Volume Title: National Income and Its Composition, 1919-1938, Volume II

Volume Author/Editor: Simon Kuznets, assisted by Lillian Epstein and Elizabeth Jenks

Volume Publisher: NBER

Volume URL: http://www.nber.org/books/kuzn41-3

Publication Date: 1946

Chapter Title: Characteristics of the Data and Procedures

Chapter Author: Simon Kuznets, Lillian Epstein, Elizabeth Jenks

Chapter URL: http://www.nber.org/chapters/c5549

Chapter pages in book: (p. 475 - 500)

Characteristics of the Data and Procedures

1 The Underlying Data

ON THE simplest possible basis the information that underlies our estimates may be classified into three broad groups: (a) Data that result from an attempt by the collecting agency to cover exhaustively the area to which our estimates refer; e.g., Interstate Commerce Commission data on income originating in the steam railroad industry; Census reports on wages and salaries of employees in manufacturing, mining, and trade; Bureau of Internal Revenue data on corporate incomes. (b) Data that explicitly cover only a part of the area measured in the given industry type of income cell; and since this partial coverage is intentional and is recognized by the collecting agency, it is usually well defined. To this category belong statistics for several states on payrolls in various industries; information gathered from sample collections of corporate reports; results of questionnaire surveys necessarily partial in coverage. (c) Data that do not relate directly to the industry type of income cell for which the estimate is being made, but whose magnitude or changes are assumed to be similar to the magnitude or changes that are to be estimated.

Within each category of underlying data some are more complete and reliable than others. The collecting agencies are not always successful in attaining complete coverage and accurate reports. The degree of success depends upon the number and size of reporting units in the field, the power of

475

regulation and control exercised by the collecting agency, and the intensity of the factors that make for bias in reporting. Within the first broad category some data are complete and reliable, e.g., employee compensation reported by steam rail-roads to the Interstate Commerce Commission. Some are incomplete in certain minor respects, because it was not deemed advisable to spend the labor and time necessary to obtain exhaustive coverage; e.g., the *Census of Mines and Quarries* does not segregate the labor cost included in contract work or cover individual placer miners. Some Census reports were intended to be complete but, because of the large number of small units in the field and lack of authority to enforce reporting, they are manifestly incomplete, e.g., the *Census of Business* for 1935. Finally, some data are complete in terms of the number of reporting units, but may suffer from a bias, varying in direction or magnitude from year to year and difficult to measure; e.g., reports on income submitted by corporations to tax authorities.

In the second category, the degree of admittedly partial coverage varies. A given sample may be based on reports for a few states, on information from the larger economic units that publish their income accounts and balance sheets, or on a questionnaire study conducted under specific conditions of sampling, editing of returns, etc. The magnitude and direction of the probable bias and the basis of the adjustment by which complete coverage can be attained differ from one group of data to another.

Even in the third category, information *not* relating to the particular area to be measured, data may vary in pertinence. If for two related industries we have estimates that are similar in magnitude or in fluctuations for some years, we base estimates for the missing years for the one industry upon those for the other. If we lack such a quantitative foundation we may base an estimate upon qualitative knowledge of kinship between the cell to be measured and the cells for which we have estimates. This qualitative knowledge of kinship, espe-

cially with reference to industrial divisions, may be specific, relating estimates for one industry to those for another, or general, relating estimates for one industry or homogeneous combination of industries to those for a fairly heterogeneous combination of industries.

Most estimates are based upon data that belong to several of these three broad categories and the groups within them. Estimates of manufacturing wages for intercensal years are based upon Census reports (first category) and sample data on payrolls (second category). Entrepreneurial withdrawals in some industries for some years are estimated from (a) complete data on number in one year; (b) sample data on changes in number; (c) average withdrawals based upon average salaries, thus combining data belonging to all three broad categories. For only a few cells, notably employee compensation in manufacturing for Census years, dividends for most industries in recent years, and income originating in some public utilities, are our estimates based entirely upon data belonging to the first broad category.

The reason is obvious. No final estimates can rest upon data belonging exclusively either to the second or third category: by definition, they are either incomplete or do not relate directly to the area under measurement. And from the first category only data that are complete in coverage and unbiased can be used directly for our final estimates. Such data are to be had for most cells for only a few years, and often not at all. So far as they can be obtained for at least one year, the incomplete and biased information for other years must be adjusted to the complete data in order to derive the final estimate. In other words, the partial data in the second category, the indirectly related data in the third, and data in the first category that have gaps or biases cannot be used by themselves. They are either adjusted to compensate for their bias or incompleteness or used as indicators of fluctuations but not of magnitude, i.e., as indexes for interpolating and extrapolating beyond the years for which complete data are to be had.

2 *Adjustment, Interpolation, Extrapolation*

A ADJUSTMENT

By adjustment we mean that a given total or ratio, reported in
the data and considered either incomplete or excessive, is re-
vised in an attempt to approximate the true figure. It is not
intricate and presents no technical difficulties if the shortage
or excess is known; otherwise, it is almost impossible. The
overall controlling totals, effective as they are in adjusting the
combined coverage of the specific cells to a national income
total, are of no use in adjusting estimates within industrial
divisions and type of income categories.

