Journal of Official Statistics, Vol. 21, No. 4, 2005, pp. 675–699

Price Indexes for Elementary Aggregates: The Sampling Approach

Bert M. Balk[1]

At the lowest level of aggregation of a CPI or PPI quantity information is usually unavailable and matched samples of prices must be used for the index computation. Familiar indexes at this level of aggregation are those of Dutot, Carli, and Jevons. An important, yet often overlooked characteristic of these and similar indexes is that they are sample statistics, whose properties can be studied from the sampling point of view. This article provides a systematic study of this topic and concludes with a number of recommendations for statistical practice.

Key words: CPI; PPI; elementary aggregate; price index.

1. Introduction

Mainstream (bilateral) index number theory applies to aggregates consisting of a finite set of commodities. Two basic assumptions are that the set of commodities does not change between the two periods compared, and that all the price and quantity data, which are necessary for the computation of an index are available to the statistician. In this article I am concerned with what to do when the second of these assumptions is not, or cannot be fulfilled. There are, of course, various kinds of unavailability of data. The situation I will consider in particular in the article is that nothing but price data are available for a sample of commodities and/or respondents.

Since such a situation materializes at the very first stage of the computation of any official price index, such as a Consumer Price Index (CPI) or a Producer Price Index (PPI), we are dealing here with an issue of great practical significance.

The usual approach to the problem of unavailable quantity data is to consider price indexes which are functions of prices only. The main formulas discussed in the literature and used in practice are:

- the ratio of arithmetic average prices (the formula of Dutot),
- the arithmetic average of price relatives (the formula of Carli),
- the geometric average of price relatives = the ratio of geometric average prices (the formula of Jevons).

[1] Statistics Netherlands, P. O. Box 4000, 2270 JM Voorburg, The Netherlands, and Erasmus Research Institute of Management, Erasmus University Rotterdam, The Netherlands. Email: bblk@cbs.nl and bbalk@rsm.nl
Acknowledgments: The views expressed in this article are those of the author and do not necessarily reflect any policy of Statistics Netherlands. An earlier version has been published as Research Paper No. 0231 of the Methods and Informatics Department of Statistics Netherlands, as well as in the Proceedings of the Seventh Meeting of the International Working Group on Price Indices (Ottawa Group), May 2003. Over the years, the author acknowledges helpful comments from Jörgen Dalén, Erwin Diewert, Jan de Haan, Peter Hill, Paul Knottnerus, Mick Silver, and some unknown referees.

The suitability of these formulas has been studied by various methods. Following the early contribution of Eichhorn and Voeller (1976), Dalén (1992) and Diewert (1995) studied their properties from an axiomatic point of view. Additional insights were obtained by deriving (approximate) numerical relations between these formulas, and by combining these relations with more or less intuitive economic reasoning. Balk's (1994) approach was to see which assumptions would validate these formulas as estimators of true but unknown population price indexes, which by definition are functions of prices and quantities. An overview of the state of affairs can be found in Chapter 20 of the recently completed *CPI Manual* and *PPI Manual* (2004).

This article develops the sampling approach. In Section 2 it is argued that, although not known to the statistician, all the detailed price and quantity data of the commodities and respondents pertaining to the aggregate under consideration exist in the real world. Section 3 then argues that the first task faced by the statistician is to decide on the nature of the aggregate (homogeneous or heterogeneous) and on the target price index (the unit value index or some superlative or nonsuperlative price index). Next the sampling design comes into the picture. With the aid of these two pieces of information, one can judge the various estimators with respect to their performance. This is the topic of Section 4, which is on homogeneous aggregates, and Sections 5–7, which are on heterogeneous aggregates and superlative target price indexes. Section 8 adds to this topic with some micro-economic considerations on the choice of a sample price index. Section 9 discusses the not unimportant case where, for operational reasons, a nonsuperlative price index was chosen as target. Section 10 surveys the behaviour of the various sample price indexes with respect to the Time Reversal Test, and reviews the (approximate) numerical relations between them. Section 11 summarizes the key results and concludes with some practical advice.

2. Setting the Stage

The aggregates covered by a CPI or a PPI are usually arranged in the form of a tree-like hierarchy (according to some international classification such as COICOP or NACE). Any *aggregate* is a set of economic transactions pertaining to a set of commodities. Commodities can be goods or services. Every economic transaction relates to the change of ownership (in the case of a good) or the delivery (in the case of a service) of a specific, well-defined commodity at a particular place and date, and comes with a quantity and a price. The price index for an aggregate is calculated as a weighted average of the price indexes for the subaggregates, the (expenditure or sales) weights and type of average being determined by the index formula. Descent in such a hierarchy is possible as far as available information permits the weights to be decomposed. The lowest-level aggregates are called *elementary* aggregates. They are basically of two types:

- those for which all detailed price and quantity information is available;
- those for which the statistician, considering the operational cost and/or the response burden of getting detailed price and quantity information about all the transactions, decides to make use of a representative sample of commodities and/or respondents.

Any actual CPI or PPI, considered as a function that transforms sample survey data into an index number, can be considered as a stochastic variable, whose expectation ideally equals its population counterpart. The elementary aggregates then serve as strata for the sampling procedure. We are of course particularly interested in strata of the second type.

The practical relevance of studying this topic is large. Since the elementary aggregates form the building blocks of a CPI or a PPI, the choice of an inappropriate formula at this level can have a tremendous effect higher up in the aggregation tree.

The detailed price and quantity data, although not available to the statistician, nevertheless – at least in principle – exist in the outside world. It is thereby frequently the case that at the respondent level (outlet or firm) some aggregation of the basic transaction information has already been executed, usually in a form that suits the respondent's financial and/or logistic information system. This could be called the basic information level. This, however, is in no way a naturally given level. One could always ask the respondent to provide more disaggregated information. For instance, instead of monthly data one could ask for weekly data; whenever appropriate, one could ask for regional instead of global data, or one could ask for data according to a finer commodity classification. The only natural barrier to further disaggregation is the individual transaction level (see Balk 1994 for a similar approach).

Thus, conceptually, for all well-defined commodities belonging to a certain elementary aggregate and all relevant respondents there exists information on both the quantity sold and the associated average price (unit value) over a certain time period. Let us try to formalize this somewhat. The basic information – which in principle exists in the outside world – is of the form $\left\{(p_n^t, q_n^t); n = 1, \ldots, N\right\}$ where t denotes a time period; the elements of the population of (non-void) pairs of well-defined commodities and respondents, henceforth called elements, are labelled from 1 to N; p_n^t denotes the price, and q_n^t denotes the quantity of element n at time period t. It will be clear that N may be a very large number, since even at very low levels of aggregation there can be thousands of elements involved. We repeat that it will be assumed that the population does not change between the time periods considered. Of course, in reality the population changes more or less continuously. It is important, however, to study the properties of the price index estimators in a controlled environment.

It is assumed that we must compare a later Period 1 to an earlier Period 0. The later period will be called comparison period and the earlier period base period. The conceptual problem is to split the value change multiplicatively into a price index and a quantity index,

$$\sum_{n=1}^{N} p_n^1 q_n^1 / \sum_{n=1}^{N} p_n^0 q_n^0 = P(p^1, q^1, p^0, q^0) Q(p^1, q^1, p^0, q^0) \qquad (1)$$

where $p^t \equiv (p_1^t, \ldots, p_N^t)$ and $q^t \equiv (q_1^t, \ldots, q_N^t)$ $(t = 0,1)$. This is traditionally called the index number problem. Both indexes depend on the prices and quantities of the two periods.

3. Homogeneity or Heterogeneity

There is now an important conceptual choice to make. In the statistician's parlance this is known as the "homogeneity or heterogeneity" issue. Although in the literature a lot of words have been devoted to this issue, at the end of the day the whole problem can be

reduced to the rather simple looking operational question:
(2) Does it make (economic) sense to add up the quantities q_n^t of the elements $n = 1, \ldots, N$?

If the answer to this question is "yes," then the elementary aggregate is called "homogeneous" and the appropriate, also called target, price index is the unit value index

$$P_U \equiv \frac{\sum_{n=1}^N p_n^1 q_n^1 / \sum_{n=1}^N q_n^1}{\sum_{n=1}^N p_n^0 q_n^0 / \sum_{n=1}^N q_n^0} \tag{3}$$

that is, the average comparison period price divided by the average base period price. Balk (1998) shows that the unit value index satisfies the conventional axioms for a price index, except the commensurability axiom and the proportionality axiom. However, when the elements are commensurate, the commensurability axiom reduces to $P(\lambda p^1, \lambda^{-1} q^1, \lambda p^0, \lambda^{-1} q^0) = P(p^1, q^1, p^0, q^0)(\lambda > 0)$, which clearly is satisfied. The corresponding quantity index is the simple sum or Dutot index

$$Q_D \equiv \sum_{n=1}^N q_n^1 \Big/ \sum_{n=1}^N q_n^0 \tag{4}$$

When the quantities are additive, we are obviously dealing with a situation where the same commodity during a time period is sold or bought at different places and/or at different subperiods at different prices. Put otherwise, we are dealing with pure price differences. These can be caused by market imperfections, such as price discrimination, consumer ignorance, or rationing. Economic theory seems to preclude this possibility since it states that in equilibrium "the law of one price" must hold. In reality, however, market imperfections are the rule rather than the exception. But also physical restrictions can play a role. Although, for instance, the "representative" consumer is assumed to be fully informed about all the prices and to have immediate and costless access to all the outlets throughout the country, the sheer physical distance between the outlets precludes "real" consumers from exploiting this magical possibility. Thus price differences exist where they, according to a representative-agent-based theory, are not supposed to exist.

