# Statistics of Income Studies of Business Income and Taxes

**by Dan Rosa and Dorothy Collins***

This article on business income and taxes is the fourth in a series on the Statistics of Income (SOI) program [1]. Previous articles presented the international statistics program, domestic special studies such as those on tax-exempt organizations and estate taxes, and most recently the studies of individual income and taxes. The present article covers the annual corporation, partnership, and sole proprietorship programs and related studies.

The primary users of studies on business income and taxes in the Federal Government have traditionally been and continue to be the Office of Tax Analysis in the Office of the Secretary of the Treasury, the Congressional Joint Committee on Taxation, and the Bureau of Economic Analysis in the Department of Commerce. Since 1980, however, the Small Business Administration has sponsored periodic studies of employment and payroll associated with corporations, partnerships and sole proprietorships.

The SOI programs for three entity types (corporations, partnerships, and sole proprietorships) are commonly referred to as the "business income tax returns" programs, although there are other vehicles for conducting a trade or business such as farmers' cooperatives, tax-exempt organizations that have "unrelated business" activity, and estates and trusts that have sole proprietorship activity [2]. However, most U.S. business activity is covered on corporation income tax returns, Form 1120 series; partnership returns of income, Form 1065; and individual income tax returns, sole proprietorship Schedules C (for nonfarm businesses and professions) and F (for farming businesses), attached to Forms 1040. Figure A shows the number of business returns by industry for 1985. Nonfarm sole proprietorships are the most common business type by far, numbering almost 12 million for 1985, followed by corporations with 3.3 million and then partnerships with 1.7 million [3].

Figure B presents a much different view. In terms of total receipts, corporations account for 90 percent of the total, reporting $8.4 trillion compared to $5.4 billion for nonfarm sole proprietorships and $3.7 billion for partnerships.

## CORPORATE STUDIES

Beginning with the Revenue Act of 1916, the tax law has required the publication of annual "facts deemed pertinent and valuable" with respect to the operation of the income tax law. The first Statistics of Income (SOI) report which fulfilled these requirements was published in 1918 and contained data for both corporation and individual income tax returns for 1916 [4]. For corporations, the 1916 SOI also contained information for 1909–1915 obtained from the *Annual Reports of the Commissioner of Internal Revenue.* A separate volume presenting corporate statistics that began with the SOI report for 1934 continues through to the present day [5].

### Scope of Corporation Studies

*Program content.*—For all years, the SOI corporate statistics have, in general, included corporations of all types, that are organized for profit [6]. These data are the only source of financial information about *all* corporations. Other sources include only the large or publicly-held, or those in certain industries.

In the beginning, the data items and classifications used in SOI were extremely limited [7]. Their primary purpose was to measure how the taxpayer responded to both the tax law and tax administration system, so that in addition to industry the emphasis was on geography showing where the returns were being filed. Gradually, requests for additional data from tax policymakers and estimators of future tax revenue and from numerous Congressional, Federal, State, and private economic research agencies, resulted in additional data items and size classifications being introduced.

By 1922, data for the complete income statement were published; balance sheet data appeared starting with 1926 following a change in the tax return form. Total assets, the basic size classifier used for corporation statistics, was introduced for 1931 along with the amounts distributed to stockholders. By 1934, a 251-page report was produced presenting income, deductions, assets, and liabilities cross-classified by major industry and size of total assets. Additional classifiers were introduced for returns with net income and returns "with balance sheets" (inasmuch as not all corporations filed them). The 1934 format was retained through the 1957 publication.

Separate tables were added for "Small Business Corporations" (now called "S Corporations") in the 1958 volume, the first year for this new corporate tax entity which allowed certain closely-held corporations to be taxed through their shareholders. (The tax code subsequently created other new types of corporate tax entities for which separate

*Dan Rosa is with the Corporation Statistics Branch and Dorothy Collins is with the Individual Statistics Branch. Significant contributions were also made by Keith Gilmour, Kimm Bates, David Paris, Raymond Wolfe, and Alan Zempel.

**Figure A.—Business Income Tax Returns by Industry, Income Year 1985**

[Numbers of returns are in thousands]

| Industrial division | Corporations | | Partnerships | | Nonfarm sole proprietorships | |
|---|---|---|---|---|---|---|
| | Number[1] | Percentage of total | Number | Percentage of total | Number | Percentage of total |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| All industries | 3,282 | 100.0% | 1,714 | 100.0% | 11,929 | 100.0% |
| Agriculture, forestry, and fishing | 103 | 3.1 | 136 | 7.9 | 319[2] | 2.7[2] |
| Mining | 41 | 1.3 | 62 | 3.6 | 176 | 1.5 |
| Construction | 318 | 9.7 | 57 | 3.3 | 1,454 | 12.2 |
| Manufacturing | 277 | 8.4 | 30 | 1.8 | 326 | 2.7 |
| Transportation and public utilities | 138 | 4.2 | 25 | 1.5 | 546 | 4.6 |
| Wholesale and retail trade | 917 | 28.0 | 201 | 11.7 | 2,289 | 19.2 |
| Finance, insurance, and real estate | 518 | 15.8 | 844 | 49.2 | 1,014 | 8.5 |
| Services | 939 | 28.7 | 341 | 19.9 | 5,532 | 46.4 |
| Nature of business not allocable | 24 | 0.7 | 18 | 1.1 | 272 | 2.3 |

[1] Includes parent corporations filing consolidated returns for affiliated groups of companies.

[2] Represents only businesses engaged in agricultural services. For 1982, the most recent year that *Statistics of Income* data are available on farms, there were nearly 2.7 million farm proprietorships.

statistics were produced, most notably Domestic International Sales Corporations, or DISC's, starting with 1972, and Foreign Sales Corporations, or FSC's, starting with 1985. However, compared to S Corporations, these entities were small in number.)

Beginning with 1959, balance sheet data were published that represented *all* corporations (as asset- and liability estimates were imputed and included in the statistics for the first time for the small number of corporations that did not file balance sheets with their returns). Clearly defined totals for current assets and liabilities could now be derived for the first time (because of another tax form change) and size of business receipts was introduced as a classifier [8].

"Income subject to tax," which in contrast to net income is the base on which tax was computed, was first included in SOI for 1959 [9]. In addition, the net income per books of account was added for 1963 to facilitate comparison with the net income computed under the Internal Revenue Code.

At various times subjects of particular interest were also included in the statistics, many of them on tax-related computations. Examples of such computations involved depreciation; inventories; foreign tax, investment, and certain other tax credits; and net gain or loss from sales of business assets. Examples of other kinds of statistics included from time to time are returns of "controlled group" members and related to these, consolidated returns filed for "affiliated groups" of corporations.

In most respects, the format and content of the most recent report, The *Statistics of Income for 1985,* is similar to that for 1974. It contains the complete balance sheets and income statements, as well as the tax and tax-related items, including tax and payment credits. The classifiers include industry, total assets, business receipts, and the presence of net income. Separate tables are presented for S Corporations. However, some of the more detailed tax-oriented statistics have been discontinued, and somewhat less industry or size information is now published for certain categories. In part, this reflects changing needs of the principal users of the data.