Most of the adjustments we made were upward, since the
common defect of reported data is a shortage in coverage. We
made no qualitative adjustments; i.e., we made adjustments
solely for areas for which a definite quantitative basis could be
found. This means that in several industrial divisions the
data as reported in the basic source were used unchanged, even
though there were grounds for suspecting incompleteness of
coverage. But since in almost all cases a quantitative basis for
adjustment was available for only a few years or a single year
in the period, and in several cases only for an industry as a
whole, not for the various types of income originating in it,
we had to apply the same relative adjustment to all years in
the period or to all types of income in the industry. We did
this, however, only when the adjustments were relatively
minor, and when there was no evidence that they would differ
from year to year or be substantially different within the in-
dustry for different types of income.

Adjustments are illustrated in the procedures by which,
from Census totals of contract construction, we derived a more
comprehensive total for the construction industry; by which
the reported balance sheet totals of long term interest-bearing
securities of corporations in *Statistics of Income* were raised to
the more complete coverage of the income accounts; or by
which the totals for interstate pipe lines reporting to the In-

terstate Commerce Commission were raised to include all pipe lines in the country.

B   INTERPOLATION

Interpolation is used when final and complete estimates for a given cell are available for more than one year in a period, but not for successive years. By it partial but directly related data, indirectly related data, or other information are used to derive estimates for the intervening years.

Interpolation may be based upon specific data, whether directly or indirectly related to the area estimated, or upon general assumptions concerning the character of changes during the intervening years. Since we deal with irregularly changing quantities, rather than with paths of mathematical functions, we avoided interpolation based upon general assumptions concerning the character of changes between the two terminal years. When neither direct nor indirect data were available, straight line interpolation was used, usually to derive not the final magnitudes themselves but subsidiary ratios or numbers. We used the straight line procedure because in the absence of specific information concerning the character of changes during the intervening years, it was most convenient to assume the simplest type of movement.

When specific data were available, interpolation was most frequently by the simple ratio procedure:

Let A and E be the complete totals available, and B, C, and D the estimated totals for the intervening years; let a, b, c, d, and e be the partial direct, or indirect, data available for all years. Then

$$A = a \, \frac{A}{a} = A$$

$$B = b \left(\frac{A}{a} + \frac{\frac{E}{e} - \frac{A}{a}}{4} \right) = b \left(\frac{3A}{4a} + \frac{E}{4e} \right)$$

$$C = c \left(\frac{A}{2a} + \frac{E}{2e} \right)$$

$$D = d \left(\frac{A}{4a} + \frac{3E}{4e} \right)$$

$$E = e \frac{E}{e} = E$$

This method apportions any change in the relative disparity between the partial or indirect data and the complete direct data in the two terminal years along an arithmetic straight line. It might be more consistent with the relative character of the disparity to apportion any change along a logarithmic straight line. Then

$$B = b \frac{A}{a} \sqrt[4]{\frac{E}{e} \div \frac{A}{a}} = b \sqrt[4]{\frac{E}{e} \times \left(\frac{A}{a} \right)^3}$$

$$C = c \sqrt[4]{\left(\frac{E}{e} \right)^2 \times \left(\frac{A}{a} \right)^2}$$

$$D = d \sqrt[4]{\left(\frac{E}{e} \right)^3 \times \left(\frac{A}{a} \right)}$$

$$E = e \sqrt[4]{\left(\frac{E}{e} \right)^4} = E$$

But the logarithmic straight line would have entailed more laborious calculations, and since the assumption as to the way any change in the relative disparity at the two terminal years should be apportioned among the intervening years is necessarily arbitrary, it was considered justifiable to choose the procedure that required fewer calculations.

In a few instances an interpolation that may be designated proportional was used.

$$B = A + \left[(E-A) \times \frac{(b-a)}{(e-a)} \right] = E - \left[(E-A) \times \frac{(e-b)}{(e-a)} \right]$$

$$C = A + \left[(E-A) \times \frac{(c-a)}{(e-a)} \right] = E - \left[(E-A) \times \frac{(e-c)}{(e-a)} \right]$$

$$D = A + \left[(E-A) \times \frac{(d-a)}{(e-a)} \right] = E - \left[(E-A) \times \frac{(e-d)}{(e-a)} \right]$$

$$E = A + \left[(E-A) \times \frac{(e-a)}{(e-a)} \right] = E - \left[(E-A) \times \frac{(e-e)}{(e-a)} \right] = E$$

The advantage of proportional interpolation is that it does *not* assume, as does the simple ratio method, a progressive movement over the intervening years of the change from one terminal year to the other in the relative discrepancy between the basic and the partial (or indirect) data. It assumes that the proportional distribution among the intervening years of the total change from one terminal year to the other is portrayed accurately by the partial or indirect data upon which the interpolation is based. Such an assumption is preferable for short periods marked by a sustained cyclical rise or decline. Over such periods the total rise or decline in the sample data may be smaller or larger than in the universe, but the change in the disparity from one terminal year to the other need not be at the same rate per intervening year. It is assumed that the annual pattern of the cyclical expansion or contraction, as far as its proportional distribution among the intervening years is concerned, is faithfully revealed by the sample data.[1]

[1] The difference between the ratio and the proportional methods may be illustrated by the following simple example. Let us assume the values for the universe at two terminal years to be 100 and 200. The values for the sample for the same years are 50 and 90. The value for the intervening year in the sample is 70. Then, the value for the intervening year in the universe will be interpolated as follows:

By the ratio method:

$$70 \left[\left(\frac{1}{2} \times \frac{100}{50} \right) + \left(\frac{1}{2} \times \frac{200}{90} \right) \right] = 70 \times 2.1111 = 147.7777$$