If the answer to Question (2) is "no," which in practice will mostly be the case, then the elementary aggregate is called "heterogeneous" and there are various options available for the target price index. First of all, the axiomatic as well as the economic approach to index number theory leads to the conclusion that the target price index should be some superlative index. According to the theoretical surveys in the recent *CPI Manual* and *PPI Manual* (2004), three price indexes appear to be particularly relevant. The first is the Törnqvist price index

$$P_T \equiv \prod_{n=1}^N (p_n^1 / p_n^0)^{(s_n^0 + s_n^1)/2} \tag{5}$$

where $s_n^t \equiv p_n^t q_n^t / \sum_{n=1}^N p_n^t q_n^t$ $(t = 0,1)$ is element n's value share in Period t. This price index is a weighted geometric average of the price relatives, the weights being average value shares. The corresponding quantity index is defined as

$$\tilde{Q}_T \equiv \left(\sum_{n=1}^{N} p_n^1 q_n^1 \Big/ \sum_{n=1}^{N} p_n^0 q_n^0 \right) \Big/ P_T \tag{6}$$

The second superlative price index is the Fisher index,

$$P_F \equiv \left(\frac{\sum_{n=1}^{N} p_n^1 q_n^0}{\sum_{n=1}^{N} p_n^0 q_n^0} \right)^{1/2} \left(\frac{\sum_{n=1}^{N} p_n^1 q_n^1}{\sum_{n=1}^{N} p_n^0 q_n^1} \right)^{1/2} = (P_L P_P)^{1/2} \tag{7}$$

which is the geometric average of the Laspeyres and the Paasche price indexes. In this case the quantity index is given by

$$Q_F \equiv \left(\frac{\sum_{n=1}^{N} p_n^0 q_n^1}{\sum_{n=1}^{N} p_n^0 q_n^0} \right)^{1/2} \left(\frac{\sum_{n=1}^{N} p_n^1 q_n^1}{\sum_{n=1}^{N} p_n^1 q_n^0} \right)^{1/2} = (Q_L Q_P)^{1/2} \tag{8}$$

which is the geometric average of the Laspeyres and the Paasche quantity indexes. The third superlative price index is the Walsh index, defined as

$$P_W \equiv \frac{\sum_{n=1}^{N} p_n^1 \left(q_n^0 q_n^1 \right)^{1/2}}{\sum_{n=1}^{N} p_n^0 \left(q_n^0 q_n^1 \right)^{1/2}} \tag{9}$$

in which case the quantity index is defined by

$$\tilde{Q}_W \equiv \left(\sum_{n=1}^{N} p_n^1 q_n^1 \Big/ \sum_{n=1}^{N} p_n^0 q_n^0 \right) \Big/ P_W \tag{10}$$

The Walsh price index is a member of the class of so-called basket price indexes, that is, price indexes that compare the cost of a certain basket of quantities in the comparison period to the cost in the base period. The Laspeyres and Paasche price indexes are typical examples: the first employs the base period basket and the second the comparison period basket. The basket of the Walsh price index is an artificial one, consisting of the geometric averages of the quantities of the two periods.

Many statistical offices, however, are forced for operational reasons to define implicitly or explicitly a nonsuperlative price index as target. In general their target appears to have the form of a Lowe price index

$$P_{Lo} \equiv \frac{\sum_{n=1}^{N} p_n^1 q_n^b}{\sum_{n=1}^{N} p_n^0 q_n^b} \tag{11}$$

where b denotes some period prior to the base period 0. The corresponding quantity index is then defined by

$$\tilde{Q}_{Lo} \equiv \left(\sum_{n=1}^{N} p_n^1 q_n^1 \Big/ \sum_{n=1}^{N} p_n^0 q_n^0 \right) \Big/ P_{Lo} \tag{12}$$

Notice that the five price indexes considered above all satisfy the Time Reversal test, that is, using the notation introduced above, $P(p^1, q^1, p^0, q^0) = 1/P(p^0, q^0, p^1, q^1)$.

It could be that the statistician is unable to decide between a simple "yes" and "no" in reply to Question (2); that is, he or she finds that for certain subsets of the elementary aggregate $\{1, \ldots, N\}$ it makes sense to add up the quantities whereas for the remainder it does not. Then the aggregate should be split into subsets to which either the "yes" or the "no" answer applies. If this splitting appears not to be feasible then the "no" answer should take precedence over the "yes" answer. Thus, conceptually, we have to deal with but two cases. The decision, however, is not always simple. See Silver and Webb (2002) for considerations and empirical evidence regarding the so-called unit value bias. This bias emerges when a heterogeneous aggregate is treated as being homogeneous.

Having defined the target price (and quantity) index, the statistician must face the basic problem that not all the information on the prices and quantities of the elements is available. The best he or she can obtain is information $\{p_n^0, q_n^0, p_n^1, q_n^1; n \in S\}$ for a sample $S \subset \{1, \ldots, N\}$. More realistic, however, is the situation where the information set has the form $\{p_n^0, p_n^1; n \in S\}$, that is, only a matched sample of prices is available. From this sample information the population price index (or quantity index) must be estimated. This is the point where the theory of finite population sampling will appear to be helpful for obtaining insight into the properties of the various estimators.

At the outset we must notice that in practice the way in which the sample S is drawn usually remains hidden in a sort of darkness. The main problem is that there is no such thing as a sampling frame. Knowledge about the composition of the elementary aggregate, in the form of an exhaustive listing of all its elements, is usually absent. There is only more or less ad hoc evidence available about particular elements belonging or not belonging to this aggregate. In order to use the theory of finite population sampling, however, we must make certain simplifying assumptions about the sampling design.

In the remainder of this article we will consider two scenarios. Each of these is believed to be more or less representative of actual statistical practice. The first scenario assumes that S is drawn as a simple random sample without replacement, which means that each element of the population has the same probability of being included in the sample. This so-called (first-order) inclusion probability is $\Pr(n \in S) = \varsigma(S)/N$, where $\varsigma(S)$ denotes the (prespecified) sample size.

In the second scenario the more important elements of the population have a correspondingly larger probability of being included in the sample than the less important elements. This will be formalized by assuming that the elements of S were drawn with probability proportional to size (pps) and without replacement, where size denotes some measure of importance. If the size of element n is denoted by a positive scalar a_n ($n = 1, \ldots, N$), then the probability that element n is included in the sample S is $\Pr(n \in S) = \varsigma(S)a_n / \sum_{n=1}^N a_n$. Without loss of generality, it can be assumed that $\Pr(n \in S) < 1$ for $n = 1, \ldots, N$. (Elements for which initially this probability would turn out to be larger than or equal to 1 are selected with certainty and from the remaining set of elements a sample is drawn.) Notice that in both scenarios it is the case that $\sum_{n=1}^N \Pr(n \in S) = \varsigma(S)$.

Usually the sample S has been drawn at some period prior to the base period 0, say Period b. In particular this means that in the case of pps sampling the size measure, which is either based on relative quantities (for homogeneous aggregates) or relative values (for heterogeneous aggregates), refers to period b. Consider now the target indices P_U, P_T, P_F, and P_W. All these indices are based on population price and quantity data of the two periods 0 and 1. This implies immediately that any estimator that is based on sample data of the form $\{p_n^0, q_n^0, p_n^1; n \in S\}$ or $\{p_n^0, p_n^1; n \in S\}$ will be biased, since the two sampling designs do not compensate for the missing quantity data. Put otherwise, in order to get (approximately) unbiased estimators of the target indexes we must either work with estimators based on sample data $\{p_n^0, q_n^0, p_n^1, q_n^1; n \in S\}$ or relax the requirement that the size measure used in pps sampling refers to the prior period b. The last alternative leads of course to sampling designs that look unrealistic from a practical point of view. The author is very well aware of this. However, it is considered important to study the behaviour of index estimators in somewhat idealized circumstances, in order to get at least an idea about their behaviour in more realistic situations.

4. Homogeneous Aggregates

Suppose we deal with a homogeneous aggregate. Then the target (or population) price index is the unit value index P_U. If the total base period value $\sum_{n=1}^N p_n^0 q_n^0$ as well as the total comparison period value $\sum_{n=1}^N p_n^1 q_n^1$ is known, the obvious route to take – see Expression (3) – is to estimate the Dutot quantity index Q_D. A likely candidate is its sample counterpart

$$\hat{Q}_D \equiv \sum_{n\in S} q_n^1 \Big/ \sum_{n\in S} q_n^0 \tag{13}$$

Suppose that S is a simple random sample. Then one can show (detailed in Section 12) that

$$E(\hat{Q}_D) \approx Q_D \tag{14}$$

which means that $\hat{Q}_D$ is an approximately unbiased estimator of the population Dutot quantity index Q_D. The bias tends to zero when the sample size increases.

