Volume Title: The Price Statistics of the Federal Goverment

Volume Author/Editor: Report of the Price Statistics Review Committee

Volume Publisher: NBER

Volume ISBN: 0-87014-072-8

Volume URL: http://www.nber.org/books/repo61-1

Publication Date: 1961

Chapter Title: Sampling Considerations in the Construction of Price Indexes with Particular Reference to the United States Consumer Price Index

Chapter Author: Philip J. McCarthy

Chapter URL: http://www.nber.org/chapters/c6493

Chapter pages in book: (p. 197 - 232)

STAFF PAPER 4

SAMPLING CONSIDERATIONS IN THE CONSTRUCTION OF PRICE INDEXES WITH PARTICULAR REFERENCE TO THE UNITED STATES CONSUMER PRICE INDEX

Philip J. McCarthy, Cornell University

I. Introduction

Theoretical discussions of price indexes have been concerned primarily with an economic approach to problems arising out of such questions as:

1. Is it a price index or a cost-of-living index that is needed?
2. How should a price index or a cost-of-living index be constructed? (Indifference curve approach, Laspeyres, Paasche, etc., and the relations between these various forms.)
3. How does one choose a base period and a period in which weights are determined?
4. What methods are to be used in dealing with quality changes, with the disappearance of old items and the appearance of new items, and with related problems?
5. How does one aggregate over consumers, over cities, over States, and so on?

The problems suggested by these questions exist whether one is dealing with "samples" or with complete sets of data, and an excellent summary of the literature on many of these problems has been given by von Hofsten (*1*, Chap. 13).[1]

Leaving aside for the moment the problems of sampling, most index numbers start from some fixed formula. That is, given a single consumer and the complete universe of prices and quantities, some basic way of computing the index is chosen. For example,

$$R^{(1)}=\frac{\Sigma q_{i0}p_{i1}}{\Sigma q_{i0}p_{i0}}$$

is the Laspeyres formula which serves as a model for practically all of the currently computed price indexes. In this formula q_{i0} represents the quantity of commodity i consumed in the base period, designated as time zero, p_{i0} represents the price per unit of this commodity at time zero, and p_{i1} represents the price of the same commodity at time one. Thus $R^{(1)}$ is the Laspeyres index for time one with time zero as base. Presumably the principal reason for choosing the Laspeyres form is that it uses base quantity weights (which are all that are usually available), and can therefore be easily explained as the relative cost of a fixed market basket of goods and services (see Jaffe, *2*, p. 7). Even after deciding that a price index, as opposed to a cost-

[1] Italic numbers in parentheses refer to bibliography at end of paper.

of-living index, is needed and that a fixed formula such as Laspeyres can be taken as a starting point, the theoretical and practical problems mentioned under points (3), (4), and (5) above still remain. Comments on some of these problems as they influence sampling considerations will be made later in this report.

Because many individuals seem to desire a cost-of-living index rather than a price index, and because it does not appear possible to translate the indifference curve approach into a form that has practical applications, there have been attempts to develop formal statistical and economic models that will provide cost-of-living indexes. Thus Stone (*3*) describes a linear expenditure system for consumers, and gives references to papers that show how this system can lead to a cost-of-living index; Brady and Hurwitz (*4*) refer to additive or multiplicative models that have been used to explore the international comparisons of food costs; and Neiswanger (*5*) recommends the use of varying weights in the Laspeyres formula, the weights being determined from current prices and the estimated direct and cross elasticities between commodities included in the index. In these situations, one would be using data (even for the entire universe of consumers and of prices and quantities) to estimate parameters of a model rather than for direct substitution into a formula. Model considerations would then provide the index.

No matter how one chooses to resolve the questions and problems that have thus far been mentioned, the actual construction of a price index will always be based upon samples of data rather than upon data derived from complete enumerations of the pertinent universe. The quality of the index will, therefore, depend to some extent upon how these samples are obtained. This fact has long been recognized, but the sampling aspects of index number construction have never been accorded the attention that has been devoted to their economic aspects. The purpose of this report is to present some comments and observations concerning the sampling problems that arise in the computation of the more or less traditional Laspeyres-oriented indexes, with special reference to the United States Consumer Price Index.

II. A Brief Description of Where Sampling Enters the Consumer Price Index

There are clearly many points at which sampling is used in the determination of a value of the Consumer Price Index. These are:

A. DETERMINATION OF THE ITEM WEIGHTS (CONSUMER EXPENDITURE SURVEYS)

1. Selection of points or intervals in time at which consumer expenditure surveys will be made.
2. Selection of a sample of cities in the United States.
3. Selection of a sample of consuming units in the selected cities.

B. SELECTION OF THE SAMPLE OF CITIES FOR THE INDEX

1. This sample is customarily selected from among the cities used in the Consumer Expenditure Surveys, although other index designs would be possible. Thus one could impute weights to cities not originally included in the consumer expenditure survey.

2. In making up a national index, the individual city results are evidently averaged with weights proportional to both the city wage-earner and clerical-worker population that each city represents and to the value of the city market basket in the base period.

C. SELECTION OF A SAMPLE OF ITEMS OF EXPENDITURE THAT IS TO BE PRICED IN COMPUTING THE INDEX

1. The items of expenditure are divided into major groups (food, housing, apparel, etc.), then into subgroups, sub-subgroups, etc.
2. Ultimately the subdivision process leads to a group of items that are "somewhat similar." One or more specific items are selected to represent this subdivision.
3. Finally, one or more "specified-in-detail" items are chosen to represent the specific items and also represent the "somewhat similar" group of items.

D. DETERMINATION OF THE POINTS IN TIME AT WHICH PRICE QUOTATIONS FOR THE "SPECIFIED-IN-DETAIL" ITEMS ARE TO BE OBTAINED

1. The index is published monthly. Thus one or more points in time during the month must be chosen for pricing.
2. Some items in some cities are not priced each month, and so a scheme for sampling months, etc., is integrated into the program.

E. SELECTION OF A SAMPLE OF PRICE REPORTERS FROM WHOM PRICE QUOTATIONS ARE OBTAINED

1. The sampling problem here varies with the item. Thus samples of families are used for rent reports, samples of stores for food reports, etc.

It will not be possible in this brief report to examine each of these components in the overall sampling design of the Consumer Price Index, or even to list the other possible sources of random variation that may to some extent influence a reported value of the index (e.g., variability among enumerators, random errors in the reporting of prices, and the like). Rather, attention will be focused primarily on the sampling of items of expenditure, since this is an area in which there appears to have been a great deal of controversy, and on the problem of measuring the total sampling error of the Consumer Price Index. It should be noted that a detailed description of the methodology of the 1950 Survey of Consumer Expenditures has been provided by Lamale (*6*).

III. The Computation of the Consumer Price Index

Once the individual price quotations have been obtained for the "specified-in-detail" items, the next problem is that of combining these into the various desired indexes. At this point we shall only be concerned with the computational details of this combination process. The relationship of this computational procedure to the design used in selecting a sample of index cities will be discussed in Section VI.

Suppose we have c index cities. Let the population (city wage-earner and clerical-worker population) weight assigned to the i-th city be W_i, $i=2, \ldots, c$, where $\Sigma W_i = 1$. Within the i-th city, let the base weight assigned to a specific item, but not a "specified-in-detail" item, be ${}_iw_{jklm}^{(0)}$ . . ., where i represents the city, the superscript (0) represents the base time period, and the subscripts j, k, l, m, etc., represent successively finer classifications or subdivisions of the items. Thus j represents major expenditure groups (food, housing, apparel, etc.), k represents a subgroup within a major expenditure group, and so on until the individual items are reached. Thus a man's nylon business shirt might be a specific item, although it would not yet be a "specified-in-detail" item for which prices would be obtained. The number of subgroup classifications depends upon j. The quantities ${}_iw_{jklm}^{(0)}$ . . . are relative weights so that $\sum_{j,k,\ldots i} {}_iw_{jklm}^{(0)} \ldots = 1$ for each i.

These weights for specific items are essentially relative expenditure weights. That is, they can be viewed as the average proportion of total expenditures, for consumer expenditure survey families in a city, that was spent on these specific items in the base period. (Actually, it is very difficult to determine the exact steps that were followed in arriving at the weights since published accounts—e.g., BLS Bulletin 1168 (*8*, p. 3)—are not very explicit on the details of the procedure.) Once weights have been determined for specific items, weights for the various classes of the subdivision process are obtained by summing these item weights. Thus we have ${}_iw_j{}^{(0)}$, ${}_iw_{jk}{}^{(0)}$, etc.

For the present purposes, let us assume that ${}_iw_{jklm}^{(0)}$ represents the weight for a specified item or for a "price family" of related items. There can be more or fewer subscripts for a given j, but assuming an equal number of subgroups for each major group makes notation easier and the general argument goes through no matter what the number of subscripts. Then a single specified-in-detail item is selected out of the jkl-th subdivision. Denote this sampled item by $jklx$.

For item $jklx$, in city i, at time t, a number of price quotations are obtained by sampling households, food stores, etc. Call these price quotations ${}_ip_{jklx_1}^{(t)}$, ${}_ip_{jklx_2}^{(t)}$, ... By some appropriate averaging process, we end up with an average price ${}_i\bar{p}_{jklx}^{(t)}$. Then the price relative for this specified-in-detail item is:

$$ {}_iR_{jklx}^{(t)} = \frac{{}_i\bar{p}_{jklx}^{(t)}}{{}_i\bar{p}_{jklx}^{(0)}} $$

Not only does this procedure give a price relative for item $jklx$, but it also gives an index for the jkl-th subgroup since item x was selected to represent the entire group. That is, ${}_iR_{jklx}^{(t)}$ is taken as an approximation to:

$$ {}_iR_{jkl}^{(t)} = \frac{\sum_m {}_iw_{jklm}^{(0)} \frac{{}_i\bar{p}_{jklm}^{(t)}}{{}_i\bar{p}_{jklm}^{(0)}}}{{}_iw_{jkl}^{(0)}} $$

where the summation is taken over all specified-in-detail items contained in the jkl-th subgroup. Note that this is essentially of Laspeyres form, since, approximately,

$$ {}_iw_{jklm}^{(0)} = K \cdot {}_i\bar{q}_{jklm}^{(0)} \cdot {}_i\bar{p}_{jklm}^{(0)} $$

Actually, the expenditure surveys do not obtain individual quantities for specified-in-detail items. Rather, quantities are determined for specific items, and these are translated into the fraction of an average family's expenditures that goes for each of these specific items. Of course, base period prices must be obtained for the specified-in-detail items that go into the index.

If one now wishes to make up a city index for the jk-th subgroup, or for all items combined, the following procedure is used:

$$ {}_tR^{(t)}_{jk} = \frac{\sum_{l} {}_tw^{(0)}_{jkl} \frac{{}_t\bar{p}^{(t)}_{jklx}}{{}_t\bar{p}^{(0)}_{jklx}}}{{}_tw^{(0)}_{jk}} $$

$$ {}_tR^{(t)} = \sum_{j,k,l} {}_tw^{(0)}_{jkl} \frac{{}_t\bar{p}^{(t)}_{jklx}}{{}_t\bar{p}^{(0)}_{jklx}} $$

Note that no divisor is required for the all-item index since the ${}_tw^{(0)}_{jkl}$ sum to one when all major groups, subgroups, etc., are considered.