But the proportional is to be preferred to the ratio method solely under three conditions. The first, already mentioned, is that the period of interpolation is brief; for when it is long,

By the proportional method:

$$100 + \left[(200-100) \times \frac{70-50}{90-50}\right] = 150$$

Of course, if there is no change in the relative disparity between the sample and the universe from one terminal year to the other, the ratio and the proportional methods are bound to yield the same results, as can be demonstrated:

Assume that $\dfrac{A}{a} = \dfrac{E}{e}$

Then $\dfrac{a}{A} = \dfrac{e}{E}$ $e = \dfrac{Ea}{A}$ $a = \dfrac{Ae}{E}$

By the ratio method: $B = b\left(\dfrac{3A}{4a} + \dfrac{E}{4e}\right) = b\dfrac{A}{a}$

By the proportional method: $B = A + \left[(E-A) \times \dfrac{b-a}{e-a}\right]$

$$= A + \left[(E-A) \times \frac{b - A\dfrac{e}{E}}{E\dfrac{a}{A} - A\dfrac{e}{E}}\right]$$

$$= A + \left[(E-A) \times \frac{b - A\dfrac{a}{A}}{(E-A)\dfrac{a}{A}}\right]$$

$$= A + \frac{b - A\dfrac{a}{A}}{\dfrac{a}{A}} = \frac{A\dfrac{a}{A} + b - A\dfrac{a}{A}}{\dfrac{a}{A}}$$

$$= \frac{b}{\dfrac{a}{A}} = b\dfrac{A}{a}$$

one should assume some progressive change in the relative disparity between the sample and the universe from one terminal year to the other. Second, the total change in the sample and in the universe must be substantial and in the same direction. For unless it is substantial, the percentage distribution in the proportional method is likely to be erratic; and unless the change in the sample and the universe are in the same direction, there is no basis for assuming similarity in the percentage distribution of the change. Finally, the changes in the sample during the intervening years should all be in the same direction, i.e., either positive or negative, since the percentage distributions of totals whose components are different in sign are likely to be erratic.

These three conditions limited severely the use of the proportional method. As a result, it was applied in our estimates largely to interpolate manufacturing wages and salaries for intercensal years; and even then only for those items in which the changes conformed to the conditions just described. For all other interpolations the ratio method was used.

C   EXTRAPOLATION

Extrapolation is used when the period for which the estimates are to be derived on the basis of partial or indirect data or assumptions concerning the character of change has an open end. The need for this device arises when the basic and complete figures cover only part of the period and *all* preceding or *all* succeeding years must be derived with the help of samples or on some other basis.

Like interpolation, from which it differs solely by the absence of a second basic terminal value, extrapolation can be carried through by assuming a general pattern of change over the missing years. This assumption can be based on the behavior of the complete figures available for part of the period; e.g., one may assume that the relative change from year x to year x $+$ 1 is the same as that from x $-$ 1 to x. Or a mathematical formula can be fitted to the period covered by the

complete figures and values for earlier or later years extra-
polated. But for obvious reasons such assumptions were
avoided and an effort was made to find specific data, either
partial or indirectly related, upon which extrapolation could
be based. In only a very few instances, and for relatively minor
quantities, was extrapolation based on the assumption that
values for a missing (usually earlier) year were equal to the
figures available for the nearest year.

When direct data were used, extrapolation was based upon
the assumption that the relative change in them from the
terminal year to the years to be estimated accurately portrayed
the relative change in the values to be estimated. This was,
of course, one of many assumptions that might have been
made. But lacking information that would lead to a choice of
any other procedure, we used the simplest. Its implication is
that the relative disparity in the terminal year between the
complete figure and the partial or indirect data remained con-
stant over the years for which values were extrapolated. Since
such an assumption is valid for only a short period, we tried to
confine such extrapolation to periods not exceeding two years.
Most uses of extrapolation in our estimates are, as a matter of
fact, for only the most recent year or two.

3 Data, Procedures, and Margins of Error

Originally we intended to classify the estimates for the various
industry type of income cells according to the character of the
underlying data and the procedure used. Such a classification
might have spared us the delicate task of assigning specific
values to the error margins of the various estimates. An esti-
mate based on complete Census totals is subject to a narrower
relative error than one based on incomplete data or data not
directly related to the given industry type of income cell.
Similarly, the very reason for adjustment, interpolation, or
extrapolation—lack of complete and directly related data—
means that the estimate is subject to a wider margin of error
than an estimate obtained by direct use of comprehensive

figures. Furthermore, differences in the reliability of the groups within the three broad categories of underlying data may spell differences in the reliability of estimates derived from them. Similarly, all other conditions being equal, an interpolation is likely to yield more reliable estimates than an extrapolation; an interpolation based upon a directly related and large sample will yield more reliable estimates than a straight line interpolation; and so on.

Had our experiment proved successful we might have presented the classification as an adequate indication of the accuracy of the estimates. It would have served as an effective summary of our detailed notes in Part Four. But we could not classify the estimates according to the reliability of the data and procedures, for two reasons. First, the number of classes that could and had to be made was unmanageably large. As already indicated, for few cells are the estimates based directly and exclusively upon comprehensive data: for most, data from more than one category are used, and the combinations vary considerably. In some cases a single adjustment or extrapolation of basic Census totals is used; in others, dollar values are derived through estimates of the number of persons engaged and of per capita averages, each of these in turn based upon complex combinations of data in various categories. The attempt to classify the interpolation and extrapolation procedures, whenever these were used, was no more successful, since they may be applied to derive the final estimate directly or to obtain one or several subsidiary quantities from which in turn the final estimate is derived. A complete description of this variety of combinations of data and procedures yields a complex classification having slight advantage over the detailed notes to the basic tables in Part Four.

