

INCOME AND TAX DISTRIBUTION STATISTICS, 1916-1986

Janet McCubbin and Fritz Scheuren
Internal Revenue Service

The modern Federal income tax is now seventy-five years old and for all of those years the Statistics of Income (SOI) Program of the Internal Revenue Service has compiled income and tax data from the returns which have been filed. For this anniversary we are preparing a number of historical series that look at how what we do has changed [1]. This paper fits into our "celebration" efforts by providing a re-examination of size distributions of individual income and tax shares for 1916 through 1986. These estimated distributions and the methodological innovations employed to produce them are discussed in the context of changes in the tax law and the economy over the period.

Organizationally, the paper is divided into five sections: The first provides some background information, including the motives for doing the research and references to similar studies being conducted elsewhere. The second section calls attention to various conceptual issues which complicate the interpretation of the data. The statistical methods used in the manipulation and analysis of the data are described in the third section and some preliminary results of the study are presented in Section Four. We conclude in the last section by discussing our plans to continue to study these data.

1. BACKGROUND

Many researchers at IRS and elsewhere are interested in using tax data to measure the distribution of income and wealth [2]. It is now possible (thanks to advances in statistical methodologies) to begin looking at the entire set of published SOI cross-sections as if it were a time series of income and tax distribution data.

A word of caution may be necessary to those who would hope to draw broad conclusions from these data. This research is ongoing and the estimates presented here are preliminary. More importantly, the phenomenon we are studying, changes in the distributions of income and tax shares, is so complex that it is impossible to obtain from the data any unqualified conclusions about the equality of the income distribution or tax system in the U.S. (The complexities are discussed further in Section Two.)

The scope of our work differs markedly from that of similar efforts, in that it emphasizes the upper tail of the income distribution and covers a period of seventy-one years [3]. Some of our results can, however, be compared with those of other researchers [4].

2. CONCEPTUAL ISSUES

As we have noted, it is difficult to interpret time series information about the distribution of income and tax shares because of the intricacies of the U.S. economy. The complexities of our data are intensified by the myriad of changes to the Federal income tax laws which have occurred over the period. The most impor-

tant changes in the components of income included in the data occurred in 1944. Through 1943, Statistics of Income data are classified by size of net income. For 1944 and thereafter, Statistics of Income data are classified by size of adjusted gross income.

These concepts are briefly defined below:

- o Net income is the sum of all positive amounts of income (which generally includes wages, salaries, business and partnership profits, some portion of profits from the sale of assets, dividends and rents and royalties) less deductions (which generally include losses incurred in trade or business, some taxes paid, interest paid, some losses from the sale of assets, losses from casualty or theft, bad debts, charitable contributions and medical expenses).

- o Adjusted Gross Income (AGI) is gross income less allowable deductions, which generally include expenses related to business and employment and losses from the sale of property.

Because of this change, income data tabulated for years prior to 1944 are not precisely comparable to data tabulated for 1944 and thereafter.

Other conceptual issues exist and these should be kept in mind when looking at our results:

- o changes in the realization of income;
- o the increase in non-monetary income;
- o supply-side effects of shifts in taxation,
- o the growth in real incomes over time;
- o the sensitivity of our measure to "taste" factors; and
- o income size versus economic well-being.

One of the most important of these issues is the relationship between realized and unrealized income. The tax system and traditional survey measures focus on money income. For most individuals income is on a cash, rather than on an accrual, basis [5]. Changes in wealth due to the appreciation of assets are captured in the data only when they result in increased income flows, such as greater dividend incomes, or in the form of profits from the sale of assets. Unrealized gains are not taxed or reported.

This distinction is important because a great deal of evidence [6] suggests that the way income is taxed determines, in part, the portion of total income which is realized. Lower tax rates should generate larger realized incomes. Individuals who have the ability to control the portion of income that they realize (generally individuals at the upper end of the income distribution) will exhibit the most variation in the relationship of realized to unrealized income. Thus, the effect of a particular tax law revision can be to change the shape of the realized income distribution; however, the effect on the total income distribution seems less clear.