These expressions for city indexes are often expressed in terms of the current relatives,

$$ \frac{{}_t\bar{p}^{(t)}_{jklx}}{{}_t\bar{p}^{(t-1)}_{jklx}} $$

Thus algebraic manipulation leads to:

$$ {}_tR^{(t)} = {}_tR^{(t-1)} \sum_{j,k,l} {}_tw^{(t-1)}_{jkl} \frac{{}_t\bar{p}^{(t)}_{jklx}}{{}_t\bar{p}^{(t-1)}_{jklx}} $$

where

$$ {}_tw^{(t-1)}_{jkl} = \frac{{}_tw^{(0)}_{jkl} \frac{{}_t\bar{p}^{(t-1)}_{jklx}}{{}_t\bar{p}^{(0)}_{jklx}}}{\sum_{j,k,l} {}_tw^{(0)}_{jkl} \frac{{}_t\bar{p}^{(t-1)}_{jklx}}{{}_t\bar{p}^{(0)}_{jklx}}} $$

The quantities ${}_tw^{(t-1)}_{jkl}$, which also sum to one when all major groups, subgroups, etc., are considered, are what are ordinarily referred to as "current value" or "current importance" weights. They are, of course, only approximations to value weights because of the previously noted restrictions on ${}_tw^{(0)}_{jkl}$, and because a single item, here denoted by x, is used to represent all items in the class jkl. They can also be expressed in terms of current relatives and the current value weights for the preceding time period.

The foregoing would appear to be a faithful representation of the computation of Consumer Price Indexes at the city level. An alternative computational procedure that uses "hypothetical" base quantities is sometimes described in the literature (e.g., 7). That is, consumer expenditure survey data would be used to determine an average dollar expenditure for items in group jkl. Call this value ${}_t\bar{v}^{(0)}_{jkl}$. If the average price of item $jklx$ at time zero is, as before, ${}_t\bar{p}^{(0)}_{jklx}$, then a hypothetical base quantity can be obtained to associate with item *jklx* that will account for all expenditures in group *jkl*. That is:

$$ {}_i\bar{q}^{(0)}_{jklx} = \frac{{}_i\bar{v}_{jkl}}{{}_i\bar{p}^{(0)}_{jklx}} $$

Then we would have, for example,

$$ {}_iR^{(t)} = \frac{\sum\limits_{j,k,l} {}_i\bar{q}^{(0)}_{jklx}\, {}_i\bar{p}^{(t)}_{jklx}}{\sum\limits_{j,k,l} {}_i\bar{q}^{(0)}_{jklx}\, {}_i\bar{p}^{(0)}_{jklx}} $$

where the hypothetical quantities now appear explicitly. This expression can also be written in terms of price relatives and the index for the preceding time period.

The final problem in the computation of the Consumer Price Index is that of describing how the city price relatives and city indexes are combined into United States indexes. Here the published accounts of the Bureau leave much to be desired. The most detailed published account appears in BLS Bulletin 1168 (*8*) and consists of the following sentences:

> Weighting of price relatives to calculate the average price change for groups of goods and services and for all items combined is carried out for each city separately. In combining the cities into the United States all city index, each city is given an importance or weight proportionate to the wage-earner and clerical-worker population it represents in the index. . . . The importance of cities in the index is now based on the Census figures for 1950. As new Census population figures become available, the Bureau will adjust the city weights accordingly.
>
> In the actual calculation of the index, population and expenditure weights are combined, so that index value weights are the product of three factors—base year quantities, population, and current prices. Aggregates for the United States index can therefore be calculated by a simple summation of value weights for the individual cities.

There would appear to be many possible interpretations of this quotation, but the general descriptive material given by Mudgett (*7*) and a passing comment (in *9*, p. 129) would suggest that the following procedure is used for obtaining the United States all-item index. Using the hypothetical quantity form for ${}_iR^{(t)}$, we define

$$ R^{(t)} = \frac{\sum\limits_i W_i \sum\limits_{j,k,l} {}_i\bar{q}^{(0)}_{jklx}\, {}_i\bar{p}^{(t)}_{jklx}}{\sum\limits_i W_i \sum\limits_{j,k,l} {}_i\bar{q}^{(0)}_{jklx}\, {}_i\bar{p}^{(0)}_{jklx}} $$

Thinking of the W_i as actual population figures, we see that this is essentially the value of the contents of all consumers' market baskets (the contents of the basket differ from city to city) at time *t* relative to the value at time zero. This expression can be rewritten to show its relation to the individual city indexes in the following manner:

$$R^{(t)} = \sum_i W'_{i\,i}R^{(t)}$$

$$= \sum_i W'_i \sum_{j,k,l} {}_iw^{(0)}_{jklx} \frac{{}_i\bar{p}^{(t)}_{jklx}}{{}_i\bar{p}^{(0)}_{jklx}}$$

where

$$W'_i = \frac{W_i \sum_{j,k,l} {}_i\bar{q}^{(0)}_{jklx}\, {}_i\bar{p}^{(0)}_{jklx}}{\sum_i W_i \sum_{j,k,l} {}_i\bar{q}^{(0)}_{jklx}\, {}_i\bar{p}^{(0)}_{jklx}}$$

Thus a city index is weighted in proportion to the population it represents and to the value of the city market basket in the base period. It is assumed that the W'_i are what are referred to as "relative cost-population weights Dec. 1952" in Table 1 of BLS Bulletin 1168 (*8*). This behavior of a Laspeyres index, where the aggregation is over individual consumers rather than over cities, has been noted by von Hofsten (*1*, p. 123).

The foregoing expression for the all-item United States Consumer Price Index can also be written in terms of average U.S. prices and quantities if we define

$$\bar{q}^{(0)}_{jklx} = \sum_i W_i\, {}_i\bar{q}^{(0)}_{jklx}$$

and

$$\bar{p}^{(0)}_{jklx} = \sum_i \left(\frac{W_i\, {}_i\bar{q}^{(0)}_{jklx}}{\sum_i W_i\, {}_i\bar{q}^{(0)}_{jklx}} \right) {}_i\bar{p}^{(0)}_{jklx}$$

$$\bar{p}^{(t)}_{jklx} = \sum_i \left(\frac{W_i\, {}_i\bar{q}^{(0)}_{jklx}}{\sum_i W_i\, {}_i\bar{q}^{(0)}_{jklx}} \right) {}_i\bar{p}^{(t)}_{jklx}$$

Then,

$$R^{(t)} = \frac{\sum_{j,k,l} \bar{q}^{(0)}_{jklx}\, \bar{p}^{(t)}_{jklx}}{\sum_{j,k,l} \bar{q}^{(0)}_{jklx}\, \bar{p}^{(0)}_{jklx}}$$

$$= \sum_{j,k,l} \left(\sum_i W'_i\, {}_iw^{(0)}_{jklx} \right) \frac{\bar{p}^{(t)}_{jklx}}{\bar{p}^{(0)}_{jklx}}$$

where W'_i is as previously defined. The quantity $\sum_i W'_i\, {}_iw^{(0)}_{jklx}$ is the relative importance of item *jklx* in the national all-item index. Thus this national relative importance figure is merely a weighted average of the city relative importance figures, where the city weights are proportional to the city population and to the value of the city market basket in the base period.

It should be observed that this last form of the national CPI, in terms of average U.S. prices and quantities, is essentially the form used by the Agricultural Marketing Service in computing the Index of Prices Paid by Farmers (*10* and *11*). However, the following differences between the CPI and the Living Component of this index might be noted for present purposes:

1. The 1956 Farm Expenditure Survey, on which the weighting pattern of the 1959 revision of the index was based, covered a national sample of farmers located in 306 primary sampling units (counties or pseudo-counties) and the results were expanded to U.S. totals on a basis representing all farms. Similarly, the AMS obtains price reports widely throughout the United States. Thus the BLS emphasis on a relatively small sample of cities, for many of which a city index is actually published, does not have its counterpart in the Index of Prices Paid by Farmers.

2. In obtaining an average U.S. price, an unweighted average price for a commodity is computed for each State (*10*, p. 36) and a weighted average of the state values then gives the national average. Apparently the state average prices are weighted by estimates of *current* purchases. This is at variance with the practice already described for the CPI where city average prices are essentially weighted by estimates of *base year* quantities.

3. The AMS does not price a "specified-in-detail" item but prices a specified item. Furthermore, reporters are requested to report prices for the item commonly bought by farmers, that is, "the volume sellers." This practice follows from the desire of the AMS to obtain estimates of the prices that would be secured if the total amount of money spent by all farmers in the United States for the commodity under discussion as of the 15th of a given month were to be divided by the number of items bought.

In concluding this discussion of the computation of the national CPI, we observe that an alternative way of combining city indexes would be to use weights W_i instead of W'_i. That is, the price change occasioned by a group of individuals would be weighted only in proportion to their number, and not additionally in proportion to their expenditures in the base year. This is essentially a problem in aggregation, as has been observed by von Hofsten (*1*, p. 125), and is in some respects related to the work of H. Theil (*12*) on "Linear Aggregation of Economic Relations." This in turn carries us back to the comments and references made in the first section of this report to formal statistical and economic models.

The foregoing account of the mechanical details of computing the Consumer Price Index represents a synthesis of bits and pieces of information culled from a wide variety of sources. I would like to suggest that it is the responsibility of the Bureau of Labor Statistics to publish technical materials that will describe in precise detail at least the major outlines of the actual procedures used by the Bureau. One has only to examine descriptions given in statistics textbooks of the city weighting procedure of the CPI in order to appreciate the need for such descriptions. Similar comments could undoubtedly be made about the Index of Wholesale Prices and the Indexes of Prices Paid and Received by Farmers.

It should be apparent from the discussion of this section that the procedures used by the Bureau of Labor Statistics lead not only to an all-item United States Consumer Price Index but also to a wide variety of subindexes (e.g., U.S. indexes for subgroups of items such as food, dairy products, housing, and the like; city all-item indexes; and in some instances city indexes for subgroups of items.). The remainder of this report will be devoted almost entirely to sampling problems as they relate to the all-item U.S. index, although many of the ideas and comments could be applied in a somewhat changed form to the subindexes. An extensive discussion of sampling for the food-at-home portion of the index, with particular reference to city indexes, has recently been published by Kruskal and Telser (*13*).

IV. Views on Sampling Variability and Price Indexes

It is clear from the foregoing discussion, as well as from publications of the BLS (e.g., *8*), that the data used in computing a value of the CPI are derived almost entirely from samples—samples of consumer families, samples of cities, samples of commodities, samples of points in time, and samples of price reporters. Furthermore, we have seen that these extremely numerous bits and pieces of sample data are combined in a most complex manner in order to arrive at a value of the CPI for a given month.

No individual can quarrel with the fact that a value of the CPI depends in some way on the particular samples from which the index data are obtained. It must therefore follow that a possibly different value of the index would result if a different sample were used at any stage in the process, and that it would be desirable to be able to attach a measure of sampling variability to a particular value of the index. Once this general area of agreement is reached, however, many diverse views have been expressed concerning the sampling variability of price indexes.

