Volume Title: Analysis of Wisconsin Income

Volume Author/Editor: Hanna, Frank A., Joseph A. Pechman, and Sidney M. Lerner

Volume Publisher: UMI

Volume ISBN: 0-870-14164-3

Volume URL: http://www.nber.org/books/hann48-1

Publication Date: 1948

Chapter Title: PART III,The Accounting Period and the Distribution of Income, Chapter 5

Chapter Author: Frank A. Hanna, Joseph A. Pechman, Sidney M. Lerner

Chapter URL: http://www.nber.org/chapters/c1016

Chapter pages in book: (p. 229 - 256)

Chapter 5

Relations between Annual Accounting Periods

THE DATA essential for investigating, by means of Lorenz curves, the error involved in using annual accounting periods to indicate the distribution of income for longer accounting periods have been exhausted. This error is directly associated with the reranking of families from one year to another. If the reranking is negligible, we may safely conclude that the annual Lorenz curves fluctuate about an average that adequately represents the distribution in a single accounting period composed of the years included in the average. But if the reranking is considerable, the Lorenz curve for the longer accounting period will lie closer to the line of equal distribution than the average of the Lorenz curves for annual periods. In general, the more reranking the greater the difference between the two Lorenz curves covering the same period. This Chapter is devoted to measuring the reranking between annual distributions.

In *Wisconsin Individual Income Tax Statistics,* incomes in each year 1930-35 are cross-classified with those of 1929; 1935 incomes with those in each year 1929-34; and incomes in each pair of consecutive years 1929-35. The items covered are economic income, net taxable income, wages and salaries, interest, dividends, business profits, net rents, capital gains, and capital losses.[1] From

[1] *Changes in Income of Identical Taxpayers, 1929-35.* All the data in this volume are family data, i.e., the income of husbands and wives filing separate returns were combined before tabulation.

a glance at these tables some reranking is evident. Some families move into higher brackets, some into lower from one year to another. But to measure the extent of reranking resort must be had to correlation techniques, which yield only approximate results. And these, unfortunately, cannot be translated into terms of the shape and position of Lorenz curves.[2]

The linear correlation coefficient is a measure of shifting that includes two independent attributes: reranking and nonlinear regression.[3] However, reranking alone affects the relation of the income distribution for several years taken as a single accounting period to the average of income distributions for the years that it comprises.

A Coefficients of Correlation

In computing the coefficient of correlation, obtained by approximate methods, for each pair of years for which basic tables are available, extreme returns were omitted, after it was found that their inclusion gave rise to a strong upward bias and often yielded meaningless results (Table 15).[4] Failure to exclude the one ex-

[2] The basic tables were designed for correlation techniques. However, by assuming (1) a uniform distribution of the income of the earlier year throughout the later year groups (since the income of only one of the two years is cross-classified, and only a simple distribution is provided for the other year), and (2) a uniform distribution or some other simplifying assumption yielding an array from the resulting irregular and overlapping groups, the distribution for the two years cross-classified in each table can be approximated by combining incomes.

This method is confined to combining only two years' incomes; when one of the years is 1929 and the other 1934, it could hardly be called a two-year accounting period. At most it would furnish a basis for judging the extent of reranking between the two years. The coefficient of correlation also provides such a basis. While it is subject to some chance influences and, where the regression is nonlinear, to a decided bias, no effort has been made to ascertain whether its deficiencies would lead to more or to less error than the irregular and overlapping groups that could be obtained by combining the grouped incomes in cross-classification tables.

[3] If higher degree curves were fitted to the data, part of this influence might be overcome. But such a task is arduous and subject to limitations of its own. Furthermore, linear regression provides a good basis for acertaining the differential treatment among groups, as will be shown below.

[4] The line of regression was based on each of the paired group means weighted by the frequencies in the group. Then the formula $r = b \div \frac{\sigma_y}{\sigma_x}$ was applied, where σ_x and σ_y are the standard deviations of the entire distribution of each year correlated, the subscript x referring to the earlier, the subscript y to the

treme return from the distribution of interest gives coefficients approximating unity: a cursory glance at the basic tables suffices to show the absurdity of such a value. While the values for net taxable income and dividends were appreciably reduced by the omission of a few extreme returns, the coefficients are still quite high, and all are positive. The omission of extreme returns from the computations of capital gains and capital losses changed the entire behavior pattern of the coefficients; without these omissions they fluctuated greatly from one pair of years to another, and many were quite high; with the omissions, they were consistently small, and one was negative.

The process by which certain returns were isolated and omitted from the calculation of the correlation coefficients was largely subjective, and consequently unsatisfactory. Undoubtedly each series contains other returns which, if excluded, would further reduce the value of the coefficients. But it is desirable to rid the distribution of only the extreme returns that lead to absurdly high values, and when there was doubt or empirical evidence that this was not the case, a return was included.

The problem raised by the fact that individuals who receive an income of a specific type in one year may not receive the same type of income every year was discussed, so far as it related to capital gains, in the preceding Chapter. Except for the two over-all items, economic and net taxable income, there was a problem as to which returns to include in the computation of the correlation coefficients. The alternatives seemed to be: (a) to include all 13,184 families, (b) to exclude from the 13,184 families all that did not receive the specific receipt at any time 1929-35, or (c) to include only families that received the income during one of the two years correlated. While there would be some justification for adopting (a), it was rejected on grounds of convenience. (c) leads to logical difficulties in comparing the coefficients for several in-

later year. Since economic income was used merely as a basis of classification and was not tabulated, the product of the mid-points of the group intervals and the frequency in the group had to be used to approximate its volume. For the open class at the top of the distribution, economic income was approximated from the data in the footnotes to the basic tables which list complete data for each return. The incomes reported on the returns omitted in computing the correlation coefficient are given in notes to Table 15.

TABLE 15

Coefficients of Correlation Between Incomes for Various Years
Sample of Identical Taxpayers, 1929-1935

	ECONOMIC INCOME	NET TAXABLE INCOME	WAGES	INTEREST	DIVIDENDS	BUSINESS PROFITS	RENTS	CAPITAL GAINS	CAPITAL LOSSES
Correlation with 1929									
1930 and 1929	.86	.66	.89	.98	.78	.74	.79	.40	.35
1931 and 1929	.78	.60	.81	.89	.55	.60	.65	.13	.26
1932 and 1929	.70	.51	.72	.89	.53	.54	.50	.14	.20
1933 and 1929	.64	.50	.67	.85	.51	.50	.35	.09	.22
1934 and 1929	.65	.55	.68	.78	.57	.54	.43	.03	.07
1935 and 1929	.69	.57	.66	.76	.61	.47	.26	.21	.05
Correlation with 1935									
1935 and 1934	.86	.78	.94	.92	.93	.81	.55	.15	.21
1935 and 1933	.79	.70	.87	.84	.79	.73	.79	.15	.18
1935 and 1932	.81	.65	.79	.86	.77	.64	.42	.09	.15
1935 and 1931	.78	.64	.76	.82	.74	.59	.30	.07	.15
1935 and 1930	.75	.63	.72	.76	.69	.55	.31	−.00	.16
1935 and 1929	.69	.57	.66	.76	.61	.47	.26	.21	.05
Consecutive Pairs of Years									
1929 and 1930	.86	.66	.89	.98	.78	.74	.79	.40	.35
1930 and 1931	.85	.76	.90	.89	.84	.75	.85	.12	.57
1931 and 1932	.83	.78	.89	.99	.72	.79	.77	.25	.38
1932 and 1933	.85	.76	.90	.96	.87	.78	.48	.06	.35
1933 and 1934	.84	.73	.94	.71	.81	.79	.55	.47	.26
1934 and 1935	.86	.78	.94	.92	.93	.81	.55	.15	.21
Number of Returns	13,184	13,175	11,517	4,922	4,801	2,795	3,082	1,467	1,848

