heavily subsidised (in some countries) public services like health, educational or day care services.

Conceptual problems are not really errors but rather part of the process of establishing the target for statistical measurement. But when comparing inflation figures for different countries their role is much like the kind of errors discussed above. Also, starting from the CLI, there should be only one right answer and also from this point of view, one might speak of errors. The size of this "error" could, most reasonably, be assessed by comparing the outcomes from different approaches with each other. We will often refer to measures like this as bias risks rather than bias estimates, since there is no consensus on the right answer.

3. A Formal Error Model

In this section we present an attempt to formulate an error model which puts the various types of errors in their place and, in principle, adds them up to a measure of total error. An essential ingredient of the model is a distinction between errors within all the separate surveys and errors 'between surveys," i.e., those that influence only the all products estimate.

3.1. The components and steps of the model

Using the above line of reasoning, we start from the concept of a fixed weight index. This is defined in terms of true weights, w^*_{hj}, and true subindices, I_{hj}, for product groups

$$I^* = \sum_{h \in U} \sum_{j \in U_h} w^*_{hj} I_{hj} \quad \text{with} \quad \sum_{h \in U} \sum_{j \in U_h} w^*_{hj} \equiv 1 \tag{1}$$

where U is the whole population of product groups which is divided into superstrata (more about this concept below) U_h of product groups $j \in U_h$.

Next, we replace the w^*_{hj} with estimated weights w_{hj}, with their sum also standardised to unity and obtain

$$I = \sum_{h \in U} \sum_{j \in U_h} w_{hj} I_{hj} \quad \text{with} \quad \sum_{h \in U} \sum_{j \in U_h} w_{hj} \equiv 1. \tag{2}$$

The difference between (2) and (1) is the (high level) *weight error*. Thus we define our first error component, a bias, as

$$B_1 = I - I^*. \tag{3}$$

The process of estimating I could be viewed as consisting of two steps. Step 1 is the selection of product groups reflecting the fact that some groups are excluded. For reasons given above we introduce a quasi-randomisation model in which the covered groups are considered a stratified simple random sample of product groups from each superstratum h. In practice we will have a certainty superstratum $h = C$, where all product groups are covered and a set of sampled superstrata $h \in H$, so that $C \cup H = U$. The "sample" of covered product groups in superstratum h is denoted S_h and $S = \bigcup_{h \in H} S_h$.

In step 2 we estimate the product group indices I_{hj} with $\hat{I}_{hj}$. In this step we take samples of price observations within the product groups U_h according to various sampling designs. We end up with the estimator

$$\hat{I} = \sum_{h \in U} \sum_{j \in S_h} \tilde{w}_{hj} \hat{I}_{hj}, \quad \text{where} \tag{4}$$

$$\tilde{w}_{hj} = \begin{cases} w_{hj} & \text{if} \quad h = C \\ w_{hj} \dfrac{w_h}{w_h^S} & \text{if } h \in H, \end{cases} \tag{5}$$

$$w_h = \sum_{j \in U_h} w_{hj} \quad \text{and} \quad w_h^S = \sum_{j \in S_h} w_{hj}.$$

Expectation of $\hat{I}$ is taken with respect to both the quasi-sampling of product groups and to the sampling of price observations. Since $\hat{I}$ has the form of a sum of stochastic ratios (with w_h^S in the denominator) there is no exact expression for $E(\hat{I})$. We denote by $E_R(\hat{I})$ its usual ratio approximation.

In step 2 we allow for a nonsampling bias so that

$$E(\hat{I}_{hj}) = I_{hj} + b_{hj} = I^B_{hj} \tag{6}$$

where b_{hj} is the nonrandom bias induced by the measurement procedure for product group j in stratum h due to, e.g., item and outlet substitutions, imperfect quality adjustments or errors in the price capturing process.

