Sampling Methods in Price Indices for Calculating Elementary Aggregates

Ebrahim Nasirifar

Central Bank of the Islamic Republic of Iran, e-mail: Nasiri_e@yahoo.com

Abstracts

Choosing the appropriate method to calculate the exact price index is important. Use the proportional method of nature of data can reduce the error. At the lowest level of aggregation of price index 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. In this paper, the properties of these indices for calculating elementary aggregates have been studied of point of view various sampling methods.
Key words: elementary price index; sampling survey; Aggregation, quantity

1. Introduction

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 paper we are concerned with what to do when the second of these assumptions is or cannot be fulfilled .There are, of course, various kinds of unavailability of data. The situation we will consider in particular in this paper is that nothing but price data is available for a sample of commodities and/or respondents.
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).

2. Aggregate

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 sub aggregates, 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 allows 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.

Suppose that $\{(p_n^t, q_n^t); n = 1,...N)\}$ is amounts related to commodity price and weight 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 labeled 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 may 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 times periods considered.
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 into a price index and a quantity index:

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

This is traditionally called the index number problem. Both indexes should depend only on the prices and quantities of the two periods.

3. Homogeneity or heterogeneity

If the price index formula is sensitive to add up the quantities q_n^t of the elements n=1…,N state of the "homogeneous" is called. Target price index for the elementary aggregate is the unit value index which is defined as follows:

$$P_U = \frac{\sum_{n=1}^{N} p_n^1 q_n^1 / \sum_{n=1}^{N} q_n^1}{\sum_{n=1}^{N} p_n^o q_n^o / \sum_{n=1}^{N} q_n^o} \qquad (2)$$

that is, the average comparison period price divided by the average base period price. The corresponding quantity index is the simple sum or Dutot index

$$Q_D = \sum_{n=1}^{N} q_n^1 / \sum_{n=1}^{N} q_n^o \qquad (3)$$

If the price index formula isn't sensitive to add up the quantities q_n^t of the elements n=1…, N state of the "heterogeneous" is called, then 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. Three price indexes appear to be particularly relevant. The first is the Törnqvist price index

$$P_T = \prod_{n=1}^{N} (p_n^1 / p_n^o)^{(S_n^o + S_n^1)/2} \qquad (4)$$

Where $s_n^t = p_n^t q_n^t / \sum_{n=1}^{N} p_n^t q_n^t$ (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

$$Q_T = (\sum_{n=1}^{N} p_n^1 q_n^1 / \sum_{n=1}^{N} p_n^o q_n^o) / P_T \qquad (5)$$

The second superlative price index is the Fisher index,

$$P_F = \left(\frac{\sum_{n=1}^{N} p_n^1 q_n^o}{\sum_{n=1}^{N} p_n^o q_n^o} \right)^{1/2} \left(\frac{\sum_{n=1}^{N} p_n^1 q_n^1}{\sum_{n=1}^{N} p_n^o q_n^1} \right)^{1/2} = (P_L P_P)^{1/2} \qquad (6)$$

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

$$Q_F = \left(\frac{\sum_{n=1}^{N} p_n^o q_n^1}{\sum_{n=1}^{N} p_n^o q_n^o}\right)^{1/2} \left(\frac{\sum_{n=1}^{N} p_n^1 q_n^1}{\sum_{n=1}^{N} p_n^1 q_n^o}\right)^{1/2} = (Q_L Q_P)^{1/2} \qquad (7)$$

That is the geometric average of the Laspeyres and the Paasche quantity indexes.
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 maximum he or she can obtain is information $\{p_n^o, p_n^1, q_n^o, q_n^1; n \in S\}$ for a sample $S \subset \{1,..., N\}$.More realistic, however, is the situation where the information set has the form $S \subset \{1,..., N\}$, that is, nothing but 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 the sample *S* is drawn usually remains hidden in certain 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 assumptions about the sampling design.
In the remainder of this paper 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 a simple random sample, 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 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 were drawn with probability proportional to size, where size denotes some measure of importance. If the size of element *n* is denoted by a positive scalar *n a* (*n*=1,…,*N*), then the probability that element *n* is included in the sample S is $\Pr(n \in S) = \varsigma(S) a_n / \sum_{n=1}^{N} a_n$.
Now if it is not possible to access all the information on prices and quantities, the price indices P_U, P_T, P_F will be studied as parameters. All these indices are based on data from the two periods 0 and 1. The parameter estimators obtained based on the data samples and the performance of the sampling method is investigated.

