Volume Title: Price and Quantity Trends in the Foreign Trade of the United States

Volume Author/Editor: Robert E. Lipsey

Volume Publisher: UMI

Volume ISBN: 0-870-14154-6

Volume URL: http://www.nber.org/books/lips63-1

Publication Date: 1963

Chapter Title: Sampling Characteristics and Accuracy of Index Numbers

Chapter Author: Robert E. Lipsey

Chapter URL: http://www.nber.org/chapters/c1891

Chapter pages in book: (p. 110 - 127)

CHAPTER 5

Sampling Characteristics and Accuracy of Index Numbers

Sampling Problems in the Construction of Price Indexes[1]

QUESTIONS of sampling procedure almost always arise in the construction of price and other index numbers, but are rarely treated explicitly. They are, in fact, obscured by the use of index number terminology. Our consideration of sampling problems first arose in setting up standards for commodity classification. We also wished to say something about the accuracy of our indexes beyond the usual warnings that they must be used with care. Many of the decisions to be made in designing the indexes and the questions to be answered in appraising them were closely analagous to problems of sampling design and the measurement of sampling error. We have attempted, therefore, to translate our problems into a simplified sampling terminology.

A THEORETICAL DESCRIPTION OF SAMPLING FOR A PRICE INDEX

Suppose that, ignorant of the vast index number literature and unable to collect every price, one set out to measure the average change in prices between two dates.

The first procedure to come to mind might be to list all the commodities, choose from among them in some random fashion and strike an average of the price ratios, weighting them all equally. But this method is clearly unsatisfactory—the classification of commodities is arbitrary, and, there-

[1] The first part of this chapter is an expanded version of a paper on "Some Sampling Problems in the Construction of Price Indexes" read at the Annual Meeting of the American Statistical Association, December 1955. Several substantial discussions of this subject have since appeared, each treating it from a slightly different viewpoint and, in some cases, giving evidence that random sampling is a practical possibility. The following are some of the main contributions: Irma Adelman, "A New Approach to the Construction of Index Numbers," *The Review of Economics and Statistics*, August 1958; K. S. Banerjee, "Calculation of Sampling Errors for Index Numbers" *Sankhya*, January 1960, and "A Comment on the Sampling Aspects in the Construction of Index Numbers," *The Review of Economics and Statistics*, May 1960; two staff papers of the NBER Price Statistics Review Committee: Philip J. McCarthy, "Sampling Considerations in the Construction of Price Indexes with Particular Reference to the United States Consumer Price Index," and Victor Zarnowitz, "Index Numbers and the Seasonality of Quantities and Prices," published in *The Price Statistics of the Federal Government*, New York, National Bureau of Economic Research, 1961.

fore, the frequency with which any group of commodities is represented in such a selection depends on the fineness with which the group has been broken down, rather than on its importance. Each commodity would have an equal chance of being represented, but not each dollar of trade. If each commodity is thought of as a cluster of transactions, this procedure is one in which samples of equal size are drawn from each commodity cluster, even though some clusters are much larger than others. The probability of inclusion in the sample for a given dollar of trade, as well as the sampling fraction, would be inversely proportional to the size of the cluster.

What is needed is a method by which we can dip at random into the stream of trade, giving each dollar of transactions an equal opportunity to be represented in the sample, and, therefore, giving each commodity or group of commodities representation in proportion to the value of its trade. This might be achieved if the number of times a commodity appeared on the list was proportional to its importance (as measured by base-year value, given-year value, or some combination of the two, the choice depending on the type of index number used). Such a method would be equivalent to choosing from a list of dollars of trade, rather than commodities, and it would give each dollar of trade an equal chance of inclusion.

Of course this would be even more impractical than our first list of commodities. The same results could be achieved by selecting commodities from the first list and then weighting each price ratio by the importance of the commodity it represents. If we assume that all of the price ratios for a given commodity are identical (or that the sample of dollars of trade in that commodity would give an unbiased estimate of the mean or index for that commodity), the weighting achieves the same result as taking equal sampling fractions for each commodity. The equality of sampling fractions insures equal probability of inclusion for each dollar of trade.

The size (or importance) measure can be easily described for the Paasche, Laspeyres, Marshall-Edgeworth, and several other indexes. In the case of the Laspeyres price index, for example, it is the base-year value of (trade or exports in) the commodity. For the Paasche index it is the base-year price multiplied by the quantity in the year being compared with the base year. And for the Marshall-Edgeworth index it is the average of the Paasche and Laspeyres weights. Each of these can be put in the form: index $= \Sigma ab$, where a is the weight, the ratio of the size (e.g., value) of the commodity to the total for all commodities, and b is the ratio of given-year price to base-year price. The Fisher index cannot be represented in this way,[2] but its weights can be approximated by those of the Marshall-Edgeworth index.