Almost from the beginning the tax before credits (or more specifically the tax after all credits *except* the foreign tax credit) has been the amount emphasized in the corporate

**Figure B.—Total Receipts Reported on Business Income Tax Returns, Income Year 1985**

[Money amounts are in billions of dollars]

| Industrial division | Corporations | | Partnerships | | Nonfarm sole proprietorships | |
|---|---|---|---|---|---|---|
| | Amount[1] | Percentage of total | Amount[1] | Percentage of total | Amount | Percentage of total |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| All industries | $8,398 | 100.0% | $367 | 100.0% | $540 | 100.0% |
| Agriculture, forestry, and fishing | 70 | 0.8 | 9 | 2.6 | 12[2] | 2.2[2] |
| Mining | 142 | 1.7 | 23 | 6.4 | 13 | 2.3 |
| Construction | 387 | 4.6 | 22 | 6.1 | 71 | 13.1 |
| Manufacturing | 2,831 | 33.7 | 23 | 6.3 | 17 | 3.1 |
| Transportation and public utilities | 772 | 9.2 | 12 | 3.2 | 26 | 4.8 |
| Wholesale and retail trade | 2,474 | 29.5 | 70 | 19.1 | 205 | 38.0 |
| Finance, insurance, and real estate | 1,182 | 14.1 | 92 | 25.1 | 31 | 5.8 |
| Services | 535 | 6.4 | 113 | 30.7 | 157 | 29.1 |
| Nature of business not allocable | 5 | 0.1 | 2 | 0.5 | 8 | 1.5 |

[1] Total receipts for corporations and partnerships include income from investments, in addition to income from sales and operations.

[2] Represents only businesses engaged in agricultural services. For 1982, the most recent year that *Statistics of Income* data are available on farms, total receipts for farm proprietorships amounted to $99 billion.

statistics because it comes closer to measuring the total worldwide tax burden of U.S. corporations and corresponds to the total worldwide net income reported on the tax return. Looked at this way, tax after credits represents that part of the worldwide tax that is payable to the U.S. Government [10].

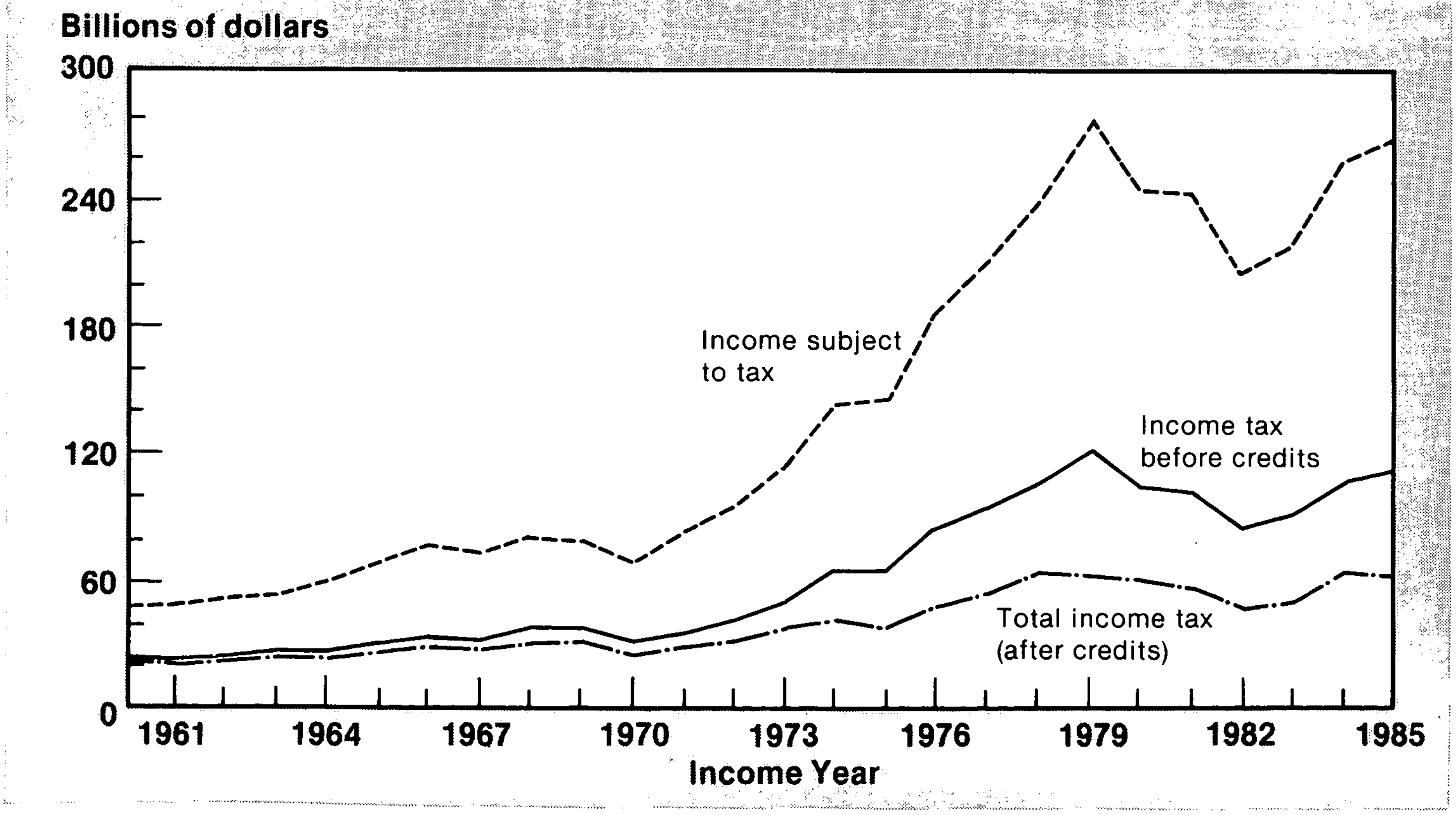
Total income tax before and after credits and the associated income subject to tax are presented for the period 1960–1985 in Figure C. While income tax before credits grew by 410 percent, income tax after credits increased by only 209 percent, from $20.6 billion to $63.7 billion. One reason for this relatively slow growth in the tax after credits is the foreign tax credit, which grew from $1.2 billion for 1960 to $36.8 billion for 1979 [11].

Geographic data were eliminated beginning with the 1970 report [12]. The geographic distributions had become increasingly misleading for economic analyses because they were based strictly on the place of filing or on the mailing address of the corporate headquarters. Often these locations bear little relationship to the place or places where business operations are conducted. This is especially so in the case of the larger corporations. Thus, there is no way of determining from income tax returns alone the amount of income originating in a specific State or the amount of tax payable on that income.

*Source Book.*—Source Books of unpublished data starting with 1926 have been compiled to preserve the detailed cross-classifications of a maximum number of data items (which are summarized for publication in each annual SOI report). As part of a Work Project during the depression years of the 1930's, extensive permanent public records for 1926–1936, in ink, were prepared for each of these years. From 1937 to the present (except for 1952) the Statistics of Income Division has continued the annual production of the Corporation Source Book [13].

*Population Coverage and Item Content.*—Beginning with 1959, there have been over one million corporation returns filed for each income year. The total number has grown steadily since World War II. As shown in Figure D, by 1962 there were over three times as many corporation returns as for 1945. In the 20 years after 1962, the number doubled to approximately 3.3 million. Corporations are projected to continue to grow in number; by Income Year 1990 it is estimated that there will be 4.5 million corporate returnsfiled [14]. While these increases provide a measure of economic growth, they also reflect changes in the tax law, and may

Figure C
**Corporate Income Subject to Tax, Income Tax Before Credits, and Total Income Tax (After Credits), Income Years 1960 – 1985**



also indicate an increasing preference for the corporate form of organization by previously unincorporated businesses [15].

In contrast to the gradual increase in the total number of returns, there has been a wide variation in the number of returns with or without net income over the years, particularly during the period 1916–1944. The depressions of 1921 and the 1930's resulted in more returns without income than with income, so that for 1932, for example, corporations with net income comprised only 18 percent of all those in business.

