PRICE INDEXES AND SAMPLING

By ERLAND v. HOFSTEN

Statistical Section, National Social Welfare Board, Stockholm
and
Indian Statistical Institute, Calcutta

SUMMARY : Some practical difficulties of obtaining sampling errors for price index numbers have been discussed in this note.

In recent years it has become more and more widely accepted that statistical estimates should be accompanied by error estimates, and that information about the statistical error can only be obtained, if the estimate is based on a probability sample. There is however, one exception from this general rule, namely the field of index numbers, where very little is known and stated about the precision of the computed figures. However, two authors have recently given solutions of standard errors for price index numbers. In the view of the present author a nearer scrutiny of this aspect of the problem demonstrates that there is a certain inescapable controversy and inconsistency as regards price index numbers.

The two authors are Banerjee (1956) and Adelman (1958). Banerjee points out that prices are normally collected only for a few of the items to be covered by the index. He then gives the unbiased estimate of the index as well as the variance. Banerjee's solution assumes that only two points of time 0 and 1 are compared and that Laspeyres' formula is used. This implies that prices are assumed available at period 1 for all articles which were available at period 0 and that new items are ignored.

Adelman completely overlooks Banerjee's paper, although it was published in a widely circulated journal two years earlier. She does not specify the index formula used; her index concept is one of price relatives between two periods not too far apart, and she states that the weights applied must not be out of date. For comparisons over longer intervals she arrives at the chain index solution.

To start with let us consider the problem of comparing two periods only. During the first period we have, with usual notations, certain quantities, q_0, and prices, p_0, of all articles on the market.

In order to take a probability sample we must define the universe properly and contruct a frame, from which to draw the sample. If we consider *one* period only, our universe may consist of all the purchases which have taken place during the period (possibly purchases by some properly delimited population category, such as working class families, etc.). A sampling frame should consist of all these purchases; in this connection we ignore the practical difficulties to obtain such a frame.

It is important to note that not only the quantities but also the prices refer to a period and that the prices are "paid prices" not "demanded prices." This distinction is of importance, not so much because of bargaining, discounts, sales, etc., but because of the definition of the universe.

It is not quite clear which definition Irma Adelman employs. Her statement "since loss leaders and similar sub-normal price situations often exist on Thursday, Friday and Saturday, all the pricing was done during the early part of the week (p.245)" seems to imply that she uses the demanded price as definition.

For the index computation we cannot be satisfied with having a sample referring to one period only; the index implies a *comparison* between at least two periods. If we accept the Laspeyres' solution, this implies that we base the sample of items on the conditions prevailing during period 0 (= quantities purchased and prices actually paid). For period 1 we will want to ascertain the amount of money required in order to buy the same quantities in the new price situation. This implies a hypothetical question. Thus if an item is available but not at all purchased in situation 1, it will nevertheless enter into the index computation. Prices for situation 1 will not be paid prices but demanded prices, and it will not be possible for situation 1 to construct any sampling frame, which corresponds to the one in situation 0. This is not very satisfactory, but is a necessary consequence of the Laspeyres' approach.

The Paasche index, implying the reverse of the Laspeyres' index, of course does not solve the problem.

The indifference defined index also tries to give an answer to a hypothetical question, i.e., what is the amount of money required to attain an unchanged indifference level (= generally less amount of money than required for the Laspeyres' solution)? Information about the actual position of the consumer in situation 1, does not solve the problem if the index is based on the indifference level in situation 0.

It remains to be seen whether a universe can be defined where "paid prices" can be used for both periods 0 and 1. But then only articles actually bought during both periods 0 and 1 will be included, because no price relatives can be formed for items purchased only during one of the periods. And what about the weights, shall they be an average for the two periods and then why? This solution will be rather vague and unsatisfactory.

And finally, is it possible to envisage "demanded prices" as the price definition right through? Clearly not, in any case it seems difficult to find any universe then, and where do the quantities come in?

The differences between demanded prices and actually paid prices is also very clear, if we consider the problem raised in Stone (1956), when the price per unit is a function of the quantity purchased. This is often the case for electricity and telephone charges, where a basic payment is made, but also occurs regarding other items of expenditure, where bulk purchases may lead to a lower price per unit. If the Laspeyres' solution is employed, the index will not show any change, as not more money is required in order to keep the consumption pattern unaltered. But if paid prices are used, account must be taken of the price change which has been a consequence of the altered consumption.