We have discussed, so far, only simple sampling procedure, but we know, from such studies as those of Mitchell and Mills,[3] that prices can be divided into groups which show distinctly different cyclical or trend characteristics. For this reason, a stratified rather than a simple random sample would improve the accuracy of our estimate of the mean. We should distinguish, to cite Mitchell's classification, crude from manufactured, agricultural from nonagricultural, animal from vegetable and from mineral, and consumer from producer goods. It is advisable to make even finer distinctions if groups within these strata differ significantly in the characteristics which interest us.

Stratification involves breaking the universe into several subuniverses, sampling within each as before, and then giving each mean (that is, price index) the weight of the subuniverse, or stratum, to which it refers, instead of the weight of the commodities selected. Stratification will increase the precision of our estimate of the mean even if we take a proportional sample (which, on the average, produces the same sampling fractions as a simple random sample) by insuring the proper weight for each stratum in each sample, instead of only on the average among all samples. Stratification also opens another avenue towards increased precision: the more variable groups can be sampled more heavily than the less variable ones. Proportionate sampling can be described as that in which $\frac{n_h}{\Sigma n_h}=\frac{N_h}{\Sigma N_h}$, where n_h is the number in the sample from a stratum and N_h is the total number in the stratum. Optimum sampling, cost factors aside, is such that $\frac{n_h}{\Sigma n_h}=\frac{N_h S_h}{\Sigma N_h S_h}$, where S_h is the standard deviation for the stratum.[4] An optimum allocation shifts the sample from the less to the more variable strata.

ACTUAL SAMPLING PROCEDURES IN PRICE INDEX CONSTRUCTION

It is obvious that the preceding paragraphs are not a description of the way in which price indexes are presently computed. In particular, the

[2] The Fisher index is a square root and can therefore be irrational. But Σab must be rational, because the a's and b's are fractions, and their products and the sums of their products must therefore be rational.

[3] Wesley C. Mitchell, "Index Numbers of Wholesale Prices in the United States and Foreign Countries," BLS *Bulletin 284*, 1921; and Frederick C. Mills, *The Behavior of Prices*, New York, NBER, 1927.

[4] Morris H. Hansen, William H. Hurwitz, and William G. Madow, *Sample Survey Methods and Theory*, New York, 1953, Vol. I, p. 209.

selection of prices for inclusion in the indexes is not made by random methods. Instead commodities are chosen to obtain the greatest coverage at the least cost. A selection may be made, for example, of a number of the most important items,[5] or of those in which trade is greater than a given amount, or perhaps of a sufficient number of items to reach a specified portion of the total.

Such methods may rest on the assumption that the value of trade in a commodity is not correlated with price behavior. Unfortunately, this is not true. Most of the commodities of large value are crude or semimanufactured materials or foodstuffs. Commodity classes for manufactures tend to be relatively small.[6] Since the price behavior of manufactured goods differs from that of foods and materials, selection by amount of trade tends to bias the index towards the behavior of crude products.

Random selection is hampered, even for those agencies which collect their own price data, by ignorance of those properties of the universe which would be needed to guide sampling procedure.[7] For those working with already collected data such as foreign trade reports, the problem of nonresponse is the main obstacle. That is, for the great majority of commodities listed in the U.S. customs returns, either no data on quantities (and unit values) are given at all, or the commodity titles are amalgamated into groups so heterogeneous that the unit values cannot be treated as prices. Because most commodity categories give no information on price changes, index number compilers are often led to use whatever is available without worrying about possible biases.[8]

[5] This was the case, for example, with the import and export price indexes computed by Theodore J. Kreps, "Import and Export Prices in the United States and the Terms of International Trade, 1880–1914" *Quarterly Journal of Economics* August 1926. The Department of Commerce indexes are described as including directly "all leading commodities for which quantities are available and which show a reasonable degree of homegeneity . . .," U.S. Department of Commerce, Bureau of Foreign and Domestic Commerce, *Foreign Trade of the United States, 1936–49*, GPO, 1951, note to Table 10, p. 6. See also Dorothy S. Brady and Abner Hurwitz, "Measuring Comparative Purchasing Power" *Problems in the International Comparison of Economic Accounts*, Studies in Income and Wealth, Volume Twenty, Princeton University Press for the NBER, 1957.

[6] Since the commodity classification is arbitrary, these manufactured goods categories could be amalgamated into larger classes only at the cost of grouping together dissimilar articles. These groups would be so heterogeneous that changes in unit values could not be interpreted as price changes. Thus the selection problem would have been solved by producing what could be described in sampling terminology as a nonresponse problem. The large manufactured goods classes so created would not yield any meaningful price data. One reason for this difficulty is that in the manufacturing process a few types of raw cotton, for example, can be made into many types of cloth and these into uncountable varieties of clothing.