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 $\sum_{n=1}^{N} p_n^o q_n^o$ 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 is to estimate the Dutot quantity index Q_D . A likely candidate is its sample counterpart

$$\hat{Q}_D = \sum_{n \in S} q_n^1 / \sum_{n \in S} q_n^o \qquad (8)$$

Suppose that S is a simple random sample and recall that the inclusion probabilities are $\Pr(n \in S) = \varsigma(S)/N$, where $\varsigma(S)$ denotes the sample size. Then the expected value of the sample Dutot quantity index is

$$E(\hat{Q}_D) \approx Q_D \qquad (9)$$

Expression (9) means that $\hat{Q}_D$ is an approximately unbiased estimator of the population Dutot quantity index Q_D . The bias is of technical nature and will approach zero when the sample size gets larger.
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 $q_n^o / \sum_{n=1}^{N} q_n^o \ (n = 1,..., N)$. Thus the probability that element n is included in the sample is equal to $\Pr(n \in S) = \varsigma(S) q_n^o / \sum_{n=1}^{N} q_n^o$ .Then the expected value of the sample Carli quantity index is equal to

$$E(\hat{Q}_C) = \frac{1}{\varsigma(S)} \sum_{n=1}^{N} (q_n^1 / q_n^o) \Pr(n \in S) = \sum_{n=1}^{N} (q_n^o / \sum_{n=1}^{N} q_n^o)(q_n^1 / q_n^o) = Q_D \qquad (10)$$

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

$$\hat{P}_U = \frac{\sum_{n \in S} p_n^1 q_n^1 / \sum_{n \in S} q_n^1}{\sum_{n \in S} p_n^o q_n^o / \sum_{n \in S} q_n^o} \qquad (11)$$

This presupposes that the sample is of the form $\{(p_n^t q_n^t, q_n^t); t = 0,1; n \in S\}$, that is, for every sampled element one disposes of its value and its quantity in both periods. Then one can show, in much the same way as was done in expression (9) that under simple random sampling the sample unit value index is an approximately unbiased estimator of the target unit value index P_U .
Suppose next that only sample prices are available, that is, the sample is of the form $\{(p_n^o, p_n^1); n \in S\}$, and consider the sample Dutot price index, defined as

$$\hat{P}_D = \frac{\sum_{n \in S} P_n^1}{\sum_{n \in S} P_n^o} = \frac{(1/\varsigma(S)) \sum_{n \in S} P_n^1}{(1/\varsigma(S)) \sum_{n \in S} P_n^o} \qquad (12)$$

Under probability proportional to size sampling, whereby again the size of element n is defined as its base period quantity share, it is easily seen that, apart from a technical bias,

$$E(\hat{P}_D) \approx \frac{\sum_{n=1}^{N} p_n^1 q_n^o / \sum_{n=1}^{N} q_n^o}{\sum_{n=1}^{N} p_n^o q_n^o / \sum_{n=1}^{N} q_n^o} \qquad (13)$$

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

$$\frac{E(\hat{P}_D)}{P_U} \approx \frac{\sum_{n=1}^{N} p_n^1 q_n^o / \sum_{n=1}^{N} q_n^o}{\sum_{n=1}^{N} p_n^1 q_n^1 / \sum_{n=1}^{N} q_n^1} \qquad (14)$$

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 which is independent of the sample size. This structural part is given by the right hand side of expression (14). 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.

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 to be the target and consider its sample analogue

$$\hat{P}_T = \prod_{n \in S} (p_n^1 / p_n^o)^{(\hat{S}_n^o + \hat{S}_n^1)/2} \qquad (15)$$

Where $\hat{s}_n^t = p_n^t q_n^t / \sum_{n \in S} p_n^t q_n^t \ (t = 0,1)$ is element *n*'s sample value share. Prove that Balk (2000),

$$E(\ln \hat{P}_T) \approx \ln P_T \qquad (16)$$

This means that $\ln \hat{P}_T$ is an approximately unbiased6 estimator of $\ln P_T$. Employing Jensen's Inequality, one obtains

$$E(\hat{P}_T) \geq P_T \qquad (17)$$

That is, the sample Törnqvist price index has an upward bias relative to its population counterpart. However, this bias is of technical nature and will approach zero when the sample size gets larger.