For 1916–1950 data were extracted for SOI from each return filed. Beginning with 1951, universe estimates were obtained from statistical samples [16]. Over the years the size of the samples has decreased while the population of returns increased, as illustrated in Figure E. (Figure E also shows the "certainty" sample, i.e., returns in the population sampled at the 100 percent rate.) The sample for Income Year 1951 comprised 41.5 percent of the population, or 285,000 of the 687,000 returns filed. For 1985, the sample proportion had decreased to 2.6 percent, or 94,000 returns from a population of over 3.5 million.

Stratification of the 1951 sample was by size of total assets and industry. For 1952-1965, the stratification was by size only—volume of business for 1953–1958 and total assets for 1952 and 1959–1967 with no industry stratification. In the 1960's, the Internal Revenue Service (IRS) began to computerize its manual administrative processing and the SOI sample designation, heretofore a manual operation, was soon also computerized, based on the Master File system which contained accounts for all taxpayers. This new system enabled more efficient, sophisticated, and effective sample designs to be used than under manual sampling. Since the samples could be smaller in size, economies as well as improved data were realized. The first computerized design, for 1968, employed total assets and size of net income or deficit as the major stratifiers.

The use of the employer identification number (EIN), which is permanently and uniquely assigned to each corporation, was also used for the first time to select the 1968 sample [17]. Because a corporation uses the same EIN each year, use of this identifier to select the samples over several years can facilitate selection of returns of the same corporations over time. The advantage here lies in the

Figure D
**Number of Active Corporation Income Tax Returns, Income Years 1916 – 1985**

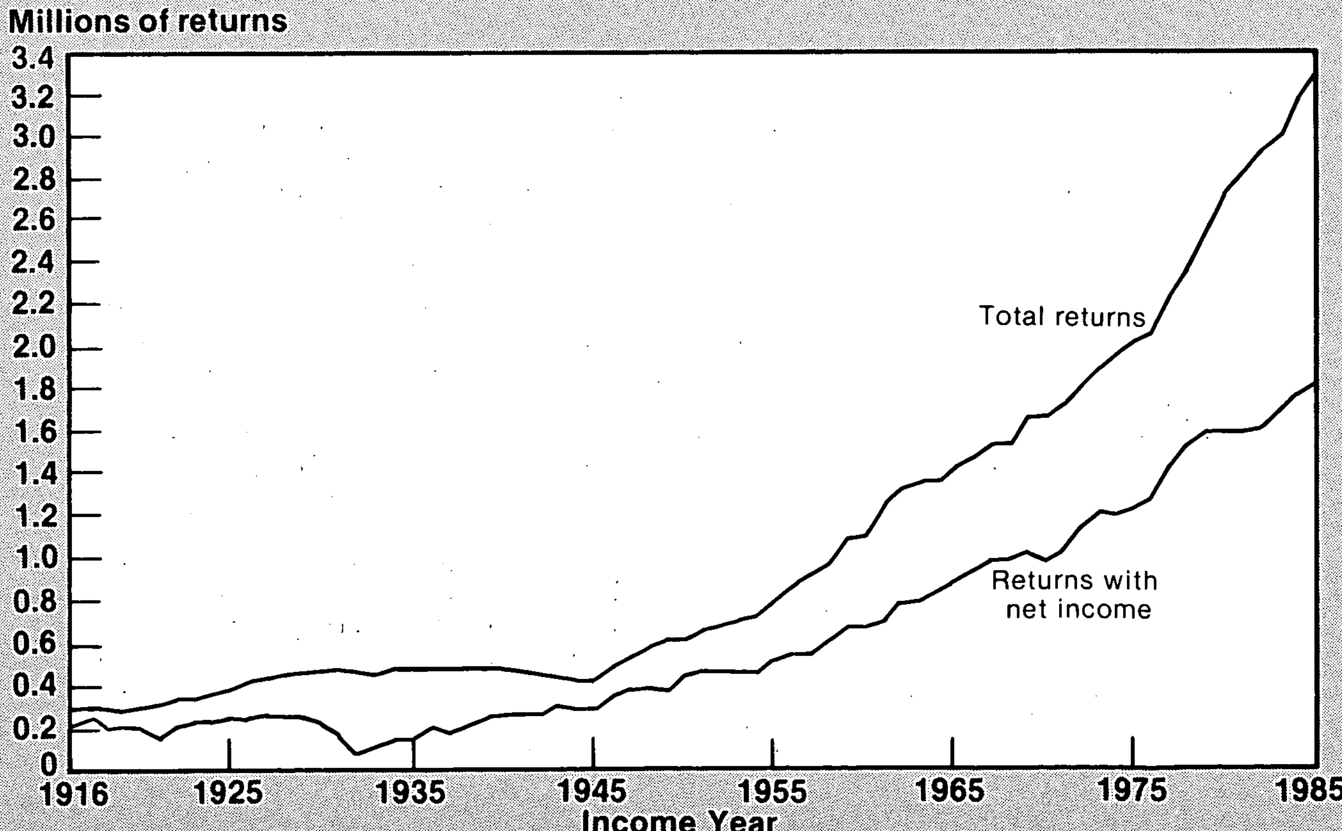
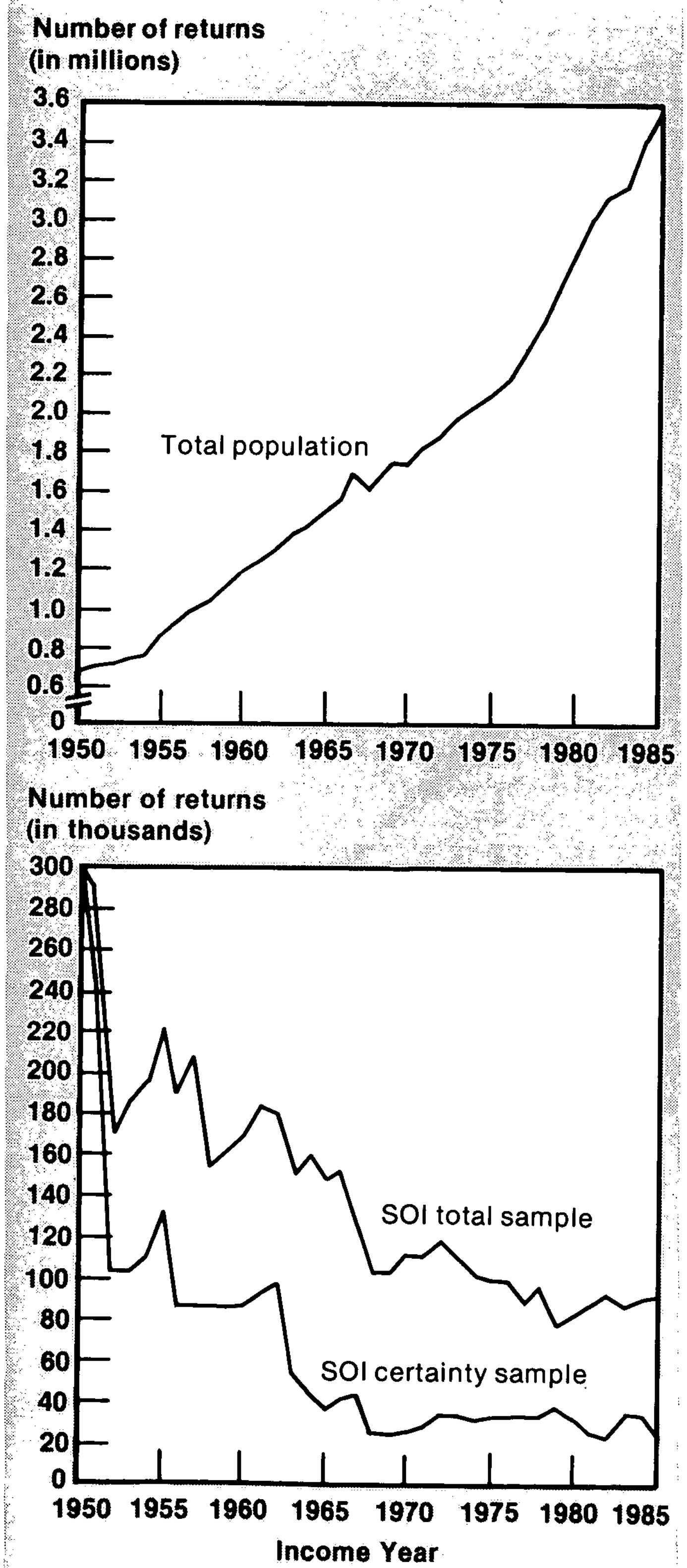


Figure E
**Corporation Income Tax Returns: Number in Population, SOI Total Sample, and SOI Certainty Sample, Income Years 1950-1985**




resultant reduction of sampling variance for the estimates of year-to-year change. Although the SOI sampling scheme has been refined over the past 20 years, the concepts have remained essentially the same as for 1968.