The second and even more important difficulty was that the tentative classification by character of data and procedures did not represent definite classes of error margins. Even relatively complete data differ in the relative undercoverage or bias to which they are subject; the coverage of partly complete

data may range all the way from 1 to 99 per cent; and there is
no inherently uniform relation between the coverage of a
sample and the reliability of estimates based upon it. Straight
line interpolation may mean one range of error when applied
directly to obtain the estimate itself, and another when used
to derive a subsidiary quantity from which the estimate is
made; and similar variations in the margin of error may ac-
company differences in the duration and character of the
period covered. As a consequence, a classification based on
characteristics of data and procedures would have to have sub-
divisions for the differences in reliability among and within
its groups. In other words, it proved difficult to evaluate mar-
gins of error on the basis of classes of underlying data and pro-
cedures, since *within* each class the various estimates still
differed in reliability, and *among* some there were no apparent
differences in reliability. Therefore, we had to evaluate the
margins of error *directly,* fully aware that they represent at
best merely an informed opinion. Sample classifications, one
for all 1929 estimates and the other for estimates based on
interpolation and extrapolation only, for several industries
for all years, demonstrated clearly the difficulties discussed
above and proved of small use (compared to the detailed de-
scription of procedures in Part Four) in judging the margins of
error in the final estimates. The procedure we finally adopted
and the results are described in detail in Chapter 12. But first
we give briefly the results of a preliminary test that was ap-
plied to some of the interpolations and extrapolations, a test
which, like the comparisons in Chapter 10, suggests the mar-
gins of error to which our estimates are subject.

4 Test of Selected Interpolations and Extrapolations
For several industries we have for the period studied at least
three Census values, as well as directly related samples upon
which we base interpolation for non-Census years. It is then
possible to test our estimates by comparing, with the Census
figure for the intermediate year, an estimate that is derived for

the intermediate Census year by interpolation. To illustrate: the *Biennial Census of Manufactures* reports salaries in 1919, 1921, and 1923. We use sample data to interpolate salaries for the intercensal years 1920 and 1922. The same sample data can be used to estimate salaries for 1921 by interpolating between the Census values for 1919 and 1923; and this interpolated value for 1921 can then be compared with the Census figure for 1921.

A somewhat similar procedure can be used when only two Census values are reported, but then the reliability of an extrapolation alone can be tested. To illustrate: if Census values are reported for 1929 and 1935, and we have interpolated the intervening years on the basis of sample data, we can test the goodness of this sample by extrapolating, from the 1929 Census value, a 1935 estimate for comparison with the 1935 Census value (or by extrapolating the 1935 Census value back to 1929 for comparison with the 1929 Census value). This procedure tests the sample data as a basis for an extrapolation, not an interpolation, index.

Neither procedure is an infallible test. Both necessarily exaggerate the errors in our estimates, since the first assumes a gap in the data wider than that actually filled; and the second tests a sample as a basis for an extrapolation index whereas it is actually used for an interpolation. Yet, when such tests can be applied to only one rather than several time units, they may yield accidentally a favorable showing that is not necessarily valid for other years in which the sample has been applied to derive estimates. Finally, the test can be applied to merely a few extrapolations and interpolations, since for many others *no* basic comprehensive data are reported for more than one time unit. The results of these tests are presented to indicate those cells for which interpolation rests upon several comprehensive totals; and to suggest the margin of error that may arise from the use of sample data in estimating values for non-Census years.

Table 95 presents tests of interpolation when at least three

TABLE 95

Tests of Selected Indexes used for Interpolation during Periods containing at least Three Census Values

INDUSTRY AND TYPE OF INCOME OR PERSONS ENGAGED (1)	YEAR FOR WHICH TEST IS MADE (2)	BASIC DATA (3)	ESTIMATED DATA (4)	% DIF. (5)	TERMINAL YEARS (6)		INDEX TESTED (7)	INDEX DESCRIBED IN NOTES TO Table	Col.
Anthracite coal									
Wage earners	1929	144,761	144,585	−0.12	1919	1935	Wage earners, sample, 1919–29; BLS, employment, 1929–35	Q 9	1
Salaried workers	1929	8,493	8,277	−2.54	1919	1935	Salaried workers, sample	Q 9	7
Wages ($000)	1929	233,123	235,670	+1.09	1919	1935	Wages, sample, 1919–21; FRB, wages, 1921–29; BLS, wages, 1929–35	Q 4	1
Salaries ($000)	1929	21,626	20,556	−4.95	1919	1935	Salaries, sample	Q 4	7
Bituminous coal									
Wage earners	1929	459,532	473,520	+3.04	1919	1935	Our employment index, 1919–29; BLS, employment, 1929–35	Q 9	2
Salaried workers	1929	23,724	21,690	−8.57	1919	1935	Ratio *	Q 9	8
Salaries ($000)	1929	58,751	57,388	−2.32	1919	1935	Avg. salary, sample	Q 4	8
Men's clothing mfg.									
Wage earners	1921	245,661	244,774	−0.36	1919	1923	BLS, men's clothing combined with shirts & collars	M 23	6
Wages ($000)	1921	263,732	261,897	−0.72	1919	1923		M 6	6
Other wearing apparel mfg.									
Wage earners	1921	48,988	59,777	+22.02	1919	1923	BLS, wearing apparel	M 23	9
Wages ($000)	1921	49,570	62,074	+25.22	1919	1923	BLS, wearing apparel	M 6	9
Woolen goods mfg.									
Wage earners	1921	194,300	196,245	+1.00	1919	1923	BLS, woolen & worsted goods combined with carpets & rugs	M 23	11
Wages ($000)	1921	213,682	214,996	+0.61	1919	1923		M 6	11
Other textile fabrics mfg.									
Wage earners	1921	77,876	82,306	+6.37	1919	1923	BLS, textile fabrics	M 23	15
Wages ($000)	1921	75,406	77,737	+3.09	1919	1923	BLS, textile fabrics	M 6	15