Consider next the sample Carli quantity index, defined as

$$\hat{Q}_C \equiv \frac{1}{\varsigma(S)} \sum_{n\in S} (q_n^1/q_n^0) \tag{15}$$

Assume that the elements were drawn with probability proportional to size, whereby the size of element n is defined as its base period quantity share $q_n^0/\sum_{n=1}^N q_n^0$ $(n = 1, \ldots, N)$. Thus the probability that element n is included in the sample is equal to $\Pr(n \in S) = \varsigma(S) q_n^0/\sum_{n=1}^N q_n^0$. Then the expected value of the sample Carli quantity index is equal to

$$\begin{aligned} E(\hat{Q}_C) &= (1/\varsigma(S)) \sum_{n=1}^N \left(q_n^1/q_n^0\right) \Pr(n \in S) = \sum_{n=1}^N \left(q_n^0 \Big/ \sum_{n=1}^N q_n^0\right) \left(q_n^1/q_n^0\right) \\ &= Q_D \end{aligned} \tag{16}$$

Put otherwise, under this sampling design, the sample Carli quantity index is an unbiased estimator of the population Dutot quantity index.

Let the total comparison period value now be unknown to the statistician and consider the sample unit value index

$$\hat{P}_U \equiv \frac{\sum_{n\in S} p_n^1 q_n^1 \Big/ \sum_{n\in S} q_n^1}{\sum_{n\in S} p_n^0 q_n^0 \Big/ \sum_{n\in S} q_n^0} \tag{17}$$

This presupposes that the sample is of the form $\left\{(p_n^t q_n^t, q_n^t); t = 0, 1; n \in S\right\}$, that is, for every sampled element one knows its value and its quantity in the two periods. This situation will typically occur when one has access to electronic transaction data (so-called scanner data). Then one can show, in much the same way as was done in the case of Expression (14), that under simple random sampling the sample unit value index is an approximately unbiased estimator of the target unit value index P_U. Likewise, by mimicking the proof of (14), one can show that

$$\left(\sum_{n=1}^N p_n^0 q_n^0\right) \frac{\sum_{n\in S} p_n^1 q_n^1}{\sum_{n\in S} p_n^0 q_n^0} \tag{18}$$

is an approximately unbiased estimator of the aggregate's total comparison period value $\sum_{n=1}^N p_n^1 q_n^1$. Notice that (18) has the form of a ratio estimator.

Suppose next that only sample prices are available, that is, the sample is of the form $\left\{p_n^0, p_n^1; n \in S\right\}$, and consider the sample Dutot price index, defined as

$$\hat{P}_D \equiv \frac{\sum_{n\in S} p_n^1}{\sum_{n\in S} p_n^0} = \frac{(1/s(S))\sum_{n\in S} p_n^1}{(1/s(S))\sum_{n\in S} p_n^0} \tag{19}$$

The second part of this expression reflects the familiar interpretation of the sample Dutot price index as a ratio of unweighted average sample prices. Clearly, taking the average of prices is the counterpart of the adding-up of quantities, i.e., the first makes sense if, and only if, the second does. Under pps sampling, whereby again the size of element n is defined as its base period quantity share, it can be shown (detailed in Section 12) that, apart from a nonlinearity bias, which tends to zero when the sample size increases,

$$E(\hat{P}_D) \approx \frac{\sum_{n=1}^N p_n^1 q_n^0 / \sum_{n=1}^N q_n^0}{\sum_{n=1}^N p_n^0 q_n^0 / \sum_{n=1}^N q_n^0} \tag{20}$$

The denominator of the right-hand side ratio is the same as the denominator of the unit value index P_U. The numerators, however, differ: the target index uses comparison period quantity shares as weights, whereas $E(\hat{P}_D)$ yields base period quantity shares as weights. Thus the sample Dutot price index will in general be a biased estimator of the unit value index. The relative bias amounts to

$$\frac{E(\hat{P}_D)}{P_U} \approx \frac{\sum_{n=1}^N p_n^1 q_n^0 / \sum_{n=1}^N q_n^0}{\sum_{n=1}^N p_n^1 q_n^1 / \sum_{n=1}^N q_n^1} \tag{21}$$

The relative bias of the sample Dutot price index thus consists of two components, a technical part, which vanishes, as the sample size gets larger, and a structural part that is independent of the sample size. This structural part is given by the right-hand side of Expression (21). It vanishes if the (relative) quantities in the comparison period are the same as those in the base period, which is unlikely to happen in practice. The result, expressed by (20), goes back to Balk (1994, p. 139); see also Diewert (2002, Section 7.4).

5. Heterogeneous Aggregates and the Törnqvist Price Index

We now turn to the more important situation where we deal with a heterogeneous aggregate. Suppose that the Törnqvist price index P_T is decided on as the target and consider its sample analogue

$$\hat{P}_T \equiv \prod_{n\in S}\left(p_n^1/p_n^0\right)^{(\hat{s}_n^0+\hat{s}_n^1)/2} \tag{22}$$

where $\hat{s}_n^t \equiv p_n^t q_n^t / \sum_{n\in S} p_n^t q_n^t$ $(t=0,1)$ is element n's sample value share. It is clear that the sample must be of the form $\left\{(p_n^t q_n^t, p_n^t); t=0,1; n\in S\right\}$, that is, for each sample element we must know its value and its price in the two periods. Under the assumption of simple random sampling it can be shown that

$$\begin{aligned}
E(\ln\hat{P}_T) &= \frac{1}{2}E\left(\frac{\sum_{n\in S}p_n^0q_n^0\ln\left(p_n^1/p_n^0\right)}{\sum_{n\in S}p_n^0q_n^0}+\frac{\sum_{n\in S}p_n^1q_n^1\ln\left(p_n^1/p_n^0\right)}{\sum_{n\in S}p_n^1q_n^1}\right)\\
&\approx \frac{1}{2}\left(\frac{E\left((1/\mathfrak{s}(S))\sum_{n\in S}p_n^0q_n^0\ln\left(p_n^1/p_n^0\right)\right)}{E\left((1/\mathfrak{s}(S))\sum_{n\in S}p_n^0q_n^0\right)}+\frac{E\left((1/\mathfrak{s}(S))\sum_{n\in S}p_n^1q_n^1\ln\left(p_n^1/p_n^0\right)\right)}{E\left((1/\mathfrak{s}(S))\sum_{n\in S}p_n^1q_n^1\right)}\right)\\
&= \frac{1}{2}\left(\frac{(1/N)\sum_{n=1}^N p_n^0q_n^0\ln\left(p_n^1/p_n^0\right)}{(1/N)\sum_{n=1}^N p_n^0q_n^0}+\frac{(1/N)\sum_{n=1}^N p_n^1q_n^1\ln\left(p_n^1/p_n^0\right)}{(1/N)\sum_{n=1}^N p_n^1q_n^1}\right)\\
&= \ln P_T
\end{aligned} \tag{23}$$

This means that $\ln\hat{P}_T$ is an approximately unbiased estimator of $\ln P_T$. But what can be said about the estimator $\hat{P}_T$ itself? Using the Taylor series expansion of $f(x)=\exp\{x\}$, one obtains

$$E(\hat{P}_T) = E(\exp\{\ln\hat{P}_T\}) = \exp\{E(\ln\hat{P}_T)\}(1+R) \tag{24}$$

The leading term of the remainder R, $(1/2)E(\ln\hat{P}_T - E(\ln\hat{P}_T))^2$, is positive, but can be shown to tend to zero when the sample size increases towards N. (Expression (24) is an instance of Jensen's Inequality, which says that $E(f(x)) \geq f(E(x))$ when $f(x)$ is convex and the expectation $E(x)$ exists.) Combining (23) and (24) one obtains

$$E(\hat{P}_T) \approx P_T \exp\{R_1\}(1+R) \tag{25}$$

where R_1 denotes the bias that corresponds to (23). It is difficult to predict the direction of the entire bias of the sample Törnqvist price index. However, in any case the bias tends to zero for increasing sample size.