[7] See Brady and Hurwitz in, *International Comparison of Economic Accounts*, pp. 310–311. Their discussion relates mainly to international comparisons of price levels, but could apply almost as well to comparisons over time.

[8] See, however, the articles by Adelman, Banerjee, and McCarthy, referred to in Note 1.

The problem of nonresponse would not be troublesome if it were spread evenly over the commodity universe. But we find differences in price behavior between manufactured goods and crude materials, and between goods whose method of production is changing technologically and those whose technology is stable. The former of each pair are likely to show high rates of nonresponse which threaten to bias the index.[9]

STRATIFICATION TO MINIMIZE SELECTION AND NONRESPONSE BIAS

The possibilities of bias inherent in nonrandom sampling methods and in extensive nonresponse cannot be eliminated completely, but we can attempt to minimize their effects. As in reducing sampling error, the method is to stratify the universe by those attributes of commodities which we know to be related to price behavior. In addition, stratification by attributes which are related to nonresponse or selection bias, would eliminate some bias due to differences in nonresponse among strata, although not bias due to within-strata differences.[10]

There is no way of being agnostic with regard to the price behavior of any commodity. If the stratification has any validity, every commodity should be placed within some stratum. Omitting a commodity from the price index is equivalent to assuming that its behavior is that of the average of all included commodities. It would be illogical, for example, to treat machinery, which we know to be a durable, nonagricultural, producers' good as behaving like the average of all commodities if we have a durable vs. nondurable or a producers' vs. consumers' or an agricultural vs. nonagricultural product classification which reveals significant differences in price behavior.

[9] Some of these shortcomings in the BLS Wholesale Price Index of that period are discussed in Morris A. Copeland, "Some Suggestions for Improving our Information on Wholesale Commodity Prices," and Robert W. Burgess, "The General Structure of Wholesale Prices," both in *Proceedings of the Ninety-second Annual Meeting of the American Statistical Association*, 1931.

[10] The sampling problems involved in the construction of price indexes from data collected for other purposes are similar to those dealt with in Appendix G of *Statistical Problems of the Kinsey Report*, by William G. Cochran, Frederick Mosteller, and John W. Tukey (American Statistical Association, Washington, 1954). In both cases, the sample has not been drawn randomly, and it is therefore difficult to know exactly what the parent population is. The stratification described here is parallel, if it is performed after the sample has been drawn, to the process of "adjustment" of sample means described in that report. It can be thought of as a process by which the characteristics of the sample are compared with those of the population and the sample mean reweighted in accordance with the characteristics of the population. The constructor of price indexes has one advantage: there have been studies of the price universe which give some guidance as to which characteristics are significant for pre-sampling stratification or post-sampling adjustment.

It would be ideal to design the stratification scheme in advance, using knowledge about the behavior of prices gained from other studies. Such a stratification would reveal many empty classes, classes containing only commodities for which we have no price data, and would illuminate the areas where bias is most likely. We have usually made the best guess possible by amalgamating many such classes with those which seemed most closely related.

THE MEASUREMENT OF THE PRECISION OF PRICE INDEXES

Published price indexes have rarely been accompanied by estimates of sampling error, but some independent estimates have been attempted. With the exception of those in the articles by Adelman and Banerjee mentioned earlier, they have probably exaggerated the accuracy of the indexes.

A. L. Bowley, in 1924[11] made some measurements of the sampling error of Sauerbeck's index, published in the *Statist*. His method indicated coefficients of variations (standard error ÷ mean), of 1.6 to 3.4 per cent for the 1899-1913 period (forty "independent" price series), and 4.6 to 6.0 per cent for the 1913, 1919-22 period (thirty-nine "independent" price series).[12] Frederick C. Mills[13] made more extensive investigations of this subject, estimating coefficients of variation for eight of his own index numbers. The coefficients for the fixed-base indexes, which were in every case larger than those for the corresponding link relatives, had the following ranges:

	1891–1913	1914–26
Unweighted arithmetic mean	.8-2.1	.7-4.7
Unweighted geometric mean	.8-1.8	.6-1.8
Weighted arithmetic mean	1.4-3.4	.9-3.0
Weighted geometric mean	1.4-3.4	1.0-3.1

If the confidence interval is measured by twice the coefficient of variation, these figures indicate ranges of error of 3 to 12 per cent for the *Statist* index. For the Mills indexes, the ranges are 1.5 to 7 per cent in the prewar period and 1.2 to 9.5 per cent in the later years (even though the series covers 200 to 400 commodities).

11 *Relative Changes in Price and Other Index-Numbers.* London and Cambridge Economic Service, Special Memorandum No. 5, Feb. 1924, pp. 6–8.

12 Bowley computed probable errors of the means for only one year. We extended the computation to the remaining years using his method and his data, and increased the probable errors by 50 per cent to approximate standard errors.