Lumber mfg.									
Wage earners	1925	585,327	577,836	1923	1927	−1.28	BLS, sawmills, combined with millwork	M 23	18
Wages ($000)	1925	603,308	602,495	1923	1927	−0.13	BLS, sawmills, combined with millwork	M 6	18
Stone, clay, & glass mfg.									
Wage earners	1925	336,232	338,050	1923	1927	+0.54	BLS, brick, tile, & terra cotta, combined with glass & cement	M 23	19
Wages ($000)	1925	444,079	448,206	1923	1927	+0.93		M 6	19
Heating apparatus mfg.									
Wage earners	1925	95,860	95,151	1923	1927	−0.74	BLS, steam & hot water apparatus & steam fittings, combined with stoves	M 23	20
Wages ($000)	1925	138,350	137,450	1923	1927	−0.65		M 6	20
Other construction materials mfg.									
Wage earners	1925	115,248	106,989	1923	1927	−7.17	BLS, structural & ornamental metalwork, combined with cast iron pipe	M 23	21
Wages ($000)	1925	169,389	156,208	1923	1927	−7.78		M 6	21
Printing									
Wage earners	1925	314,021	314,001	1923	1927	−0.01	BLS, book & job printing, combined with newspaper & periodical printing	M 22	5
Wages ($000)	1925	540,826	542,078	1923	1927	+0.23		M 5	5
Iron & steel mfg.									
Wage earners	1925	588,021	542,300	1923	1927	+0.80	BLS, blast furnaces, steel mills, & rolling mills	M 23	24
Wages ($000)	1925	846,256	849,227	1923	1927	+0.35	BLS, blast furnaces, steel mills, & rolling mills	M 6	24
Other transp. equipment mfg.									
Wage earners	1925	88,244	81,862	1923	1927	−1.66	BLS, cars, combined with locomotives	M 23	27
Wages ($000)	1925	120,281	116,898	1923	1927	−2.81	BLS, cars, combined with locomotives	M 6	27
Non-ferrous metal mfg.									
Wage earners	1927	272,667	267,696	1925	1929	−1.83	BLS, brass, bronze, & copper products, combined with stamped & enamelled ware	M 23	31
Wages ($000)	1927	388,185	371,519	1925	1929	−3.04		M 6	31

Table 95 (cont.)

INDUSTRY AND TYPE OF INCOME OR PERSONS ENGAGED (1)	YEAR FOR WHICH TEST IS MADE (2)	BASIC DATA (3)	ESTIMATED DATA (4)	% DIF. (5)	TERMINAL YEARS (6)		INDEX TESTED (7)	INDEX DESCRIBED IN NOTES TO Table	Col.
Misc. excl. rubber mfg.									
Wage earners	1921	341,108	335,595	−1.62	1919	1923	BLS, all mfg.	M 23	35
Wages ($000)	1921	371,115	345,619	−6.87	1919	1923	BLS, all mfg.	M 6	35
Food & tobacco mfg.									
Salaried workers	1921	149,400	154,780	+3.60	1919	1923	Ratio *	M 24	1
Salaries ($000)	1921	290,528	291,479	+0.33	1919	1923	Avg. salary, sample	M 7	1
Wearing apparel mfg.									
Salaried workers	1921	101,878	121,174	+18.94	1919	1923	Ratio *	M 25	3
Salaries ($000)	1921	209,640	225,558	+7.59	1919	1923	Avg. salary, sample	M 8	3
Textile fabrics mfg.									
Salaried workers	1921	50,441	51,218	+1.54	1919	1923	Ratio *	M 25	4
Salaries ($000)	1921	129,491	132,811	+1.03	1919	1923	Avg. salary, sample	M 8	4
Leather mfg.									
Salaried workers	1921	8,058	9,657	+19.84	1919	1923	Ratio *	M 25	5
Salaries ($000)	1921	19,927	19,855	−0.36	1919	1923	Avg. salary, sample	M 8	5
Lumber & furniture mfg.									
Salaried workers	1921	51,309	54,309	+5.85	1919	1923	Ratio *	M 25	6, 11
Salaries ($000)	1921	114,058	114,673	+0.54	1919	1923	Avg. salary, sample	M 8	6, 11
Stone, clay, & glass mfg.									
Salaried workers	1921	28,919	32,859	+13.62	1919	1923	Ratio *	M 25	7
Salaries ($000)	1921	65,258	63,680	−2.42	1919	1923	Avg. salary, sample	M 8	7
Heating apparatus mfg.									
Salaried workers	1921	12,923	14,445	+11.78	1919	1923	Ratio *	M 25	8
Salaries ($000)	1921	30,797	25,600	−16.88	1919	1923	Avg. salary, sample	M 8	8