Many of the writers who have dealt with the problems of index number construction have assumed, either explicitly or implicitly, that "good" data are available. They have then simply not concerned themselves with the problem of attaching a measure of sampling variability to a computed value of the index. Thus Stone (*3*, p. 118) does not directly discuss problems of sampling but does say the following:

> The quality of index-numbers must depend to a large extent on the quality of the statistical data available, empirical information about product heterogeneity and the factors with which it is associated, and a skillful use of resources in making the innumerable adjustments and approximations that arise in practice.

When the expressed views of individuals who have been intimately associated with the actual production of index numbers are examined, we find a recognition of the need for "good" sampling and "large" samples, but, at the same time, a feeling of doubt that one can or should measure the sampling precision of an index number. Their arguments run somewhat as follows: The fixed market basket concept underlying the Laspeyres index formula is impossible to realize completely in the market place because some items disappear between two points in time, new items are continually introduced into the stock of existing items, and many of the items that are nominally available at both points in time have changed in quality and are thus, in reality, different items. The combined effect of these factors (including the manner in which one chooses to deal with them) is so large that it overwhelms the sampling effects. The conclusion drawn from this line of reasoning is that it is impossible and/or unnecessary to discuss the sampling precision of an index number. Some typical comments along these lines are von Hofsten's (*1*, p. 42):

> It may thus seem as if the important thing would be to select the items to be priced in such a way as to guarantee a good sample. (*Footnote:* This point of view is stressed by Mudgett (1951). He does not, however, consider the remaining problems which will be discussed presently.) One finds in practice, however, that the selection of items in current price index series is based on common sense and not on proper sampling methods. A consideration of the remaining problems, which will be undertaken presently, will show that this is no serious drawback. Moreover, the use of sampling would be expensive, as it would require complete lists of commodities.
>
> There is also another sampling problem involved here, viz, the selection of retail outlets where the prices shall be collected. In a large city only a few shops may be visited and the price of a single article may vary considerably from shop to shop, or, at least, from district to district. Although the difference in price *change* may not be so important, the selection of retail outlets will have a certain influence on the index. To be satisfactory the price collection should be based on an efficient sample of retail outlets. The construction of such a sample cannot be too difficult.

Jaffe (*2*, pp. 10–12) says:

> Statisticians who ask how well the CPI measures the price movements of the wage-earner's basket of purchases often have in mind the precision of the index in terms of its sampling error. I must regretfully assure them that while we believe the CPI provides a measurement of price change sufficiently accurate for practical uses, we are unable to supply a statistical measure of its precision. Before going on with the reasons for this, I would like to state further that I don't consider this lack terribly important. The idiosyncracies of the price data are far more significant in determining the character and accuracy of a price index. I am afraid that a measure of sampling error that ignored the problems of price measurement and comparison would, by giving the wrong impression of accuracy, defeat its own purpose.
>
> Since the probability sampling is so generally accepted as desirable, its honoring in the breach calls for some explanation. Given unlimited resources it would probably be possible to establish probability sampling procedures for all components of the Consumer Price Index. However, because of the wide scope of the index, the diversity of elements that

must be sampled, and the complexity of the marketing situations in which prices must be gathered, there is no practical probability sampling approach that can be applied with present resources. This does not mean that we at the Bureau ignore the statistical principles of sampling. They are applied to the extent that is practical and are always held forth as guides to our day-to-day sampling decisions.

In sharp contrast to these views, a number of authors have implied or stated that it is possible to take existing sampling theory—as is set forth, for example, by Hansen, Hurwitz, and Madow (*14*) or by Cochran (*15*)—and apply it more or less directly to a Laspeyres-type index, particularly with reference to the sampling of commodities. Thus, Mudgett (*7*, p. 51) says:

> The sampling error of index numbers arises from the fact that calculations are based on a set of n commodities found in the two periods of the comparison and this set is used to represent the whole list of N common commodities. * * * The index based on n, however, is an estimate. * * * In the usual statistical sense it is a variable, and, therefore, for all possible samples of n that could be taken from N there is a frequency distribution of these errors. We need only a knowledge of some of the properties of this distribution in order to gain needed insight into the accuracy of any determination of I_{01}. This knowledge is readily available from modern statistical theory. * * *

Mudgett does not apply these ideas to the actual computation of sampling errors for index numbers, nor does he cite instances where others have performed such computations. A somewhat similar, but more detailed, account of this view has been given by Banerjee (*16*), who states:

> Whereas it was necessary to construct the True Index in the precise estimation of CLI, and whereas, instead, Laspeyres' formula is being used at the cost of precision, it would, at least, only be reasonable to make sure that Laspeyres' Index be precisely calculated. This aspect of precision does not appear to have been paid the attention it deserves, so much so that it sometimes causes an embarrassment, when different organizations, while calculating the CLI for the same area and the same economic stratum of population, come out with different figures for the same index. Difference in the figures for the same index could have been appreciated if the coverage (the sample, or the way the sample is selected) and the error of estimation were made available. In absence of such information, controversies arise causing difficulties at administrative levels. With a view to systematizing the study, the concept of standard error in index number calculation was introduced in an earlier note (Banerjee, 1956a) where it was shown that it would be possible to calculate the standard error for an estimated CLI under certain assumptions.

64846—61——14

A somewhat more extreme view of the sampling of commodities for index construction has recently been set forth by Adelman (*17*). She essentially espouses the approach given by Mudgett and Banerjee, but suggests the use of a more or less continually changing probability sample of items. To some extent at least, this approach is advocated as a solution to the dilemma of a continually changing universe of commodities. An appropriate quotation from the Adelman article is the following (*17*, p. 240):

> The construction of an index number is normally associated with the selection, on an *a priori* basis, of the sample of commodities which is to be utilized in the evaluation of the index. The use of such a judgment sample precludes the determination of the extent to which an observed difference in two indexes can be ascribed to sampling errors, rather than to real causes. This defect is extremely important, since index numbers are generally employed for intertemporal, interregional, or intersectoral comparisons, where differences are often quite small, and their significance correspondingly uncertain. Furthermore, the use of an arbitrary fixed sample permits neither changes in product quality nor the introduction or disappearance of consumer products readily to be incorporated into the standard type of index. Any attempt to take such effects into account must of necessity impair the continuity of the index through time.
>
> At present, the author knows of no method in use which will allow the realistic evaluation of the statistical errors associated with an index number. In view of the practical significance of this problem, it is suggested in this paper that the items used in the computation of an index be chosen in a statistical manner. The use of a probabilistic sample would, in principle at least, remove all the above mentioned deficiences inherent in the normal method of sample selection. And, while the proposed procedure would not solve the problem of appropriate weighting, it would have the further advantage of being in conformity with the modern statistical trend towards the replacement of judgment samples by probability samples.

As a conclusion to this brief and *purposive* selection of views on the sampling variability of index numbers, we observe that the appearance of the Banerjee and Adelman papers led von Hofsten (*18*, p. 403) to reply in the following words:

> 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.

The preceding views seem to offer three mutually exclusive choices for treating the sampling variability of a price index, namely: (1) ignore it, (2) determine it by a more or less direct application of existing sampling theory, or (3) modify the definition of the price index so that existing theory will apply. If any one of these choices could be adopted in its entirety, then it would simply be a case of carefully setting forth the full consequences of the choice. Unfortunately, there would appear to be elements that cannot be ignored in each of the views, and the real situation can be described only in composite terms. An attempt will now be made to present such a composite view.

V. The Sampling of Commodities

It is clear that the feelings of doubt that have been expressed concerning the possibility and desirability of attempting to measure the sampling precision of an index number arise primarily from difficulties encountered in maintaining a fixed market basket of goods and services when the universe of commodities available to the consumer is continually changing. The accepted approach to this problem by the producers of index numbers, as discussed, for example, by von Hofsten (*1*) and Stone (*3*, pp. 47–59), has been to maintain the fixed market basket as nearly as possible, but to make a variety of adjustments for the disappearance of old items, for the changing quality of continuing items, and for the appearance of new items. Adelman (*17*), as indicated in the previous citation, not only questions this method of constructing an index number but also states that it is impossible to attach a measure of sampling precision to an index so determined. Finally, von Hofsten (*18*) seems to accept Adelman's statement about the impossibility of computing sampling precision, but is unwilling to accept her solution of a continually changing sample of commodities.

It is not the purpose of this report to argue the meaningfulness of the Adelman approach to index number construction, or to justify or criticize the techniques that are being used to adapt a Laspeyres-type index to situations where there is a continually changing universe of commodities. Rather, we shall argue that it is quite reasonable to talk about the sampling precision of an index determined by the latter method, provided (1) that a very general view of sampling precision, similar to that described by Stephan and McCarthy (*19*, Chap. 10), is adopted, (2) that sampling theory is not asked to take over a task of which it is incapable, namely that of specifying the form of a "true" index, and (3) that one does not always expect to measure this precision by the application of more or less standard formulas from the theory of sampling. Furthermore, we shall argue that it is necessary to talk about and measure the sampling precision of such an index.

For present purposes, assume that a price index of Laspeyres' type is to be computed under circumstances (e.g., for an individual consumer or for a single city) where sampling variability arises *only* from the fact that a sample of items is selected at time zero. That is, base year weights (or other appropriate weights) are known without error; base year prices for specified-in-detail commodities are known without error; and given year prices are known without error for any specified-in-detail item. Furthermore, we ask that there exist a well-defined set of procedures for making adjustments for quality change, and for introducing new items into the index. In other words, we require that these procedures be set forth in such detail that any two individuals or organizations who start with the same sample of commodities at time zero and who independently follow these procedures through successive time intervals will arrive at time t with indexes that are identical in all respects—items, weights, price relatives, and value of the index. It may well be that it is impossible to devise a set of procedures that will uniquely determine the entire process of index number construction, but comments on this aspect of the problem will be deferred until later.

At this point, it is assumed that some well-defined sampling procedure will be used to select a sample of specified-in-detail items from the universe of such items as it exists at time zero. If complete generality is desired, then the only requirement is that this sampling procedure be so specified that repeated and independent applications of the procedure can be made. For example, one might think of a population of teams of experts in consumer price index construction. If each of a number of teams independently chose a sample of items on the basis of their expert judgment, subject possibly to some general set of instructions relating to such sample features as size form of stratification, and the like, then this would conform to the present requirements. Actually, there are strong arguments for using some form of probability sampling at this stage in order to obtain a "good" sample of items and this matter will be discussed later. (The suggestions of Adelman (*17*) and Banerjee (*16*) are therefore pertinent.) Suppose now that one thinks of drawing an indefinitely large number of independent samples in accordance with the defined sampling procedure, and of following each of these samples of items through to time t as specified by the quality and new item adjustment procedure. The values of the index, say $\hat{R}_1^{(t)}, \hat{R}_2^{(t)}, \hat{R}_3^{(t)}, \ldots$, that result from these and successive independent applications of the sampling procedure will undoubtedly differ among themselves, and will define the sampling distribution of the index with respect to the sampling of items. The variance $V(\hat{R}^{(t)})$ of this distribution, if it were known, would provide a perfectly acceptable measure of sampling precision for the index. Furthermore, it is quite clear that an estimate of $V(\hat{R}^{(t)})$ can actually be obtained in this situation by the simple expedient of drawing two or more independent samples of commodities, following each of them through time in accordance with the defined adjustment procedures, and computing the variance among the resulting estimates. Such estimates of variance will, of course, be very "poor" (i.e., their values will be subject to a large amount of sampling variability) if they are based on only a small number of repetitions, but one may perhaps justifiably argue that a "poor" estimate of sampling variability is better than no estimate.