13 *Behavior of Prices*, pp. 240–274.

Fisher did not publish any extensive calculations on actual index numbers, although he recognized the existence of sampling problems. For a 200-commodity index he compiled from *Dun's Review,* he suggested a probable error of 1.5 per cent,[14] which would imply a standard error of slightly over 2 per cent. Mudgett[15] presents the formulas for the standard error of the mean (i.e., the index), both weighted and unweighted, with and without the finite sampling correction, and for stratified as well as unstratified sampling. He points out that stratification can be effective in reducing the sampling variability of the average, but he does not discuss its use to minimize the effects of bias in selection.[16] He is therefore led to say of the BLS Wholesale Price Index, which has for some years contained over 800 items, "It might even be possible to say that such a comprehensive index is practically devoid of sampling error."[17] Since Mudgett mentions the total number of items, it would appear that for this purpose he is treating the BLS index as if it were constructed from a simple random sample.

STRATIFICATION AND THE MEASUREMENT OF SAMPLING ERROR

We suspect that most of the preceding estimates of sampling error are too low because they assume simple random sampling, and, therefore, probability of representation proportional to size. In fact, there are serious differences in representation, and the groups which are poorly represented are not necessarily those with low dispersion.[18] The total number of items included in an index is clearly not significant without some information about the distribution (consider, for example, a 100-item index where ninety-eight of the items were drawn from one identifiable half of the population and only two from the other).

The error caused by combining in the same stratum groups which differ in the extent of coverage (or nonresponse) can be illustrated by the following example. Suppose that we can stratify a population into two groups that are equal in size (N_h) but differ in the extent of coverage (or probability of inclusion in the sample). Let us say that they differ to the extent that the number of commodities in the sample from one group (Kn_h) is K times the number from the other group (n_h).

[14] Irving Fisher, *The Making of Index Numbers,* Boston, 1922, p. 340.

[15] Bruce D. Mudgett, *Index Numbers,* New York, 1951, pp. 51–54.

[16] Mudgett does observe that it is often exceedingly difficult to draw a random sample. *Ibid,* p. 53.

[17] *Ibid.,* p. 54.

[18] It might be that the poorly covered groups, since they are frequently manufactured products, have a large proportion of sticky prices and therefore small dispersion of price changes over short periods. But this would not be likely for price trends over longer periods.

The variance of the sample mean $(\sigma_{\bar{x}}^2)$ from a stratified sample can be written as $\frac{1}{N^2}\Sigma\left[N_h^2\frac{S_h^2}{n_h}\right]$ where $N = \Sigma N_h$ and S_h^2 is the variance within a stratum.[19] In our example, with the two strata described above, this variance $(\sigma_{\bar{x}}^2)$ becomes

$$\frac{1}{(2N_h)^2}\left[N_h^2\frac{S_h^2}{n_h} + N_h^2\frac{S_h^2}{Kn_h}\right]$$

which reduces to $\frac{S_h^2}{n_h}\cdot\frac{1+K}{4K}$.

But suppose we had combined these two strata into a single one and had treated the stratified sample as if it were a simple random sample. Our estimate of the variance of the mean would have been $\frac{S^2}{n}$ where

$$S^2 = \frac{n_h S_h^2 + Kn_h S_h^2}{n_h + Kn_h} \text{ and } n = n_h + Kn_h$$

This estimate of the variance reduces to $\frac{S_h^2}{n_h}\cdot\frac{1}{1+K}$. The ratio of the first, correct, estimate of the variance of the mean to the second, incorrect, one is $\frac{(1+K)^2}{4K}$. Or, in other words, the valid estimate of the standard error of the mean (or index) would be $\frac{K+1}{2\sqrt{K}}$ times the estimate derived by treating the sample as random, as was done by Bowley and Mills and, implicitly, by some of the others mentioned above.

For small values of K the understatement of the standard error is not large; at $K=2$ it is about 6 per cent. It rises to 14 per cent for $K=3$, 20 per cent for $K=4$, and 40 per cent for $K=9$.

This ratio would be higher if it took into account the case where n in one stratum is so small that it should be treated as a small sample.

Thus another important reason for stratification emerges : without it we cannot make any reasonable estimate of the sampling error of the index. It is true that the stratification which would be optimum for increasing the precision of the estimate of the mean and for reducing bias in that estimate (one based on homogeneity with respect to the mean, or price behaviour) would not be the optimum stratification for estimating the sampling error of the mean. The latter would be one which revealed the greatest differences in coverage (probability of inclusion) among strata;

19 Hansen, Hurwitz, and Madow, *Sample Survey Methods*, p. 189.

that is, which grouped together types of commodities whose degree of coverage was similar. But a detailed stratification for the former purpose is likely to reveal many of the differences in coverage relevant to the latter.