The previous result critically depends on the availability of sample quantity or value information. Suppose that we cannot obtain these data and consider the sample Jevons price index

$$\hat{P}_J = \prod_{n \in S} (p_n^1 / p_n^o)^{1/(\varsigma(S)} \qquad (18)$$

Under probability proportional to size sampling, whereby the size of element *n* is now defined as its base period value share s_n^o, it is easily seen that

$$E(\ln \hat{P}_J) = \sum_{n=1}^{N} s_n^o \ln(p_n^1 / p_n^o) = \ln\left(\prod_{n=1}^{N} (p_n^1 / p_n^o)^{s_n^o}\right) \qquad (19)$$

By employing Jensen's Inequality, this leads to the result that

$$E(\hat{P}_J) \geq \left(\prod_{n=1}^{N} (p_n^1 / p_n^o)^{s_n^o}\right) = P_{GL} \qquad (20)$$

At the right hand side 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

$$\frac{E(\hat{P}_J)}{P_T} \geq \prod_{n=1}^{N} (p_n^1 / p_n^o)^{(s_n^o - s_n^1)/2} (1 + R) \qquad (21)$$

Instead of defining the size of element *n* as its base period value share s_n^o, one could as well define its size as being $(s_n^o + s_n^1)/2$, the arithmetic mean of its base and comparison period value share. Then we obtain, instead of (20),

$$E(\hat{P}_J) \geq \prod_{n=1}^{N} (p_n^1 / p_n^o)^{(s_n^o + s_n^1)/2} = P_T \quad (22)$$

which means that the sample Jevons price index is an approximately unbiased estimator of the population Törnqvist price index. The bias will now vanish when the sample size gets larger. This result was mentioned by Diewert (1995).

6. Heterogeneous aggregates and the Fisher price index

Suppose that instead of the Törnqvist price index one has decided that the Fisher price index (6) 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

$$\hat{P}_F = \left(\frac{\sum_{n\in S} p_n^1 q_n^o \sum_{n\in S} p_n^1 q_n^1}{\sum_{n\in s} p_n^o q_n^o \sum_{n\in s} p_n^o q_n^1} \right)^{1/2} = \left(\frac{(1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^o (1/\varsigma(S))\sum_{n\in S} p_n^1 q_n^1}{(1/\varsigma(S))\sum_{n\in s} p_n^o q_n^o (1/\varsigma(S))\sum_{n\in s} p_n^o q_n^1} \right)^{1/2} \quad (23)$$

Prove that Balk (2000),

$$E(\ln \hat{P}_F) = \ln(P_F) \quad (24)$$

Again applying Jensen's Inequality, we see that

$$E(\hat{P}_F) \geq P_F \quad (25)$$

which means that under simple random sampling the sample Fisher price index has an upward bias relative to its population counterpart. This bias, however, will approach zero when the sample size gets larger.

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

$$\hat{P}_C = \frac{1}{\varsigma(S)} \sum_{n\in S} (p_n^1 / p_n^o) \quad (26)$$

Under probability proportional to size sampling, whereby the size of element *n* is defined as its base period value share s_n^o, we immediately see that

$$E(\hat{P}_C) = \sum_{n=1}^{N} s_n^o (p_n^1 / p_n^o) = \frac{\sum_{n=1}^{N} p_n^1 q_n^o}{\sum_{n=1}^{N} p_n^o q_n^o} = P_L \quad (27)$$

Thus the expected value of the sample Carli price index appears to be equal to the population Laspeyres price index. This result was already mentioned by Balk (1994); see also Diewert (1995). The relative bias of the sample Carli price index with respect to the population Fisher price index appears to be

$$\frac{E(\hat{P}_C)}{P_F} = \frac{P_L}{P_F} = \left(\frac{P_L}{P_P}\right)^{1/2} \quad (28)$$

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

7. Conclusion

In this paper for calculating the elementary aggregate relationship between target price index, sample price index and sampling methods was investigated. 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) (pps).
The following table presents the key results in the order of their appearance.

Sample price index	**Target price index**	**Sampling method**	**Expected value of sample index**
Unit value	Unit value	srs	Unit value
Dutot	Unit value	pps-qo	Biased estimate of target index
Törnqvist	Törnqvist	srs	Törnqvist
Jevons	Törnqvist	pps-so	Geometric Laspeyres
Fisher	Fisher	srs	Fisher
Carli	Fisher	pps-so	Laspeyres

References

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

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

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

4- Dalen, J. (1992) “Computing Elementary Aggregates in the Swedish Consumer Price Index”; ***Journal of Official Statistics***, 8, 129–147.

5- 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.

6- 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.

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

8- PPI Manual. (2004) “*Producer Price Index Manual”: Theory and Practice*, Published for ILO, IMF, OECD, UN, the World Bank by IMF, Washington, DC;

9- Reinsdorf, M, Letter to the Editor. (1994) “A New Functional Form for Price Indexes’ Elementary Aggregates”; ***Journal of Official Statistics***, 10, 103–108.