From the foregoing discussion, where it has been assumed that the quality and new item adjustment procedure gives a unique index result for any given sample of commodities, we see that the existence of a continually changing universe of commodities is in itself no reason for arguing that one can neither define nor estimate a measure of sampling precision to associate with the index. However, this measure of sampling precision is obviously defined about the mean of the sampling distribution of $\hat{R}^{(t)}$, namely $E(R^{(t)})$. This is consistent with ordinary sampling theory usage. Thus, Cochran (*15*, p. 10) says: "Accuracy usually refers to the size of deviations from the true mean μ, whereas precision refers to the size of deviations from the mean m obtained by repeated application of the sampling procedure." If the "true" value of the index at time t is denoted by $R_P^{(t)}$, where $R_P^{(t)}$ would be obtained by applying the quality and new item adjustment procedures to the complete universes of commodities as they exist at times 0, 1, 2, —, t, then the difference $E(\hat{R}_P^{(t)})-\hat{R}_P^{(t)}$ is the bias of the estimate arising from the sampling and estimation procedures. If the selection were based on expert judgment, then such bias might arise because all the experts might consciously or unconsciously eliminate from the selection process items having a different form of price behavior from those items that were considered for selection. This would be not unlike the bias of "self-selection" that was of vital concern in the evaluation of the Kinsey investigation (*20*).

As a final point, we note that one usually questions the procedures that are used to adjust for quality and to introduce new items into the index and therefore views $R_P^{(t)}$ as only an approximation to a *true* index say $R_T^{(t)}$. Thus the overall bias in the estimate $R_p^{(t)}$ is composed of two parts: $E(\hat{R}_P^{(t)})-R_P^{(t)}$, which arises from the sampling and estimation procedures and $R_P^{(t)}-R_T^{(t)}$ which arises from the quality and new item adjustment procedure. (Of course, $R_T^{(t)}$ represents a Laspeyres-type index computed from all commodities in accordance with "perfect" quality and new item adjustment procedures. This might still differ from the index that is really desired, but this problem will not be considered here.)

On the basis of the foregoing description, the total error in a single estimate, say $R_P^{(t)}$, can be written as:

$$[R_P^{(t)}-R_T^{(t)}]=[R_P^{(t)}-E(R_P^{(t)})]+[E(R_P^{(t)})-R_P^{(t)}]+[R_P^{(t)}-R_T^{(t)}]$$

The first term on the right represents the error or variability which arises from the use of sampling, and the sampling precision of an index refers only to the magnitude of this error. The second term represents the bias arising from the sampling and estimation procedures, and the third term represents the bias that arises through the use of imperfect quality and new item adjustment procedures.

It would appear that at least some of the differences in opinion that have already been cited can be traced to a failure to distinguish carefully among these three components of error and, in particular, to distinguish between the first and third components. All writers agree that it is unlikely that anyone will ever be able to devise a "perfect" set of rules for treating quality changes and for introducing new items into the index. In other words, the exact value of the quantity $(R_P^{(t)}-R_T^{(t)})$ is unknown and will remain so, although it is to be expected that a continuing program of basic research would lead to improved sets of procedures that would reduce the magnitude of this difference. But, as argued earlier, these facts in no way lead one to the conclusion that it is impossible to estimate the value of the first component, namely $\hat{R}_P^{(t)} - E(\hat{R}_P^{(t)})$. Neither do they lead to the conclusion that it is unnecessary to estimate the value of this component, and some observations will now be set forth in relation to this aspect of the problem.

Clearly the designer of a procedure for constructing an index number of prices is making some sort of judgment about the magnitude of the possible errors due to the sampling of commodities when he states, as von Hofsten does (*1*, p. 74) that ". . . the interpretation of the price development for the different items constitutes a problem which is of much greater numerical importance than the selection and the weighting of the items." This conclusion of von Hofsten's was based upon an investigation in which he used two different procedures to determine price relatives (between December 1946 and December 1949) for 189 of the 236 items that made up the Swedish cost-of-living index at that time. Items that appeared in the index for only a portion of this period were not included in the investigation. Furthermore, items falling into the two groups "rent" and "fuel and light" were treated as in the official computations and the indexes for these two groups were weighted with results obtained from the 189 experimental items by the two experimental procedures, say P_1 and P_2. Von Hofsten obtained an index value by P_1 of 109.2 and an index value by P_2 of 106.6, a difference of 2.6 index points. This difference was due mainly to differences arising in the clothing group. However, we are not here concerned with the reasons for such differences but only with their magnitudes. An examination of the price relatives used in this investigation appears to substantiate von Hofsten's previously quoted general conclusion, but it is also of interest to attempt to obtain an actual numerical estimate of the possible error due to the sampling of commodities.

The 189 price relatives used by von Hofsten, together with their base weights, were classified into four major groups—food, clothing, shoes, and miscellaneous items. Let us regard these four major groups as four strata where the individual price relatives are denoted by R_{ij}, i being the stratum subscript and j being the item subscript within a stratum. Then the variance among price relatives within the i-th stratum is given by

$$\sigma_i^2 = \frac{\sum_j w_{ij}(R_{ij} - \mathrm{R}_i)^2}{\sum_j w_{ij}}$$

where w_{ij} is the base weight for the ij-th item and

$$R_i = \frac{\sum_j w_{ij} R_{ij}}{\sum_j w_{ij}}$$

The values of R_i are given in von Hofsten's book, but it has been necessary to compute the values of the σ_i^2. The results are given in Table 1.

For present purposes, let us imagine that we are in a somewhat more ideal situation than was outlined in the beginning of this section.

TABLE 1.—*Variances of Price Relatives for 189 Items in the Swedish Cost-of-Living Index* [a]

[Relatives are for December 1949 on December 1946 as base]

Group of items	Number of items	Weight, Kr., December 1946	"Index figure actually used"	
			R_i	σ_i^2
Food	69	2,406	109.5	184.89
Clothing	23	829	101.7	57.05
Shoes	13	182	99.6	14.18
Miscellaneous	84	1,326	108.0	154.13
Rent	(b)	634	103.6	(b)
Fuel and light	(b)	306	101.9	(b)
Total		5,683	106.6	

[a] The relatives used for these computations are given in Tables 4.1, 4.2, 4.3, and 4.4 of (*1*) under the heading "Index Figure Actually Used."
[b] Not given in (*1*).

That is, the data of Table 1—and the data of von Hofsten from which they are derived—will be regarded as the complete population of price relatives from which samples are to be drawn. With reference to the model for the components of error, $R_P^{(t)}-R_T^{(t)}$ is unknown but we will accept von Hofsten's conjecture that $R_{P_1}^{(t)}-R_{P_2}^{(t)}=2.6$ places an upper bound on this error. It is quite likely that the actual value of the error is smaller than 2.6. Furthermore, we will assume a sampling model is used such that $E(\hat{R}_P^{(t)}-R_P^{(t)})$ is known to be zero and such that the value of $\hat{R}_P^{(t)}-E(\hat{R}^{(t)})$ can be estimated. A reasonable form for such a model, following Adelman (*17*), would appear to be:

1. From within the i-th item group, or stratum, a sample of n_i relatives are drawn with replacement and with probability proportionate to w_{ij}. Then it can be shown, if we take

$$\hat{R}_i=\frac{\sum\limits_{j=1}^{n_i} R_{ij}}{n_i},$$

that

$$E(\hat{R}_i)=R_i$$

and

$$V(\hat{R}_i)=\frac{\sigma_i^2}{n_i}$$

where R_i and σ_i^2 are as previously defined.

2. If the individual strata indexes are now combined in accordance with strata weights, we have

$$\hat{R}=\frac{\sum\limits_{i}(\sum\limits_{j} w_{ij})\hat{R}_i}{\sum\limits_{i}\sum\limits_{j} w_{ij}}$$

and

$$V(\hat{R})=\sum_i\left(\frac{\sum\limits_j w_{ij}}{\sum\limits_i\sum\limits_j w_{ij}}\right)^2\frac{\sigma_i^2}{n_i}$$

Applying this variance formula to the data of Table 1, we obtain

$$V(\hat{R}_P^{(t)}) = (.423)^2 \frac{184.89}{69} + (.146)^2 \frac{57.05}{23} + (.032)^2 \frac{14.18}{13} + (.233)^2 \frac{154.13}{84}$$

$$= .63$$

This result can be viewed as a crude approximation to the first component of error in the model for components of error since

$$V(\hat{R}_P^{(t)}) = E[\hat{R}_P^{(t)} - E(\hat{R}_P^{(t)})]^2$$

It is obviously only a crude approximation to the first component of error for a wide variety of reasons, among which are the following:

1. We have assumed a fixed population of items for which the variance among price relatives, for a three-year period, could be computed. The effects on sampling precision of the procedures used by von Hofsten in following through the quality changes of the items are certainly mirrored to some extent in the computed value. However, there would be no way of knowing whether this type of analysis catches the full effects of a complex quality and new item adjustment procedure without actually following through with independent samples of items as outlined previously.
2. The sampling model which has here been applied to the data was almost certainly not used in the original selection of items for the Swedish Cost-of-Living Index. Nevertheless, this would appear to be a reasonable type of model if one were going to obtain a sample on a probability basis.
3. The strata used in the present computations are much larger than would be found in practice, and thus the observed value of .63 is too large. As a matter of fact, references *2* and *11* would seem to indicate that a stratum would often be defined in terms of a single specified item and would be composed of different qualities of this item, i.e., different specified-in-detail items. If we assume 200 strata of equal weight, with a single item to be drawn out of each stratum, and a within-stratum variance of 20 (somewhat low as far as Table 1 is concerned, but perhaps still too high for actual situations), then

$$V(R_P^{(t)}) = \sum_{i=1}^{200} (.005)^2\, 20$$
$$= .10$$

It should not be too difficult to obtain actual variance estimates to employ in such crude computations as this. For example, Adelman (*17*, Table 2) reports some variances computed within rather narrow food groups on the basis of data gathered in several food stores in the Berkeley, California area. (Her time period between quotations was 14 weeks while we are here dealing with a period of three years, and the variance will, of course, be a function of time.) Also, Staff Paper No. 2 reports some research based on Sears catalogues. Some illustrative variances were computed among three-year price relatives for men's cotton work shirts, using only items that were identical for the three-year period. One set of 10 items (1950–52) gave a variance of 33; another set of 12 (1953–55) gave a value of 15; and a third set of 15 (1956–58) gave a value of 15. These values, and those of Adelman, do not appear out of line with the value of 20 that was inserted in the preceding computation.

4. The sampling of items has been viewed here as occurring with replacement. Under certain circumstances it might be appropriate to regard it as occurring without replacement, and the effect of this would be to make the value of .63 too large. This distinction would disappear if items were stratified to the point where only a single specified-in-detail item were drawn out of each stratum, assuming, of course, that there was more than one specified-in-detail item in each stratum.