It is beyond the scope of this brief progress report on our statistical results to discuss any of these conceptual issues in depth. We are, however, very interested in the views of those

who read this paper (or who heard us speak in New Orleans). The key questions that we think need to be dealt with include: How well does our varying measure of "income" represent the economic well-being of our population? and, Is even the concept of economic well-being stable?

3. STATISTICAL METHODS

This paper is based entirely on previously published income and tax data for 1916 through 1986. In generating the results discussed here, we proceeded as follows:

- o First, we obtained copies of the original Statistics of Income tables (showing the number of individual income tax returns filed, income amounts and tax amounts by income size class).
- o Second, we transferred the original tabulations to an electronic medium. The result of this work was a set of cross-sectional data with different fixed dollar size classes expressed in current dollars--a very difficult data set to interpret.
- o Third, to overcome some of the problems inherent in using annual cross-sections in a time series context, we subjected the data to a number of adjustments. In particular, these adjustments included transforming the data from current to constant dollars, standardizing the coverage for each year and standardizing the income size classes.

Constant Dollar Adjustment.--We used the Consumer Price Index (CPI) to derive constant dollar amounts. The CPI, probably the most widely used source of information on changes in the prices of consumer goods and services, is the only readily available series covering the years prior to 1929. We do, however, plan to analyze the sensitivity of our estimates to the use of the other price indices, such as the CPI-U-X1. Unless otherwise noted, all money amounts presented here are in 1986 dollars [7].

Standardizing Coverage.--Next we addressed the variation in the data caused by the fluctuation of the income tax filing requirement. The

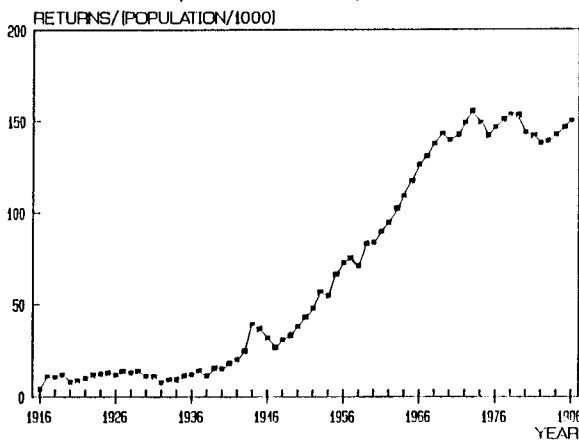
threshold for married filers has varied (in current dollars) from \$5,000 of gross income (in 1921 through 1939) to \$500 (in 1944 through 1947). During the early years of the income tax, only fairly affluent individuals were required to file income tax returns. It really was not until World War II that the income tax became a mass tax. To adjust for the effect of variations in the filing requirement, we have confined our attention to returns with \$25,000 or more of net income (or adjusted gross income for 1944 and thereafter) [8]. Figure A shows the ratio of returns with \$25,000 or more to the total U.S. population for each year. The increase in this ratio reflects the growth in real incomes over the period.

Standardizing Size Classes.--Finally, we employed an osculatory interpolation procedure described in Oh and Scheuren [9] to adjust for variations in the current dollar income size classes used in the original SOI publications. This procedure provided a means for creating data classified by uniform constant dollar size classes or constant percentiles for every year.

4. INITIAL ANALYSIS

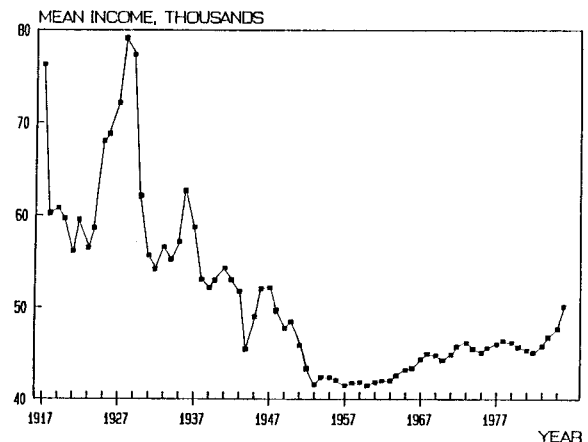
While the number of returns with incomes of \$25,000 or more in constant dollars has increased, the real mean income of the group has generally declined over the 71 year period (see Figure B). Mean income dropped very sharply just prior to and during World War II and then leveled off. It has increased slightly in recent years. One reason for this is that returns above the \$25,000 cutoff have not remained a stable percentile of the total income distribution.