The aggregate bias in step 2 is called B_2 and we have

$$B_2 = \sum_{h \in U} \sum_{j \in U_h} w_{hj} b_{hj} = E_R(\hat{I}) - I \approx E(\hat{I}) - I. \tag{7}$$

Our interest now focuses upon the total error of the estimate in (4). As our measure of total error we choose the mean squared error, i.e.,

$$E(\hat{I} - I^*)^2 = E_1 E_2 (\hat{I} - I^*)^2 \tag{8}$$

where E_1 and E_2 stand for expectation in step 1 and 2, respectively. We may decompose (8) into

$$E(\hat{I} - I^*)^2 = V(\hat{I}) + B^2(\hat{I}) \tag{9}$$

with V denoting total variance and B total bias. Further decomposition, conditioning on the sample S of quasi-sampled product groups, gives

$$\begin{aligned} V(\hat{I}) &= V_1 + V_2, \quad \text{where} \\ V_1 &= V_1\{E_2(\hat{I}|S)\} \quad \text{and} \\ V_2 &= E_1\{V_2(\hat{I}|S)\} \end{aligned} \tag{10}$$

and

$$B(\hat{I}) = B_1 + B_2 (+\text{ratio estimator bias}). \tag{11}$$

After calculations given in detail in Dalén (1993, appendix 1), we obtain the following expressions for our four error components

$$\begin{aligned} V_1 &\approx \sum_{h \in H} \frac{N_h(N_h - n_h)}{n_h} \sigma_h^2 \quad \text{with} \quad \sigma_h^2 = \frac{1}{N_h - 1} \sum_{j \in U_h} w_{hj}^2 (I_{hj}^B - I_{h.}^B)^2 \\ &\text{and } I_{h.}^B = \frac{\sum_{j \in U_h} w_{hj} I_{hj}^B}{w_h} \end{aligned} \tag{12}$$

$$V_2 \approx \sum_{h \in H} \frac{N_h}{n_h} \sum_{j \in U_h} w_{hj}^2 V_{hj} \tag{13}$$

$$B_1 = \sum_{h \in U} \sum_{j \in U_h} (w_{hj} - w_{hj}^*) I_{hj} \tag{14}$$

$$B_2 = \sum_{h \in U} \sum_{j \in U_h} w_{hj} b_{hj} \tag{15}$$

where
N_h is the total number of product groups in superstratum h,
n_h is the number of covered (quasi-sampled) groups in superstratum h and
$V_{hj} = V(\hat{I}_{hj})$ is the sampling variance in step 2 for product group $j \in h$.

3.2. *Estimators of error components*

Next we seek the best possible estimates of the four error components defined in (12)–(15) above. In the practical applications we will use the following four estimates.

$$\hat{V}_1 = \sum_{h \in H} \frac{(w_h - w_h^S)}{w_h} \sum_{j \in S_h} \tilde{w}_{hj}^2 (\hat{I}_{hj} - \hat{\hat{I}}_{h.})^2 \quad \text{with} \quad \hat{\hat{I}}_{h.} = \frac{\sum_{j \in S_h} w_{hj} \hat{I}_{hj}}{\sum_{j \in S_h} w_{hj}} \tag{16}$$

$$\hat{V}_2 = \sum_{h \in U} \sum_{j \in S_h} \tilde{w}_{hj}^2 \hat{V}_{hj}, \text{ where } \hat{V}_{hj} \text{ is an unbiased estimate of } V_{hj} \tag{17}$$

$$\begin{aligned} \hat{B}_1 &= \sum_{h \in U} \sum_{j \in S_h} (\tilde{w}_{hj} - \tilde{w}_{hj}^*) \hat{I}_{hj} \text{ with } \tilde{w}_{hj}^* = \begin{cases} w_{hj}^* & \text{if } \quad h = C \\ w_h^* w_{hj}^* / w_h^{S*} & \text{if } \quad h \in H, \end{cases} \\ & w_h^* = \sum_{j \in U_h} w_{hj}^* \quad \text{and} \quad w_h^{S*} = \sum_{j \in S_h} w_{hj}^* \end{aligned} \tag{18}$$

$$\hat{B}_2 = \sum_{h \in U} \sum_{j \in S_h} \tilde{w}_{hj} b_{hj}. \tag{19}$$

In Dalén (1993, appendix 2) the properties of these estimated error components are derived. It turns out that $\hat{V}_2$ and $\hat{B}_2$ are approximately unbiased (first order Taylor linearisation). However, $\hat{B}_1$ and $\hat{V}_1$ are biased and we have the following relations