Notes to Table 15:

In computing these correlation coefficients extreme returns were omitted. The incomes (in dollars) reported on these omitted returns are shown below (one return on each line):

1929	*1930*	*1931*	*1932*	*1933*	*1934*	*1935*
		NET TAXABLE INCOME				
34,180	19,770	29,380	20,570	9,500	15,950	39,560
		40,370		35,740	34,040	29,300
13,610	35,020	28,650	33,390	40,770	44,320	21,510
17,540	21,280	6,440	33,650	25,880	67,910	37,000
438,540	470,830	497,310	333,620	254,480	478,350	363,430
149,310	11,890	76,850		14,900	15,850	19,430
115,520	76,500	44,930			22,370	20,750
	115,690					122,160
94,490	266,030	4,460	2,010	1,230	12,380	38,840
		INTEREST				
549,990	564,800	546,360	443,790	300,200	256,710	263,360
		DIVIDENDS				
145,880	100,030	58,680	15,220	12,510	16,390	44,130
133,700	140,630	85,450	71,190	54,370	74,900	75,680
132,260	66,930	70,080	52,540	67,560	90,180	12,960
107,080	104,370	66,400	28,270	59,300	46,040	41,170
258,080	38,790	15,350	4,570	4,050	159,400	8,620
		RENTS				
18,420	21,980	17,930		8,090		
			950	26,130	1,210	30,430
	13,360	19,030	6,340	160	30	
		CAPITAL GAINS				
3,100	8,120	10,710	710	2,490	111,210	195,870
	4,170				56,940	
85,150	5,600	10,210		14,730	35,550	36,500
258,080	41,820	21,220	1,790	19,300	4,040	30,610
267,300	230,740	11,950	88,910	19,670	9,690	26,690
31,910	3,860	1,040	860	13,660	33,380	102,500
79,230						
91,990	29,420	29,740		650		22,280
73,980	20,240	960	210	4,460	840	1,190
56,720	51,410	11,100	7,560	1,620	25,510	
23,080	1,280	470	2,850	64,390	1,850	22,890
	70,930					
		CAPITAL LOSSES				
38,230	30,900	43,890	84,710	150,490	6,400	5,600
	88,900	23,440	36,960	204,880	119,110	69,810
338,670	277,690	33,760	104,470	16,700	9,840	2,400
165,510		1,360		360		
27,250	97,120	25,420	117,070	14,110	670	7,510
226,970	233,430	33,500		270,010	19,250	38,070
184,060	16,780	106,770	102,340	337,620	182,620	25,300
10,040	54,130	3,700	8,090			
52,700	28,000	91,120				
6,980	80	13,420	4,990	99,860	32,790	14,370
	4,700	700	1,330	11,900	92,090	1,430
26,490	59,850	11,750	5,380	9,350	530	2,740
	9,530	23,110	61,070			
	9,060	3,510	56,590			
920	18,590	28,960	65,080	21,180	43,860	11,270
4,220	3,960	47,040	70,990	16,030		1,810
57,960		3,480		150		

terrelated years.[5] The adoption of (b) must be justified largely on grounds of expediency, since it is partly a matter of convenience and does not answer satisfactorily all the questions involved.[6] The coefficients for each type of receipt (Table 15) were based upon the number of families that received the specified type of receipt in at least one of the seven years.

Correlation with 1929

The correlation coefficients for 1930-35 income with 1929 show considerable diversity. As years more remote from 1929 are correlated with it, the magnitude of the divergence widens. In general, values decrease to 1933, and increase in 1934 and 1935 (Chart 12). The correlation coefficients thus tend to follow the trend of the volume of income, although interest decreases continuously throughout the period (the sharpest drops occurring in 1931 and 1934) and wages, business profits, and net rents, after showing increased correlation with 1929 in 1934, show less correlation (with 1929) in 1935 than in any other year.

The behavior of these coefficients suggests that the decreases in the volume of income after 1929 were a disturbing influence that gave rise to considerable shifting among individuals, but that the rising income after 1933 tended to return individuals to their former position on the income scale. The upturn in the values of the coefficients in 1934 is all the more remarkable since it might have been expected that the further removed in time any two years are, the less correlated the annual incomes. The pattern of interest would more nearly fit our expectations than that of economic or net taxable income. The similarity of the changes in the coefficients and in the volume of income ends with this broad

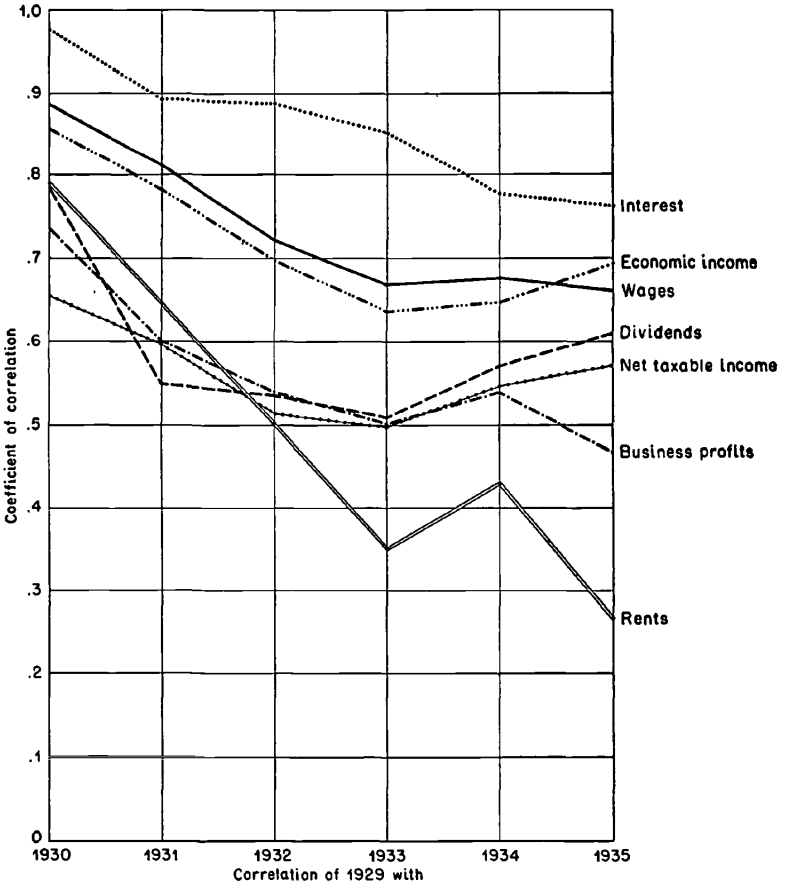
[5] An effort is made below to ascertain whether there is a greater tendency for incomes that decreased from 1929 to 1930 to increase from 1930 to 1931 than there is for incomes that increased from 1929 to 1930. Logically each coefficient should be based upon the same number of persons.

[6] In part this question involves a matter of fact concerning which we have no data. It seems reasonable that as long as a person continues to own a single share of stock he should be included in the distribution of dividend recipients, even though he may actually have received no dividends during the entire seven years. This may be important, especially in the case of the closely held corporation used as a savings vehicle. But income tax returns afford evidence of ownership of securities only indirectly, through interest and dividend receipts.

comparison. For example, a large decline in dividends from 1931 to 1932 is accompanied by a negligible decline in the coefficient of correlation. On the 1929 base, there seems to be no correspondence between the size of the changes in the volume of income and in the coefficient of correlation. A small change in the volume of income may disarrange incomes more than a large change.