FIGURE A.--RATIO OF RETURNS TO RESIDENT POPULATION, 1916 TO 1986.



For size of resident population, see Statistical Abstract of the United States, 1987, U.S. Department of Commerce, p. 8.

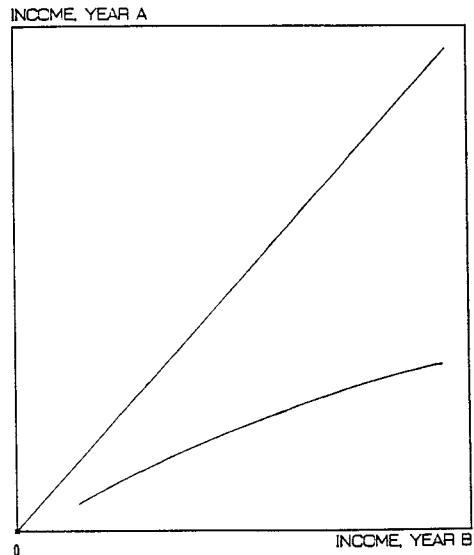
FIGURE B.--CONSTANT DOLLAR MEAN INCOME, 1917 TO 1986.



We have found it useful to employ quantile-quantile analysis [10] to study changes in income and tax shares between years. Quantile-quantile analysis allows us to compare two distributions at a series of common percentiles to evaluate similarities and differences in the shape, location and scale of the distributions. If two distributions are exactly the same, the Q-Q plot will be a straight line passing through the origin with a slope of 1 (see Figure C). If the Q-Q plot is not linear, then the shapes of

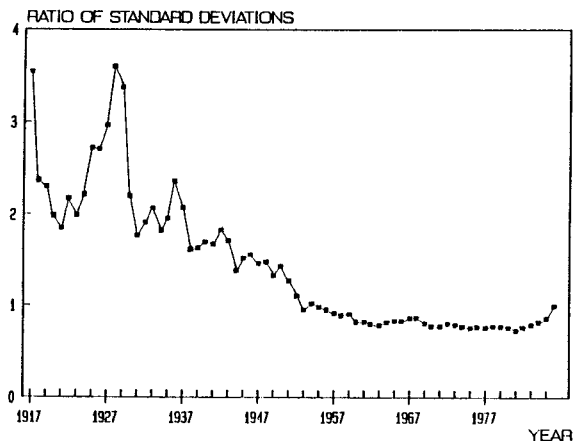
the distributions are different; if the plot does not pass through the origin, then the locations of the distributions are different. The slope of the plot is the ratio of the standard errors of the distributions. We can use this ratio as a measure of relative inequality [11].

FIGURE C.--SAMPLE QUANTILE-QUANTILE CHART.



We have compared the standard deviation of the income distribution for 1986 to the standard deviation for each prior year to evaluate the change in relative income inequality (before tax) over the period (see Figure D). Various economic conditions, such as the boom in the 1920's followed by the Great Depression, contributed to greater income inequality and greater variation from year to year, in the earlier part of this century. After a period of transition in the 1940's, the degree of income inequality has remained fairly stable from year to year. Other data series show similar periods of volatile inequality followed by decreasing inequality and then fairly constant inequality [12].

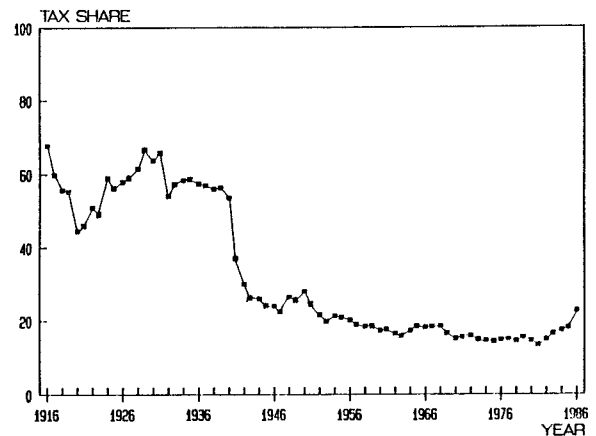
FIGURE D.--STANDARD DEVIATIONS, ALL YEARS RELATIVE TO 1986.