The previous result critically depends on the availability of sample quantity or value information. Suppose now that we cannot obtain such data and consider the sample Jevons price index (see also Bradley 2001, p. 379)

$$\hat{P}_J \equiv \prod_{n\in S} \left(p_n^1/p_n^0\right)^{1/\varsigma(S)} \tag{26}$$

Under pps sampling, whereby the size of element n is now defined as its base period value share s_n^0, resulting in $\Pr(n \in S) = \varsigma(S)s_n^0$, it is easily seen that

$$E(\ln \hat{P}_J) = E\left(\frac{1}{\varsigma(S)}\sum_{n\in S}\ln\left(p_n^1/p_n^0\right)\right) = \sum_{n=1}^{N} s_n^0 \ln\left(p_n^1/p_n^0\right) = \ln\left(\prod_{n=1}^{N}\left(p_n^1/p_n^0\right)^{s_n^0}\right) \tag{27}$$

By employing (24), with $\hat{P}_T$ substituted by $\hat{P}_J$, we obtain that

$$E(\hat{P}_J) = \prod_{n=1}^{N}\left(p_n^1/p_n^0\right)^{s_n^0}(1+R) \equiv P_{GL}(1+R) \tag{28}$$

Apart from the remainder term, we have obtained the so-called Geometric Laspeyres population price index, which in general will differ from the Törnqvist population price index. The relative bias of the sample Jevons price index with respect to the Törnqvist population price index is

$$\frac{E(\hat{P}_J)}{P_T} = \prod_{n=1}^{N}\left(p_n^1/p_n^0\right)^{(s_n^0-s_n^1)/2}(1+R) \tag{29}$$

The relative bias of the sample Jevons price index thus consists of two components, a technical part, which vanishes, as the sample size gets larger, and a structural part that is independent of the sample size. This structural part is given by the first part of the right-hand side of Expression (29). It vanishes when base period and comparison period value shares are equal, which is unlikely to occur in practice.

Instead of defining the size of element n as its base period value share s_n^0, one could as well define its size as being $(s_n^0 + s_n^1)/2$, the arithmetic mean of its base and comparison period value share. Then we obtain, instead of (28),

$$E(\hat{P}_J) = \prod_{n=1}^{N}\left(p_n^1/p_n^0\right)^{(s_n^0+s_n^1)/2}(1+R) \equiv P_T(1+R) \tag{30}$$

and instead of (29)

$$\frac{E(\hat{P}_J)}{P_T} = 1 + R \tag{31}$$

that is, the sample Jevons price index is an approximately unbiased estimator of the population Törnqvist price index. The bias will vanish when the sample size gets larger. This result goes back to Diewert (2002, Section 7.4).

6. Heterogeneous Aggregates and the Fisher Price Index

Suppose that instead of the Törnqvist price index one has decided that the Fisher price index (7) should be the target. Suppose further that our sample information consists of prices and quantities. The sample analogue of the population Fisher price index is

$$\hat{P}_F \equiv \left(\frac{\sum_{n\in S} p_n^1 q_n^0 \sum_{n\in S} p_n^1 q_n^1}{\sum_{n\in S} p_n^0 q_n^0 \sum_{n\in S} p_n^0 q_n^1} \right)^{1/2} = \left(\frac{(1/s(S))\sum_{n\in S} p_n^1 q_n^0 \,(1/s(S))\sum_{n\in S} p_n^1 q_n^1}{(1/s(S))\sum_{n\in S} p_n^0 q_n^0 \,(1/s(S))\sum_{n\in S} p_n^0 q_n^1} \right)^{1/2} \tag{32}$$

Then one can show (detailed in Section 12) that, under simple random sampling,

$$E(\ln \hat{P}_F) \approx \ln P_F \tag{33}$$

By repeating the argument concerning Expressions (24–25), one may then also conclude that the sample Fisher price index itself is an approximately unbiased estimator of its population counterpart. The bias tends to zero when the sample size increases.

Suppose now that only sample prices are available, and consider the sample Carli price index,

$$\hat{P}_C \equiv \frac{1}{s(S)} \sum_{n\in S} p_n^1 / p_n^0 \tag{34}$$

Under pps sampling, whereby the size of element n is defined as its base period value share s_n^0, we immediately see that

$$E(\hat{P}_C) = \sum_{n=1}^{N} s_n^0 \left(p_n^1/p_n^0\right) = \frac{\sum_{n=1}^{N} p_n^1 q_n^0}{\sum_{n=1}^{N} p_n^0 q_n^0} = P_L \tag{35}$$

Thus the expected value of the sample Carli price index appears to be equal to the population Laspeyres price index. This result goes back to Balk (1994, p. 139); see also Diewert (2002, Section 7.4). The relative bias of the sample Carli price index with respect to the population Fisher price index follows immediately from (35) and appears to be

$$\frac{E(\hat{P}_C)}{P_F} = \frac{P_L}{P_F} = \left(\frac{P_L}{P_P} \right)^{1/2} \tag{36}$$

which is the squared root of the ratio of the population Laspeyres price index and the population Paasche price index. Notice that this bias is of a structural nature, i.e., will not disappear when the sample size gets larger.

The population Fisher price index can also be written as

$$P_F = \left(\sum\nolimits_{n=1}^{N} s_n^0 \left(p_n^1/p_n^0\right)\right)^{1/2} \left(\sum\nolimits_{n=1}^{N} s_n^1 \left(p_n^1/p_n^0\right)^{-1}\right)^{-1/2} \tag{37}$$

We now consider whether, following a suggestion of Fisher (1922, p. 472, Formula 101), the Carruthers-Sellwood-Ward (1980)-Dalén (1992) sample price index

$$\hat{P}_{CSWD} \equiv \left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)\right)^{1/2} \left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)^{-1}\right)^{-1/2} \tag{38}$$

under some sampling design might be a suitable estimator of the population Fisher price index. The CSWD sample price index is the geometric average of the sample Carli price index (34) and the sample Harmonic (or Coggeshall) price index

$$\hat{P}_H \equiv \left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)^{-1}\right)^{-1} \tag{39}$$

Thus, consider

$$\ln \hat{P}_{CSWD} = \frac{1}{2} \ln\left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)\right) - \frac{1}{2} \ln\left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)^{-1}\right) \tag{40}$$

Under pps sampling, whereby the size of element n is defined as its base period value share s_n^0, and again using the Taylor series expansion of $f(x) = \ln x$, we find that

$$E\left(\ln\left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)\right)\right) = \ln\left(\sum\nolimits_{n=1}^{N} s_n^0 \left(p_n^1/p_n^0\right)\right) + R_1 = \ln P_L + R_1 \tag{41}$$

Similarly,

$$\begin{aligned} E\left(\ln\left(\frac{1}{\varsigma(S)} \sum_{n \in S} \left(p_n^1/p_n^0\right)^{-1}\right)\right) &= \ln\left(\sum\nolimits_{n=1}^{N} s_n^0 \left(p_n^1/p_n^0\right)^{-1}\right) + R_2 \\ &= \ln(1/P_{HL}) + R_2 \end{aligned} \tag{42}$$

where P_{HL} is called the population Harmonic Laspeyres price index. Combining (40), (41), and (42), one obtains

$$E(\ln \hat{P}_{CSWD}) = \frac{1}{2}(\ln P_L - \ln(1/P_{HL}) + R_1 - R_2) = \ln(P_L P_{HL})^{1/2} + \frac{1}{2}(R_1 - R_2) \tag{43}$$

The leading term of $(R_1 - R_2)/2$ is

$$-\frac{1}{4}\left(cv\left((1/\varsigma(S)) \sum_{n \in S} \left(p_n^1/p_n^0\right)\right)\right)^2 + \frac{1}{4}\left(cv\left((1/\varsigma(S)) \sum_{n \in S} \left(p_n^1/p_n^0\right)^{-1}\right)\right)^2 \tag{44}$$

both parts of which tend to zero when the sample size increases. Thus, $\ln \hat{P}_{CSWD}$ will be an approximately unbiased estimator of $\ln (P_L P_{HL})^{1/2}$, and, repeating a by now familiar argument,

$$E(\hat{P}_{CSWD}) = (P_L P_{HL})^{1/2} \exp\{(R_1 - R_2)/2\}(1 + R) \tag{45}$$

where R also tends to zero when the sample size increases. The main right-hand side term clearly differs from the population Fisher price index. The relative bias of the CSWD sample price index with respect to the population Fisher price index is

$$\frac{E(\hat{P}_{CSWD})}{P_F} = \left(\frac{P_{HL}}{P_P}\right)^{1/2} \exp\{(R_1 - R_2)/2\}(1 + R) \tag{46}$$

Notice that the relative bias consists of two components, a technical component, which vanishes, as the sample size gets larger, and a structural component, which is independent of the sample size.