The problems discussed above refer to the fact that the universe is changing. Such changes are most marked as regards clothing and so-called miscellaneous items, whereas they are less marked for food items. Incidentally most authors on index numbers overlook these problems, because they choose examples among food items. However, looking upon the budget as a whole, the problems of the changing universe are severe; for evidence see Hofsten (1952).

If the periods compared are near each other, it is tempting to state that then the changes of the universe must be so small that they can be overlooked. In order to make possible comparisons over long intervals, we must then resort to the chain index solution.

The chain index implies an integral solution of the index problem [cf. Divisia (1925).] If this solution is chosen, then *it is necessary that the infinitesimal expression for the index, i.e. the index for each separate link, is correct*. If this is not the case, the chain index only implies a comfortable technique, by which the intrinsic problems of comparing two distant periods are avoided. Incidentally the chain index solution is not available for geographical comparisons, at least not between different countries.

Adelman states about the chain index that "the ease with which new products or qualities can be incorporated into our scheme (and obsolete items eliminated) provides a significant advantage over the current system" (p.243). This advantage, in my mind implies a great danger, because it violates the principle that the infinitesimal expression for the index must be correct.

There is one additional problem of a partly practical character. A computation of a standard error for an index will in the first hand refer to a comparison between two periods only [as in Banerjee (1956) and Adelman (1958)]. But in actual practice indexes are most often given in the form of long regular series. If the series is computed as a chain index, what standard error formula shall then be used? And as the consumer will desire to compare any single index figure with any other figure, what standard errors shall be given?

My conclusion from the above arguments is that there is no such thing as a statistical precision for a price index. Attempts to define the index in a statistical way, applying modern theory of sampling, only demonstrate that there is no satisfactory solution available. We may, therefore, just as well keep to the old practice and define the price index in an operational way and abstain from giving standard errors. This, of course, does not exclude the usefulness of applying the chain index solution or of basing the selection of items on probability sampling and making analyses of the precision of price measurements. But when applying the chain index solution we must not allow the substitution of some items against others without making quality adjustments; see Hofsten (1952) and Stone (1956).

References

ADELMAN, IRMA (1958): A new approach to the construction of index numbers. *The Review of Economics and Statistics*, XL, No. 3, 240-249.

BANERJEE, K. S. (1956): A Note on the optimal allocation of consumptive items in the construction of cost of living index. *Econometrica*, 24, 294-295.

DIVISIA, FRANCOIS (1925): L' indice monetaire et la theorie de la monnaie. *Revue d'economie politique*, Paris.

HOFSTEN, E. v., (1952): *Price Indexes and Quality Changes*, Allen & Union, Stockholm and London.

STONE, R., (1956): *Quantity and Price Indexes in National Accounts*. OEEC, Paris.

Paper received : April, 1959.

CORRIGENDA

Bias in Estimation of Serial Correlation Coefficients : By A. Sree Rama Sastry, *Sankhyā*, **11**, 281-296.

Formula (11) on page 283 should be read

$$g_k = -\frac{\sum_{i=1}^{T-k-1} (T-k-i) \left\{\mu_2(k+i)+\mu_2(|k-i|) - \frac{2\mu_2(k)\mu_2(i)}{\mu_2(0)}\right\}}{(T-k-1)(T-k)\mu_2(0)-2\sum_{i=1}^{T-k-1}(T-k-i)\mu_2(i)} \quad \ldots \quad (11)$$

The author wishes to thank Mr. E. G. Phadia for pointing out the printing error.

Expressions for The Lower Bound to Confidence Coefficients : By Saibal Kumar Banerjee, *Sankhyā*, **21**, 127—140.

1. $\frac{t\hat{\lambda}}{n}$ occurring in (i) expression (2.4.3), (ii) first sentence of para 2.5 and (iii) table heading of Table 1, all at page 129, should be read as $t\sqrt{\frac{\hat{\lambda}}{n}}$

2. B_2 occurring in para 3.7, page 132, is

$$\bar{B}_2 = \frac{\sum_1^k \lambda_i^2 B_{2i}}{\sum_1^k \lambda_i^2}$$

ACKNOWLEDGEMENT

Thanks are due to the following for their kind help in the editing of the papers published in this issue.

Dr. R. R. Bahadur	Professor D. B. Lahiri
Dr. K. S. Banerjee	Dr. A. Matthai
Dr. D. Basu	Dr. Sujit Kumar Mitra
Dr. I. M. Chakravarti	Dr. C. R. Rao
Dr. E. v. Hofsten	Dr. J. Roy