CHART 12

Correlation of 1929 Incomes with 1930–1935 Incomes
Sample of Identical Taxpayers



Interest shows, in every year, more correlation with 1929 than does any other receipt or the two over-all measures, economic and net taxable income. Arrayed in the order of their average

correlation coefficients, interest is followed by wages and economic income, dividends, net taxable income, and business profits; net rents show less correlation than any receipt yielding significant coefficients.[7] The largest differences in these averages are between interest and wages, between economic income and dividends, and between business profits and rents. In several cases, the size of these averages is not at all clear from an examination of Chart 12. The coefficients for wages and economic income change order in 1935. For 1931 dividends show less correlation with 1929 than net taxable income and business profits, and for 1932, less correlation than business profits. The average coefficients for business profits and net taxable income are almost equal, although the coefficients for the former exceed those for the latter in every year except 1934 and 1935. In 1930-31 rents yielded larger coefficients than dividends, net taxable income, or business profits, but their continued rapid decline carried them far below these items in 1933-35.

The greater correlation with 1929 observable in 1934, and for a few items in 1935, may be more apparent than real. Quite possibly it is due to the method of selecting the sample, since only those who showed a tendency to maintain a relatively high position in the array of incomes were required to file year after year.

Correlation with 1935

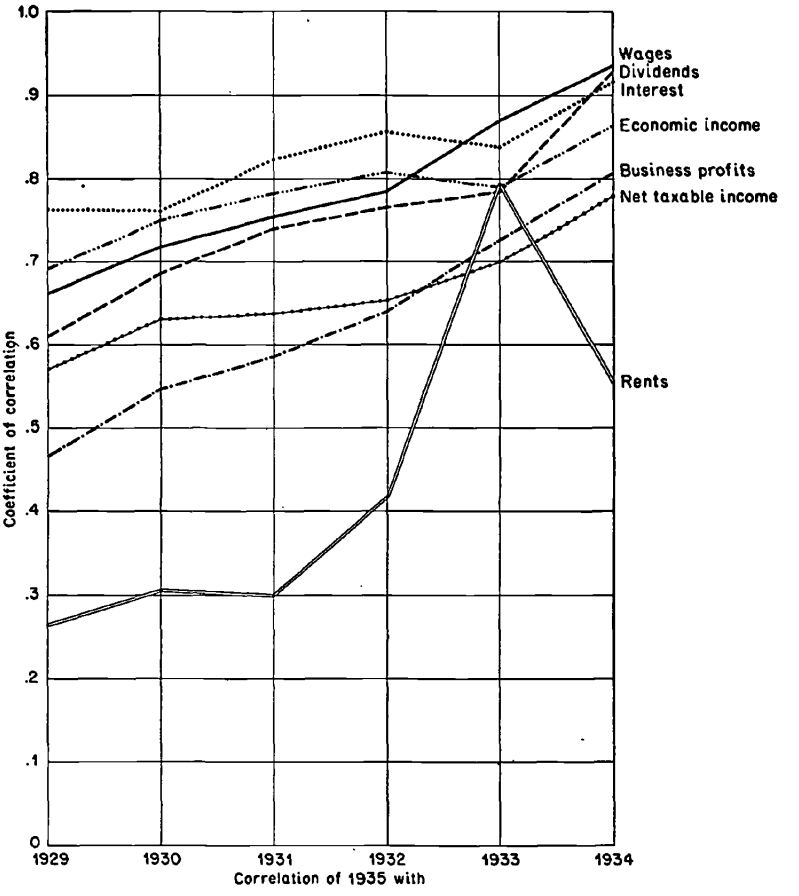
Like the coefficients of correlation based on 1929, those based on 1935 (Chart 13) show considerable diversity. In general, there is more correlation with 1935 than with 1929; and there is less correlation the longer the interval between 1935 and the year correlated with it. Net rents is the one receipt showing an average correlation with 1929 greater than the average correlation with 1935. The coefficients for interest and economic income are lower for 1933 than for 1932 and, except for these two items, there is no evidence that the size of the coefficient is associated with changes in the volume of income. The array of receipts according to their coefficients is the same as that based on 1929, and

[7] The coefficients yielded by capital gains and capital losses are so low that any correlation is doubtful. Consequently, the coefficients yielded by the capital items are discussed below.

is marked by considerable shifting of order, particularly in 1933 and 1934. Except for net rents, the amount of correlation of each receipt with 1935 is more nearly the same than with 1929.

CHART 13

Correlation of 1935 Incomes with 1929–1934 Incomes
Sample of Identical Taxpayers



The higher (average) coefficients based on 1935 than on 1929 would suggest either (1) that for any given time segment the influences making for a new array of annual incomes are more persistent and powerful than those making for the continuance of a given array of incomes, or (2) that the method of selecting the families to be included in the sample operated to eliminate

incomes that showed little or no correlation with those of 1929. The two hypotheses are not inconsistent, but since it is certain that the second influence is present, the coefficients of correlation afford no adequate test of the first.[8]

The scant evidence of an association between changes in the volume of income and the amount of correlation contrasts sharply with the broad outlines of the association between these two variables when the correlation is with 1929. In the present series, the length of the interval between the two years correlated seems to exert the dominating influence. Even interest and economic income, which show more correlation between 1932 and 1935 than between 1933 and 1935, continue to decline as more and more years intervene.[9] As in the case of 1929, there is no close correspondence between the size of the changes in the coefficient of correlation and in the volume of income. Rather there are sharp drops from 1934 to 1933, and again from 1930 to 1929.

Throughout the series based on 1929 and on 1935, the relation between the correlation coefficients and the volume of income becomes confused with the effect of the interval between the years correlated. The latter influence, however, is absent in the coefficients for pairs of consecutive years.

Correlations of Pairs of Consecutive Years

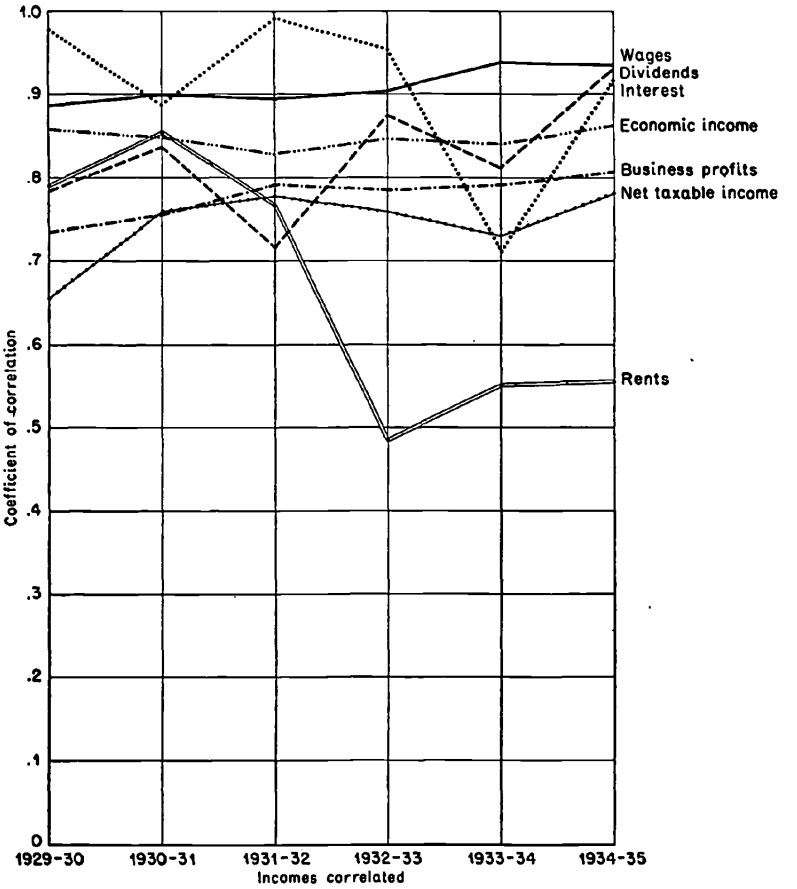
The correlation coefficients between receipts in pairs of consecutive years for three items—economic income, wages, and business profits—are represented, on Chart 14, by almost horizontal lines. For each the lowest coefficient is more than 90 percent of the highest. The coefficients for net taxable income, except that for 1929-30, fluctuate little. Net rents fluctuate most, and both interest and dividends vary considerably from one pair of years

8 Some, though far from conclusive, evidence on the validity of the first hypothesis is given by the coefficients of partial correlation; see below.