Instead of defining the size of element n as its base period value share s_n^0, one could as well define its size as being $(s_n^0 + s_n^1)/2$, the arithmetic mean of its base and comparison period value share. Then we obtain, instead of (41),

$$\begin{aligned} E\left(\ln\left(\frac{1}{s(S)}\sum_{n\in S}(p_n^1/p_n^0)\right)\right) &= \ln\left(\sum_{n=1}^{N}\frac{1}{2}\left(s_n^0 + s_n^1\right)\left(p_n^1/p_n^0\right)\right) + R_1 \\ &= \ln((P_L + P_{PAL})/2) + R_1 \end{aligned} \tag{47}$$

where

$$P_{PAL} \equiv \sum_{n=1}^{N} s_n^1 \left(p_n^1/p_n^0\right) \tag{48}$$

is the population Palgrave price index. Similarly, instead of (42) we get

$$\begin{aligned} E\left(\ln\left(\frac{1}{s(S)}\sum_{n\in S}\left(p_n^1/p_n^0\right)^{-1}\right)\right) &= \ln\left(\sum_{n=1}^{N}\frac{1}{2}\left(s_n^0 + s_n^1\right)\left(p_n^1/p_n^0\right)^{-1}\right) + R_2 \\ &= \ln\left(\left(P_{HL}^{-1} + P_P^{-1}\right)/2\right) + R_2 \end{aligned} \tag{49}$$

Combining these two equalities with (40), we get

$$E(\ln \hat{P}_{CSWD}) = \ln\left(\frac{P_L + P_{PAL}}{P_{HL}^{-1} + P_P^{-1}}\right)^{1/2} + \frac{1}{2}(R_1 - R_2) \tag{50}$$

with, again, a remainder term that tends to zero when the sample size increases. Finally,

$$\begin{aligned} E(\hat{P}_{CSWD}) &= \left(\frac{P_L + P_{PAL}}{P_{HL}^{-1} + P_P^{-1}}\right)^{1/2} \exp\{(R_1 - R_2)/2\}(1 + R) \\ &= P_F\left(\frac{1 + P_{PAL}/P_L}{1 + P_P/P_{HL}}\right)^{1/2} \exp\{(R_1 - R_2)/2\}(1 + R) \end{aligned} \tag{51}$$

Notice that the population ratio P_P/P_{HL} is the temporal antithesis of P_{PAL}/P_L. There is numerical evidence (see Vartia 1978) that these ratios are each other's reciprocal. Thus, under the pps sampling design defined immediately before Expression (47), the CSWD sample price index turns out to be an approximately unbiased estimator of the population Fisher price index.

Let us finally consider the following modification of the CSWD sample price index:

$$\hat{P}_B \equiv \left(\frac{1}{s(S)}\sum_{n\in S}\left(p_n^1/p_n^0\right)\right)^{1/2}\left(\frac{1}{s(S)}\sum_{n\in S}\left(q_n^1/q_n^0\right)\right)^{-1/2}\left(\frac{1}{s(S)}\sum_{n\in S}\left(p_n^1q_n^1/p_n^0q_n^0\right)\right)^{1/2} \tag{52}$$

This is the product of a sample Carli price index, a sample Harmonic quantity index, and a sample Carli value index. It is straightforward to show, using the same reasoning as in the previous paragraphs, that under pps sampling, whereby the size of element n is defined as its base period value share s_n^0,

$$E(\ln\hat{P}_B) \approx \frac{1}{2}\left[\ln P_L - \ln Q_L + \ln\left(\frac{\sum_{n=1}^N p_n^1q_n^1}{\sum_{n=1}^N p_n^0q_n^0}\right)\right] = \ln P_F \tag{53}$$

and thus

$$E(\hat{P}_B) \approx P_F \tag{54}$$

where the bias tends to zero for increasing sample size. However, it is clear that the computation of $\hat{P}_B$ requires more information than the computation of $\hat{P}_{CSWD}$, namely all sample quantity relatives. If one has access to scanner data, however, this should not be a problem.

7. Heterogeneous Aggregates and the Walsh Price Index

Suppose that the Walsh price index (9) were chosen as the target and that our sample information consists of prices and quantities. The sample analogue of the population Walsh price index is

$$\hat{P}_W \equiv \frac{\sum_{n\in S}p_n^1\left(q_n^0q_n^1\right)^{1/2}}{\sum_{n\in S}p_n^0\left(q_n^0q_n^1\right)^{1/2}} \tag{55}$$

Suppose again that S is a simple random sample. Then we find, in the same way as detailed earlier, that

$$E(\hat{P}_W) = E\left(\frac{(1/s(S))\sum_{n\in S}p_n^1\left(q_n^0q_n^1\right)^{1/2}}{(1/s(S))\sum_{n\in S}p_n^0\left(q_n^0q_n^1\right)^{1/2}}\right) \approx \frac{E\left((1/s(S))\sum_{n\in S}p_n^1\left(q_n^0q_n^1\right)^{1/2}\right)}{E\left((1/s(S))\sum_{n\in S}p_n^0\left(q_n^0q_n^1\right)^{1/2}\right)}$$

$$= \frac{(1/N)\sum_{n=1}^N p_n^1\left(q_n^0q_n^1\right)^{1/2}}{(1/N)\sum_{n=1}^N p_n^0\left(q_n^0q_n^1\right)^{1/2}} = P_W \tag{56}$$

which means that the sample Walsh price index is an approximately unbiased estimator of the population Walsh price index.

Suppose now that only sample prices are available. The population Walsh price index can be written as a quadratic mean of order 1 index:

$$P_W = \frac{\sum_{n=1}^{N} \left(s_n^0 s_n^1\right)^{1/2} \left(p_n^1/p_n^0\right)^{1/2}}{\sum_{n=1}^{N} \left(s_n^0 s_n^1\right)^{1/2} \left(p_n^1/p_n^0\right)^{-1/2}} \tag{57}$$

which suggests the following sample price index

$$\hat{P}_{BW} \equiv \frac{\sum_{n\in S} \left(p_n^1/p_n^0\right)^{1/2}}{\sum_{n\in S} \left(p_n^1/p_n^0\right)^{-1/2}} \tag{58}$$

Since there appears to be no name attached to this formula in the literature, Expression (58) will be baptized the Balk-Walsh sample price index. Under a pps sampling design, whereby the size of element n is defined as $\left(s_n^0 s_n^1\right)^{1/2}$, the geometric mean of its base and comparison period value share, we find that

$$E(\hat{P}_{BW}) = E\left(\frac{(1/\varsigma(S))\sum_{n\in S} \left(p_n^1/p_n^0\right)^{1/2}}{(1/\varsigma(S))\sum_{n\in S} \left(p_n^1/p_n^0\right)^{-1/2}}\right) \approx \frac{E\left((1/\varsigma(S))\sum_{n\in S} \left(p_n^1/p_n^0\right)^{1/2}\right)}{E\left((1/\varsigma(S))\sum_{n\in S} \left(p_n^1/p_n^0\right)^{-1/2}\right)}$$

$$= \frac{(1/N)\sum_{n=1}^{N} \left(s_n^0 s_n^1\right)^{1/2} \left(p_n^1/p_n^0\right)^{1/2}}{(1/N)\sum_{n=1}^{N} \left(s_n^0 s_n^1\right)^{1/2} \left(p_n^1/p_n^0\right)^{-1/2}} = P_W \tag{59}$$

Thus, under this sampling design, the Balk-Walsh sample price index appears to be an approximately unbiased estimator of the population Walsh price index. The bias will approach zero when the sample size increases. It is easy to demonstrate that, if the size of element n had been defined as its base period value share, s_n^0, the expectation of the Balk-Walsh sample price index would be unequal to the population Walsh price index.

8. Heterogeneous Aggregates: Micro-economic Considerations Regarding the Choice of the Sample Price Index

The previous three sections demonstrated that, when nothing but sample prices are available and the sampling design is restricted to one that uses only base period value share information, it is impossible to estimate any of the population superlative price indexes unbiasedly. Basically, there remain a number of second-best alternatives, namely the sample Jevons (26), Carli (34), Harmonic (39), Carruthers-Sellwood-Ward-Dalén (38) and Balk-Walsh (58) price indexes. Is any one of these to be preferred?

To assist with the choice, let us consider the sample Generalized Mean (GM) price index, which is defined as

$$\hat{P}_{GM}(\sigma) \equiv \left(\frac{1}{s(S)} \sum_{n \in S} \left(p_n^1 / p_n^0 \right)^{1-\sigma} \right)^{1/(1-\sigma)} \quad (\sigma \neq 1)$$

$$\equiv \prod_{n \in S} \left(p_n^1 / p_n^0 \right)^{1/s(S)} (\sigma = 1) \tag{60}$$

It is immediately seen that $\hat{P}_J = \hat{P}_{GM}(1)$, $\hat{P}_C = \hat{P}_{GM}(0)$, and $\hat{P}_H = \hat{P}_{GM}(2)$, whereas $\hat{P}_{CSWD} = [\hat{P}_{GM}(0)\hat{P}_{GM}(2)]^{1/2}$, and $\hat{P}_{BW} = [\hat{P}_{GM}(1/2)P_{GM}(3/2)]^{1/2}$. However, since the GM price index is a monotonous function of σ, it appears that, to the second order, $\hat{P}_{CSWD} \approx \hat{P}_{BW} \approx \hat{P}_{GM}(1)$ (see also Section 10). Thus these five sample price indexes are members of the same family.