5. Taking the population of items as given, it has been necessary to leave out some sources of variation since no data were given for the "rent" and "fuel and light" strata. This would make the observed value of .63 too small.

6. The relatives reported by von Hofsten represent averages over cities and localities and over outlets within cities and localities. This would make the observed value of .63, as it refers to the sampling of commodities, too large.

Putting together all of these bits and pieces of information, it seems reasonable to guess that the variance due to the sampling of commodities of the Swedish Cost-of-Living Index as described by von Hofsten is something of the order of .1 to .6. The actual value is apt to be near the lower end of this range since most of the stated reservations appear to place the observed value of .63 on the high side. For present purposes, let us assume that $V(\hat{R}_P^{(t)}) = .2$. This means that the standard deviation of $\hat{R}_P^{(t)}$ would be approximately .5 and that a large sample, 95 percent confidence interval for $R_P^{(t)}$ would have total width of about 2.

The foregoing estimates of .2 for $V(\hat{R}_P^{(t)})$ and 2.6 for $R_P^{(t)} - R_T^{(t)}$, where t in this instance represents a period of three years, can be regarded as nothing more than crude approximations to the true values. Nevertheless, if they are fully recognized as such, it is instructive to examine their relative order of magnitude. For example, it is customary to measure the total error of an estimate by its Mean Square Error, which is defined as the sum of its variance and the square of its systematic error or bias. Thus in this example, where $E(\hat{R}_P^{(t)}) - R_P^{(t)}$ is assumed to be zero,

$$\begin{aligned} MSE(R_P^{(t)}) &= V(R_P^{(t)}) + (R_P^{(t)} - R_T^{(t)})^2 \\ &= .2 + (2.6)^2 \\ &= .2 + 6.76 \\ &= 6.96 \end{aligned}$$

The magnitude of the MSE is almost completely determined by the procedural differences and this confirms von Hofsten's previously cited observation.

But this result does not lead to the conclusion that the sampling precision of the index can be ignored. If it is assumed that the main goal of index number construction is to measure accurately the value of $R_T^{(t)}$, as opposed possibly to the measurement of time-to-time changes in $R_T^{(t)}$ (which will be discussed shortly) or to the production of indexes for subgroups of commodities, then this result strongly suggests that too great a fraction of available resources is being spent on maintaining a relatively large sample of commodities and too small a fraction on basic research aimed at reducing the magnitude of $(R_P^{(t)} - R_T^{(t)})$. For example, if the sample sizes given in Table 1 were each reduced by a factor of about two-thirds—69 foods to 23, 23 clothing items to 8, 13 shoe items to 4, and 84 miscellaneous items to 28—then $V(R_P^{(t)})$ would be increased from .63 to 1.89. Reducing 1.89 by a factor of one-half for the reasons given previously, we might expect $V(R_P^{(t)})$ to be roughly .95, or say 1.00. The *MSE* of $\hat{R}_P^{(t)}$ now becomes

$$\begin{aligned} MSE(\hat{R}_P^{(t)}) &= 1.00 + (2.6)^2 \\ &= 1.00 + 6.76 \\ &= 7.76 \end{aligned}$$

The procedural error still dominates the *MSE*, accounting for 87 percent of its value, even though the sample of commodities has been reduced by almost two-thirds.

Naturally, it would always be possible to reduce the sample size to a point where the variance of $\hat{R}_P^{(t)}$ would become much larger than the procedural error. However, the practical problem is determining an economic balance between sampling precision and procedural error and then allocating resources so as to reduce the magnitude of the one which dominates. This can only be accomplished if "decent" estimates of $V(\hat{R}_P^{(t)})$ and of $(R_P^{(t)} - R_T^{(t)})$ are available. In this respect, there would appear to have been too much effort placed on expanding the number of specified items included in index computations (the usual procedure being to include all specified items which have more than some minimum base weight) and too little effort placed on estimating the variability of price relatives, say among specified-in-detail items within a specified item, and on estimating the value of $R_P^{(t)} - R_T^{(t)}$. Such investigations can and should be carried out and published, at least for the benefit of the scientific community.

As a final point in this discussion of $V(\hat{R}_P^{(t)})$ and $R_P^{(t)} - R_T^{(t)}$, it should be emphasized that both of these quantities are functions of time. Since all relatives are equal to 100 at $t=0$, $V(\hat{R}_P^{(t)})$ will be extremely small for values of t close to zero. Furthermore, as t increases there will be opportunity for the relatives of different items to "spread apart" and thus ordinarily one would expect $V(\hat{R}_P^{(t)})$ to be an increasing function of time. $R_P^{(t)} - R_T^{(t)}$ will also be very close to zero for values of t close to zero since there will not need to be many quality adjustments in a short period of time. However, the manner in which this procedural error changes with time is not as easy to forecast as for $V(\hat{R}_P^{(t)})$. Under most circumstances one would expect procedural "bias" to increase with time, but this is another problem that needs investigation.

Thus far emphasis has been placed on the estimation of $R_T^{(t)}$ by means of $\hat{R}_P^{(t)}$. The problem becomes slightly different if the goal is to estimate short-term changes in $R_T^{(t)}$, say to estimate $R_T^{(t)}-R_T^{(t-1)}$. Suppose that $\hat{R}_P^{(t)}-\hat{R}_P^{(t-1)}$ is employed as an estimate of this quantity. Then, using the model for the error in $\hat{R}_P^{(t)}$ and still assuming no bias in the sampling and estimation procedure, we have for the error of this estimate:

$$[\hat{R}_P^{(t)}-\hat{R}_P^{(t-1)}]-[R_T^{(t)}-R_T^{(t-1)}]=$$
$$[\hat{R}_P^{(t)}-E(\hat{R}_P^{(t)})]-[\hat{R}_P^{(t-1)}-E(\hat{R}^{(t-1)})]+$$
$$[R_P^{(t)}-R_T^{(t)}]-[R_P^{(t-1)})-R_T^{(t-1)}]$$

It seems reasonable that if the time between t and t-1 is short, say, of the order of a month, then the difference between the last two terms will be extremely small. That is, the procedural error will not change much from month to month. Thus

$$[\hat{R}_P^{(t)}-\hat{R}_P^{(t-1)}]-[R_T^{(t)}-R_T^{(t-1)}]\doteq[\hat{R}_P^{(t)}-E(\hat{R}_P^{(t)})]-[\hat{R}_P^{(t-1)}-E(\hat{R}_T^{(t-1)})]$$

Therefore the error in $\hat{R}_P^{(t)}-R_P^{(t-1)}$ will be due almost entirely to sampling error and the standard formula for the variance of a difference gives:

$$V(\hat{R}_P^{(t)}-\hat{R}_P^{(t-1)})=V(\hat{R}_P^{(t)})+V(\hat{R}_P^{(t-1)})-2\rho\sqrt{V(\hat{R}_P^{(t)})V(\hat{R}_P^{(t-1)})}$$

where ρ is the correlation between $\hat{R}_P^{(t)}$ and $\hat{R}_P^{(t-1)}$. Under these circumstances it would seem absolutely essential to have an estimate of sampling precision for the difference of $\hat{R}_P^{(t)}$ and $\hat{R}_P^{(t-1)}$ since it is not even possible to argue that this sampling precision is overshadowed by the procedural error.

An estimate of V $(\hat{R}_P^{(t)})$, when t is equal to three years, has already been obtained. The value of $V(\hat{R}_P^{(t)}-\hat{R}_P^{(t-1)})$, for a month-to-month change, can therefore be estimated if an appropriate approximation to ρ can be found. It is possible to obtain a very crude estimate of ρ from data published by the Bureau of Labor Statistics (*21*) through the following line of reasoning:

1. Let us view $\hat{R}_P^{(t)}$ as an unweighted average of a random sample of price relatives.

2. Then $\hat{R}_P^{(t-1)}$ is the unweighted average of price relatives for *exactly the same sample of items* for the preceding month.

3. The correlation between the means of two variables, each variable being measured on exactly the same random sample of elements, is the same as the correlation between the values of the variables for the individual elements in the population from which the samples are drawn.

4. Therefore the correlation between $\hat{R}_P^{(t)}$ and $\hat{R}_P^{(t-1)}$ can be approximated by the correlation between the price relatives for a sample of items in month (t–1) and month t.

5. It seemed unnecessary, at the level of approximation being discussed here, to go to the individual item price relatives and thus we have computed month-to-month correlations using the United States city average subgroup index for the following 20 subgroups: cereals and bakery products; meats, poultry, and fish; dairy products; fruits and vegetables; other foods at home; rent; gas and electricity; solid fuels and fuel oil; house furnishings; household operation; men's and boys' apparel; women's and girls' apparel; footwear; other apparel; private transportation; public transportation; medical care; personal care; reading and recreation; and other goods and services. These monthly indexes (with 1947–49 as base) are published in von Hofsten's Tables B–2 and B–3 (*18*), and the values of the correlation coefficients are given in Table 2 below. These correlations are undoubtedly overestimates of the true ρ's because of the use of grouping and also possibly because of the imputation process used when individual items are not priced each month in each city.

TABLE 2.—*Month-to-Month Correlations for 20 United States City Average Subgroup Indexes* [a]

(1947–49=100)

Months	*Correlation coefficient*
January 1947–February 1947	0.956
January 1948–February 1948	.799
January 1950–February 1950	.983
June 1950–July 1950	.983
January 1953–February 1953	.992
January 1958–February 1958	.998

[a] Original data are given in *18*, Tables B–2 and B–3.

It will be noted from an examination of this table that ρ is also a function of time. The smallest value occurs in the middle of the base period—i.e., the lowest value occurs for the comparison January 1948–February 1948 while the base period is 1947–49—and ρ increases as one moves from this base period. The reason is quite simple. As one moves from the base period, individual price relatives spread out in terms of magnitude. Yet month-to-month changes for the same specified-in-detail item are small. Therefore the greater the dispersion in price relatives, the greater will be the value of the correlation coefficient.

Returning now to the numerical example, let us consider t to be about three years. Then $V(\hat{R}_P^{(t)}) \doteq 0.2$, $V(\hat{R}_P^{(t-1)}) = 0.2$, $\rho^{(t)} \doteq 0.98$, and $V(\hat{R}_P^{(t)} - \hat{R}_P^{(t-1)}) = 0.2 + 0.2 - 2(0.98)\sqrt{0.2 \times 0.2} = 0.008$. Thus the standard error of the estimate of the difference is $\sqrt{0.008} = 0.09$, or approximately 0.1. Since a month-to-month change in the U.S. Consumer Price Index of 0.1 or 0.2 of a percentage point is ordinarily regarded as of some practical significance, at least by the newspapers and by parties to collective bargaining agreements, a standard error of 0.1 is not particularly small in this context. It would, therefore, seem important to have better estimates of the standard error of such changes than have been produced by the rough methods being used here.

Still leaving aside any discussion of interregional or intersubgroup comparisons of price indexes, there is yet a further argument that leads to the conclusion that it is absolutely essential to have a measure of sampling precision relating to the sampling of commodities. As outlined in Section II, the U.S. Consumer Price Index involves not only the sampling of commodities but also the sampling of cities and the sampling of retail outlets within cities. (Other indexes, such as those prepared by the Agricultural Marketing Service, do not have the BLS emphasis on cities but the same problems arise in other ways.) Just as there must be a balance between the procedural error of the index and the error due to the sampling of commodities, so also must there be a balance between these errors and the errors due to the sampling of cities and of retail outlets. Again it is not possible to discuss such a balancing operation unless some attempt is made to measure these components of error.