9 The behavior of the coefficient for net rents in 1933 is not readily explained. The high correlation indicated by the coefficient is in evidence throughout the distribution. The one extreme return, on which $26,130 was reported in 1933 and $30,430 in 1935, was omitted in computing this coefficient. When this return is included, the coefficient is .90. There is no apparent reason why the net rents of 1933, rather than those of any other year, should be most closely correlated with those of 1935.

CHART 14

Correlation of Incomes in Pairs of Consecutive Years
Sample of Identical Taxpayers, 1929–1935



to another. If it were not for the low coefficient of the correlation for 1933-34, interest would be the most highly correlated; as it is, the average coefficient based upon wages is slightly higher than that for interest. Business profits, quite close to net taxable income in the two sets of coefficients previously considered, is definitely above it in this series. Otherwise, the average coefficients for each item are arrayed in the same order as those based on 1929 and 1935.

For every item the changes in the volume of income seem to be entirely independent of the size of the correlation coefficient.

A large change in income may disrupt the array of individual incomes less than a small change.

Do the lower coefficients for 1929-30 and 1933-34 than for the following pairs of consecutive years mean that the turning points of the business cycle are more disturbing to the income hierarchy than the cyclical phases they initiate? The series of correlations with 1935 seem to be consistent with this hypothesis; the correlation with 1935 declines sharply from 1934 to 1933, and again from 1930 to 1929. The first drop may be explained by the proximity of the years to 1935—it seems reasonable that there would be a sharp decline in the correlation as we move from a contiguous year to a year one year further removed. But the decline from 1930 to 1929 in the correlation with 1935 is more than can be explained by the additional intervening year. On the other hand, the correlation coefficients with 1929 contribute little to this hypothesis. The decline in the correlation coefficients as more years intervened is arrested in 1934, but whether because correlation with 1933 was unduly low, the change in the direction of the movement of income, or the disturbing influence of declining income had been removed is a matter for conjecture. The relation between the 1929-30 and 1929-31 coefficients would suggest that the proximity of the years correlated was more potent than the change in the direction of the movement of income.

Except for this slight piece of evidence of a cyclical pattern, the correlation of a particular type of receipt seems to be about the same for each pair of consecutive years; although the variations in the coefficients for dividends and interest are sizable, even they seem to fluctuate around an average that could be described by a straight line; in any case, there is no cyclical pattern.[10] Net rents, which are appreciably less correlated in consecutive pairs of years after than before 1932, seem to be an exception for which no explanation readily suggests itself.

Capital Gains and Capital Losses

In the preceding Chapter much of the change in the shape of the Lorenz curve as the accounting period was lengthened from one

[10] Large items in only one or two years can cause quite a change in a coefficient; witness the fluctuations in dividends and interest.

to seven years was attributed to the nonrecurring character of capital gains. Partly because these data are available for longer accounting periods, and partly because the coefficients of correlation yielded by capital gains are so low that little confidence can be placed in them (the corresponding coefficients for capital losses are only slightly higher), these coefficients have not yet been discussed.

Since capital gains and capital losses are nonrecurring items, it is to be expected that they will yield low coefficients of correlation for successive years, for correlation presumes recurrence. Furthermore, there is little evidence of a correlation between capital gains and size of income when income does not include capital gains. For 1936 the correlation of capital gains with total income (excluding capital gains) yields a coefficient of +.51; and even this value seems to be biased upward by a few extreme items. Correlation of the capital gains received by families included in the sample of identical taxpayers with economic income yields coefficients ranging from .11 to .23.[11] Similar data are not available

The 917 families reporting only 1 capital gain during the 7 years, 1929-35	.23
The 70 families reporting 4 capital gains during the 7 years, 1929-35	.17
The 15 families reporting capital gains each year 1929-35	.11

for capital losses, but since they also result from capital transactions similar results might be expected.

B Is the Regression Linear?

The line of regression, fitted by ordinary methods of least squares, expresses the average relation between the size of receipts in a given pair of years. If the income histories of all individuals followed the same pattern, this line would suffice. Such a pattern need not be a straight line; a decline in the volume of income might cause bigger decreases in the incomes at both extremes of the distribution than in those in the middle. In such a case the regression would be curvilinear. A linear regression describes the situation when all incomes regress proportionally toward the mean. Deviations of the averages of particular groups from this

11 Oscar F. Litterer computed these coefficients from unpublished tabulations of the Wisconsin Income Tax Study and has kindly permitted us to publish them.

line can be interpreted as meaning that the behavior of these groups differed from the average. A group average above the line of regression indicates that the group fared better than the average; a group average below the line of regression indicates that the group fared less well than the average.

Since the line of regression is fitted to particular data, all group averages could hardly be expected to lie exactly on it. Chance variations would suffice to cause slight deviations from it. If, however, chance variations were solely responsible for the variations of group averages from it, they might be expected to be distributed randomly about it. But if all the averages for the groups from, e.g., $500 to $6,000, are above it, while the averages for the other groups at both ends of the distribution are below it, they can hardly be described as randomly distributed and their deviations attributed to chance. Rather, it would seem to indicate that a linear regression did not fit the data. Similarly, if the average of the $1,000-1,500 group was above the line of regression for all pairs of years correlated, it could be said to deviate 'significantly' from the line of regression.

In Chart 15 the deviations of the group averages from the line of regression are summarized for each pair of years correlated. The heavy line indicates the groups above the line of regression, the light line, groups with averages below the line of regression.[12]

While the groups vary from one type of receipt to another, panel A shows a very definite blocking of the groups that are above the line of regression. For economic income the groups below $500, from $3,000 to $5,000, and from $15,000 to $20,000 have averages above the line of regression, on 1929, in each year 1930-35. The net taxable income groups $2,500-5,000 are the only ones that meet this rigid condition, although there is a strong tendency for all groups in the ranges $2,000-8,000 and $10,000-20,000 to be above average. Wages are above average in the $3,000-8,000 range. Dividends are consistently above average

12 Charted in this way, the magnitudes of the deviations are not shown. Although these magnitudes vary from group to group, from one type of income to another, and from one pair of years to another, this method seemed preferable to presenting only a part of each distribution or reducing the scale so as to show the entire regression. Though some information is omitted, the information contained in charts of this scale is more readily usable.

in only one group, $1,000-1,500. Both dividends and interest behave differently in 1930-32 and 1933-35. In the case of dividends the wide segment of the middle groups; which held up better than average during the first years of decreasing income, was considerably narrowed as dividends reached bottom and began to rise again. The opposite situation is true for interest; a larger segment of the middle classes was above the line of regression after 1933 than before. Although business losses are not shown on Chart 15, when they are taken in account, both ends of the distribution of business profits were above the line of regression throughout the period, while all except a few scattered groups in the middle of the distribution were below average. The middle brackets of rents, like those of dividends and interest, tended to be above the line of regression.