Our data and other data also indicate that income inequality has increased again during the last few years. Some of this increase may be cyclical and a result of the prosperity enjoyed in the U.S. since 1982. Some of the apparent increase in inequality may be a result of changes in the relationship of realized to unrealized income, motivated in part by changes in such things as the tax treatment of capital gains mandated by the Tax Reform Act of 1986 [13]. Another important factor that we have not yet accounted for is the variation in the income concept from year to year, especially between the 1916 through 1943 period and the post-1943 period.

The income tax share of the top 1 percent of returns with incomes of \$25,000 or more follows a similar pattern, particularly for the post-World War II period (see Figure E). The percentage of taxes paid by the top 1 percent of returns with \$25,000 or more was about 20 percent for thirty years or so, but recently increased to about 26 percent.

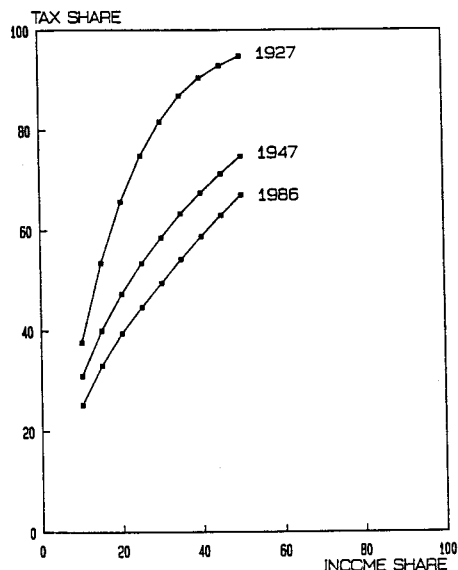
FIGURE E.--INCOME TAX SHARE OF TOP 1 PERCENT OF RETURNS, 1916 TO 1986.



The whole distribution of the tax share of our subgroup of returns with income of \$25,000 or more has, in fact, shifted. Figure F shows the portion of taxes paid by returns accounting for the top 10 through 50 percent of income, for 1927, 1947 and 1986, three years chosen because they are typical of the three periods we have identified. From 1916 through 1937, relative tax burdens were higher for the higher income groups. (Absolute tax shares were, however, much smaller, since effective tax rates were low.) The 1940's, typified by 1947, comprise a period of transition. Effective tax rates increased for the whole subgroup, but tax shares became more evenly distributed, as tax rates increased less for returns with the highest incomes. The year 1986 is typical of the later period (from 1950 on), although it exhibits the highest relative tax shares since the early 1950's.

Finally, we return to quantile-quantile analysis to evaluate the shapes of the before-tax income distributions for a few selected years: 1917 and 1937 are compared in Figure G. Since the function is nearly linear, we can conclude that the two distributions are from the same family (other evidence indicates the Pareto

FIGURE F.--TAX AND INCOME SHARES: 1927, 1947 AND 1986.



family [14]). The locations are nearly equal, but 1937 has a markedly smaller scale parameter. This is consistent with Figure D, which shows income inequality peaking in 1928 and falling sharply thereafter.

The next two graphs conform to the same pattern of decreasing inequality. The locations of the distributions for 1947 and 1967, as well as the scale parameters, are different. The plot of 1967 versus 1986 has a slope greater than 1, which reflects the recent increase in relative income inequality that we have already noted.

5. PLANS FOR FURTHER RESEARCH

Our plans for further work with these data include publishing articles in the SOI Bulletin, enhancing the data base and disseminating the data to researchers outside of IRS. The first Bulletin article, scheduled for publication in the Winter 1988-89 issue, will describe the distribution of tax shares among returns with at

least \$25,000 of income for 1916 through 1950. We will do a parallel analysis for the remaining years in the Spring issue. After having time to deal with at least some of the definitional issues which make year-to-year comparisons difficult to interpret, we will begin to analyze the distribution of income. This analysis will be presented in a summer Bulletin article.