The next section of this report will present some computations which suggest that the variance of the U.S. Consumer Price Index, due to the sampling of cities and retail outlets, and for a month some three years after the base period, is something of the order of 0.01. Thus we have, using previous estimates:

Procedural bias squared	6.76
Variance due to sampling of commodities	.20
Variance due to sampling of cities and retail outlets	.01

These results suggest that not only is the sampling error due to the sampling of commodities overshadowed by the procedural bias, but that this sampling error in turn completely dominates the sampling error arising from the sampling of cities and of retail outlets. The efficient allocation of resources, as far as the overall U.S. Consumer Price Index is concerned, would therefore call for a reduction in the size of city sample and size of retail outlet sample and the assignment of these resources to work on the procedural error. But again this cannot be done unless "decent" estimates of error are available.

The preceding discussion of the sampling of commodities for a consumer price index, where the index is constructed on the basis of a fixed market basket of goods and services, has set forth the argument that it is possible and necessary to define and estimate the sampling precision of an index determined on such a basis. In concluding this discussion, we should like to mention a few points which have been touched on lightly or omitted entirely.

1. The relationship between the sampling of commodities and the adjustment procedures for quality change and new items has been discussed in extremely general terms. Under these circumstances the only satisfactory approach to the measurement of sampling precision would appear to be through the use of two or more independent samples of commodities. However, if these adjustment procedures are defined in a somewhat more restrictive fashion, then it should be possible to apply standard sampling theory more or less directly to the problem. For example, suppose that one starts out at time zero with a population of N items, where the i-th item has weight w_i. Furthermore, let us suppose that this list of items and their associated weights remains unchanged throughout the period for which the index is to be constructed. The price relative for a particular item at time t, $R_i^{(t)}$, will reflect quality changes in the original item, or may even take into account the fact that the original item has disappeared from the market and that a new item has been substituted for it.

(The only types of situations which are excluded are those in which an original item disappears without a direct substitute appearing, and those in which a new item appears which is not a direct substitute for a previous item.) Then any probability mechanism used at time zero to select a sample of commodities can also be viewed as having selected a sample from the population of price relatives $R_i^{(t)}$ as it exists at time t. Standard sampling theory can thus be used to determine the precision of the estimate made at time t. The estimates of sampling precision obtained at different times are of course correlated since they are based on the same sample of commodities, but this is another problem.

2. One of the terms in the components of error model has essentially been ignored in the preceding discussion, namely the term $E(\hat{R}_P^{(t)}) - R_P^{(t)}$. This is the bias arising from the sampling and estimation procedure. There is no satisfactory way of estimating the magnitude of this component from empirical data derived from repeated applications of a single nonprobability model sampling procedure, although conceivably it could be as large as, or larger than, the procedural error if the judgment approach used in the selection of specified-in-detail items were badly at fault. The only real way of controlling this error is to use some form of probability sampling in the original selection of items for the index, whether or not estimates of sampling precision are to be obtained by independent samples or through the use of the probability model, or to estimate the magnitude of the error through experimental studies.

3. It has been assumed that the quality change and new item adjustment procedure, designated by P, can be set forth in such detail that any two individuals or organizations who start with the same sample of commodities at time zero and who independently follow these procedures through successive time intervals will arrive at time t with indexes that are identical in all respects. In actual practice, it is probably impossible to achieve this uniqueness. There will always be cases where borderline decisions are required which could sometimes go one way and sometimes another. The effect of this lack of uniqueness would be to add still another component of random error to the model used in this section, and the only way to evaluate the magnitude of such an error would be through some type of empirical investigation. It could even happen that one might wish to build certain elements of randomness into the rules of procedure. For example, if it were impossible or too costly to decide among, say, three alternative ways of treating a certain quality adjustment problem, then one might choose to use each procedure one-third of the time. The choice on any particular occasion would be made on the basis of some random device.

4. As a final point, we emphasize again that the discussion has been directed at the error in an overall index. If one is concerned with city, or regional, or subgroup indexes, or with comparisons among such indexes, then it is still necessary that one define, study, and measure the components of error for each such index. However, it may be that requirements on accuracy at this level produce, as a byproduct, greater accuracy than is actually needed at the level of the overall index.

VI. Estimates of Sampling Error Arising From the Sampling of Cities and Retail Outlets for the U.S. Consumer Price Index

As was noted in Section II, the sampling of commodities is only one of the many sampling problems which must be faced in index number construction. In particular, it is necessary to select a sample of localities in which current prices are to be collected and, within these, a sample of retail outlets from which these prices will actually be obtained. In the case of the U.S. Consumer Price Index, localities are synonomous with cities, but the same problems exist whether the emphasis is on cities, on counties, or on some other type of local unit. The error in the final index will be partially determined by the manner in which these sampling problems are resolved. All of the reasons set forth in the preceding section concerning the necessity for measuring the error arising from each of the several sources apply equally well here, and we shall now describe an empirical investigation concerning the combined sampling error in the U.S. Consumer Price Index due to the sampling of cities and retail outlets.

In recent years the U.S. Consumer Price Index has been based upon a national sample of 46 cities. (Reference *8*, pp. 70–71, lists these 46 cities. Complete pricing in one of these cities, Ravenna, Ohio, was discontinued in 1956.) The Bureau of Labor Statistics made available to the Price Review Committee monthly indexes (for all items and for a number of subgroups) for each of these cities for the period 1953–59 with 1953 equal to 100. These city indexes were reported only for those months in which the full list of goods and services was priced in a given city. Thus all-item indexes were available each month for the twelve cities in the largest size class, every third month for the eighteen cities in the next two size classes, and either every fourth month or every third month (1957–59) for the fifteen cities in the smallest size class. These are the basic data that will be used in this section.

The present sample of 46 cities was selected as a preliminary to the 1950 Consumer Expenditures Survey (*22*). Cities were first stratified into four size groups. All cities in the largest size group were drawn into the sample, while the samples in each of the other three size groups were selected by application of a so-called Latin square design. No formal analysis of this sampling design has been published by the Bureau and no attempt will be made here to develop such an analysis. Rather, we shall, as an approximation, view the sample as the result of a much more straightforward type of design, namely, as the result of selecting a single city with probability proportionate to size from each of 34 strata, the 12 largest cities being self-representing.

Before actually presenting the design and results of the present empirical investigation, it is of interest to examine briefly the formal properties of the sampling design just mentioned since they illuminate some of the results that have already been given in Section III and serve as a guide to the estimation procedure. Suppose one has a population of N cities which are divided into L strata where the h-th stratum contains N_h cities. For the i-th city in the h-th stratum, let

$R_{hi}{}^{(t)}$ be the city index at time t,[2] W_{hi} be the fraction of the total population contained in this city. Thus

$$\sum_{h=1}^{L} \sum_{i=1}^{N_h} W_{hi} = 1$$

Let V_{hi} be the cost or value of this city's market basket of goods and services in the base period.
Then the all-item, all-city index is given by

$$R^{(t)} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} (W_{hi}V_{hi})R_{hi}^{(t)}}{\sum_{h=1}^{L} \sum_{i=1}^{N_h} (W_{hi}V_{hi})}$$

This is the same as

$$R^{(t)} = \frac{\sum_{h=1}^{L} \sum_{i=1}^{N_h} W_{hi}\cdot(\text{cost of market basket in city } hi \text{ at time } t)}{\sum_{h=1}^{L} \sum_{i=1}^{N_h} W_{hi}\cdot(\text{cost of market basket in city } hi \text{ at time } 0)}$$

or, essentially the ratio of the cost of all market baskets at time t to the cost of all market baskets at time 0.

The foregoing expression for $R^{(t)}$ can be expressed in terms of the strata indexes $R_h{}^{(t)}$ as

$$R^{(t)} = \sum_{h=1}^{L} \left(\frac{\sum_{i=1}^{N_h} W_{hi}V_{hi}}{\sum_k \sum_j W_{kj}V_{kj}} \right) R_h{}^{(t)}$$

but the above expression cannot be used for the estimation of $R^{(t)}$ since the strata weights will not be known. That is the value of V_{hi} will be known only for those cities that are actually drawn into the sample. What must be done is to estimate separately the numerator and denominator of the original expression for $R^{(t)}$.

With the foregoing stratification setup, the most reasonable approach would appear to be to select n_h cities with probability proportionate to size, i.e., W_{hi}, and with replacement from the h-th stratum.

[2] In Section III the subscripts following $R^{(t)}$ identified items and groups of items. In this section it will not be necessary to indicate indexes for items and groups of items and so subscript positions following $R^{(t)}$ will always identify cities.

Then

$$\hat{R}^{(t)}=\frac{\sum_{h=1}^{L}\left(\sum_{j=1}^{N_h}W_{hj}\right)\frac{\sum_{i=1}^{n_h}V_{hi}R_{hi}^{(t)}}{n_h}}{\sum_{h=1}^{L}\left(\sum_{j=1}^{N_h}W_{hj}\right)\frac{\sum_{i=1}^{n_h}V_{hi}}{n_h}}$$

If each $n_h=1$, as would ordinarily be the case in practice, this becomes very simple, namely,

$$\hat{R}^{(t)}=\frac{\sum_{h=1}\left(\sum_{j=1}^{N_h}W_{hj}\right)V_{hi}R_{hi}^{(t)}}{\sum_{h=1}^{L}\left(\sum_{j=1}^{N_h}W_{hj}\right)V_{hi}}$$

where the subscript hi now represents the single city drawn from the h-th stratum. This is the form of estimate that was given at the end of Section III, where the quantity

$$\left(\sum_{j=1}^{N_h}W_{hj}\right)V_{hi}\Big/\sum_{h=1}^{L}\left(\sum_{j=1}^{N_h}W_{hj}\right)V_{hi}$$

was symbolized by W'_i. These quantities are evidently the "relative cost-population weights Dec. 1952" given in Table 1 of BLS Bulletin No. 1168(*8*).

Let us now take these observed city or stratum weights as approximations to the true stratum weights. That is

$$W'_h=\frac{\left(\sum_{j=1}^{N_h}W_{hj}\right)V_{hi}}{\sum_{h=1}^{L}\left(\sum_{j=1}^{N_h}W_{hj}\right)V_{hi}}\doteq\frac{\sum_{j=1}^{N_h}W_{hj}V_{hj}}{\sum_{h=1}^{L}\sum_{j=1}^{N_h}W_{hj}V_{hj}}$$

Then

$$\hat{R}^{(t)}=\sum_{h=1}^{L}W'_h\hat{R}_h^{(t)}=\sum_{h=1}^{L}W'_h\hat{R}_{hi}^{(t)}$$

where, as before, the subscript i represents the one particular city in the h-th stratum that was drawn into the sample. Finally,

$$V(\hat{R}^{(t)})=\sum_{h=1}^{L}W_h'^2V(\hat{R}_{hi}^{(t)})$$

In order to estimate this quantity it is necessary to obtain estimates of $V(\hat{R}_{hi}{}^{(t)})$.