The six sets of regressions for each receipt cannot be considered independent since the basis of comparison is the regression of a later year on 1929. In each regression, a person's position is determined by the income he received in 1929. Extreme classifications, although due wholly to chance factors obtaining in 1929, are continued throughout panel A. A person whose annual income, typically, is $15,000 but who happened to receive only $4,000 in 1929 would give an upward bias to the $4,000 group average during each subsequent year if his income returned to and stayed at its usual level. In the presence of such forces, little reliance can be placed on the consistent pattern of deviations found in this panel. In one sense, each regression may be said to represent all the regressions on 1929.

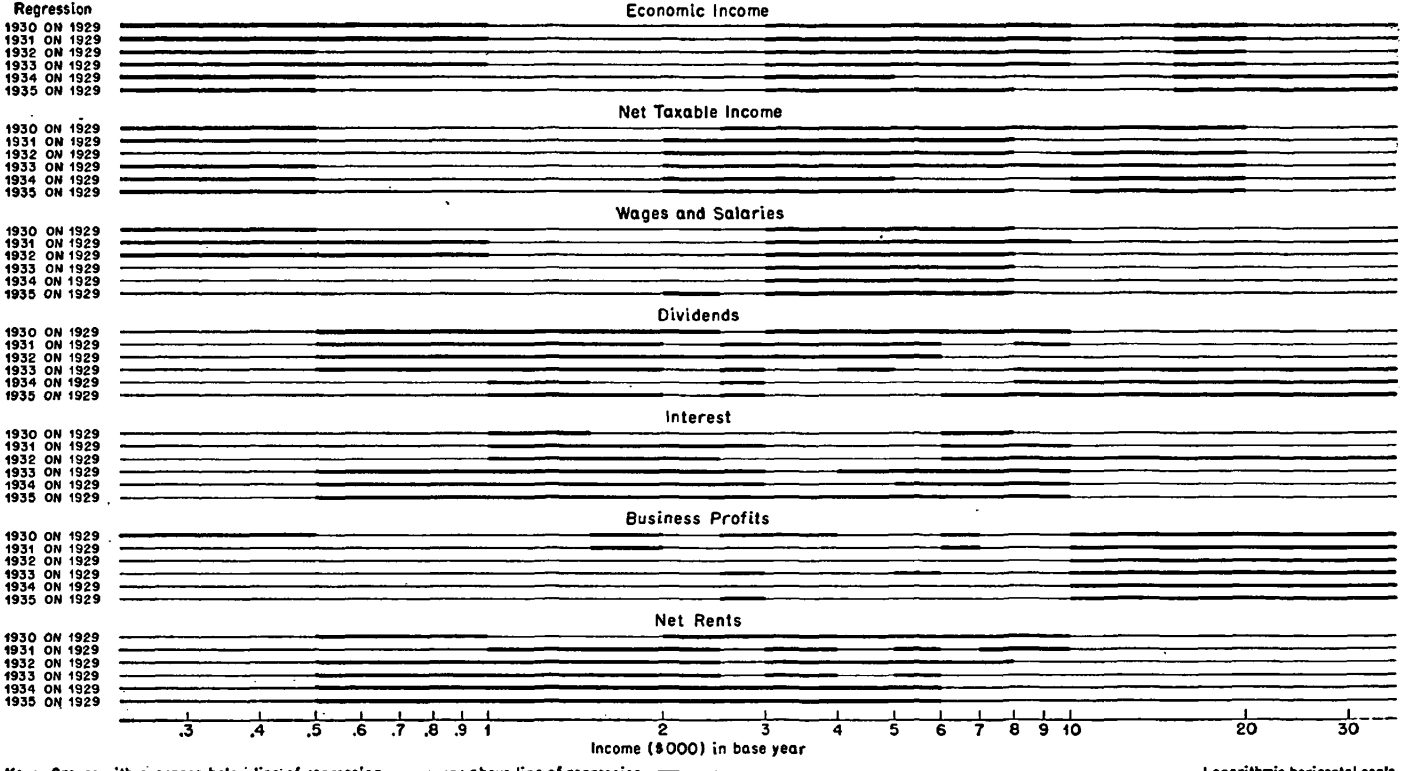
While panel B also has one year common to all regressions, the groups along the X-axis vary with the year correlated with 1935; for example, a person with $4,000 in 1930 and $3,000 in 1931 would be in the $4,000 group in the line for 1930 and in the $3,000 group for 1931, even though he had $8,000 in 1935. His 1935 income, being greater than in earlier years, would not serve to raise the average of the group to which he belonged in the earlier years.

Except for interest, and possibly rents for 1929-32, panel B shows a less distinct pattern than panel A. Interest within the $500-10,000 range averaged higher in 1935 than the line of re-

CHART 15

Relation of Group Averages to Regression Line
Sample of Identical Taxpayers, 1929–1935

Panel A: Regressions of 1930–35 on 1929

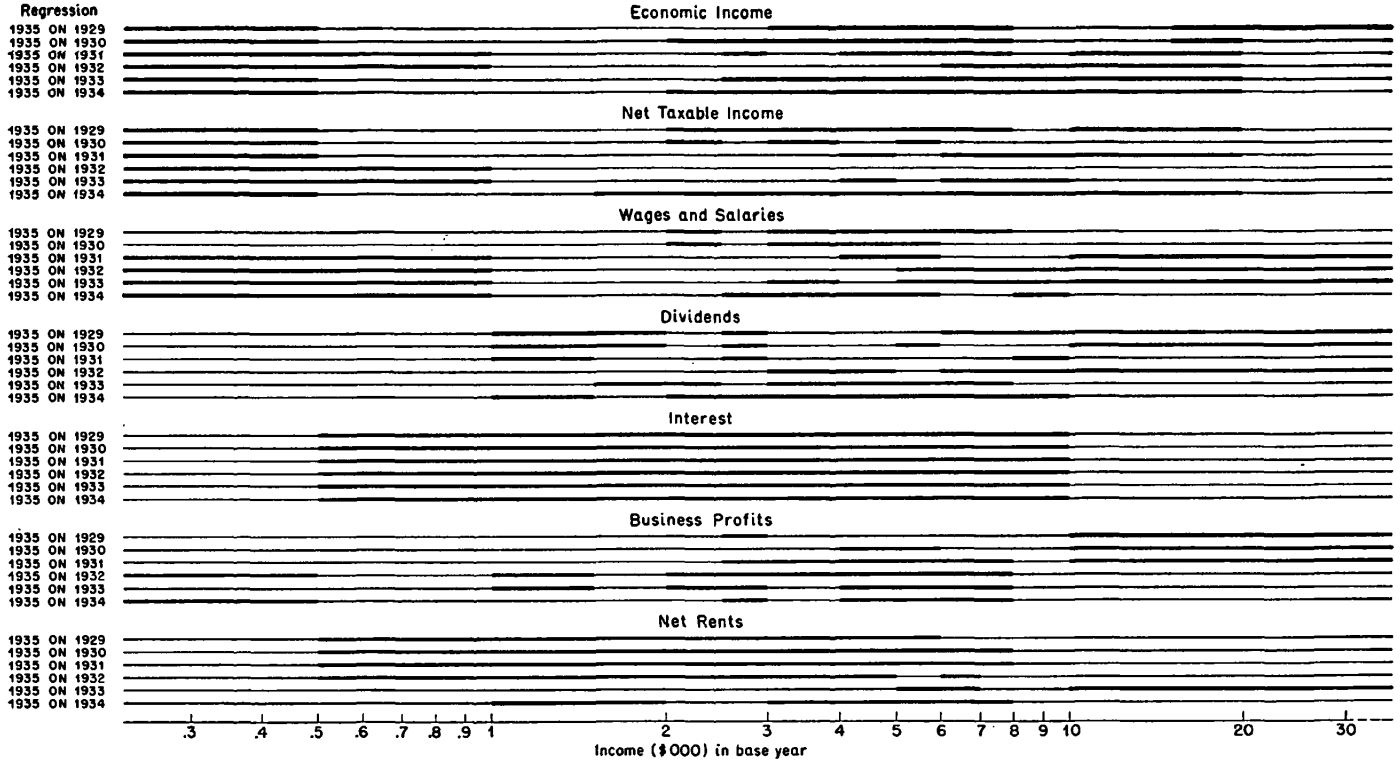


Key: Groups with averages below line of regression ———— ; above line of regression ————.