Under pps sampling, whereby the size of element n is defined as its base period value share s_n^0, one obtains that

$$E(\hat{P}_{GM}(\sigma)^{1-\sigma}) = \sum\nolimits_{n=1}^{N} s_n^0 \left(p_n^1 / p_n^0 \right)^{1-\sigma} \tag{61}$$

To apply Jensen's Inequality, a distinction must be made between two cases. If $\sigma \leq 0$ we obtain

$$E(\hat{P}_{GM}(\sigma)) \leq \left(\sum\nolimits_{n=1}^{N} s_n^0 \left(p_n^1 / p_n^0 \right)^{1-\sigma} \right)^{1/(1-\sigma)} \equiv P_{LM}(\sigma) \tag{62}$$

whereas if $\sigma \geq 0$ we obtain

$$E\left(\hat{P}_{GM}(\sigma) \right) \geq \left(\sum\nolimits_{n=1}^{N} s_n^0 \left(p_n^1 / p_n^0 \right)^{1-\sigma} \right)^{1/(1-\sigma)} \equiv P_{LM}(\sigma) \, (\sigma \neq 1) \tag{63}$$

$$E\left(\hat{P}_{GM}(1) \right) \geq \prod\nolimits_{n=1}^{N} \left(p_n^1 / p_n^0 \right)^{s_n^0} \equiv P_{LM}(1)$$

where $P_{LM}(\sigma)$ is the Lloyd-Moulton (LM) population price index. Thus, for $\sigma \leq 0$ the sample GM price index has a negative bias, and for $\sigma \geq 0$ a positive. The bias tends to zero when the sample size increases.

Economic theory teaches us that the LM index is exact for a Constant Elasticity of Substitution (partial) revenue function (for the producers' output side) or (partial) cost function (for the producers' input side or the consumer; see Balk 2000). The parameter σ is thereby to be interpreted as the (average) elasticity of substitution within the aggregate. On their output side, producers are supposed to *maximize* revenue, which implies a nonpositive elasticity of substitution. Producers on their input side and consumers, however, are supposed to *minimize* cost, which implies a nonnegative elasticity of substitution. In particular, the conclusion must be that, under the pps sampling design here assumed, the sample Jevons, Harmonic, CSWD, and Balk-Walsh price indexes are inadmissible for the producer output side since the expected value of each of these indexes would exhibit positive substitution elasticity. The sample Carli price index is admissible, even unbiased, but would imply a zero substitution elasticity.

9. Heterogeneous Aggregates and the Lowe Price Index

Let us now turn to the more realistic case in which the Lowe price index (11) is defined as the target. The population Lowe price index can be written as a ratio of two Laspeyres price indexes

$$P_{Lo} = \frac{\sum_{n=1}^{N} p_n^1 q_n^b / \sum_{n=1}^{N} p_n^b q_n^b}{\sum_{n=1}^{N} p_n^0 q_n^b / \sum_{n=1}^{N} p_n^b q_n^b} = \frac{\sum_{n=1}^{N} s_n^b \left(p_n^1/p_n^b\right)}{\sum_{n=1}^{N} s_n^b \left(p_n^0/p_n^b\right)} \tag{64}$$

where s_n^b is element n's value share in period b ($n = 1, \ldots, N$), which is assumed to be some period prior to the base period. This suggests the following sample price index (see also Bradley (2001, p. 377); note that he uses the name "modified Laspeyres index" instead of "Lowe index")

$$\hat{P}_{Lo} \equiv \frac{\sum_{n \in S} p_n^1/p_n^b}{\sum_{n \in S} p_n^0/p_n^b} \tag{65}$$

which is the ratio of two sample Carli price indexes. Indeed, under a pps sampling design, whereby the size of element n is defined as s_n^b, that is its period b value share, it is easily demonstrated that

$$E(\hat{P}_{Lo}) = E\left(\frac{(1/\mathfrak{s}(S))\sum_{n \in S} p_n^1/p_n^b}{(1/\mathfrak{s}(S))\sum_{n \in S} p_n^0/p_n^b}\right) \approx \frac{E\left((1/\mathfrak{s}(S))\sum_{n \in S} p_n^1/p_n^b\right)}{E\left((1/\mathfrak{s}(S))\sum_{n \in S} p_n^0/p_n^b\right)}$$

$$= \frac{(1/N)\sum_{n=1}^{N} s_n^b \left(p_n^1/p_n^b\right)}{(1/N)\sum_{n=1}^{N} s_n^b \left(p_n^0/p_n^b\right)} = P_{Lo} \tag{66}$$

The bias tends to zero when the sample size increases.

Alternatively and perhaps more consistent with practice, one could consider the so-called price-updated period b value shares, defined as

$$s_n^{b(0)} \equiv \frac{s_n^b \left(p_n^0/p_n^b\right)}{\sum_{n=1}^{N} s_n^b \left(p_n^0/p_n^b\right)} = \frac{p_n^0 q_n^b}{\sum_{n=1}^{N} p_n^0 q_n^b} (n = 1, \ldots, N) \tag{67}$$

Under a pps sampling design, whereby the size of element n is now defined as $s_n^{b(0)}$, that is its price-updated period b value share, it is immediately seen that

$$E(\hat{P}_C) = \sum_{n=1}^{N} s_n^{b(0)} \left(p_n^1/p_n^0\right) = P_{Lo} \tag{68}$$

that is, the sample Carli price index is an unbiased estimator of the population Lowe price index. However, if the size of element n were defined as s_n^b, that is its period b value share itself, one would have obtained

$$E(\hat{P}_C) = \sum_{n=1}^{N} s_n^b \left(p_n^1/p_n^0\right) \tag{69}$$

which, unless the prices have changed between the periods b and 0, differs from the population Lowe price index.

10. The Time Reversal Test and Some Numerical Relations

When there is nothing but sample price information available, that is, the sample has the form $\{p_n^0, p_n^1; n \in S\}$, the menu of sample price indexes appears to be limited. For a *homogeneous aggregate* only the sample Dutot price index (19) is available. Note that this index, like the population unit value index, satisfies the Time Reversal test, that is, using obvious notation,

$$\hat{P}_D(p^1, p^0)\hat{P}_D(p^0, p^1) = 1 \tag{70}$$

However, as has been shown, under a not unreasonable sampling design the sample Dutot price index is a biased estimator of the target unit value index.

For a *heterogeneous aggregate* one has, depending on the definition of the target price index, the choice between the sample Carli price index (34), the sample Jevons price index (26), the sample Harmonic price index (39), the sample CSWD price index (38), the sample Balk-Walsh price index (58) and the sample Lowe price index (65). The first three indexes are special cases of the sample GM price index (60), respectively for $\sigma = 0, 1, 2$. Since the GM price index is monotonously increasing in $1 - \sigma$, we obtain the general result that

$$\hat{P}_{GM}(p^1, p^0; \sigma)\hat{P}_{GM}(p^0, p^1; \sigma) \geq 1 \text{ for } \sigma < 1 \tag{71}$$

$$\hat{P}_{GM}(p^1, p^0; \sigma)\hat{P}_{GM}(p^0, p^1; \sigma) \leq 1 \text{ for } \sigma > 1 \tag{72}$$

which means that the GM price index fails the Time Reversal Test. In particular, the Carli price index and the Harmonic price index fail the Time Reversal test, that is,

$$\hat{P}_C(p^1, p^0)\hat{P}_C(p^0, p^1) \geq 1 \tag{73}$$

and

$$\hat{P}_H(p^1, p^0)\hat{P}_H(p^0, p^1) \leq 1 \tag{74}$$

The Jevons price index, the CSWD price index and the Balk-Walsh price index satisfy the Time Reversal test, as one verifies immediately. As has been shown in Section 8, under a not unreasonable sampling design these three sample price indexes are (approximately) unbiased estimators of the LM population price index with $\sigma = 1$. The sample Lowe price index also satisfies the Time Reversal Test. This index is, under a not unreasonable sampling design, an (approximately) unbiased estimator of the population Lowe price index.

I now turn to numerical relations between all these indexes. It is well known that

$$\hat{P}_H \leq \hat{P}_J \leq \hat{P}_C \tag{75}$$

and thus we might expect that $\hat{P}_{CSWD} = (\hat{P}_H \hat{P}_C)^{1/2}$ will be close to $\hat{P}_J$. The magnitudes of the differences between all these indexes depend on the variance of the price relatives

p_n^1/p_n^0. When all the price relatives are equal, the inequalities (75) turn into equalities. In fact, Dalén (1992) and Diewert (1995) showed that, to the second order, the following approximations hold (only their main results are presented here; an additional one is to be found in Reinsdorf 1994):

$$\hat{P}_J \approx \hat{P}_C\left(1 - \frac{1}{2}\operatorname{var}(\varepsilon)\right) \quad (76)$$

$$\hat{P}_H \approx \hat{P}_C(1 - \operatorname{var}(\varepsilon)) \quad (77)$$

$$\hat{P}_{CSWD} \approx \hat{P}_C\left(1 - \frac{1}{2}\operatorname{var}(\varepsilon)\right) \quad (78)$$

where $\operatorname{var}(\varepsilon) \equiv (1/\varsigma(S))\sum_{n\in S}\varepsilon_n^2$ and $\varepsilon_n \equiv (p_n^1/p_n^0 - \hat{P}_C)/\hat{P}_C$ $(n \in S)$. In the same way one can show that

$$\hat{P}_{BW} \approx \hat{P}_C\left(1 - \frac{1}{2}\operatorname{var}(\varepsilon)\right) \quad (79)$$

The method of proof is to write the ratio of $\hat{P}_{BW}$ to $\hat{P}_C$ as a function *f(ε)* and expand this function as a Taylor series around 0. Notice thereby that $\sum_{n\in S}\varepsilon_n = 0$. Hence the sample Jevons price index, the sample CSWD price index and the sample Balk-Walsh price index approximate each other to the second order. From the point of view of simplicity, the sample Jevons price index obviously gets the highest score.