The variance of $\hat{R}_{hi}^{(t)}$, as an estimate of $R_h^{(t)}$ for a fixed sample of commodities, depends upon the variability among indexes for cities in the h-th stratum and upon the sampling precision of the estimates of the average prices for commodities within cities, i.e., upon the within-city samples of retail outlets. Since the design under discussion assumes that only a single city is drawn from each stratum, it is impossible to obtain a direct estimate of this within-stratum variance. However, an overestimate of this variance can be obtained by the method of "collapsed strata," as described by Cochran (*15*, pp. 105–106), Hansen, Hurwitz, and Madow (*14*, pp. 399–401), or Sukhatme (*23*, pp. 339–404). Roughly speaking, one takes two strata which are as nearly alike as possible, "collapses" these two into a single stratum, estimates the variance in this stratum from the two observations, and then uses this variance estimate for each of the two original strata.

The foregoing procedure was applied to the 46 cities for which BLS supplied monthly indexes. Nothing could be done about the 12 largest cities that were drawn into the sample with certainty, but the remaining cities were paired within size classes as nearly as possible by geographic closeness. These pairings are given in Table 3, together with the values of W'_h. Many of these pairings are of necessity far from ideal, but the effect of this should be to inflate further the variance estimates. Note that the 46th city, Ravenna, Ohio is also included in this table. Even though monthly all-item indexes were not provided for this city, it can be treated just like the other "unpaired" cities as far as its weight is concerned.

TABLE 3.—*Pairings of Cities for the Empirical Variance Computations*

Paired Cities	*Values of W'_h*	
Size Class B:		
Kansas City, Mo., and Minneapolis, Minn	0.024,	0.025
Portland, Oreg., and Seattle, Wash	.024,	.027
Houston, Tex., and Atlanta, Ga	.024,	.021
Cincinnati, Ohio, and Youngstown, Ohio	.022,	.021
Scranton, Pa. (unpaired)		.021
Size Class C:		
Canton, Ohio, and Charleston, W. Va	.020,	.024
Lynchburg, Va., and Huntington, W. Va	.022,	.019
Evansville, Ind., and Middletown, Conn	.020,	.025
Madison, Wis., and Newark, Ohio	.023,	.019
San Jose, Calif. (unpaired)		.024
Size Class D:		
Grand Forks, N. Dak., and Rawlins, Wyo	.011,	.012
Madill, Okla., and Shawnee, Okla	.010,	.011
Camden, Ark., and Grand Island, Nebr	.009,	.011
Garrett, Ind., and Laconia, N.H	.013,	.010
Anna, Ill., and Shenandoah, Iowa	.010,	.011
Glendale, Ariz., and Lodi, Calif	.012,	.015
Middlesboro, Ky., and Pulaski, Va	.008,	.010
Sandpoint, Idaho (unpaired)		.010
Ravenna, Ohio (unpaired)		.012

For a particular pair of cities and for a month in which the all-item index is available in each city, the foregoing procedure leads to the following estimate of variance for each of the strata from which the cities were drawn

$$\hat{V}=\frac{1}{2}(\hat{R}_1^{(t)}-\hat{R}_2^{(t)})^2$$

where $\hat{R}_1^{(t)}$ is the index for the first city of the pair and $\hat{R}_2^{(t)}$ is the index for the second city. This assumes that the strata are of the same "size." This is not quite the case here, but no attempt was made to use a more precise form of estimate since the present way of viewing the sampling procedure is only a very rough approximation to the true situation and since $\hat{V}$ is known to be an overestimate of the true variance, even if all the proper assumptions did hold. As an example of the application of this formula, in January 1958 (with 1953=100) the index for Cincinnati was 107.2 and the index for Youngstown, Ohio, was 108.5. Therefore

$$\hat{V}=\frac{1}{2}(107.2-108.5)^2$$

$$=\frac{1.69}{2}=0.845$$

It should be clearly recognized that this estimate of the within-stratum variance includes not only the effect of the sampling of cities but also the effect of the sampling of retail outlets within the cities. It does not include any appreciable effect due to the sampling of commodities since essentially the same sample of commodities is used in each of the cities.

Two difficulties were encountered in applying this procedure to the cited data. First, because the quarterly pricing cycle was not the same for all cities, it was sometimes necessary to use the index for one city in a pair with the other city's index for either the preceding or succeeding month. This would have a tendency to inflate the variances. Second, no computations could be made for the unpaired cities. In this instance, the average of the variance estimates obtained from pairs of cities in the same size class was arbitrarily assigned to these unpaired cities.

The outlined computations were performed for each of either three or four months in the years 1953–59, and the resulting between-two-cities estimates of variance were combined in accordance with the formula

$$\hat{V}(\hat{R}^{(t)})\doteq\sum_{h=1}^{34}W_h'^2\hat{V}(\hat{R}_{hi}^{(t)})$$

The values obtained are given in Table 4. These values are the contribution to the variance of $\hat{R}^{(t)}$ of the 34 cities in the B, C, and D size class strata, representing some 58 percent of the total strata weight. The remaining 42 percent is allocated among the 12 cities in the largest size class. Since these cities are self-representing, their contribution to the variance is only in terms of the within-city variance and it is not possible to estimate this with the paired-city approach.

TABLE 4.—*Between-Two-Cities Estimate of the Variance of $\hat{R}^{(t)}$, Ignoring the Within-City Contribution from the Twelve Cities in the Largest Size Class (1953=100)*

Month:	$\hat{V}(\hat{R}^{(t)})$	Month—Continued	$\hat{V}(\hat{R}^{(t)})$
January 1953	0.0018	January 1957	0.0086
April	.0004	April	.0076
October	.0011	July	.0054
January 1954	.0015	October	.0106
April	.0012	January 1958	.0143
October	.0037	April	.0214
January 1955	.0038	July	.0140
April	.0107	October	.0157
October	.0038	January 1959	.0192
January 1956	—	April	.0249
April	.0079	July	.0255
July	.0034	October	—
October	.0084		

As has already been noted, there are a number of factors which tend to make these values overestimates of the true variances, and a single major factor (neglect of the within-city component of variance for the 12 largest cities) which tends to make them underestimates. For present purposes we shall simply regard these as counteracting effects and take the computed values as being roughly of the correct order of magnitude. It should be noted that these estimates are themselves subject to large and unknown sampling fluctuations.

There are two features of these data which stand out. First, there is a definite tendency for the values to increase over time and this is to be expected. The price relatives for all items in all cities essentially start out at 100 in the base period and there are increasing opportunities for them to spread out as the time period under study deviates from the base period. This effect was mentioned in connection with the sampling of commodities, but illustrative data were not presented at that point. Second, the actual magnitudes of these variances are small, particularly in comparison with the estimates given in the last section for the procedural error and for the sampling error due to the sampling of commodities. This comparison was made in Section V and its implications for the U.S. Consumer Price Index were discussed at that point. (The value of 0.01 used for the variance due to sampling of cities and retail outlets was obtained by rounding up the variance figures given in Table 4 for 1956, some three years after the base period.)

The analyses of this section have been carried out with two goals in mind. The first was simply that of obtaining a crude estimate of sampling error due to the sampling of cities and the sampling of retail outlets within cities, which could then be compared with and added to the estimates obtained in the preceding section for procedural error and for the sampling error arising from the sampling of commodities. The second goal was that of indicating an approach to the sampling of cities that would lead to a relatively easy way of estimating sampling variability, recognizing, of course, that the original sample of cities was not selected in accordance with this scheme. These analyses could have been extended to subgroups of commodities in the U.S. Consumer Price Index (e.g., food), but such extensions were deemed outside the purview of this investigation. Although further comments concerning choice of a city sample will be found in the next section, we should like to close this discussion with the observation that the selection of the city sample should be in accordance with some form of probability model in order that no systematic error or bias enter the all-item all-city index from this source.

VII. The Rough Outlines of a Sample Design for Estimating the Total Sampling Error of the U.S. Consumer Price Index

The analysis of the preceding section had many shortcomings, among which were: (1) the original sample of cities was not selected in accordance with the design that dictated the analysis; (2) the estimate of error contained no component for the sampling of retail outlets within the twelve strata of self-representing cities; and (3) the estimate of error contained no component for the sampling of commodities. Nevertheless, this analysis did provide an indication of the magnitude of the error due to the sampling of cities and to the sampling of retail outlets and it did illustrate the type of design that might be expected to lead to "simple" estimation of the total sampling error. The necessity for designing a complex sampling operation so that "simple" estimates of error can be obtained has long been recognized and has been discussed by many authors (e.g., *14*, p. 440, and *17*, pp. 220–229) under such titles as "replicated samples," "ultimate clusters," and "random groups." This need becomes overwhelming in the case of a price index where the number of commodities entering the index is large and where the quality adjustment procedure makes it difficult to apply variance estimating procedures derived from sampling theory to all components of the design. Furthermore, these estimates have to be made more or less continuously since the sampling errors can be expected to increase with the length of time from the base period. Some of the considerations that might apply in this instance will now be outlined.

Since the present emphasis of the Consumer Price Index is on a city sample approach (and some comments on this will be made in the next section), cities will be regarded as the ultimate clusters and the discussion will be built around this city sample. The most appropriate type of design would appear to be that outlined in the preceding section where the cities are first grouped into strata—probably on the basis of size and geographic location and possibly on the basis of additional variables—and then one or more cities are drawn with probability proportionate to size and with replacement from those strata containing more than one city. We are not here concerned with the details of this operation, but do assume the following: (1) that the relationships between city indexes and strata indexes and between strata indexes and the U.S. index are clearly specified in formal terms, and (2) that a probability model be used to select cities within strata which is consistent with the specified form of index, which provides unbiased or "nearly" unbiased estimates of the stratum and U.S. indexes, and which permits a within-strata estimate of variance either by the drawing of two or more cities within each strata or through the use of collapsed strata.

As was emphasized in the preceding section, this type of approach provides no contribution to the variance estimate from the sampling of commodities if the same sample of commodities is used in each sample city. It would seem therefore that it is absolutely imperative that the index be based upon at least two—and this is probably also the maximum number that would be considered—independently selected samples of commodities. These two samples would be selected at the time of index revision and, in accordance with present practice, would be followed through time with the best possible quality and new item adjustment procedures. The manner in which these samples would be chosen would be a matter for technical investigation, but we might make the following general observations:

1. The fact that two samples are to be selected does not mean that each must be equal in size to the desired overall sample. Rather, one would probably make each of them one-half the size of the desired overall sample. Thus a total sample of some 300 commodities would be drawn as two independent samples, each consisting of some 150 items.

2. As was argued in Section V, one would attempt insofar as possible to draw these samples in accordance with known probability models. This would probably mean that items would be highly stratified, most likely into 150 strata, and that two independent drawings would be made in each stratum with probability proportionate to weights provided by the Consumer Expenditure Surveys, and with replacement. The strata could, of course, be defined by making use of every available bit of information about substitutability, similarity of price movements, and the like, and the suggestions offered by Adelman (*17*) and Banerjee (*16*) should be thoroughly studied in making these selections from within strata. The very least that one might expect is that two groups operating in a completely independent fashion each choose a sample of 150 items from the defined strata.