Logarithmic horizontal scale

CHART 15 *(continued)*

Panel B: Regressions of 1935 on 1929-34

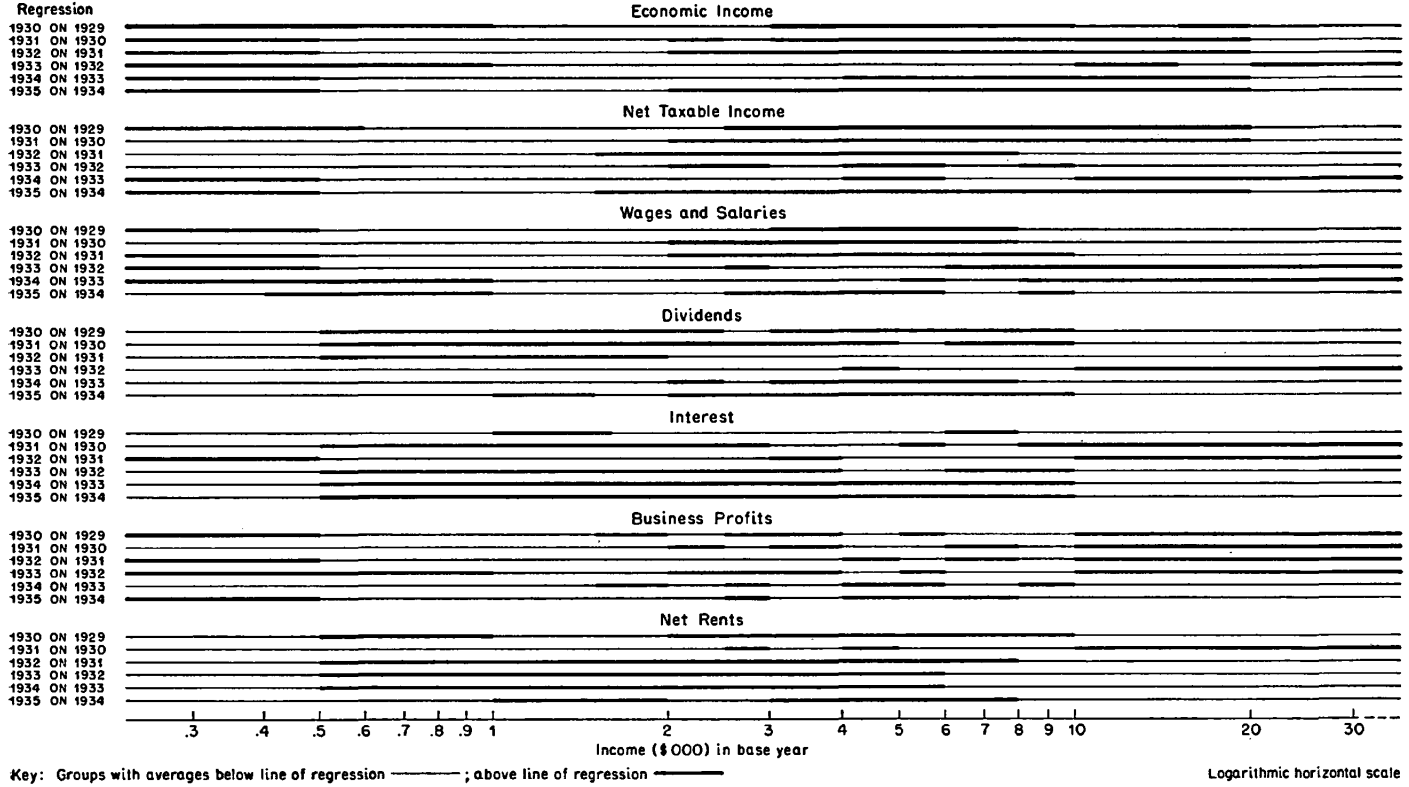


Key: Groups with averages below line of regression ———— ; above line of regression ————.

Logarithmic horizontal scale

CHART 15 *(concluded)*

Panel C: Regressions of the Later on the Earlier of Pairs of Consecutive Years



Key: Groups with averages below line of regression ———— ; above line of regression ————

Logarithmic horizontal scale

gression indicates. Both extremes of the distribution 'suffered' in comparison with the middle groups. Although consistent deviation is not so pronounced, many of the heavy lines cover too many consecutive groups to be attributed to chance alone. The distinct departure from the line of regression within these areas indicates that such a line does not describe adequately the changes in income from one year to another.

The failure of certain specific groups to deviate consistently in one direction from one regression to another and the tendency for several contiguous groups to deviate uniformly within a single regression are even more marked in panel C. Except for a few scattered groups, none shows a consistent departure from the line of regression. Nevertheless, most of the regressions show distinct departures from linearity through large segments of the distribution.

While it is tolerably clear that most of the regressions depart from linearity, it is difficult to draw conclusions concerning the relation of nonlinearity to the amount of reranking. However, its effect is probably small, as most departures of the group averages from the line of regression are small.[13] For example, the 1931 average for the 1930 $5,000-6,000 group is rarely lower than for the 1930 $4,000-5,000 group, and most of these exceptions occur in the lowest two groups.[14] They are positive evidence of reranking.

As previously noted, the coefficient of correlation is merely a rough indicator of the amount of reranking. While an effort was made to exclude extreme returns that lead to absurdly high coefficients, the other incomes are not normally distributed, and there are no satisfactory criteria for determining whether this process of exclusion was carried far enough or, in some instances, too far. Nor does a linear regression adequately describe the changes in income from one year to another. The nonlinear character of the changes from year to year affects all items, but it

[13] The charts drawn for each regression in preparing Chart 15, which omits this information, are the basis for this statement.

[14] Exclusive of the lowest two groups, there are from 176 to 192 opportunities for this to happen for each item. It happens for economic income once, for net taxable income twice, for business profits 3 times, for interest 8 times, for dividends 11 times, and 17 times for rents.

may be expected that error introduced from this source is small and present in all coefficients.[15]

The coefficients differ so much in size that an array of receipts from that with the least to that with the most reranking (measured by the average correlation coefficient) will have some meaning, despite the serious imperfections of the correlation coefficient as an indicator of reranking. Such an array starts with interest; then come wages and salaries, economic income, dividends, net taxable income, business profits, and net rents. The average coefficients for business profits and net taxable income are so close that their precise order is doubtful. The largest gaps in the array occur between dividends and these two items, and between these two items and rents. Consequently, only for rents (and the capital items) would we expect to find the spread between the Lorenz curve based on several years as a single accounting period and the 'average' Lorenz curve for the same period greater than that shown in Chapter 4 for net taxable income. Such a conclusion can refer only to the differences in area encompassed by the two curves; these coefficients cannot indicate the shape of the Lorenz curve based upon a longer accounting period.