To obtain some insight into the relation between the sample Lowe price index (65) and the sample Carli price index (34), the first can be written as

$$\hat{P}_{Lo} \equiv \frac{\sum_{n\in S}(p_n^0/p_n^b)(p_n^1/p_n^0)}{\sum_{n\in S}p_n^0/p_n^b} \quad (80)$$

Consider now the difference $\hat{P}_{Lo} - \hat{P}_C$. By straightforward manipulation of this expression one can show that

$$\hat{P}_{Lo} = \hat{P}_C(1 + \operatorname{cov}(\delta, \varepsilon)) \quad (81)$$

where $\operatorname{cov}(\delta,\varepsilon) \equiv (1/\varsigma(S))\sum_{n\in S}\delta_n\varepsilon_n$, $\delta_n \equiv \left(p_n^0/p_n^b - \hat{P}_C(p^0,p^b)\right)/\hat{P}_C(p^0,p^b)$ and $\varepsilon_n \equiv \left(p_n^1/p_n^0 - \hat{P}_C(p^1,p^0)\right)/\hat{P}_C(p^1,p^0)$ $(n \in S)$. Thus the difference between these two sample price indexes depends on the covariance of the relative price changes between the periods b and 0 and of those between the periods 0 and 1. Whether this difference is positive or negative, large or small, is an empirical matter.

Although it was argued that the (sample) Dutot price index only makes sense in the case of homogeneous aggregates, it appears that this index is rather frequently used also for heterogeneous aggregates. Therefore it might be of some interest to discuss the relation between this index and the sample Jevons index. The first is a ratio of arithmetic average prices whereas the second can be considered as a ratio of geometric average prices. In order to see their relation, the Jevons index is written as

$$\ln \hat{P}_J = (1/\varsigma(S)) \sum_{n \in S} \ln \left(p_n^1/p_n^0\right) \tag{82}$$

and the Dutot index as

$$\ln \hat{P}_D = \sum_{n \in S} \left(\frac{L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)}{\sum_{n \in S} L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)} \right) \ln \left(p_n^1/p_n^0\right) \tag{83}$$

where $\bar{p}^t \equiv (1/\varsigma(S)) \sum_{n \in S} p_n^t$ $(t = 0, 1)$ are the arithmetic average prices and $L(.,.)$ denotes the logarithmic mean. This mean is, for any two positive numbers a and b, defined by $L(a, b) \equiv (a - b)/ \ln (a/b)$ and $L(a, a) \equiv a$. It is a symmetric mean with the property that $(ab)^{1/2} \leq L(a, b) \leq (a + b)/2$, that is, it lies between the geometric and the arithmetic mean (see Lorenzen 1990). Thus $L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)$ can be interpreted as the mean relative price of element n. Then

$$\ln \hat{P}_D - \ln \hat{P}_J = \sum_{n \in S} \left(\frac{L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)}{\sum_{n \in S} L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)} - \frac{1}{\varsigma(S)} \right) \ln \left(p_n^1/p_n^0\right)$$

$$= \frac{1}{\varsigma(S)} \sum_{n \in S} \left(\frac{L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)}{(1/\varsigma(S)) \sum_{n \in S} L\left(p_n^0/\bar{p}^0, p_n^1/\bar{p}^1\right)} - 1 \right) \left(\ln \left(p_n^1/p_n^0\right) - \ln \hat{P}_J\right) \tag{84}$$

which means that the (sign of the) difference between the Dutot and the Jevons index depends on the (sign of the) covariance between relative prices and price relatives. Whether this difference is positive or negative, large or small, is an empirical matter.

11. Conclusion

In this article I have considered for elementary aggregates the relation between the target index, the sample index, and the sampling design. Although the viewpoint was by and large theoretical, the arguments advanced in the previous sections lead to the following practical advice. The advice, to be practical, concerns simple random sampling (srs), sampling with probability proportional to base period quantity shares (in the case of a homogeneous aggregate), and sampling with probability proportional to base period or (price-updated) earlier period value shares (in the case of a heterogeneous aggregate). It is recognised that sampling in practice may take two stages: first the sampling of respondents (outlets or firms) and then of commodities. The discussion here was kept for simplicity in terms of single-stage sampling. It is also recognised that purposive sampling and/or sampling with cut-off rules are often used at either stage. In such circumstances there are implicit sampling frames and selection rules and some judgement will be necessary as to which theoretical sampling design most closely corresponds to the method used, and what the implications are for the choice of the sample index. (See Dorfman, Lent, Leaver, and Wegman (2001) for a comparison of two sampling designs, one pps and the other based on purposive/cut-off selection.)

Table 1 presents the key results in the order of their appearance. In the first place, it is clear that respondents should be encouraged to provide timely data on comparison and base period values and prices (or quantities). Providing a full array of scanner data would be even more helpful. Of course, in some areas this should be more feasible than in others. In such cases sample indexes, which mirror their population counterparts, should be used and respondent-commodity pairs should be sampled using simple random sampling, since each sample index would then be an (approximately) unbiased estimator of the corresponding population one.

When this approach is not feasible and the best one can obtain is a sample of (matched) prices, the sampling design should be such that important elements have a correspondingly higher probability of inclusion in the sample than unimportant elements. With respect to the sample price index to be used:

- For a homogeneous aggregate, that is an aggregate for which the quantities of the elements can be meaningfully added, one must use the sample Dutot price index. Unfortunately, under pps-q0 this index will exhibit bias, the magnitude of which depends on the dispersion of the elementary quantity changes between the two periods compared.
- For a heterogeneous aggregate not at the producers' output side, one could use the sample Jevons price index. Under pps-s0 its expected value will approximate the Geometric Laspeyres price index, which is identical to the Lloyd-Moulton price index with $\sigma = 1$.
- For a heterogeneous aggregate on the producers' output side one could use a sample Generalized Mean price index with appropriately chosen parameter $\sigma \leq 0$. Under pps-s0 the expected value of such a price index will approximate a Lloyd-Moulton price index. The limiting case ($\sigma = 0$) is the sample Carli price index. As shown above, under pps-s0 this index is an unbiased estimator of the Laspeyres price index, which corresponds to zero substitution. If this index is chosen as the target, then the sample Carli index is appropriate. (Notice that PPI Manual's (2004, par. 20.83) usage of the word "bias" refers to the fact that the Carli index does not satisfy the Time Reversal test; see Expression (73) above.)
- When the target is a Lowe price index, the sample Lowe and Carli price indexes exhibit, dependent on the sampling design (pps-sb and pps-sb(0) respectively), appropriate behaviour.

In any case the time span between the two periods compared should not become too long, since the magnitude of the bias will in general grow with the length of the time span. That is, at regular time intervals one should undertake a base period change.

There remains the practical issue as to how to decide whether an aggregate is homogeneous or not. The question posed in (2) above was:

Does it make (economic) sense to add up the quantities q_n^t of the elements $n = 1, \ldots, N$?

Table 1. Key results

Target price index	Sample price index	Sampling design	Expected value of sample index	Main equation
Unit value	Unit value	srs	Unit value	(17)
Unit value	Dutot	pps-q0	$\neq$ Unit value	(20)
Törnqvist	Törnqvist	srs	Törnqvist	(25)
Törnqvist	Jevons	pps-s0	Geometric Laspeyres = LM(1)	(28)
Fisher	Fisher	srs	Fisher	(33)
Fisher	Carli	pps-s0	Laspeyres = LM(0)	(35)
Fisher	CSWD	pps-s0	LM(1)	(45)
Walsh	Walsh	srs	Walsh	(56)
Walsh	Balk-Walsh	pps-s0	$\neq$ Walsh	
LM(σ)	GM (σ)	pps-s0	LM(σ)	(62) – (63)
Lowe	Lowe	pps-sb	Lowe	(66)
Lowe	Carli	pps-sb(0)	Lowe	(68)

For example, if the aggregate consists of 14 inch television sets, the answer must be "no." Brand differences, additional facilities such as stereo, wide screens and much more account for significant variations in price. Tins of a specific brand and type of food of different sizes similarly lack homogeneity, since much of the price variation will be due to tin size. Homogeneity is lacking when the item itself varies according to identifiable price-determining characteristics. In principle the conditions of sale need to be taken into account, since an item sold by one manufacturer may command a price premium based on better delivery, warranties, or other such features. The price at initiation should be defined to have the same specified conditions of sale, but there may be elements of trust in the buyer–seller relationship that are difficult to identify. Nonetheless for practical purposes items of the same product sold by different establishments are treated as homogenous unless there are clearly identifiable differences in the terms and conditions surrounding the sale.