For Income Year 1980 and subsequent years, a "post-stratification raking ratio estimation" procedure has been introduced to improve the industry statistics [18]. Under this procedure, estimated totals are initially developed by weighting the sample to population control totals obtained from the Master File for the two principal characteristics used in the sample design, i.e., total assets and net income (or deficit). Improved totals are then obtained by further stratifying the achieved sample by industry for each asset/net income sample class. As a result, the SOI weighted estimates published by industry are based on a two-way classification of the sample—by the original sample selection criteria using control totals for asset/net income size combinations and simultaneously by the additional criterion using control totals from the Master File by industry.

*Statistical Editing.*—The program for 1916 required obtaining only five items from each return and, by 1931, 41 items were abstracted. This number reached 344 by 1980. Seven years later (1987) the number of items more than doubled to 826. This growth is especially significant given the great amount of "statistical editing" that many SOI items require.

Statistical editing involves adjusting certain taxpayer entries based on supplemental information reported elsewhere in the return, usually in schedules that support a reported total [19]. Editing also includes the constructing of certain totals for the statistics that are reported in a format that differs from the official tax form, using information from other schedules including those improvised by the taxpayer. (IRS permits some latitude on how certain information is reported as long as it is correctly reported.)

Editing is designed to help overcome some of the limitations inherent in tax return statistics that are due to nonstandardized reporting. It also helps to achieve certain statistical definitions desired by a user. An example of the former occurs when corporations file balance sheets of their own design instead of using the balance sheet schedule that appears on the tax return form; in this case, the statistical editor must attempt to recast the taxpayer's balance sheet into the official format of the return so that uniform statistics can be produced. An example of the latter is when editors are required to examine "cost of goods sold" schedules for any depreciation reported there in order to augment the depreciation deduction on the return—an objective that is of far greater interest to tax policymakers and other economists than a cost of goods sold figure that may otherwise be correct from an accounting standpoint.

While statistical editing is minimal in producing sole proprietorship and partnership statistics (and, when it is necessary, can often be accomplished through computerized imputations), it is a major factor in producing corporation income tax return statistics because of the complexity of many of the returns, particularly those of the larger corporations which dominate the statistics.

*Industry Classifications.*—The tax return instructions request that corporations classify themselves by industry according to their principal business activity, as identified from the list of industry groups and codes contained in the instructions. Principal business activity is that which accounts for the largest percentage of total receipts. This can be a limitation because some companies, particularly the larger ones, are often engaged in multiple business activities or may be included in consolidated returns that include all members of an affiliated group, each of which may engage in different activities. In such cases, the largest percentage of total receipts may be relatively small. Year-to-year changes in the classification of specific corporations can result from mergers and other changes in organization, from filing consolidated returns, as well as from changes in the principal source of receipts.

The first Statistics of Income report (1916) contained data for over 100 industries summarized into 27 major groups (currently there are over 180 industries summarized into 60 major groups). Since 1938, the classification structure for all SOI business income tax returns has been based on the Standard Industrial Classification (SIC) system, currently issued by the Office of Management and Budget. However, the SIC is designed to classify "establishments" rather than companies which may be comprised of one or more establishments [20]. In order to apply the SIC to a legal entity or ownership basis for SOI, appropriate groups have to be combined. The resulting industry groupings that are listed in the corporation income tax return instructions and that are used for SOI are, in general, those that tend to be best represented by the corporate form of organization. Thus, for example, more industry detail is available in the corporation statistics for manufacturing or for finance, insurance, and real estate, than is available in the statistics for sole proprietorships or partnerships [21].

Although the industry definitions used for SOI conform closely to the SIC, particular provisions of the Internal Revenue Code are also taken into account. Thus, for corporations, certain types of investment and insurance companies which are defined in the Code are among the groups for which SOI industry data are provided.

*Timing of Corporate Statistics.*—A frequent misunderstanding about the corporate statistics concerns their timing. It is not always apparent to users that many corporations report for noncalendar year accounting periods. In fact, these corporations have now become the majority. Figure F shows that for Income Year 1985, only 42 percent filed for the accounting period ended December 1985.

This noncalendar year filing phenomenon is unique among income tax return filers, even among businesses (nearly all sole proprietorships and partnerships use the calendar year). This means that the corporate "income year" covered by SOI data must be uniquely defined. Figure F shows that an income year includes accounting periods ended July of one calendar year through June of the next. The December-ending accounting period is thus at the center of this span of accounting periods. This compromise has been judged to be the best means of relating the totals for all corporations to a specific calendar year. The validity of this compromise is reinforced by the fact that most of the dollar totals (including those for net income) continue to be reported by corporations that file for the December-ending accounting period (see Figure F).

In addition, many corporations, including some of the largest, request extensions of time (up to 6 months) in which to file. The combination of noncalendar year filing and filing extensions means that while returns for the last included accounting period for Income Year 1985 (ended June 1986) were first due to be received by IRS in September 1986, they could be timely filed as late as March 1987, if there were filing extensions. If these corporations were large, the administrative processing of the returns that precedes statistical processing could then also take a considerable amount of time. The SOI sample for a given year cannot be closed out while large corporation returns are unaccounted for because of their predominant effect on the statistics. As a result, the first results for 1985 were only published in the *Statistics of Income Bulletin* for the Spring 1988; the more complete data for 1985 were then published separately, later on [22].