$$\begin{aligned} & E(\hat{B}_1) = B_1 + R_B, \text{ where the discrepancy term} \\ & \quad R_B = \sum_{h \in U} \sum_{j \in h} (w_{hj} - w_{hj}^*) b_{hj} \text{ and} \end{aligned} \tag{20}$$

$$E(\hat{V}_1) = V_1' + R_V \tag{21}$$

where V_1' differs from V_1 only by stratum-wise factors $(N_h - 1)/N_h$ and R_V, a discrepancy term, depends on the step 2 sampling variances in the quasi-sampled super-strata. See Dalén (1993, appendix 2) for an exact expression for R_V.

We see that $E(\hat{B}_1)$ also depends on the survey biases b_{hj}. In principle, this dependency would make it possible to adjust $\hat{B}_1$ to obtain an unbiased estimate. This is not practicable, however, for reasons discussed below. For $\hat{V}_1$ to be useful it is necessary that R_V is either small or estimable. We will demonstrate below that both these requirements are fulfilled.

For our purpose of obtaining crude error estimates we will, in Section 4, demonstrate empirically that $\hat{B}_1$ and $\hat{V}_1$ are useful, although it is of course necessary to keep the biases in (20) and (21) in mind.

4. The Practical Application of the Error Model

4.1. *Bias due to preliminary weight estimates*

Bias estimates according to (18) were made by assuming the final National Accounts

(NA) estimates for private consumption for the years 1986, 1987, and 1988 to be the "true" weights. These final estimates are made about three years after the reference year with the use of final estimates from a multitude of primary statistical sources. In contrast, the preliminary NA figures underlying the KPI weights use preliminary estimates and projections of various kinds. Of course, the final NA estimates are not true in any real sense but they are definitely superior to the preliminary estimates and this method should therefore provide some sense of the uncertainty of the estimation method used.

The bias estimates obtained in this way were: -0.23 for 1986, -0.14 for 1987, and -0.12 for 1988 with an average of -0.16.

Here the discrepancy term R_B in (20) is potentially disturbing. It might be that the estimates of weight bias are in fact contaminated by step 2 biases. However, for R_B to have any long run influence on the bias estimates there must be a strong correlation between the step 2 biases and the weight differences. With the crude use that we make of these estimates, we believe that calculations according to (18) do not generally lead to erroneous conclusions. It is, however, necessary to look carefully at the weight differences for those products for which the step 2 bias risks are known to be significant.

4.2. *Errors due to the exclusion of product groups*

In our system, errors due to excluded product groups are evaluated according to (16). Given this equation, the next problem is the choice of a superstratum structure to represent the excluded groups which, in 1992, covered about 3.6% of private consumption. This ought to be done so that the design is "noninformative" in the intuitive sense that the actual outcome of the quasi-randomisation procedure is as likely as any other outcome. Since the nonselected product groups are small (each has a weight of less than 1%) and we have assumed simple random sampling as our quasi-sampling procedure, the selected product groups of the same superstratum should also be small.

We tried two solutions in this regard. *Method A* divides the whole CPI into two superstrata, one self-representing category of very large product groups (rents, interest for home-owners, petrol, cars and a few more) and one superstratum which covers all excluded product groups and all small covered product groups. Within method A we used three different cutoff points in terms of the weight of the product group, 2%, 1%, and 0.5%. Above these cutoff points the groups are self-representing. One drawback of method A is that some of the representing product groups will exhibit dependencies due to a common outlet sample. *Method B* selects a particular set of representing product groups for each excluded group. These sets generally consist of independently sampled product groups. Table 1 demonstrates how this was done.