MEASURES OF VARIABILITY AND SAMPLING ERROR IN THE NBER INDEXES

We have performed measurements of variability and sampling error in two ways. The first is appropriate when a weighted index is used to deflate the value of the uncovered items. It treats the covered items as if they had actually been picked with probability proportional to size. In other words, it assumes that the commodity distribution of the covered items is representative of the uncovered ones as well—that a large item represents a greater number of observations of the mean than a small one. The variance and other measures (Appendix Tables E-1 through E-3) are computed by weighting each price ratio by the size of the commodity.

There are certainly grounds for uneasiness about this method of estimation, since we are not sure of the representativeness of the sample. If, for example, the covered items in a class are dominated by a single large item which is not outstandingly important among the uncovered commodities, we are likely to have underestimated the margins of error. This danger is increased by the fact that we assume no within-commodity variance even though we know there must be some.

For these reasons, we computed, as a rough check, a second estimate of the standard error which treats each commodity, regardless of size, as a single observation. The standard error is thus estimated from an unweighted variance of the price ratios. Only the first step in these computations, the calculation of unweighted standard deviations, is shown here (Table E-1), but the relation between unweighted and weighted standard errors can be inferred from this table. The counterpart of this assumption in the index computations would be the deflation of the uncovered items by an unweighted rather than a weighted index of the covered items.

It would be possible to find from such computations that the margins of error surrounding the indexes were tolerably small even where only a small fraction of all the items were sampled, provided we were willing to assume the randomness of the sampling, and had sufficiently large numbers of items included. However, given our assumption that the covered items are free from sampling variation, these measurements exaggerate the range of error, for sampling error applies only to that part of each class which consists of uncovered items. To estimate the variability of the whole group we made a finite sampling adjustment, multiplying the variance of the mean by one minus the coverage ratio. These computations yield the adjusted measures in Appendix Tables E-2 and E-3.

The coverage ratio itself is often used as a measure of the reliability of an index.[20] The usual practice is to set a minimum level of coverage below which an index is considered too unreliable for use.[21] The logic of this criterion is that, given the degree of variation among the covered items, the standard error of the index varies directly with the noncoverage ratio.

Measures of sampling error take account of both the coverage ratio and the variability of the covered items. Thus a maximum level of error, rather than minimum coverage which is only a proxy for it, can be established as a criterion for acceptance of the index.[22] One index with a fairly low coverage may be acceptable if the price behavior is homogeneous and there are many items, while another with higher coverage may be rejected because it contains heterogeneous price behavior and few items.

Table 17 summarizes the sampling error measurements for NBER minor classes. It is evident from the coefficients of variation how important the finite sampling (or coverage) adjustment is to the reliability of the indexes. The unadjusted coefficients were frequently quite high; almost a third of the export and half of the import classes which contained more than one covered commodity showed coefficients of more than 10 per cent, and more than one out of ten had coefficients above 20 per cent. These figures exclude, however, all the classes in which there is no variability (those consisting only of one commodity) and those in which variability is unknown because none or only one of the commodities is covered.

Once the coverage adjustment is made (Columns 2 and 4) the minor class indexes appear more reliable. Of the 120 cases where unadjusted coefficients were over 10 per cent, only eight of forty-six remain on the export side and sixteen of seventy-four on the import side. If completely covered one-commodity classes are included, approximately 40 per cent of all the coefficients are zero and over half are 2 per cent or less.

The sampling variability of the five major classes which correspond to

[20] For example, in John H. Adler, Eugene R. Schlesinger, and Evelyn Van Westerborg, *The Pattern of United States Import Trade Since 1923,* Federal Reserve Bank of New York, 1952; in descriptions of the official Department of Commerce quantity and unit value indexes for U.S. exports and imports; and in Solomon Fabricant, *The Output of Manufacturing Industries,* New York, NBER, 1940.

[21] Fabricant, for example, did not accept indexes whose coverage was less than 40 per cent (*Ibid.,* pp. 34–35).

[22] Fabricant in *Output of Manufacturing Industries,* pp. 362–367, presented some calculations showing the effects on his indexes of various degrees of divergence between the price movements of covered and uncovered items, but gave only very general indications of the likelihood of each degree of divergence.