	1921				1919	1923			
Other construction materials mfg.									
Salaried workers	1921	12,888	13,310	+3.27	1919	1923	Ratio *	M 25	9
Salaries ($000)	1921	30,404	31,073	+2.20	1919	1923	Avg. salary, sample	M 8	9
Paper & printing									
Salaried workers	1921	156,500	162,374	+3.75	1919	1923	Ratio *	M 24	4, 5
Salaries ($000)	1921	317,875	321,138	+1.19	1919	1923	Avg. salary, sample	M 7	4, 5
Metal mfg.									
Salaried workers	1921	300,727	338,174	+12.45	1919	1923	Ratio *	M 24	6
Salaries ($000)	1921	721,796	659,494	−8.63	1919	1923	Avg. salary, sample	M 7	6
Chemical mfg., excl. petroleum ref.									
Salaried workers	1921	64,422	67,015	+4.03	1919	1923	Ratio *	M 25	16
Salaries ($000)	1921	143,975	140,053	−2.72	1919	1923	Avg. salary, sample	M 8	16
Petroleum refining									
Salaried workers	1921	13,705	22,494	+64.13	1919	1923	Ratio *	M 25	17
Salaries ($000)	1921	34,633	28,109	−18.84	1919	1923	Avg. salary, sample	M 8	17
Rubber tire mfg.									
Salaried workers	1921	12,526	18,924	+51.08	1919	1923	Ratio *	M 25	18
Misc. mfg., excl. rubber									
Salaried workers	1921	55,811	58,329	+5.46	1919	1923	Ratio *	M 25	19
Salaries ($000)	1921	118,719	121,408	+2.27	1919	1923	Avg. salary, sample	M 8	19
Food & tobacco mfg.									
Entrepreneurs	1921	53,406	69,279	+29.72	1919	1923	Failures, milling, bakers, liquors, & tobacco	M 28	1
Textile & leather mfg.									
Entrepreneurs	1921	35,538	38,844 ·	+7.90	1919	1923	Failures, woolens, woolen goods, cotton, etc.	M 28	2

Table 95 (concl.)

INDUSTRY AND TYPE OF INCOME OR PERSONS ENGAGED (1)	YEAR FOR WHICH TEST IS MADE (2)	BASIC DATA (3)	ESTIMATED DATA (4)	% DIF. (5)	TERMINAL YEARS (6)	INDEX TESTED (7)	INDEX DESCRIBED IN NOTES TO Table	Col.
Construction materials & furniture mfg.								
Entrepreneurs	1921	25,795	40,273	+56.13	1919 1923	Failures, lumber & lumber products, etc.	M 28	3
Printing								
Entrepreneurs	1921	22,024	26,514	+20.39	1919 1923	Failures, printing & engraving	M 28	5
Metal mfg.								
Entrepreneurs	1921	20,868	24,986	+19.73	1919 1923	Failures, iron & steel, machinery & tools	M 28	6
Chemical mfg.								
Entrepreneurs	1921	5,020	5,822	+15.98	1919 1923	Failures, chemicals, drugs, paints & oils	M 28	7
Misc. & rubber mfg.								
Entrepreneurs	1921	8,841	11,166	+26.30	1919 1923	Failures, all other mfg.	M 28	8
Elec. light & power								
Employees	1927	234,747	230,049	−2.00	1922 1932	Employment index, sample states, 1922–29; BLS, elec. light & power & gas, 1929–32	P 20	1
Wages & salaries ($000)	1927	367,632	372,074	+1.21	1922 1932	Avg. pay index, sample states, 1922–29; BLS, payrolls for elec. light & power & gas, 1929–32	P 7	1
Dividends paid ($000)	1927	338,239	338,168	−0.02	1922 1932	Dividend pay., sample	P 10	1
Mfd. gas								
Employees	1927	74,155	69,469	−6.32	1925 1929	Am. Gas Asso. employment estimates	P 20	2
Wages & salaries ($000)	1927	112,255	116,349	+3.65	1925 1929	Payrolls index, elec. light & power & gas, sample	P 7	2
Street rwy.								
Employees	1927	254,364	254,786	+0.17	1922 1932	Am. Transit Asso. employment estimates	P 20	4
Wages & salaries ($000)	1927	419,990	423,227	+0.77	1922 1932	Am. Transit Asso. payroll estimates	P 7	4

	1927	375,272	382,280	+1.87	1922	1932				
Telephone										
Employees	1927	375,272	382,280	+1.87	1922	1932	Employees, sample, 1922–27; employees reported to ICC, 1927–32	P	20	6
Wages & salaries ($ooo)	1927	463,524	400,424	+1.02	1922	1932	Wages & salaries, sample	P	7	6
Telegraph										
Employees	1927	83,483	82,988	−0.59	1922	1932	Employees, sample, 1922–27; employees reported to ICC, 1927–32	P	20	7
Wages & salaries ($ooo)	1927	102,078	105,416	+3.27	1922	1932	Wages & salaries, sample	P	7	7
Dividends paid ($ooo)	1927	14,199	13,071	−7.94	1922	1937	Div. reported to ICC, 1922–27; div. reported to ICC, plus Com'l Cable Co. div., 1927–37	P	10	7
Wholesale trade										
Employees	1933	1,136,679	1,168,554	+2.80	1929	1935	BLS, employment	T	5	1
Wages & salaries ($ooo)	1933	1,714,109	1,807,010	+5.42	1929	1935	BLS, payrolls	T	4	1
Retail trade										
Employees	1933	2,563,989	2,666,164	+3.99	1929	1935	BLS, employment	T	5	2
Wages & salaries ($ooo)	1933	2,648,704	2,783,989	+5.11	1929	1935	BLS, payrolls	T	4	2
Restaurants										
Employees	1933	388,769	470,280	+20.97	1929	1935	BLS, employment, retail trade	S	9	6

* Salaried workers to wage earners, sample.