The result of the computations for years 1980–91 according to (16) is given in Table 2. We see, that although results for a single year vary greatly between different methods, the long-run means are not much different. As a rough figure we could take $V_1 = 0.01$ as a long-term order-of-size estimate. Of course, the differences between years reflect genuine changes in that product group variations are larger in certain years. For example, in 1991 the high figures reflect differential changes in value-added

Table 1. The representation of excluded product groups in Method B

Represented group (KPI weight)	Foreign air travel (0.49%)	Child care and care of the elderly (0.99%)	Renters' repairs (0.08%)	Furniture repair (0.09%)	Other services (1.97%)
Representing groups	1 Domestic air travel 2 Packaged tours	Other product groups in the public sector: 1 Local transports 2 Pharmaceuticals 3 Medical care 4 Dental services	1 Homeowners' repairs, services 2 Homeowners' repairs, goods	1 Repairs of washing machines 2 Domestic services 3 Automobile repairs 4 Removal services 5 TV repair 6 Photo developing 7 Barbers' services	1 Vehicle inspection 2 Driver's education 3 Garage rent 4 Train fares 5 Domestic boat fares 6 Telephone services 7 Postage

Table 2. Estimates of variance contribution for excluded product groups

Year	Method A			Method B
	Cutoff point 0.5%	1%	2%	
1980	0.0067	0.0116	0.0159	0.0232
1981	0.0079	0.0119	0.0135	0.0062
1982	0.0033	0.0054	0.0090	0.0102
1983	0.0038	0.0060	0.0075	0.0016
1984	0.0038	0.0036	0.0051	0.0064
1985	0.0029	0.0038	0.0044	0.0045
1986	0.0077	0.0207	0.0184	0.0009
1987	0.0115	0.0088	0.0207	0.0022
1988	0.0050	0.0044	0.0062	0.0021
1989	0.0046	0.0055	0.0058	0.0112
1990	0.0081	0.0118	0.0170	0.0106
1991	0.0325	0.0289	0.0278	0.0248
Mean	0.0082	0.0102	0.0126	0.0087

taxes and other taxes and subsidies, which of course increases errors from omitting certain product groups from the index.

What about the discrepancy term R_V in (21)? First we note that R_V will in practice always be positive which means that (16) is likely to be an overestimate of V_1. Also, R_V is easiest to evaluate for method B, for which crude, but fairly reliable calculations indicate that it is not more than 10% of V_1. Less reliable, but not unreasonable, calculations indicate something similar for method A.

4.3. *Sampling errors in price surveys*

Sampling errors are discussed in Dalén and Ohlsson (1995). Here we will only give a summary of the results from that article. The KPI consists of many price surveys done with different methods and sampling designs including purposive sampling. It is possible to make a decomposition of the KPI such that these surveys are (at least approximately) independent. Making variance estimates for each one of these, we could in the end use (17) to arrive at an estimate of total sampling error. In practice such variance estimates have been done indicating a total sampling error for a one-year change of about

$$V_2 \approx 0,04$$

which is equivalent to a 95% confidence interval length of about 0.4. Actual computations have been done for all the larger surveys.

4.4. *Other errors in price surveys*

In principle we would like to use a formal bias estimation procedure and aggregate biases as in (19) above. In practice, this is not often possible but the above formalisation is useful as a conceptual basis for an intelligent error analysis. Let us give some examples:

Example 1. Elementary aggregate bias. In Dalén (1992, table 1) the effect of using different index formulae for elementary aggregates is demonstrated. It is argued that some formulae, such as the geometric mean or the ratio of normed mean prices (currently used at Statistics Sweden), are better than others. If one of these is taken as the better formula, the difference between the outcome using a certain formula and that using the better one could be interpreted as a formula bias. It was shown that this difference for one commonly used formula, the Sauerbeck formula (called R), was as large as 0.65 for a nine-month comparison. Moulton (1993) in a similar way compares the geometric mean with the Laspeyres type estimator (with similar properties as the Sauerbeck index) currently used in the U.S. CPI and obtains differences of the same order.

Example 2. Owner-occupied housing. In the KPI the index for owner-occupied single-family housing is calculated according to a user cost approach (user cost (1) according to Turvey et al. 1989, p. 17) covering mortgage interest, depreciation, property tax, energy, maintenance cost, etc., while the index for owner-occupied multi-family dwellings simply follows the index for (multi-family) rented dwellings. Other countries have made quite different methodological choices in this area. A natural thing is therefore to do a simple sensitivity analysis according to (19) applying an alternative method to the one used. In the following table we show the sensitivity of the KPI for all products of the following alternatives.