The Problem of Groups

The tendency for specific groups to deviate consistently from the line of regression from one pair of years to another suggests that certain groups fared better than others, whether income was decreasing or increasing. The evidence in Chart 15 tends to bear out this hypothesis. If one group of families continues to fare better than another, the distribution of their income would become increasingly less equal, and even if long spans were taken as accounting periods, this tendency would persist. Continued long enough, it would end in the polarization of incomes unless offset by graduated death taxes.

However, a particular group, defined as it is by absolute income limits, does not consist of the same families year after year. Even if every family moved the same distance toward the mean with a change in the volume of income, there would still be con-

[15] Though not necessarily to the same degree. Wide departures from linearity, such as occur in interest and rents, are often positive evidence of reranking.

siderable shifting in the grouping of families. Those grouped in the earlier year toward the extremes of the distribution would be nearer the mean in the later year. Even if such a movement were slight, there would be some reclassification of individuals from one year to the next. With nonlinear regression this tendency may be greater or smaller.

The interpretation of Chart 15, as it relates to the relative fortunes of specific groups, depends upon whether they consist largely of the same persons year after year. If they do, Chart 15 shows the income characteristics of those who fared better or worse than the average. If the populations of these groups change year after year, Chart 15 simply indicates the income levels at which families fared better or worse on the average.

The basic income tax statistics give direct information for only short periods. For any pair of years for which cross-classification tables are available, this information is direct. For example, for any pair of consecutive years 41.0-56.9 percent of the families would be grouped in the same economic income classification. In 1930-35 from 19.2 to 50.8 percent were in the same groups as in 1929, the low point being reached in 1933. Of those in specific groups in 1935, from 26.0 to 56.9 percent were in the same groups in 1929-34. These percentages follow much the same pattern as the coefficients of correlation, and the two measures are, of course, related.[16] More families were in the same economic income groups in 1929 and 1935 than in 1929 and 1933. However, the basic statistics do not tell us whether those who were, e.g., in the $8,000-10,000 group in both 1929 and 1935 include all those who were in this group in both 1929 and 1933.

Despite the shortcomings of the basic data, we may safely conclude that there is sufficient change in the population of specific income groups, defined as they are in absolute terms, to reduce greatly their value as a means of ascertaining the characteristics of families whose incomes were affected by changes in the volume of income more or less than the average. Of more value would be

[16] They comprise only a small part of the data used in computing the coefficients. If each one stayed in the same group, the coefficients would approach unity. If every one changed groups, however, it would not be inconsistent with a coefficient of unity. With wide changes in the volume of income, a high coefficient of correlation presupposes considerable regrouping (in terms of groups defined by dollar amounts) of individuals.

information concerning the extent to which families tend to maintain their relative income levels, i.e., their rank in an array of incomes, in the face of temporary reverses or monetary gains.[17] The coefficients of partial correlation, discussed in the next Section, help to supply this information.

C Coefficients of Partial Correlation

A positive coefficient of partial correlation between incomes in 1929 and 1931, with incomes in 1930 held constant, would indicate a greater tendency for incomes that declined from 1929 to 1930 to rise in 1931 than for incomes that increased from 1929 to 1930. Conversely, a negative coefficient of partial correlation would indicate a greater tendency for incomes that decreased from 1929 to 1930 to continue to decline in 1931 than for incomes that increased from 1929 to 1930. Thus a positive coefficient of partial correlation would indicate that the relative rank of a family in 1929 tends to be more useful as an indicator of its 1931 income level than the change in its income between 1929 and 1930. A negative coefficient would indicate that the change is the more significant.[18]

Two sets of these coefficients of partial correlation can be computed: for 1929-31 with 1930 held constant, and for 1933-35 with 1934 held constant.[19] Except for wages 1933-35, and divi-

[17] This is not to say that knowledge of the proportion of families who are not subject to the influence of a changing volume of income, i.e., who are able to maintain the same absolute income level year after year, is not important.

[18] These coefficients of partial correlation thus provide evidence concerning the validity of the hypothesis suggested above: "that for any given time segment the influences making for a new array of annual incomes are more persistent and powerful than those making for the continuance of a given array of incomes". While a positive coefficient indicates only that a change in income position in a particular year is not one of "the influences making for a new distribution", a negative coefficient is evidence that a new array of incomes is in the making. As will be seen, most of the coefficients are positive. However, the usefulness of the coefficient of partial correlation as a measure of relationship is reduced both by the extent to which the regressions among years are nonlinear and the annual distributions are not normal.

[19] To compute these coefficients, the correlations between a pair of years one year apart are needed in addition to the correlation coefficients between pairs of consecutive years. These are available only for 1929-31 and 1933-35.

dends and rents 1929-31, all these coefficients in Table 16 are positive. Most of them are small, less than .32, but two, rents and interest 1933-35, are exceptionally high.

Relations for longer periods can be tested only incompletely, since tables cross-classifying incomes between pairs of years with one, two, etc., years intervening are not available. Consequently, only one intervening year can be held constant, e.g., only 1933 can be held constant in testing the correlation of 1929-34. If the

TABLE 16

Coefficients of Partial Correlation
Sample of Identical Taxpayers, 1929-1935

	1929 & 1931 with 1930 constant	*1933 & 1935 with 1934 constant*
Economic income	+.20	+.24
Net taxable income	+.20	+.31
Wages & salaries	+.07	−.06
Interest	+.27	+.66
Dividends	−.32	+.14
Business profits	+.10	+.25
Net rents	−.08	+.70
Capital gains	+.09	+.09
Capital losses	+.08	+.13

Based upon the formula: $r_{91.0} = \frac{r_{91} - r_{90} \cdot r_{01}}{\sqrt{1 - r_{90}^2}\sqrt{1 - r_{01}^2}}$ where the subscripts refer to the years, and r is the coefficient of correlation.

coefficient of partial correlation is positive, the product of the two simple correlations between each terminal and a common intermediate year will be less than the coefficient of correlation between the terminal years.[20] We used this fact in Table 17, in which the sign of the coefficient of partial correlation is given for five periods based on 1929 and a like number on 1935. In each case one intervening year is held constant, and the others ignored. The positive sign of eight of these nine coefficients is not evidence that a basic change in the array of families took place.

Moreover, it provides scant evidence of the extent to which families are reranked. A high positive coefficient of partial cor-

20 This can be seen from the formula used in Table 16 to compute the coefficients of partial correlation. Whether the coefficient is positive or negative depends on the numerator, $r_{91} - r_{90} \cdot r_{01}$. If r_{91} is larger than the product of $r_{90} \cdot r_{01}$, the numerator will be positive.

TABLE 17

Signs of the Coefficients of Partial Correlation
Sample of Identical Taxpayers, 1929-1935

YEAR PARTIALLY CORRELATED	YEAR HELD CONSTANT	ECONOMIC INCOME	NET TAXABLE INCOME	WAGES & SALARIES	INTEREST	DIVIDENDS	BUSINESS PROFITS	NET RENTS	CAPITAL GAINS	CAPITAL LOSSES
1929-31	1930	+	+	+	+	−	+	−	+	+
1929-32	1931	+	+	−	+	+	+	+	+	+
1929-33	1932	+	+	+	+	+	+	+	+	+
1929-34	1933	+	+	+	+	+	+	+	−	+
1929-35	1934	+	+	+	+	+	+	+	+	+
1933-35	1934	+	+	−	+	+	+	+	+	+
1932-35	1933	+	+	−	+	+	+	+	+	+
1931-35	1932	+	+	+	−	+	+	−	+	+
1930-35	1931	+	+	+	+	+	+	+	−	+
1929-35	1930	+	+	+	+	+	+	+	+	−

Depending upon the sign of $r_{91} - r_{90} \cdot r_{01}$, where r is the coefficient of correlation and the subscripts refer to the years correlated.

relation, while indicating the importance of position in the income array in a given year, does not mean that there is no reranking. It means only that the position in an array in the earlier year is likely to have more influence on the character of that shift than the change occurring between that year and an intermediate year; that incomes that have decreased tend to increase more than those that have already increased. Furthermore, the coefficients of partial correlation for periods for which complete data are available are small and not uniform in sign. For longer periods, they are based on very sketchy data.