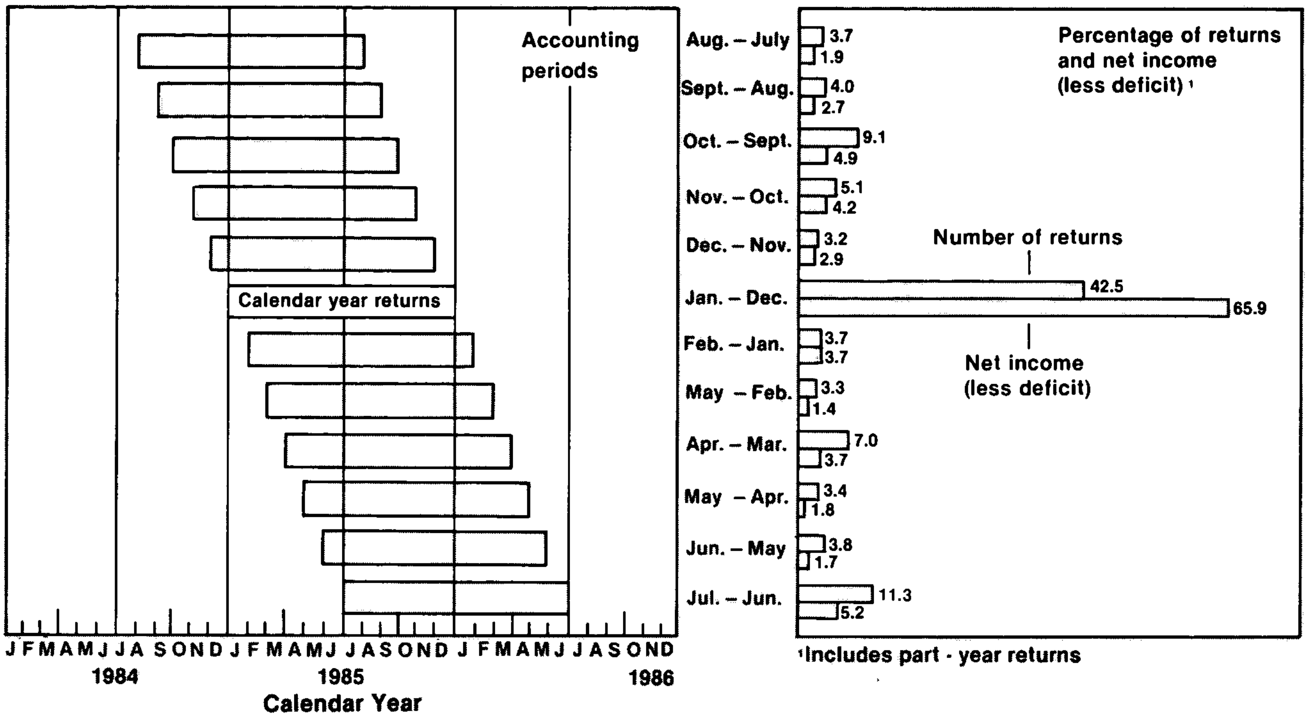
### Principal Users of the Data

Much of the previous discussion has focussed on the SOI publications. However, as is true in most Statistics of Income Division studies, the principal product of the corporate program is now the microdata (record-by-record) magnetic tape files provided to the Office of Tax Analysis and to the Congressional Joint Committee on Taxation for use in revenue estimating and tax modeling, rather than the SOI tabulations and publications that these users formerly relied upon.

As permitted by the Internal Revenue Code, the Bureau of Economic Analysis also has access to corporate records because of its role in producing the National Income and Product Accounts [23]. However, it is not routinely provided with a *complete* SOI sample microdata file in the sense that the Office of Tax Analysis and the Joint Committee staff are. Special detailed tabulations or access to the tax return

Figure F

**Corporation Income Tax Returns and Net Income (less Deficit) by Accounting Periods, Income Year 1985**




records of specific corporations or classes of corporations are usually sufficient to serve its needs.

No public access tax model file exists for corporations mainly to protect the confidentiality of taxpayer information. The very restrictive privacy statutes that prevent disclosing even the fact that a business or individual has filed a tax return require that any corporate data be released cautiously because of the great amount of company data and other information that are available in the public domain [24]. The costs inherent in reviewing statistical tabulations for possible disclosure and in the application of disclosure-avoidance techniques also limit the amount of data that can be published [25]. However, requests for special tabulations can still be produced on a reimbursable basis from the SOI data files provided disclosure can be avoided.

## Future Plans

Each year, the item content and coverage of the SOI programs are reviewed with the major users [26]. These reviews always result in changes to a particular program. Long-range planning is difficult because unforeseen changes in the tax law frequently lead to new or revised statistical requirements. Thus, for example, the new tax research needs emanating from the Tax Reform Act of 1986 are currently severely taxing the ability of the SOI statistical system to deliver timely results. These needs are evidenced in the sharp increase in the number of tax return items for which 1987 SOI data have been requested. The challenge to the Statistics of Income Division is to determine users' requirements early enough in the planning process so that they can be reflected in its budgets for the processing years concerned.

The sample size is expected to remain constant at about 90,000 returns through Income Year 1990, unless special funding by users is obtained to meet their needs. For several years, the Bureau of Economic Analysis has requested but not obtained a special appropriation in its budget to fund an increase in the SOI corporate sample size to about 120,000 returns.

A review of the sample design is needed again to determine if its composition is adequate, given the new pressures emanating from the 1986 tax reform. For example, the proportion of the total sample that is comprised of S Corporation returns (Form 1120S) may be changed given their relative significance to users.

## Special Studies

*Corporation Post-Filing Tax Adjustments.*—The statistics in the corporate program, as in most SOI programs, are derived from returns selected after initial administrative processing, but before audit examination and adjustments. In addition, adjustments such as those resulting from amended returns are excluded. Users have long been concerned about the possible effects these adjustments might have on SOI data and in 1986 a pilot effort was undertaken to determine their magnitude and structure [27].

For this purpose, the corporate tax accounts from the Master File system were obtained for the 94,000 corporations in the Income Year 1982 corporate SOI sample (the most recent sample complete at the time). Unlike the SOI statistical file, which only records data from the tax returns as originally filed for the particular year under study, the Master File maintains selected data from the entire population of returns filed, including numerous items used for tax administration purposes. Generally, these data are maintained for 5 years. Included are adjustment transactions resulting from the filing of amended returns, "carrybacks" of "unused" tax-related amounts, and results of IRS audit examination activity.

Preliminary data from this effort indicate that for corporations active in 1982, post-filing adjustments reduced the originally calculated income tax liability by at least $38 billion over the tax period 1978-1983. This reduction, which represents 11 percent of these corporations' original tax liability after credits, arose principally through unused "net operating losses" and unused investment tax credits carried back from subsequent tax years, rather than through audit examination or other compliance activities.

Current plans are to continue this research with the intent of eventually incorporating the resulting new methodology to improve the SOI statistics. Research is now focussed on issues of sample coverage and the validity for statistical purposes of data obtained from the administrative system.

*Preliminary Corporate Statistics.*—In addition to the comprehensive annual SOI report, preliminary reports were formerly produced in order to provide selected income and tax highlights before the more detailed statistics became available. The separate preliminary reports were discontinued after the 1977 statistics and replaced by more limited data now contained in annual articles published in the Winter or Spring issues of the *Statistics of Income Bulletin* [28].

For Income Years 1956 through 1977, preliminary corporate estimates were scheduled for production in April following the close of the filing period for returns with the accounting periods used for SOI, and were primarily for use by the Bureau of Economic Analysis in updating the estimates of corporate profits published as part of the annual revisions of the National Income and Product Accounts [29, 30]. By 1977, rather than just a by-product of the regular corporate SOI program (as the current and most previous published "preliminary" statistics are and were), these early estimates required so much special processing that they virtually became a separate program that could not be sustained. This special processing included a costly, customized, imputation process to compensate for the significant number of larger corporations whose returns were not yet available for the statistics by the date specified by the Department of Commerce. Shrinking resources, the increasing complexity of the overall corporate SOI program, and later filings by growing numbers of larger-size corporations finally led to the cancellation of this effort.

In order to improve the accuracy and timing of its corporate profit estimates, the Bureau of Economic Analysis has included funds in its annual budget for the last few years that would reimburse the Statistics of Income Division for developing a revised system that could again provide these early estimates. Some research into the development of a "modern" version of the "old" system is being considered in anticipation of the eventual reinstatement of this segment of the SOI program.

*International Statistics.*—Separate studies of business activity conducted abroad by U.S. corporations and of business activity conducted in the United States by foreign corporations are discussed at length in an article published in the Fall 1986 issue of the *Statistics of Income Bulletin* [31]. These studies, on foreign tax credit, Controlled Foreign Corporations, U.S. possessions corporations, Domestic International Sales Corporations, Foreign Sales Corporations, and international boycott participation, to name the major ones, are undertaken partly in acknowledgment of the increasing significance of foreign operations and of foreign corporations, and partly in response to the need for reports by the Office of Tax Analysis that are mandated by tax law. Typically, these studies include a classification of the data by country.