TABLE 96

Tests of Selected Indexes used for Interpolation during Periods containing Two Census Values

INDUSTRY AND TYPE OF INCOME OR PERSONS ENGAGED (1)	YEAR FOR WHICH TEST IS MADE (2)	BASIC DATA (3)	ESTIMATED DATA (4)	% DIF. (5)	TERMINAL YEARS (6)	INDEX TESTED (7)	INDEX DESCRIBED IN NOTES TO Table	Col.
A								
Telephone								
Interest ($000)	1937	42,806	44,190	+4.45	1922 1937	Int. paid, sample	P 12	6
'Other' mining								
Wage earners	1929	96,478	163,061	+69.01	1929 1935	BLS, employment	Q 9	4
Salaried workers	1929	11,679	11,095	−5.00	1929 1935	Ratio,* sample	Q 9	11
Oil & gas mining								
Salaried workers	1935	20,315	17,506	−13.83	1919 1935	Ratio,* selected industry	Q 9	10
Salaries ($000)	1935	41,113	46,094	+11.97	1919 1935	Avg. salary, selected industry	Q 4	10
'Other' personal service								
Employees	1935	261,011	250,274	−4.11	1933 1935	Employees, selected industry	S 9	8
Advertising								
Wages & salaries ($000)	1935	140,423	112,683	−19.75	1933 1935	Avg. salary, selected industries	S 5	6
B								
Bituminous coal mining								
Wages ($000)	1929	575,792	617,717	+7.28	1919 1929	Ratio, wages to value of product, sample	Q 4	2
Wages ($000)	1929	575,792	575,409	−0.07	1929 1935	BLS, payrolls	Q 4	2

							Q			
Metal mining										
Wage earners	1929	116,484	114,937	−1.33	1919	1929	Q	9	3	300-day workers
Wage earners	1929	115,486	155,053	+34.26	1929	1935	Q	9	3	BLS, employment
Salaried workers	1929	9,864	7,874	−20.17	1919	1929	Q	9	9	Ratio,* selected industry
Salaried workers	1929	9,732	9,008	−7.44	1929	1935	Q	9	9	Ratio,* selected industry
Wages ($000)	1929	179,498	167,761	−6.54	1919	1929	Q	4	3	Wage per day, sample
Wages ($000)	1929	178,110	222,083	+24.69	1929	1935	Q	4	3	BLS, payrolls
Salaries ($000)	1929	27,680	28,576	+3.24	1919	1929	Q	4	3	Avg. salary, selected industries
Salaries ($000)	1929	27,210	26,617	−2.18	1929	1935	Q	4	9	Avg. salary, selected industries
'Other' mining										
Wages ($000)	1929	117,534	128,351	+9.20	1919	1929	Q	4	5	Ratio, wages to value of product, sample
Wages ($000)	1929	118,922	200,456	+68.56	1929	1935	Q	4	5	BLS, payrolls
Salaries ($000)	1929	31,832	33,131	+4.08	1919	1929	Q	4	11	Avg. salary, sample
Salaries ($000)	1929	32,300	33,016	+2.22	1929	1935	Q	4	11	Avg. salary, sample

* Salaried workers to wage earners.

Census values are connected by one type of sample data. But
the table is not complete. First, when the indexes used for in-
terpolation had already been adjusted by the compiling au-
thorities to conform to Census data (e.g., the Bureau of Labor
Statistics indexes of employment and payrolls of wage earners
and the Bureau of Agricultural Economics indexes of farm
income), it seemed unnecessary to test them by the procedure
suggested. Exceptions were made, however, whenever these
indexes were combined for purposes of interpolation in such
a way that their scope did not exactly fit the industrial group
whose number or value we had to estimate. Since this occurred
in several of our manufacturing groups, there are tests in
Table 95 of BLS indexes of wage earners' employment and
payrolls. Second, for manufacturing industries the frequency
of Census values meant that the testing procedure could be
applied to more than one time unit. But it seemed unneces-
sary to test more than one year; and we selected 1921 because
it is in a period for which sample data are weakest, the gyra-
tions of the basic data greatest, and hence the possible error
in the interpolation highest. Third, for some series three
Census values were reported, but the interpolating samples
used between each pair of Census values were different. Since
such conditions were more comparable to the existence of two
sets of two Census values each, they are given in Table 96.
Finally, we have omitted from Table 95 interpolations whose
results appear in the final estimates in combination with re-
sults of other procedures that cannot be similarly tested. For
example, it is possible to test estimates of the number of entre-
preneurs for some divisions of mining. But since we combine
these in a final total for all mining with estimates for other
subdivisions for which no Census values are reported, the final
total belongs to the category that could not be tested because
only one Census value is reported. Table 96 presents tests
when only two Census values are reported (Section A), or
when three Census values are treated as two pairs of values,
each pair being connected by a different interpolation index

(Section B). It excludes interpolations whose results do not enter directly into our estimates.