TABLE 17

SIZE DISTRIBUTION OF WEIGHTED COEFFICIENTS OF VARIATION: MINOR CLASSES (EARLIEST YEAR OF EACH PERIOD)

Coefficient of Variation	*Exports* Unadjusted (1)	Adjusted (2)	*Imports* Unadjusted (3)	Adjusted (4)
	Classes Containing More Than One Covered Commodity			
0	3	47	1	36
.001 – .020	17	41	9	28
.021 – .040	23	28	20	22
.041 – .060	27	14	17	23
.061 – .080	18	9	14	13
.081 – .100	18	5	15	12
.101 – .120	6	3	16	3
.121 – .140	4	1	14	5
.141 – .160	9	1	7	2
.161 – .180	7	0	10	2
.181 – .200	4	2	9	0
.201 – .250	8	1	6	2
.251 – .300	2	0	4	0
.301 – .400	3	0	3	2
.401+	3	0	5	0
Total	152	152	150	150
	Classes Containing Only One Covered Commodity			
Complete coverage	43	43	68	68
Incomplete coverage	22	22	39	39

SOURCE: Appendix Table E-3.

Commerce Department economic classes is summarized in Table 18 (and described in greater detail in Appendix Table E-4). Coefficients of variation for imports are larger than those for the corresponding export classes —sixteen out of twenty times. The coefficients for finished manufactures are generally high; those for food classes are low, with the exception of Import Class 201 (crude foods) in 1899. The size of this coefficient is due mainly to one small minor class, Import Class 006 (spices), in which the three covered items were so divergent in behavior as to give a standard error of estimate of .44 before finite sampling adjustment and .23 even after coverage is taken into account.

On the whole, the errors seem tolerable. None of the coefficients of variation exceeds 3.5 per cent; none outside of manufactures is greater than 2.3 per cent. Seventy per cent of the total and 80 per cent of those outside finished manufactures were under 2 per cent. The coefficients are large enough, however, to suggest that it would be useful to experiment with random selection to produce more valid variability estimates.

TABLE 18

COEFFICIENTS OF VARIATION FOR SELECTED MAJOR CLASS PRICE INDEXES[a]

Economic Class	Year	Exports (%)	Imports (%)
Crude foods	1879	.1	1.0
	1889	.5	.6
	1899	.8	2.3
	1913	.8	1.1
Manufactured foods	1879	.5	1.4
	1889	.7	.9
	1899	1.2	.9
	1913	.9	.7
Crude materials	1879	1.3	1.4
	1889	.8	1.0
	1899	.8	1.2
	1913	1.3	1.5
Semimanufactures	1879	2.3	1.7
	1889	2.1	2.3
	1899	1.0	2.1
	1913	1.1	1.0
Finished manufactures	1879	1.8	2.4
	1889	2.0	3.5
	1899	2.0	3.3
	1913	2.6	2.7

SOURCE Variances from Appendix Table E-4; indexes can be calculated from Tables A-1 and A-3.

[a] The classes included are those equivalent to the five Department of Commerce economic classes.

Extent of and Changes in Coverage

Coverage ratios are interesting not only as crude measures of accuracy but also because they reflect differences, between covered and uncovered items, in price behavior and in supply and demand elasticities. Although it is rarely possible to disentangle these factors, radical changes in coverage, when the commodity list is unchanged, are grounds for suspecting heterogeneity in a commodity class. This is especially true where the changes in the coverage ratios are correlated with changes in the price index; it would appear likely in such a case that the price changes in the covered items were not duplicated in the uncovered ones. This correlation is not conclusive evidence of divergences in price behavior, however. It could result from differences in elasticity of demand. Suppose, for example, a group in which covered and uncovered commodities were identical in price behavior but the former were subject to a much more elastic demand. Coverage would then decrease every time the group's prices rose and increase every time they fell. By the same reasoning we could say that differing elasticities could conceal the expected influence of differing price behavior on coverage ratios.

COVERAGE IN NBER FOREIGN TRADE INDEXES

Coverage ratios for minor classes are summarized in Table 19 below. There are almost 6,900 class-years (numbers of classes multiplied by the number of years each is available) for which indexes might have been computed (over 5,500 indexes actually were calculated). Of the 6,900 class-years over 40 per cent consisted almost completely (more than 95 per cent) of covered commodities, and could therefore be said to suffer from virtually no sampling error. At the other extreme, for over 19 per cent of the class-years no coverage was possible or so little that no indexes were calculated. This group of empty classes was particularly important in the earliest period : 31 per cent for exports and 29 per cent for imports. Another 7.5 per cent of the class-years are of marginal quality, with coverage of less than 50 per cent. Most of these, particularly in the lowest ranges, occur in periods in which the majority of years had adequate coverage.

In every period, the proportion of classes more than 95 per cent covered was slightly higher in exports than in imports. But the better coverage in exports disappears at a somewhat lower standard : imports show a higher proportion with coverage above 60 per cent, and a smaller proportion completely uncovered in every period.

Among those groups for which indexes were calculated, over half the class-years had coverage ratios above 95 per cent. Exports had a higher proportion than imports in that class in every period, but even for imports, at least 45 per cent of the class-years had coverage ratios over 95 per cent.