D Coefficients of Variation

Although the coefficients of correlation do not provide the information necessary to determine the changes in the shape and position of the Lorenz curve as the accounting period is lengthened, a summary measure of equality, the coefficient of variation, can be computed from them.[21] This measure, like summary measures of equality based upon the Lorenz curve, e.g., the ratio of concentration, reflects only the net changes from one period to another, and provides no information as to which portions of the distribution were most affected. Nevertheless, it is a convenient summary of the scattered and indirect evidence brought together in this Chapter concerning the changes in income distributions that accompany the lengthening of the accounting period.

Coefficients of variation can be computed only for each of the seven years, 1929-35, for each of the six pairs of consecutive years, and for the two three-year accounting periods, 1929-31 and 1933-35 (Table 18). They reflect, though only approximately, the reranking of families between the two or three annual periods covered. It is the distribution of the two- or three-year totals for each family that are measured. The coefficients of correlation are used to combine distributions, and the coefficients of variation suffer from all their weaknesses.

[21] For a discussion of the coefficient of variation as a measure of equality of income, and its relation to measures based on the Lorenz curve, see Dwight B. Yntema, *The Measurement of Inequality in the Personal Distribution of Wealth or Income* (1931), an unpublished Ph. D. thesis on file at the University of Michigan Library.

TABLE 18

Coefficients of Variation for 1-, 2-, and 3-Year Accounting Periods
Sample of Identical Taxpayers, 1929-1935

ACCOUNTING PERIOD	ECONOMIC INCOME	NET TAXABLE INCOME	WAGES & SALARIES	INTEREST	DIVIDENDS	BUSINESS PROFITS	NET RENTS	CAPITAL GAINS	CAPITAL LOSSES
				ANNUAL ACCOUNTING PERIOD					
1929	1.36	1.05	.87	4.07	4.93	1.39	2.35	3.38	4.56
1930	1.40	.93	.88	3.86	5.16	1.50	2.20	7.63	4.04
1931	1.42	.84	.91	3.88	4.92	1.63	2.29	5.03	3.71
1932	1.41	.91	.99	4.03	5.11	1.88	2.22	6.39	4.04
1933	1.51	.87	1.00	3.74	6.04	1.56	2.09	5.53	3.73
1934	1.47	.92	.98	3.74	6.12	1.50	2.08	6.26	3.69
1935	1.33	.87	.94	3.36	6.18	1.41	2.04	4.04	3.66
				TWO-YEAR ACCOUNTING PERIOD					
1929-30	1.33	.90	.85	3.94	4.76	1.34	2,19	4.38	3.51
1930-31	1.35	.83	.84	3.76	4.86	1.46	2.15	6.36	3.42
1931-32	1.36	.82	.92	3.94	4.63	1.64	2.12	4.40	3.22
1932-33	1.40	.83	.97	3.86	5.32	1.62	2.06	4.38	3.23
1933-34	1.43	.83	.97	3.46	5.79	1.45	2.05	5.07	2.94
1934-35	1.35	.85	.94	3.48	5.81	1.38	1.77	3.75	2.86
				THREE-YEAR ACCOUNTING PERIOD					
1929-31	1.31	.86	.84	3.83	4.54	1.33	2.12	4.08	3.14
1933-35	1.28	.81	.94	3.39	5.68	1.39	1.99	3.59	2.56

The coefficient of variation is equal to the standard deviation divided by the mean. In computing the standard deviations for the two-year accounting periods, the variance (σ^2) was computed by using the formula:

$$\sigma^2_{\left(\frac{X_9+X_0}{2}\right)} = 1/4\left[\sigma^2_{X_9} + \sigma^2_{X_0} + 2r_{90}\,\sigma_{X_9}\,\sigma_{X_0}\right]$$

where X = income, r = the coefficient of correlation, σ = standard deviation, andthe subscripts refer to the years.

The formula given relates to 1929-30. The formula used to obtain the variance for the 3-year accounting period is:

$$\sigma^2_{\left(\frac{X_9+X_0+X_1}{3}\right)} = 1/9\left[\sigma^2_{X_9} + \sigma^2_{X_0} + \sigma^2_{X_1} + 2r_{90}\,\sigma_{X_9}\,\sigma_{X_0} + 2r_{91}\,\sigma_{X_9}\,\sigma_{X_1} + 2r_{01}\,\sigma_{X_0}\,\sigma_{X_1}\right]$$

The formula given relates to 1929-31.

The increase in equality as the accounting period is lengthened (from one to two, and from two to three years) is small; typically not more than 5 or 6 percent of the coefficient of variation. Equality seems to be increased more by the initial lengthening of the accounting period than by the subsequent addition of a third year, although we have no satisfactory measure.[22] This was to be expected, since the two-year period doubles the traditional annual period, while the three-year period is only one and a half times the two-year period. The increase in equality is about the same for each item, although the capital items show erratic changes. Except for these capital items, the increase appears to be associated with neither the equality in the annual periods nor changes in the volume of income.

In a preponderance of the cases, the reranking is sufficient to make the coefficient of variation for a two-year period less than the coefficient for either year included in it. Similarly, the coefficient of variation for a three-year period is usually less than the coefficients for any of the annual or two-year periods it covers. The few cases that deviate from this pattern represent situations in which income in one of the shorter periods is more equally distributed than in the longer, and are scattered by both period and receipt.

There are several minor changes in the array of income items from the most to the least equally distributed. For 1929-30 wages show less variation than net taxable income, while the opposite is true for the remainder of the period. The positions of the capital items also are indefinite. Capital losses seem to be paired with interest, being more equally distributed in three of the seven years. Capital gains are less equally distributed than dividends in five, but more in two years.[23] For the series of

22 When changes in the volume of income have been large, extracting the square root of a simple arithmetic average of the variances for each comparable period and dividing it by the mean yields averages larger than either of the coefficients being averaged. Consequently, a simple arithmetic average of the coefficients of variation was used in making this comparison. Any average of only two items (as is necessary for the three-year periods) is likely to be unsatisfactory.

23 The Lorenz curves for these items show that wages are consistently distributed more equally than net taxable income. Interest is equally distributed in one year, and shows about the same distribution in another. Capital gains and dividends also change positions on the Lorenz curves, although the years in which this occurs are the same for both types of measure only once (see Chart 8).

two-year accounting periods, wages and salaries are more equally distributed than net taxable income in 1929-30 but less in the other periods. Capital gains show less equality than dividends in 1930-31, but in no other period. The only shifting in the array of items between the two three-year accounting periods is between wages and salaries and net taxable income. Based upon the three-year accounting periods, the items are arrayed from the most to the least equality in practically the same order as that shown by the Lorenz curves for annual periods: dividends, capital gains, interest, capital losses, net rents, business profits, economic income, net taxable income, wages and salaries.

Lengthening the accounting period thus seems to increase the equality of the distribution of each income item sufficiently to make it yield a Lorenz curve distinct from the Lorenz curve that represents the average of several annual accounting periods. While each item is not affected equally, the differences are not as great as those between the distributions of each item. Consequently, the array of items from most to least equality is not likely to be changed by the use of accounting periods two or three times as long as the traditional calendar year.