*Secretary's Percentage.*—Foreign life insurance companies doing business in the United States must maintain a minimum surplus of assets to cover their U.S. insurance liabilities. That minimum surplus is determined by multiplying their U.S. insurance liabilities by a percentage required by the tax code that is proclaimed annually by the Secretary of the Treasury. This "percentage," based on data reported on Form 1120L, U.S. Life Insurance Company Income Tax Return, has been calculated annually since 1961 by the Statistics of Income Division.

The processing of data necessary to develop this figure takes place between September and February each year and the results are sent to the Office of Tax Analysis and the IRS Office of the Chief Counsel for review and approval by the end of each February. The Secretary's Percentage is then published in the Federal Register in March. The percentage proclaimed by the Secretary has varied from a low of 13.4 percent for 1961 to a high of 18.9 percent for 1985.

## EMPLOYMENT LINK STUDIES

The Statistics of Income Division conducts a series of periodic studies partially funded by the Small Business Administration that add employment and payroll data to the SOI files of corporations, partnerships, and sole proprietorships. Although payroll is a deductible expense on income tax returns, it is often not clearly identified as such, or it is included as a component of some other deductible expense and, in some cases, is actually capitalized and deducted over a period of years as part of the depreciation deduction.

The special study involves tabulating financial data classified by selected industry and size of business receipts, total assets, and employment. Employment and payroll are obtained from computer tape files of data from Form 941, Employer's Quarterly Federal Tax Return, and Form 943, Employer's Annual Tax Return for Agricultural Employees, that IRS regularly provides to the Census Bureau for its use in preparing its *County Business Patterns* report [32]. These employment tax returns are then linked to the business income tax returns in the SOI samples.

Work on the Income Year 1982 studies is complete [33]. Tabulations have been provided to the Small Business Administration, Office of Tax Analysis, and the Joint Committee on Taxation. The next study is scheduled for Income Year 1987, but is dependent on continued funding from the Small Business Administration or some other source. Plans are to publish the results from time to time in the *Statistics of Income Bulletin* [34].

Figure G shows employment classified by type of business entity. As would be expected, 90 percent of all corporations had employees in 1982. Most partnerships and sole proprietorships, on the other hand, were small businesses without employees. Only 13 percent of partnerships and 9 percent of sole proprietorships had at least one employee.

One interesting employment note in the corporate area is that while 835,000 or 38 percent of all corporations had 1-4 employees in 1982, the greatest number of those employed were by corporations having 500 or more employees. Over 32 million people were employed by the 7,613 largest corporations [35].

In the corporate area, a factor which complicates the linking of the files is that the reporting units for income tax returns and employment tax returns are not always the same. It is possible for a consolidated income tax return to be filed for an affiliated group, while separate employment tax returns can be filed for each group member or combination of members. In order to link records for the same reporting unit, it is first necessary to explore these relationships.

To accomplish the linkage, EIN's are transcribed for all parent corporations and their subsidiaries as reported on the Form 851, Affiliations Schedule, filed with consolidated returns included in the corporate SOI sample. Then a match is performed by computer between the Form 851 file and the file of SOI corporate income tax return records. The Form 851 data are then matched to the employment tax return file. At this point, the employment and payroll for the parent and subsidiary are aggregated so that they represent the same reporting unit as the SOI corporation income tax return.

In the partnership area, research is needed in the area of "independent contractors," i.e., those persons who appear to be providing labor to the partnership but are not being treated as employees for employment tax purposes. The analysis performed thus far indicates that many partnerships reporting a "cost of labor" in their income statements are not reporting the presence of employees through the filing of employment tax forms. Therefore, when an expected match is not made between an employment tax return and a partnership return, it is not clear whether the cause is a mismatch which should be corrected by imputation or a problem caused by an "independent contractor" issue.

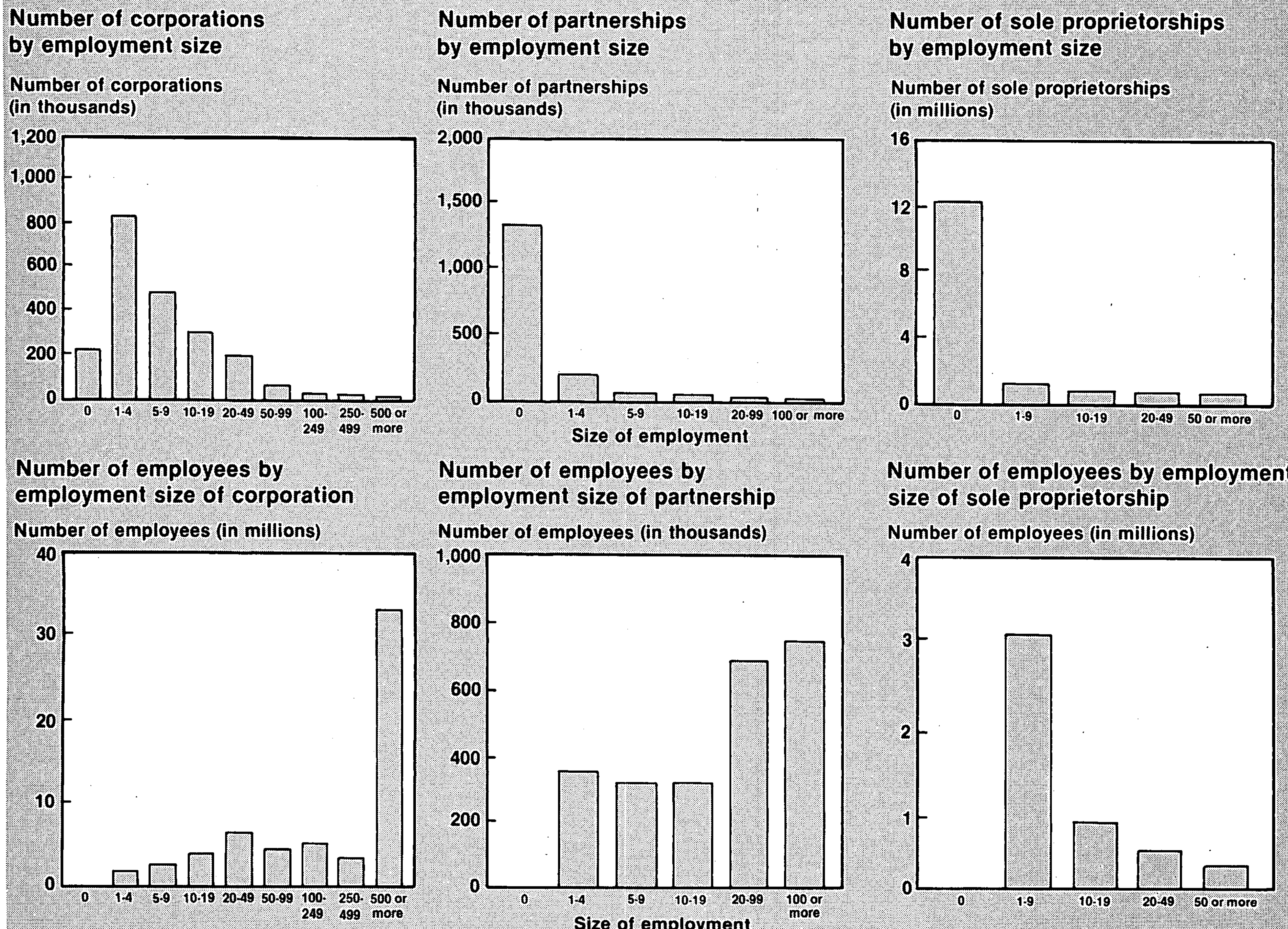
The sole proprietorship study is heavily dependent on imputation techniques to compensate for the often-missing linking variable, the EIN [36]. Nonfarm sole proprietorship activity is reported on Schedule C of the Form 1040, Individual Income Tax Return, and the identifier most important to IRS on this form is the social security number rather than the EIN. The EIN is requested on Schedule C, but only for sole proprietorships with employees and this identifier is often not reported on the return as originally filed (which is the return used for the statistics). (Farm proprietorships have been excluded from these studies.)