Measurements based on numbers of class-years do not take into account differences in the importance of individual classes. They therefore present a very conservative assessment of the indexes, since many of the largest classes (for example, cotton, grain, and tobacco exports, and coffee, tea, cocoa, and sugar imports) consist entirely or almost entirely of covered items. Measured by number or value, the coverage ratios tend to be exaggerated in classes where prices were used in place of unit values. The price series describe narrowly defined commodities but are applied here to much

TABLE 19

COVERAGE RATIOS FOR MINOR CLASSES

	Exports and Imports	*Exports*					*Imports*				
	All Periods	1879–88	1889–98	1899–1912	1913–23	All Periods	1879–88	1889–98	1899–1912	1913–23	All Periods
					PERCENTAGE DISTRIBUTION						
Coverage ratio (%)											
95 to 100	42.5	39.6	45.6	45.7	46.9	44.7	38.5	42.6	41.7	39.3	40.6
90 to 95	4.7	2.9	4.6	4.4	6.5	4.7	4.2	5.1	3.6	6.1	4.7
80 to 90	8.5	8.4	6.1	9.4	8.4	8.2	6.8	8.1	8.5	10.6	8.7
70 to 80	7.3	4.8	3.0	8.0	7.6	6.2	8.2	7.2	8.2	9.0	8.2
60 to 70	6.1	2.3	3.4	4.4	5.8	4.1	5.0	5.5	9.4	9.5	7.7
50 to 60	4.1	4.1	4.4	2.3	4.5	3.7	3.2	2.7	5.9	4.6	4.3
40 to 50	3.8	4.3	4.2	6.1	3.8	4.7	1.1	3.1	4.6	2.4	3.1
30 to 40	2.2	1.0	2.4	1.7	1.8	1.7	2.2	1.4	3.3	3.3	2.7
1 to 30	1.5	1.6	1.2	1.5	0.1	1.2	1.9	1.2	2.0	1.7	1.8
0	19.5	31.2	25.1	16.6	14.5	20.8	28.8	23.0	12.9	13.4	18.3
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
					NUMBER OF CLASS-YEARS						
Total	6,887	657	676	1,014	844	3,191	694	827	1,195	980	3,696

broader categories. For example, a BLS series on "Cattle, steers, good to choice," is used here to deflate values of an export commodity defined only as "cattle". The price series, therefore, apply only to a part of the export values, and an unknown part at that. It would be more appropriate (but much more laborious) to use a combination of several cattle series for the price index, attaching some measure of dispersion to it. Alternatively, one could count only part of the cattle series as contributing to the coverage in the class. Instead, as with the unit values, we assumed no variance within a commodity, and treated its whole value as a covered item.[23]

With these limitations in mind we may examine the coverage ratios for total exports and imports which appear in Table 20. These ratios were computed only for the earliest year in each period and for the comparable base-year figure. The earliest year of each period was used because it is generally the one with the poorest coverage.[24] For exports, coverage was above 85 per cent in each of the four periods, and for imports it fell no lower than 72 per cent.[25] Coverage of exports was highest in the earlier years and then declined as the improvement in commodity detail and in the availability of price data was offset by the decline in the importance of agricultural commodities for which both price and unit value data were plentiful. In the case of imports the shift in composition away from manufactured goods and the improvement in data led to a slight increase in coverage.

[23] This difficulty is involved in the problem of estimating from "composite commodities" discussed by Banerjee, "Calculation of Sampling Errors for Index Numbers."

[24] This may seem puzzling in view of the fact that the base year coverage shown in Table 20 is generally worse than that for the earliest year. Coverage is shown for a list of commodities that is unchanged during a period, and thus no advantage is taken of the availability of more data in later years. It is true that those commodities which were covered in 1879, for example, were a larger proportion of the total then than they were ten years later. But the commodities covered in 1888 were usually a larger proportion of the total in 1889 than were those covered in 1879. In other words, total coverage increased through time but the importance of the group of commodities covered initially usually decreased.

[25] Coverage of the Department of Commerce import indexes has been close to 70 per cent except for a fall to 60–65 per cent in 1957–59. That of the export indexes was 55–67 per cent before World War II. Since then it has ranged between 35 to 50 per cent, averaging about 45 per cent (U.S. Department of Commerce, *Business Statistics, 1957 Biennial Edition*, p. 251, and later editions). The Federal Reserve Bank indexes covered 64 to 69 per cent of the value of imports (Federal Reserve Bank of New York, *The Pattern of United States Import Trade Since 1923*, by John H. Adler, Eugene R. Schlesinger, and Evelyn Van Westerborg, May 1952, p. 64). The degree of coverage in Fabricant's output indexes ranged from 52 to 70 per cent of total value added (Solomon Fabricant, *The Output of Manufacturing Industries*, p. 602).