12. Appendix: Proofs

Proof of (14): Let S be a simple random sample without replacement and recall that the inclusion probabilities are $\Pr(n \in S) = \varsigma(S)/N$, where $\varsigma(S)$ denotes the sample size. For the expected value of the (modified) numerator and denominator of Expression (13) we obtain

$$E\left((1/\varsigma(S))\sum_{n\in S} q_n^t\right) = (1/\varsigma(S))\sum_{n=1}^{N} q_n^t \Pr(n \in S) = (1/N)\sum_{n=1}^{N} q_n^t \equiv \bar{q}^{\,t}\ (t = 0, 1) \qquad \text{(A.1)}$$

The sample Dutot index itself, however, is a nonlinear function. Expanding $\hat{Q}_D$ as a Taylor series at $(\bar{q}^{\,1}, \bar{q}^{\,0})$ and taking the expectation, one gets

$$E(\hat{Q}_D) = Q_D + R \qquad \text{(A.2)}$$

where R is the remainder. The leading term thereof is of the second order and has the form

$$\frac{\bar{q}^1}{(\bar{q}^0)^3} E\left((1/\mathfrak{s}(S))\sum_{n\in S} q_n^0 - \bar{q}^0\right)^2 - \frac{1}{(\bar{q}^0)^2} E\left(\left((1/\mathfrak{s}(S))\sum_{n\in S} q_n^0 - \bar{q}^0\right)\left((1/\mathfrak{s}(S))\sum_{n\in S} q_n^1 - \bar{q}^1\right)\right) \tag{A.3}$$

Using classical finite population sampling theory, it is easy to show (see e.g., Knottnerus 2003, p. 19) that the variance of the sample mean, occurring in the first part of Expression (A.3), equals

$$E\left((1/\mathfrak{s}(S))\sum_{n\in S} q_n^0 - \bar{q}^0\right)^2 = (1/\mathfrak{s}(S) - 1/N)(1/(N-1))\sum_{n=1}^{N}\left(q_n^0 - \bar{q}^0\right)^2 \tag{A.4}$$

It is clear that this term approaches zero when the sample size increases towards N. Similar considerations apply to the covariance term in (A.3). The entire bias R is known as small sample nonlinearity bias; empirically this bias appears to be negligible already for samples of moderate size. Instead of (A.2) we will write

$$E(\hat{Q}_D) \approx Q_D \tag{A.5}$$

and say that $\hat{Q}_D$ is an approximately unbiased estimator of Q_D.

Proof of (20): The proof proceeds in the same way as the previous one, except that now pps sampling is assumed. We find that

$$E\left((1/\mathfrak{s}(S))\sum_{n\in S} p_n^t\right) = (1/\mathfrak{s}(S))\sum_{n=1}^{N} p_n^t \Pr(n\in S) = \sum_{n=1}^{N} p_n^t q_n^0 \Big/ \sum_{n=1}^{N} q_n^0 \equiv \bar{p}^t (t=0,1) \tag{A.6}$$

and

$$E(\hat{P}_D) = \frac{\bar{p}^1}{\bar{p}^0} + R \tag{A.7}$$

The leading term of R is of the second order and has the form

$$\frac{\bar{p}^1}{(\bar{p}^0)^3} E\left((1/\mathfrak{s}(S))\sum_{n\in S} p_n^0 - \bar{p}^0\right)^2 - \frac{1}{(\bar{p}^0)^2} E\left(\left((1/\mathfrak{s}(S))\sum_{n\in S} p_n^0 - \bar{p}^0\right)\left((1/\mathfrak{s}(S))\sum_{n\in S} p_n^1 - \bar{p}^1\right)\right) \tag{A.8}$$

Knottnerus (2003; p. 71) shows that, under pps sampling without replacement,

$$E\left((1/\mathfrak{s}(S))\sum_{n\in S} p_n^0 - \bar{p}^0\right)^2 = \frac{1 + (\mathfrak{s}(S) - 1)\rho}{\mathfrak{s}(S)} \sum_{n=1}^{N}\left(p_n^0 - \bar{p}^0\right)^2 q_n^0 / \sum_{n=1}^{N} q_n^0 \tag{A.9}$$

where ρ is the sampling autocorrelation coefficient. This coefficient depends on both the population and the actual sampling design (in particular the second-order inclusion probabilities). For common sampling designs ρ appears to be of the order 1/*N*. Then $(1+(\varsigma(S)-1)\rho)/\varsigma(S)$ tends to 0 when $\varsigma(S)$ and N tend to infinity. Similar considerations apply to the covariance term in (A.8).

Proof of (31): The logarithm of the sample Fisher price index is

$$\ln \hat{P}_F = \frac{1}{2}\Bigg[\ln\left((1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^0\right) - \ln\left((1/\varsigma(S))\sum_{n\in S} p_n^0 q_n^0\right) + \ln\left((1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^1\right) - \ln\left((1/\varsigma(S))\sum_{n\in S} p_n^0 q_n^1\right) \tag{A.10}$$

Using the Taylor series expansion of $f(x) = \ln x$, and assuming simple random sampling, one obtains

$$E\left(\ln (1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^0\right) = \ln\left((1/N)\sum\nolimits_{n=1}^{N} p_n^1 q_n^0\right) + R \tag{A.11}$$

in which the leading term of R has the form

$$-\frac{1}{2}\left(cv((1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^0)\right)^2 \tag{A.12}$$

where *cv(.)* denotes the sample coefficient of variation. Similar expressions hold for the other three parts of the right hand side of Expression (A.10). Hence,

$$E(\ln \hat{P}_F) = \ln P_F + R \tag{A.13}$$

where the leading term of R has the form

$$-\frac{1}{2}\left(cv\left((1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^0\right)\right)^2 + \frac{1}{2}\left(cv\left((1/\varsigma(S))\sum_{n\in S} p_n^0 q_n^0\right)\right)^2 - \frac{1}{2}\left(cv\left((1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^1\right)\right)^2 + \frac{1}{2}\left(cv\left((1/\varsigma(S))\sum_{n\in S} p_n^0 q_n^1\right)\right)^2 \tag{A.14}$$

If all prices change proportionally, that is, $p_n^1 = \alpha p_n^0$ for $n = 1, \ldots, N$ and for a certain $\alpha > 0$, then one easily verifies that the four parts of (A.14) cancel. Moreover, each separate part holds that it approaches zero when the sample size increases towards N.

13. References

Balk, B.M. (1994). On the First Step in the Calculation of a Consumer Price Index. Papers and Final Report of the First Meeting of the International Working Group on Price Indices. Statistics Canada, Ottawa.

Balk, B.M. (1998). On the Use of Unit Value Indices as Consumer Price Subindices. Proceedings of the Fourth Meeting of the International Working Group on Price Indices, U.S. Bureau of Labor Statistics, Washington, DC.

Balk, B.M. (2000). On Curing the CPI's Substitution and New Goods Bias. Research Paper No. 0005, Department of Statistical Methods, Statistics Netherlands.

Bradley, R. (2001). Finite Sample Effects in the Estimation of Substitution Bias in the Consumer Price Index. Journal of Official Statistics, 17, 369–390.

Carruthers, A.G., Sellwood, D.J., and Ward, P.W. (1980). Recent Developments in the Retail Prices Index. The Statistician, 29, 1–32.

CPI Manual (2004). Consumer Price Index Manual: Theory and Practice. Published for ILO, IMF, OECD, UN, Eurostat, and The World Bank by ILO, Geneva.

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

Diewert, W.E. (1995). Axiomatic and Economic Approaches to Elementary Price Indexes. Discussion Paper No. 95-01. Department of Economics, The University of British Columbia, Vancouver, Canada.

Diewert, W.E. (2002). Harmonized Indexes of Consumer Prices: Their Conceptual Foundations. ECB Working Paper No. 130 (www.ecb.int).

Dorfman, A.H., Lent, J., Leaver, S.G., and Wegman, E. (2001). On Sample Survey Designs for Consumer Price Indexes. Invited Paper, 53th Session of the ISI, Seoul, Korea.

Eichhorn, W. and Voeller, J. (1976). Theory of the Price Index. Lecture Notes in Economics and Mathematical Systems 140. Berlin-Heidelberg-New York, Springer-Verlag.

Fisher, I. (1922). The Making of Index Numbers. Boston: Houghton Mifflin.

Knottnerus, P. (2003). Sample Survey Theory: Some Pythagorean Perspectives. New York, Springer Verlag.

Lorenzen, G. (1990). Konsistent Addierbare Relative Änderungen. Allgemeines Statistisches Archiv, 74, 336–344, [In German]

PPI Manual (2004). Producer Price Index Manual: Theory and Practice. Published for ILO, IMF, OECD, UN, The World Bank by IMF, Washington, DC.

Reinsdorf, M. (1994). Letter to the Editor: A New Functional Form for Price Indexes' Elementary Aggregates. Journal of Official Statistics, 10, 103–108.

Silver, M. and Webb, B. (2002). The Measurement of Inflation: Aggregation at the Basic Level. Journal of Economic and Social Measurement, 28, 21–35.

Vartia, Y.O., (1978). Fisher's Five Tined Fork and Other Quantum Theories of Index Numbers. Theory and Applications of Economic Indices, W.Eichhorn et al. (eds), Würzburg: Physica-Verlag.

Received November 2002
Revised January 2005