Future plans for this project include analysis of the methods used to impute for missing employment and payroll data and research into alternative imputation methods. This research is documented in a paper presented at the 1988 American Statistical Association meetings [37].

## PARTNERSHIP STUDIES

The partnership return of income, Form 1065, is an information return because partnerships are not taxed as

Figure G
## Summary of 1982 Employment, by Size of Employment



such. Nevertheless, an annual return is required containing an income statement; balance sheet; and schedules showing the shares of income or losses and other items, such as credits and foreign taxes, either distributed or allocated to partners. The partners are required to report the distributions or allocations from the partnership on their own income tax returns.

Counts of the number of partnerships since 1917 are available from SOI. The individual income tax return statistics also show the number of returns with income or loss from partnerships and the corresponding amounts starting with 1917.

The Statistics of Income Division did not publish financial data from the partnership returns until Income Year 1939, when they were released in a supplemental SOI report [38]. Data for 1945 and 1947 were published as press releases by the Department of the Treasury [39]. A separate SOI report for partnerships was published for Income Year 1953 [40].

Annual partnership statistics were included in the *Statistics of Income—Business Income Tax Returns* series (together with data on sole proprietorships) starting with 1957 until that series was discontinued after 1976 [41]. Thereafter, separate SOI reports on partnerships only, were released for 1977 through 1980 [42]. Because of budget constraints, the separate partnership report was then discontinued, but with the inception of the quarterly *Statistics of Income Bulletin* in 1981, an alternate publication vehicle became available in which to include some limited statistics about partnerships annually, in each Summer issue [43].

The first of a series of so-called SOI partnership compendiums, *Partnership Returns, 1978-82,* was issued in September 1985 and included tables and analyses previously published in the *Bulletin* for Income Years 1978-82, and in the separate SOI reports for Income Years 1978-80 [44]. It also included additional tables plus sections analyzing trends in the data. Future publications are now planned at 5-year intervals.

### Scope of Partnership Studies

The 1939 partnership report included data for the complete income statement and on the number of partners, classified by industry. These features were to continue for most partnership programs for the years that followed. Balance sheet data first appeared in the report for 1953 and by the late 1950's had become a biennial feature of the program. Data classifications aside from industry were mainly by size of business receipts (starting with 1945) and by size of total assets (starting with 1958). At various times over these earlier years, detailed data were also provided on such topics as the age of partnerships, depreciation and inventory practices, joint ventures, and on some of the items allocated to partners, e.g., "tax preferences" subject to the "minimum tax" and items reported in connection with the jobs or investment tax credits. Little of this information was published for more than a few years, most of it not more then once.

Currently, the annual program includes the income statement, balance sheet, the number of partners and limited partnerships, and capital gain distributions, by industry. Balance sheet data are not included in the annual *Statistics of Income Bulletin* release, but are published instead in each 5-year partnership compendium [45]. State data and data by size of business receipts are no longer compiled.

As in other areas, the partnership program content has grown in response to the needs of its users. In the early years, generally up to the early 1970's when partnerships were known primarily for their trade or service activities, most economists were interested in analyzing the partnership data along with data on corporations and sole proprietorships in order to obtain a complete picture of U.S. business activity.

As the partnership form of organization became popular as a "tax shelter" vehicle, more detail was added to the program. For example, in the early 1980's, in response to Office of Tax Analysis requests, data began to be collected on limited partnerships and from taxpayer-provided schedules on which data for real estate deductions were reported.

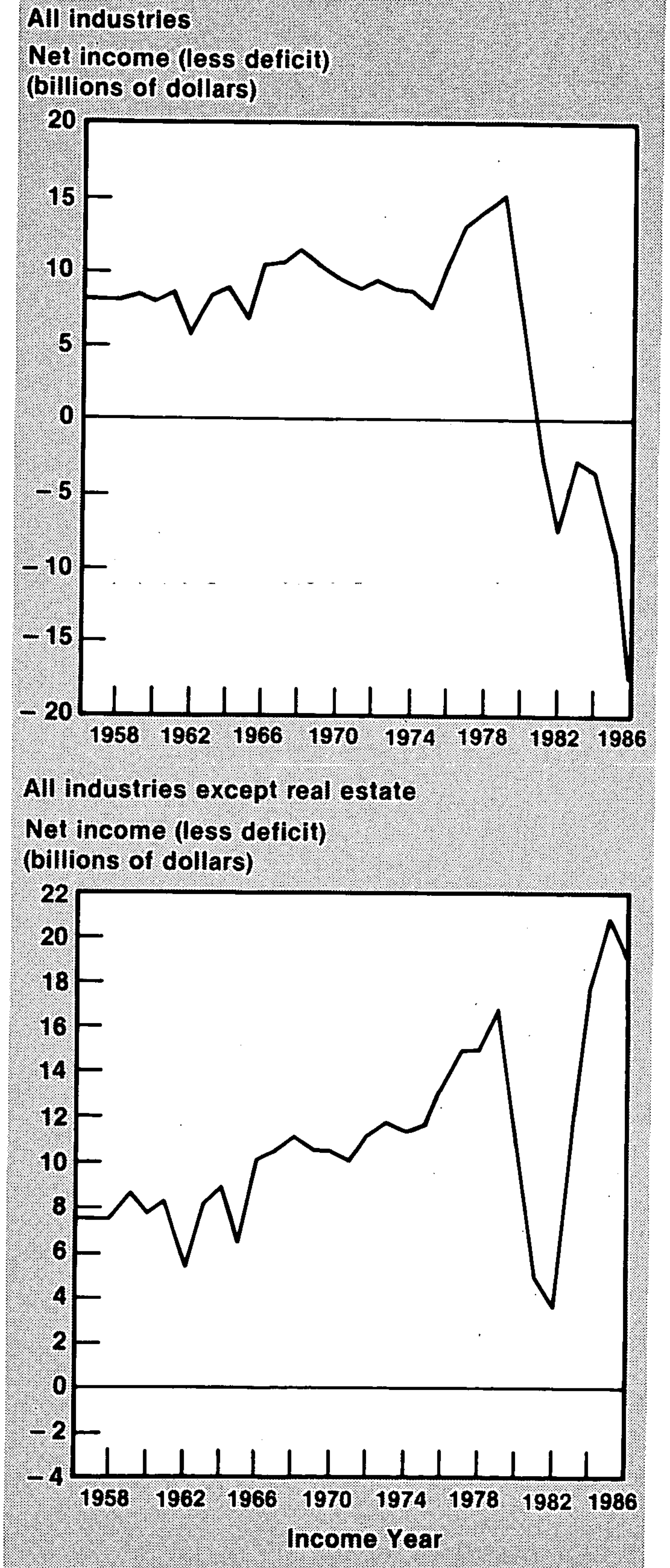
Beginning with 1981 and continuing to the present day, partnerships have reported significant overall losses primarily attributable to tax shelter activity. Most of these losses occurred as a result of real estate operations. In fact, as shown in Figure H, if real estate, which accounts for about one-third of the total number of partnerships, were removed from the statistics, partnership net income (less deficit) would be positive for all years.