3. As a final point we observe that these two samples would undoubtedly have items in common. In particular, if some strata were defined to have only a single specified-in-detail item, then this item would, of necessity, be in both samples. We shall henceforth refer to these two commodity samples as C_1 and C_2.

Now consider the two cities, say A and B, which are drawn out of a single stratum or out of the two strata which are to be collapsed into a single stratum for variance computations. There is nothing in the present procedure of combining city indexes into the U.S. index which necessitates having the same sample of commodities in each city. Let us therefore assign sample C_1 to A and sample C_2 to B and thus obtain estimates of these two city indexes, say $\hat{R}_A^{(t)}$ and $\hat{R}_B^{(t)}$, in the ordinary manner. Then an estimate of the within-stratum variance which is based upon a comparison of $\hat{R}_A^{(t)}$ and $\hat{R}_B^{(t)}$ will be influenced not only by the sampling of cities and the sampling of retail outlets within cities, but also by the sampling of commodities, and this influence will remain when one combines variance estimates across strata.

It is assumed that the sample of retail outlets in a city is chosen in accordance with a known probability model. There appears to be no reason why this cannot be done, and the mere fact that such a sample may be small is no reason for not using an appropriate probability model approach. This is the only way that one can guarantee an unbiased estimate of average prices or of average price relatives for a city.

This "half sample" approach does provide an overall estimate of the within-stratum variance, including a contribution for the sampling of commodities. It does not, however, permit one to estimate the components of this variance. If it is possible to use both samples of commodities in each city—and this will have to be recommended for the self-representing cities in the sample—then it should be possible to separate these components. For example, consider a stratum made up of an extremely large number of cities of equal weight from which two, say A and B, are selected. In each city an index estimate is prepared using commodity sample C_1 and a separate estimate using C_2, where an independent sample of retail outlets is used for each of the four indexes. The results can then be presented in a fourfold table:

Commodity Sample		City A	City B
	C_1	$\hat{R}_{1A}^{(t)}$	$\hat{R}_{1B}^{(t)}$
	C_2	$\hat{R}_{2A}^{(t)}$	$\hat{R}_{2B}^{(t)}$

A particular index, say $\hat{R}_{1A}^{(t)}$, can now be viewed as

$$\hat{R}_{1A}^{(t)}=R^{(t)}+c_1^{(t)}+s_A^{(t)}+e_{1A}^{(t)}$$

where $R^{(t)}$ is the true stratum index, $c_1^{(t)}$ is an effect due to this particular sample of commodities, $s_A^{(t)}$ is an effect due to city A, and $e_{1A}^{(t)}$ is the effect due to a particular sample of retail outlets. This is essentially an analysis of variance, random effects model, for a two-way classification without interaction. Therefore one can easily estimate from the data not only the variance of the stratum sample index, but also σ_C^2, which is the variance due to the sampling of commodities, σ_s^2, which is the variance due to the sampling of cities, and σ_ϵ^2, which is the variance due to the sampling of retail outlets within cities. It would appear that the integration of some such simple design features as this into the ongoing operations of the Consumer Price Index would provide a large amount of data concerning the accuracy of the index at relatively low cost. These estimates will naturally be subject to a large amount of sampling variability, each being based on only a single degree of freedom, but it might be possible to combine them across strata or across time periods and thus improve their reliability.

As a final point, we turn to those cities which are self-representing. Estimates of sampling error can now be obtained only by replication within each of these cities. Thus suppose that indexes are obtained from each of the two samples of commodities, C_1 and C_2, where an independent sample of retail outlets is used for each sample of commodities. Then a comparison of the two resulting indexes, $\hat{R}_1^{(t)}$ and $\hat{R}_2^{(t)}$, will provide an estimate of the variance of the city index which would be the average of these two indexes. In this instance it would not be possible to estimate the components of variance due to the sampling of retail outlets. This could only be accomplished by further replication, say by using at least two independent samples of retail outlets for each of the two samples of commodities.

The contents of this section have not been aimed at giving a detailed program of sample design for the index. Rather, they have been given as illustrations of the fact that it should be possible to obtain easily estimates of error and of the components of this error by appropriately choosing the various samples on which the index is based, without increasing the size of any of these samples.

VIII. Some Further Suggestions for Consideration

This report has taken the present form of the U.S. Consumer Price Index more or less for granted and has then argued that it is both possible and necessary to obtain and publish estimates of sampling error for the various components of the sample design, as well as for the overall U.S. index. Crude estimates of the various components of error suggest that, as far as the level of the overall U.S. index is concerned, too much effort is being expended on obtaining relatively large samples of commodities, cities, and retail outlets and too little effort on the evaluation of procedural error. Furthermore, as between commodities and other sources of sampling error, too much effort is devoted to the sampling of cities and the sampling of retail outlets within cities. But definite conclusions on these matters can come only from a program of research carried out parallel with, and yet separate from, the actual day-to-day operations of index construction.

Not only would one expect to obtain firm estimates of error from an investigation of this kind, thus leading to better allocation of resources among the components of the present design, but such an investigation might also lead to recommendations for major changes in the construction of the Consumer Price Index. Two areas which seem worthy of special attention are the following:

1. Index numbers of the Laspeyres type have traditionally been based upon a market basket of commodities which remains essentially unchanged between major weight revisions, except for adjustments which are made to account for the changing quality of items in the market basket. As noted in Section IV, Adelman has advocated drawing a completely new sample of commodities at fixed intervals, together with a chain approach for obtaining comparisions over longer periods of time, but this approach seems unlikely to be adopted by the producers of index numbers. It should, however, be possible to effect a compromise between these two extremes and thus gain some of the advantages of each. Thus one could set up a rotation schedule so that each item remains in the index for some fixed period of time, say, one, two, or three years, and so that a fixed fraction of the items are replaced each month, quarter, or year by newly selected items. This type of sampling has been successfully applied to situations where the same population is sampled on successive occasions—Coch-

ran (*15*, pp. 282–290) and Hansen, Hurwitz, and Madow (*14*, pp. 490–503)—and it might well be adaptable to commodity sampling for index numbers. In particular, this would give "new" items (i.e., not in existence at the time of the original selection) a chance to come into the index without giving up all the features of a Laspeyres index, and would also give "old" items a chance of being dropped before they became entirely obsolete.

In attempting to adapt partial replacement to the sampling of commodities for an index number, there are many problems and points to be kept in mind. Among these are the following: If a population is fixed and the goal is to estimate month-to-month changes, then the "best" procedure is to keep the same sample. However, if the goal is to estimate the actual level, then the "best" procedure is usually to replace some fraction of the sample. In the case of index numbers, the real goal is probably a mixture of the two and a compromise would be required. Replacement procedures would also depend upon cost considerations and, in view of specification problems, it would probably always be more expensive to replace an item than it would be to retain it. Furthermore, it might be necessary to make some changes in the Laspeyres concept to take account of the fact that it would be difficult to obtain base period prices and specifications for items brought in some years after the base period.

2. The Consumer Price Index is basically city-oriented. That is, indexes are computed for each city in the city sample, and these indexes are weighted to obtain the U.S. index. This emphasis on city indexes does not appear to be the most efficient way of obtaining the U.S. index. If one views the index in terms of U.S. average weights and average prices, then it is clear that quite a different sample should be used, for example, to obtain a "good" estimate of the average price of a newspaper than would be used to obtain a "good" estimate of the average price of a used car or of a woman's coat. In other words, the size of the "best" city sample for an item depends upon the cost of obtaining a price quotation and upon the variability of the item's price from city to city, and thus the size of the "best" city sample will differ considerably from item to item. It is recognized that aggregation according to a Laspeyres index calls for price quotations to be weighted in proportion to population and to value, and that a complete set of value weights could not possibly be obtained for all cities in which one would be able, for example, to collect newspaper prices. This difficulty might be overcome, for example, by deriving the Consumer Expenditure weights for the population of cities in a region rather than for a number of individual cities in the region. An added benefit of such a change in emphasis might well be that it would become more feasible for the BLS to employ selected data from other sources in the index computations, e.g., from the Monthly Retail Trade Report of the Bureau of the Census.

References

1. Erland von Hofsten, *Price Indexes and Quality Changes*, Bokförlaget Forum AB, Stockholm, 1952.
2. Sidney A. Jaffe, "The Consumer Price Index—Technical Questions and Practical Answers," a paper presented at the Annual Meeting of the American Statistical Association, 1959.

3. Richard Stone, *Quantity and Price Indexes in National Accounts*, The Organization for European Economic Co-operation, Paris, 1956.
4. Dorothy S. Brady and Abner Hurwitz, "Measuring Comparative Purchasing Power," *Studies in Income and Wealth*, Vol. 20, Princeton University Press, Princeton, 1957.
5. William A. Neiswanger, a report on U.S. Price Indexes prepared for the Bureau of the Budget and distributed to members of the Committee.
6. Helen Humes Lamale, *Methodology of the Survey of Consumer Expenditures in 1950*, University of Pennsylvania, Philadelphia, 1959.
7. Bruce D. Mudgett, *Index Numbers*, John Wiley & Sons, Inc., New York, 1951.
8. U.S. Department of Labor, *Techniques of Preparing Major BLS Statistical Series*, Bulletin 1168, 1954.
9. U.S. Government Printing Office, *Report of the President's Committee on the Cost of Living*, Washington, 1945.
10. U.S. Department of Agriculture, *Major Statistical Series of the U.S. Department of Agriculture*, Agricultural Handbook No. 118, 1957.
11. B. R. Stauber, R. F. Hale, and B. S. Peterson, "The January 1959 Revision of the Price Indexes," *Agricultural Economics Research*, XI (1959), 33–80.
12. H. Theil, *Linear Aggregation of Economic Relations*, North-Holland Publishing Company, Amsterdam, 1954.
13. William H. Kruskal and Lester G. Telser, "Food Prices and the Bureau of Labor Statistics," *Journal of Business*, XXXIII (1960), 258–279.
14. Morris H. Hansen, William N. Hurwitz, and William G. Madow, *Sample Survey Methods and Theory*, Vol. 1, John Wiley & Sons, Inc., New York, 1953.
15. William G. Cochran, *Sampling Techniques*, John Wiley & Sons, Inc., New York, 1953.
16. K. S. Banerjee, "Precision in the Construction of Cost of Living Index Numbers," *Sankhyā* 21 (1959), 393–400.
17. Irma Adelman, "A New Approach to the Construction of Index Numbers," *The Review of Economics and Statistics*, XL (1958), 240–249.
18. Erland von Hofsten, "Price Indexes and Sampling," *Sankhyā*, 21 (1959), 401–403.
19. Frederick F. Stephan and Philip J. McCarthy, *Sampling Opinions*, John Wiley & Sons, Inc., New York, 1958.
20. William G. Cochran, Frederick Mosteller, and John W. Tukey, *Statistical Problems of the Kinsey Report*, The American Statistical Association, Washington, 1954.
21. U.S. Department of Labor, *Consumer Prices in the United States, 1953–58*, Bulletin 1256, 1959.
22. Marvin Kogan, "Selection of Cities for Consumer Expenditures Survey, 1950," *Monthly Labor Review*, April 1951.
23. P. V. Sukhatme, *Sampling Theory of Surveys with Applications*, Iowa State College Press, Ames, 1954.