These Bulletin articles will include some tabulations of frequencies and income and tax amounts for constant percentiles of the U.S. Population. We will do more work along this line, especially after we obtain satisfactory estimates of the number of families per year for years prior to 1929 [15]. The data base will be augmented through the addition of items including individual sources of income (for example, wages and salaries, business income, capital gains income and dividends). We have already begun to retabulate the data using a more consistent definition of income. (For example, one way to make the income concept more consistent will be to remove capital gains from the total income amounts.)

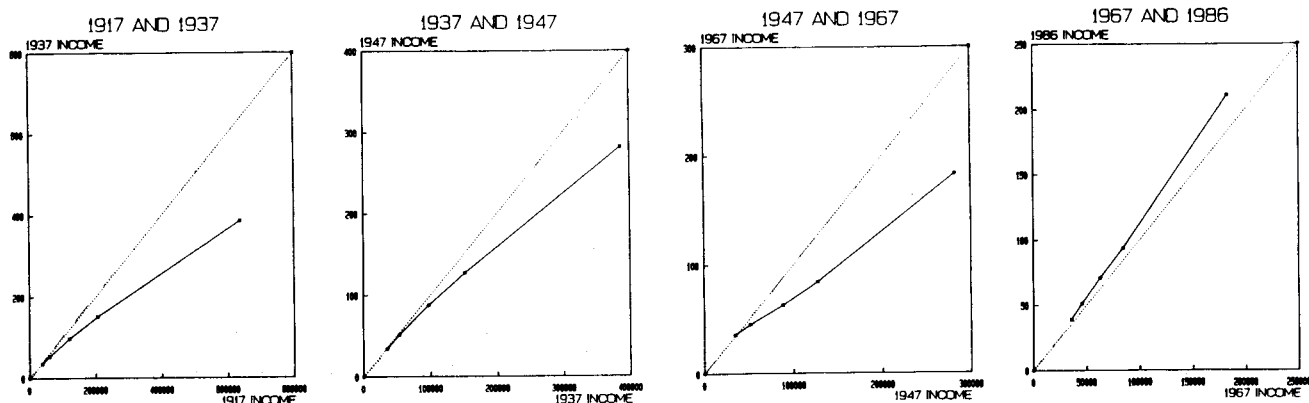
In the more distant future, we want to try to produce simulated microdata for these early years. To do this, we would adjust for changes in the tax treatment of various items over time and then re-weight the microdata available for recent years, so that they agree with the aggregates tabulated for earlier years. We could then use the re-weighted microdata as simulated data for the early years.

Finally, once we have completed the documentation of our work and developed a better understanding of the limitations and meaning of the data, we will release the data, perhaps on a floppy disk to be distributed with the SOI Bulletin. While SOI conducts some analysis of tax data in-house, an essential role of the Division is to collect data to be disseminated to other researchers. It is our hope that many researchers will use this data set to address a wide variety of economic and statistical questions.

ACKNOWLEDGMENTS

The authors thank all of their colleagues in the SOI Division and the Office of Tax Analysis

FIGURE G.--QUANTILE-QUANTILE CHARTS SHOWING INCOME FROM RETURNS ACCOUNTING FOR THE TOP 10 TO 50 PERCENT OF INCOME, SELECTED YEARS.



of the Department of the Treasury for their advice and encouragement. In particular, we are grateful to H. Lock Oh for mathematical and computer support; Wendy Alvey and Beth Kilss for editorial assistance; Nancy Dutton and Sheila Gray for copy preparation and typing support and Dorothy Wallace for data entry.