Thus the interpolation procedures can be tested for only a few industries and types of income or employment. In general, mining, manufacturing, and the public utilities are well represented, with some sprinkling of trade and service. Government, finance, construction, and the miscellaneous division are completely absent. Estimates for agriculture had already been adjusted to the Census data by the compiling authority. For the commodity producing and public utility industries alone are censuses taken in two, three, or more years during the period covered. For such important divisions as government, finance, and most of service there is either none or only one Census value during the twenty years covered by our estimates. Consequently, a large proportion of the estimates for these industrial divisions, being based on extrapolation by sample data throughout, rest upon a much less secure foundation than most of the estimates tested in Tables 95 and 96.

Tests are more numerous for estimates of the number of employees and of employee compensation than of entrepreneurs and other types of income flow because the industrial censuses, which provide the controlling figures, rarely present information on dividends or interest and practically never on net income. This does not necessarily mean that the estimates of dividends, interest, and net income are less reliable than those of employment and payrolls. For years for which *Statistics of Income* provides data on dividends or for the public utility industries covered by the Interstate Commerce Commission, annual estimates are easily derived and are of a fair degree of accuracy. But it does mean that in other cases whenever interpolation and extrapolation are based on sample data they cannot be tested by the procedures used in preparing Tables 95 and 96; and it is likely that the resulting estimates are not as reliable as those derived by the interpolation and extrapolation procedures tested in these two tables.

To ascertain the margins of errors revealed by the tests a frequency distribution of entries by classes of size of error was constructed (Table 97). All entries were included, even though there is some duplication within Table 96, and the cells to which the entries refer differ considerably in relative size.

TABLE 97

Distribution of Entries in Tables 95 and 96 by Size of Error

GROUPS OF ENTRIES	CLASSES OF SIZE OF ERROR					
	0 to 5	5 to 10	10 to 20	20 to 40	40 to 80	Total
Table 95						
Wage earners	12	2		1		15
Salaried workers	6	3	5		2	16
Employees	6	1		1		8
Entrepreneurs		1	2	3	1	7
Wages	11	2		1		14
Salaries	11	2	2			15
Empl. compensation	5	2				7
Dividends	1	1				2
Total	52	14	9	6	3	84
Table 96						
Wage earners	1			1	1	3
Salaried workers	1	1	1	1		4
Employees	1					1
Wages	1	3		1	1	6
Salaries	4		1			5
Empl. compensation			1			1
Interest	1					1
Total	9	4	3	3	2	21

By and large the preponderant number of tests suggest relatively moderate errors. Of the entries in Table 95 about two-thirds show errors of 5 per cent or less, and only slightly over one-tenth, errors in excess of 20 per cent. Even in Table 96 over 40 per cent of the entries show errors of 5 per cent or less. Table 95 has many more small errors than Table 96, and would have, even were we to omit from it tests of the number of wage earners and of wages in manufacturing (based on adjusted BLS indexes). It is obvious that the error in an inter-

polation between two Census values is likely to be much less than in an extrapolation from one Census value.

Of the seven entries for number of entrepreneurs four show errors in excess of 20 per cent. Although tests in Table 95 for number of entrepreneurs are relatively few, yet the greater error shown indicates that the estimates are less reliable than other estimates tested.

Our interpolation of the number of salaried employees is subject to a greater error than our interpolation of per capita salary. Of all entries in Tables 95 and 96 for the number of salaried employees (20), less than four-tenths show errors of 5 per cent or less; of all entries for salaries (20), over seven-tenths show errors of 5 per cent or less. The test for number is of the indexes of number of salaried employees; for salaries, of indexes of compensation per employee, the number given in the Census year being accepted as reported. Hence the test is of the reliability of the estimates of number and of per capita salary, not of our final estimates of total salaries.[2]

The evidence presented above could be expanded by including other years to which the procedure can be applied (e.g., for number of salaried employees and salaries in manufacturing); or, by studying the changes that had to be introduced into the BLS indexes of employment and payrolls for the purpose of establishing conformity to new Census figures whenever these appeared. But the results of tests of this kind, like those in the tables, would be merely illustrative and suggestive: they could not demonstrate with any precision the margins of error in the final estimates.

Two important qualifications must be noted, particularly with reference to the small error the tests reveal. First, so far

[2] The errors that would be shown by a test of estimates of salaries, regardless of number, as given in the Census are suggested by combining the errors in Tables 95 and 96 shown separately for the number of salaried employees and for salaries. These combined errors, which bear more directly upon the reliability of our final estimates of total salaries than the present entries, indicate that in most cases the error in the estimates of total salaries is greater than that in the estimates of the number of salaried employees.

as the tests are interpreted in their bearing upon the estimates derived with the help of the interpolations tested, the entries in Tables 95 and 96 are likely to be overestimates. The tests necessarily disregard the fact that the interpolations actually made are for shorter periods than is assumed in Table 95; and that they are interpolations rather than extrapolations, as is assumed in Table 96. This qualification also means that the differences in the magnitude of error shown between entries in Tables 95 and 96 are suggestive of differences between interpolations and extrapolations, but not between interpolations applying to periods of different duration. Second, the errors revealed by the tests apply only to interpolations based upon two Census values or to extrapolations based upon one Census value. They do not reveal the errors that may characterize estimates based on other than Census data; or, of course, those in which availability of direct and comprehensive data makes it possible for us to dispense with the use of samples.