ALT1: Letting the index for multi-family owner-occupied housing follow that for single-family owner-occupied housing instead of that of rented dwellings.

ALT2: Letting all indices for owner-occupied housing follow that for rented dwellings (a crude application of a rental equivalance principle, see, e.g., Early (1990) or user cost (3) in Turvey et al. (1989)).

The results for the years 1981–1991 are given in Table 3. We present the differences between the two alternative calculations and the actual KPI figure. These figures should be interpreted as bias risks rather than bias estimates, since there is no consensus on which of these alternatives is most correct. What they show is the sensitivity of

Table 3. Bias risks due to choice of method for owner-occupied housing

Year	KPI	ALT1-KPI	ALT2-KPI
1981	109,4	−0,10	0,46
1982	109,9	−0,22	1,36
1983	109,3	−0,20	1,24
1984	108,1	−0,04	0,25
1985	105,7	0,12	−0,63
1986	103,2	−0,11	0,83
1987	104,9	−0,14	0,86
1988	106,2	0,01	−0,06
1989	106,7	0,01	−0,08
1990	110,7	0,03	−0,16
1991	108,0	−0,49	2,29
1992	101,9	−0,11	0,27
Mean		−0,10	0,55

the index estimate to different, but not unreasonable, index definitions. The large bias risks for certain years are noteworthy; for 1991 this was due to changed Government rules of taxing and subsidising housing.

Example 3. Quality changes for clothing items. In the late 1980s the KPI methods of measuring price changes for clothing items were reviewed, since downward bias was suspected to exist. In the method used then, the interviewers made a subjective evaluation of the quality difference in monetary terms for (the very frequent) substitutions due to discontinuation of a particular variety. It was found that these evaluations to a large extent implied quality improvements and in particular for women's clothing, so that the products with the highest frequency of substitutions also had the smallest price increases. The difference between the index calculated with and without the quality evaluation (QE) of the interviewer was then taken as a measure of the bias risk of this method. Here, without QE means that the price of the discontinued variety is compared directly with that of the new variety. For the years 1984–86, results are given in Table 4.

Example 4. Discounts for petrol. Discounts only affect price indices insofar as their average level changes between years. If we let $\tilde{P}_t$ denote the price level before discounts and P_t the price level after discounts at time t, and d_t the average effective discount rate at time t then we have

$$I_{0t} = \frac{(1-d_t)}{(1-d_0)} \frac{\tilde{P}_t}{\tilde{P}_0} \approx [1-(d_t-d_0)]\tilde{I}_{0t} \tag{22}$$

where $\tilde{I}_{0t}$ is the index before and I_{0t} the index after discounts.

The difference between these indices would now obviously be a bias measure given that we consider inclusion of discounts to be the theoretically preferred procedure.

Applying this reasoning to petrol sales, we note that most petrol dealers apply some kind of discounts for customers with special cards. These discounts are currently not taken into consideration in the KPI. The bias incurred by not including them would be (at the time of the introduction of the discount system) the KPI weight of petrol (about 0.04) times the average discount in percent. Calculations suggest these discounts result in price reductions of 1.5% on average over all petrol customers which would entail an upward bias of 0.06% at the moment of introducing the discounts. In most years the bias would of course be much smaller, since its size would depend only on *changes* in discount rates and on the consumers' degree of taking advantage of them.

Table 4. Prices indices and bias risks for clothing 1984–86

Year	With QE	Without QE	Difference	Weight	Bias risk for KPI
1984	105.32	112.78	−7.46	0.066	−0.49
1985	105.08	109.42	−4.34	0.067	−0.29
1986	101.80	105.00	−3.20	0.072	−0.23
Mean	104.07	109.07	−5.00	0.068	−0.33

4.5. Summary and a holistic perspective

It would be nice to be able to sum up the errors according to (9)–(11) to a total error measure. Based on the empirical presentation above, this is not advisable, however. It might be possible to suggest $V \approx 0.05$ as a rough measure of total variance for a one-year link, corresponding to a 95% confidence interval of ± 0.4–0.5, since the major part of the variances reflects outlet and product sampling and these variance aggregates seem to be fairly stable over time. But the bias risks are extremely variable from year to year and could even change sign. It is also clear that these risks are often larger than the random error. This is particularly true when the effect of the choice of method for owner-occupied housing is taken into account.