TABLE 20
COVERAGE RATIOS FOR TOTAL EXPORTS AND IMPORTS

	1913–23		1899–1913		1889–99		1879–89	
	1923 (Comp. with 1913)	1913	1913 (Comp. with 1899)	1899	1899 (Comp. with 1889)	1889	1889 (Comp. with 1879)	1879
Exports	83.2	85.6	80.3	87.5	83.3	90.7	88.3	91.5
Imports	81.8	76.0	72.7	78.6	68.8	71.7	70.4	74.7

Appendix Tables E-5 to E-8 show intermediate and major class coverage ratios for the earliest year and the base year of each period. The base year coverage ratios shown include only those commodities covered in the earliest year. It is clear that the covered items are unevenly spread over the commodity universe. In exports, for example, the first twelve major classes, including all foods, crude materials, and agricultural exports, do not show a single case of coverage below 90 per cent. Import coverage was somewhat lower, but the first nine classes, consisting of foods and other agricultural products, included no cases under 86 per cent.

No major export class had less than 50 per cent coverage, and of those with between 50 and 70 per cent, twenty-six of twenty-seven cases were in classes 214, 215, 221, and 222.[26] One important component of all of these was Export Class 146 (manufactured metal products, including machinery and vehicles), whose coverage ranged between 33 and 66 per cent, mostly below 50 per cent. Among the 372 intermediate export classes listed in Table E-5, only eighteen had coverage ratios below 50 per cent (eleven among manufactured metal products) and nine others between 50 and 60 per cent.

Major import classes were more sparsely covered. There were thirteen cases below 50 per cent (as against none for exports) and nineteen between 50 and 60 per cent. But here again they were concentrated in the same area: thirty of thirty-two were in five classes.[27] Only once did coverage dip even slightly below 40 per cent.

The main sources of this poor import coverage are Import Class 150 (manufactured products of mineral origin) and its component, Import Class 147 (manufactured metal products), both of which contain very few covered items. Almost all the coverage of manufactured imports is in textiles and wood and paper products (Import Classes 064, 066, and 126).

[26] Manufactures, including tobacco products; manufactures, excluding tobacco products; mineral products; and nonagricultural products.

[27] (1) Nonagricultural products; (2) products of mineral origin; and (3) three classes of manufactured products.

Changes in the coverage ratios are of interest because they can suggest some inferences about the price behavior of uncovered items. They do this by virtue of the fact that they measure the relative rates of growth in value of covered and uncovered commodities. Where coverage is rising the covered commodities are growing more rapidly.

Especially among exports, relative value changes for major classes of commodities have tended to move in the opposite direction from relative price changes. Groups whose prices have fallen relatively have tended to gain in importance, for example, manufactured products in general and automobiles in particular. If this relationship is typical we can use these changes in coverage to draw some inferences as to the probable direction of bias in our indexes.

Some change in coverage arises from shifts in the importance of classes. For example, as we have seen, the rise within exports of the lightly covered manufactured goods class tended to lower total coverage. This change in coverage does not imply bias; it is taken account of in the construction of the index, as are any such changes arising from shifts in importance among minor classes. Shifts in importance within minor classes might suggest bias, however, because the method of constructing the indexes assumes that within each minor class prices of uncovered commodities move with those of covered commodities.

We therefore ask the following question: How does the value of covered commodities at the end of each period compare with what it would have been if the coverage in each minor class had remained constant at the earliest year's level? If actual coverage is greater, we know that covered commodities have increased in value more rapidly; if it is smaller, the uncovered items have been growing more rapidly.

Tables E-9 to E-12 show, for each intermediate and major class, actual coverage at the end of each period as a per cent of that which would have existed if there had been no changes within minor classes during the period. For total imports and total exports actual coverage is less than expected in three out of four periods, but never by more than 5 per cent. More significant lags in the growth of covered items appear among the major classes. In four major export classes, all among manufactures, non-agricultural products, and products of mineral origin, coverage within minor classes fell by more than 10 per cent. These classes, which fell in price and increased in value relative to other exports, show evidence of upward bias in the price index. That is, there is some ground for suspicion that their prices fell even more, relative to those of other classes, than is revealed by our indexes. The loss in coverage in these classes was concentrated particularly in Export Class 146 (manufactured metal products), and its main component, Export Class 143 (manufactured iron and steel products). These lost close to 50 per cent of their coverage over the four periods.

Changes in coverage among major import classes were much more scattered. There were six instances in which the growth of covered items exceeded that of uncovered items by more than 10 per cent and three over 20 per cent. (Only once did an export class show the value of covered items gaining on that of uncovered items by more than 4 per cent during one period.) All of these were among manufactured goods imports, as were three cases in which covered items fell behind by more than 20 per cent. The very low coverage in these classes left room for large increases and decreases, but in contrast to the situation on the export side, the net change in coverage was very close to zero.