NOTES AND REFERENCES

- [1] In addition to the Bulletin articles noted in Section 5, a special commemorative issue of the Statistics of Income Bulletin will be published in the Fall of 1988.
- [2] See for example:
 Scheuren, Fritz and McCubbin, Janet, "Piecing Together Personal Wealth Distributions," Statistics of Income and Related Administrative Record Research: 1986-1987, Internal Revenue Service, 1987, pp. 25-44.
 Greenwood, Daphne, "An Estimation of U.S. Family Wealth and its Distribution from Micro-Data, 1973," Review of Income and Wealth, March 1983, pp. 23-44.
 Clark, Philip R., "Estimating After-Tax Income Using Matched SOI-CPS Data," Multi-National Tax Modelling Symposium Proceedings, Revenue Canada Taxation, 1985, pp IV-1 - IV-44.
 Pechman, Joseph A., Who Paid the Taxes, 1966-1985, The Brookings Institution, 1985, Washington, DC.
 Atrostic, B.K. and Nunns, James R., "Measuring Tax Burden: A Historical Perspective," prepared for the 50th anniversary volume of the National Bureau of Economic Research Conference on Research in Income and Wealth, Review of Income and Wealth, University of Chicago Press, forthcoming.
- [3] We have presented data for only a 71-year period here, because only aggregate data, rather than data distributed by income size class, were tabulated for 1913-1915. In addition, data for 1987 are not yet available for publication.
- [4] See, for example, Money Income and Poverty in the United States: 1987, Series P-60, No. 161, Bureau of the Census, Department of Commerce, August 1988; Tax Features, Tax Foundation Inc., Vol. 32, No. 4, June 1988; and Pechman, Joseph A. (1985).
- [5] Feldstein, Martin, "A National Savings President," The Wall Street Journal, November 21, 1988, editorial page.
- [6] See, for example, Steuerle, C. Eugene, "The Relationship Between Realized Income and Wealth," SOI Bulletin, Spring 1983, pp. 29-34.
- [7] We used the Consumer Price Index developed for 1967 to derive the constant dollar amounts. For discussions of various price indices, see Triplett, Jack E., "Reconciling the CPI and the PCE Deflator," Monthly

Labor Review, September 1981, pp. 12-15; and Horrigan, Michael W. and Haugen, Steven E., "The Declining Middle-Class Thesis: A Sensitivity Analysis," Monthly Labor Review, May 1988, pp. 3-13.

- [8] Confining our attention to just one part of the income distribution is, in some ways, controversial. The difficulty in the present context is that the percentage of taxpayers above the \$25,000 threshold changes over time. As will be seen in our concluding section, some of these controversies will be topics for future study.
- [9] Oh, H. Lock and Scheuren, Fritz J., "Osculatory Interpolation Revisited," Statistics of Income and Related Administrative Record Research: 1986-1987, Internal Revenue Service, November 1987, pp. 193-202.
- [10] Wilk, M.B. and Gnanadesikan, R., "Probability Plotting Methods for the Analysis of Data," Biometrika, Vol. 55, 1968, pp. 1-17 (as described in Hoaglin, David C., "Using Quantiles to Study Shape," Exploring Data Tables, Trends and Shapes, John Wiley and Sons, 1985, pp 417-460).
- [11] To estimate the ratio of the standard deviations, we applied ordinary least squares to the function

$$F_{\text{year } n}(P_i) = \mu + \sigma F_{1986}(P_i)$$

where P_i represents the set of constant percentiles, the mean, μ , is a scaled difference between the mean of income in 1986 and income in year n and σ the ratio of the two standard deviations,

$$\sigma = \sigma_{\text{year } n} / \sigma_{1986}$$

- [12] Pechman, Joseph A. (1985).
- [13] Until November 1, 1978, 50 percent of net long-term capital gains over net short-term capital losses could be deducted from income. From November 1978 through 1986, individuals were allowed to deduct 60 percent of the excess of net long-term capital gains over net short-term capital losses. The top individual tax rate in effect during that period was 50 percent; therefore, the maximum effective tax rate on long term gains was 20 percent. Beginning in 1987, the deduction of 60 percent net long-term gains is no longer allowed. Therefore, the maximum rate applicable to long-term capital gains realized in 1987 is 28 percent, the top individual tax rate for 1987. This 40 percent increase in the maximum applicable tax rate undoubtedly led some individuals to realize more long-term capital gains income in 1986 than they would have otherwise.
- [14] Oh, H. Lock and Scheuren, Fritz (1987).
- [15] This is one of the ways in which we intend to address the problems created by using a constant dollar cutoff (see [8]).