This analysis has been concerned with errors in a one-year link. Errors for shorter time periods are, in addition, influenced by seasonal variations in consumption and prices and the way these variations are handled in the KPI. We will not attempt such an analysis here.

Errors for longer periods of, say, several years are of considerable interest. Sampling errors are affected by the correlation structure between years. This structure is very complicated and only a crude assessment is possible. In the KPI, both the product sample and the outlet sample have large degrees of overlap between years. But for large covariances to occur more factors are also necessary. One such factor would be that outlets which increase prices more than average a certain year also increase their prices more another year. Another such factor would be that a sampled/representative product which shows a price change different from the average of its product group a certain year also tends to have the same kind of price movement the next year. In general, neither of these tendencies is likely to be at hand.

Since we do not expect any significant positive correlations between years in the sampling system, sampling errors will tend to be less important for long-term comparisons. In this perspective, interest will therefore focus upon the existence of systematic errors with a tendency to have the same sign over a long period. Possible such errors are

elementary aggregate bias,
substitution biases due to the procedures for linking in new products and outlets,
errors because of incomplete or imperfect quality adjustment,
conceptual difficulties such as those for owner-occupied housing, or
failure to capture the price actually paid, due to the use of list prices, discounts, coupons, etc.

5. References

Aizcorbe, A.M. and Jackman, P.C. (1993). The Commodity Substitution Effect in CPI Data, 1982–91. Monthly Labor Review, December 1993, 25–33.

Andersson, C., Forsman, G., and Wretman, J. (1987). On the Measurement of Errors in the Swedish Consumer Price Index. Bulletin of the International Statistical Institute, 52, Book 3, 155–171.

Balk, B. and Kersten, H.M.P. (1986). On the Precision of Consumer Price Indices Caused by the Sampling Variability of Budget Surveys. Journal of Economic and Social Measurement, 14, 19–35.

Biggeri, L. and Giommi, A. (1987). On the Accuracy and Precision of the Consumer Price Indices, Methods and Applications to Evaluate the Influence of the Sampling of Households. Bulletin of the International Statistical Institute, 52, Book 3, 137–154.

Dalén, J. (1992). Computing Elementary Aggregates in the Swedish Consumer Price Index. Journal of Official Statistics, 8, 129–147.

Dalén, J. (1993). Quantifying Errors in the Swedish Consumer Price Index. Statistics Sweden, R&D Report 1993:8.

Dalén, J. and Ohlsson, E. (1995). Variance Estimation in the Swedish Consumer Price Index. Journal of Business & Economic Statistics, 13, 347–356.

Early, J. (1990). Improving Consumer Price Indexes. Journal of Official Statistics, 6, 179–204.

Edgeworth, F.Y. (1925). Papers Relating to Political Economy, 1, 298–341. London: The Royal Economic Society.

Hausman, J. (1994). Valuation of New Goods under Perfect and Imperfect Competition. NBER working paper No. 4970, National Bureau of Economic Research, Cambridge, Massachusetts.

Malmquist, S. (1958). Precisionsproblemet vid beräkning av konsumentprisindex I och II. Statistisk tidskrift, 7, 492–504 and 612–628. [In Swedish]

Moulton, B.R. (1993). Basic Components of the CPI: Estimation of Price Changes. Monthly Labor Review, December 1993, 25–33.

Pollak, R.A. (1989). The Theory of the Cost of Living Index. New York: Oxford University Press.

Ruist, E. (1953). Några synpunkter på socialstyrelsens prismaterial. In SOU 1953:23 Konsumentprisindex, Stockholm. [In Swedish]

Särndal, C.E., Swensson, B., and Wretman, J. (1992). Model Assisted Survey Sampling. New York: Springer Verlag.

Turvey, R. et al. (1989). Consumer Price Indices, An ILO manual. International Labour Office, Geneva.

Wynne, M.A. and Sigalla, F.D. (1994). The Consumer Price Index. Economic Review, Federal Reserve Bank of Dallas, Second Quarter 1994, 1–22.

Received October 1993
Revised August 1995