### Principal Users of the Data

Most of the requests for additional content in the partnership program have come from the Bureau of Economic Analysis and the Office of Tax Analysis. The Congressional Joint Committee on Taxation is the third principal user.

The Bureau of Economic Analysis uses the data for the National Income and Product Accounts. Partnership income is a minor, but not insignificant, input to the overall national accounts as well as to the personal income accounts component. In this connection, the tax return data are the only source of information about unincorporated businesses. It is thus not surprising that the Bureau of Economic Analysis was instrumental in having the Statistics of Income Division produce partnership as well as sole

Figure H
**Partnership Net Income (less Deficit), Income Years 1957-1986**




proprietorship data each year starting with 1956. Previously, it used SOI data for small corporations as a proxy for the partnership sector, but this proved to be increasingly unsatisfactory. For recent years, the Department of Commerce has also assumed a major role in funding this program.

The Office of Tax Analysis and the Joint Committee are authorized to receive microdata files which they use to model proposed tax law changes and to analyze partnership activity. As with the corporate program, the files provided to the Office of Tax Analysis and Joint Committee contain data which are not published, such as for distributions to partners, depreciation computations, and some limited detail on farm income and deductions. The Bureau of Economic Analysis is not allowed access to the microdata files; instead, it receives detailed tabulations processed to avoid disclosure, which are produced to its own specifications.

Although attempts have been made to construct public use files, confidentiality problems so far have proved too difficult to overcome [46]. As in the case of corporations, requests for special tabulations are produced on a reimbursable basis from the existing data files.

### Population Coverage

The earliest partnership studies included data from every partnership return filed. Partnership returns were sampled for the SOI studies beginning with 1953 using a three-tier sampling scheme which continued with minor modifications until 1969. Using this system, partnerships were sampled at one of three rates based on size of business receipts and total income. (Total income was defined as business receipts plus investment income minus cost of sales and operations.)

From 1970 through 1976, the sample was based on combinations of receipts and total assets with the size stratifiers being adjusted periodically. For 1977, industry was included as a stratifier in that partnerships in the real estate industry were sampled at one set of rates and all other industries were sampled at another [47]. This was done because the characteristics of the large real estate component of total partnerships are significantly different from those of other industries. As before, within each of these two major breaks the partnerships were stratified by combinations of receipts and asset size.

The sampling scheme was redesigned for 1981 to improve the sample in several strata which were poorly represented previously, resulting in many additional sample categories. However, returns were still classified by whether the partnership was in real estate or not and, in addition, by the absolute value of net income (or deficit) and total assets.

Figure I shows the population and sample size for Income Years 1956 through 1986. The significant changes in sample size reflect the users' changing levels of interest over time and their ability and desire to help fund the program.

### Future Plans

One of the major developments now in the planning stage is tied to a modification to Schedule K, Partners' Shares of Income, Credits, Deductions, etc., of the partnership return. This is the form the partnership uses to summarize its distributions to partners. The tax form change would also require partnerships to summarize their distributions by *type* of partner. This kind of information can currently be obtained from the 30,000 returns in the SOI partnership sample only by abstracting data from over 3 million Schedules K-1 filed for partners. By having this information, the Bureau of Economic Analysis will be able to reduce the double counting in the national accounts that occurs when the partner in a partnership happens to be a corporation. Presently, income to the corporate partner is counted once when it is distributed by the partnership and again when it is reported by the corporation. In addition, the Office of Tax Analysis needs to be able to distinguish between "passive" and "active" partners, an important distinction in the Tax Reform Act of 1986, in order to help evaluate the new provisions [48].

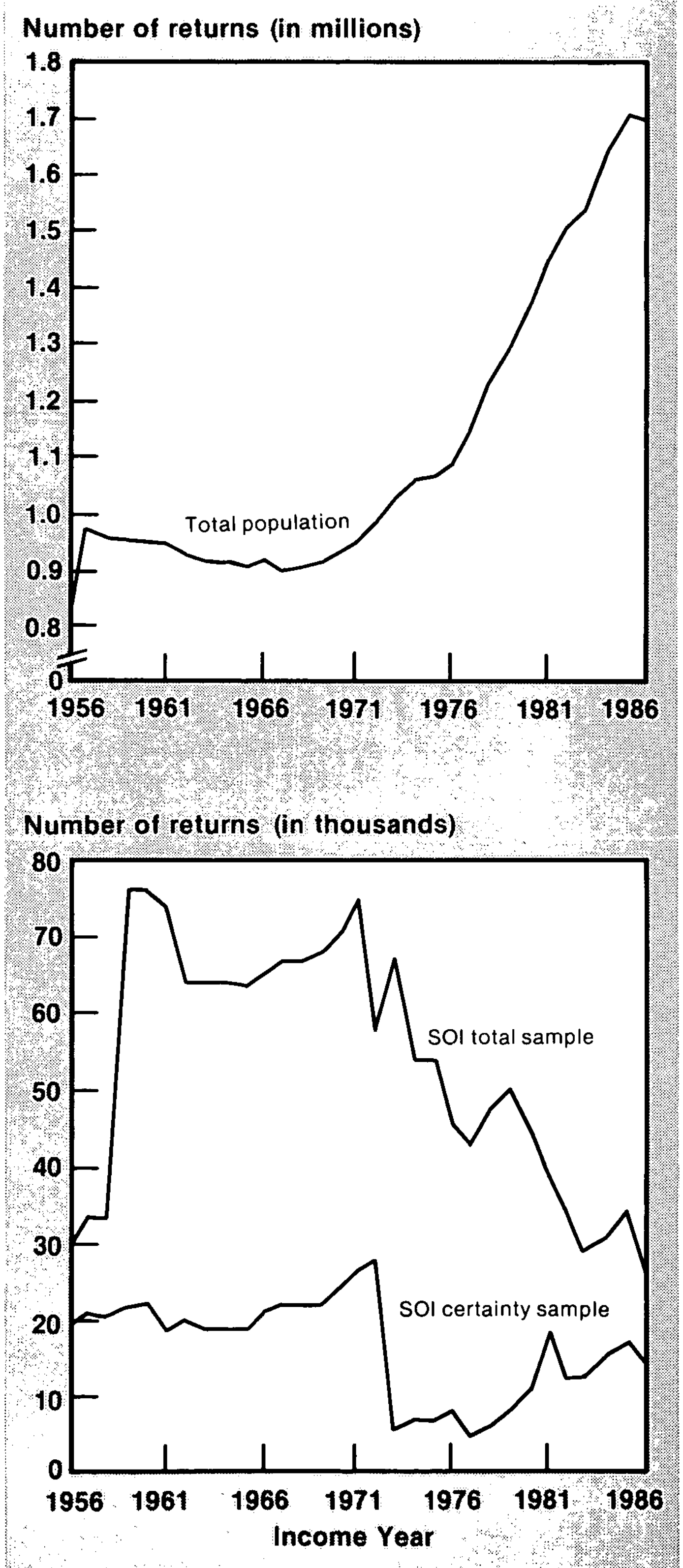
### SOLE PROPRIETORSHIP STUDIES

Information about sole proprietorship business activities is reported as part of the individual income tax return, Form 1040. The profit or loss from nonfarm businesses or professions is reported on Schedule C; that from farming, on Schedule F. In addition, for some of the more recent years, Form 4835 has been used to report the profits of farm landlords, many of whom are indirectly engaged in farm operations. Profits from each of these activities are combined with personal income from other sources on Form 1040 in order to compute "adjusted gross income" (AGI). Therefore, sole proprietorship activities are not taxed separately.

SOI data on the total profit or loss from sole proprietorship activities have been published annually as a source of personal income starting with 1917 (profits from farm proprietorships have been published as a separate source only since 1963). Statistics on at least some of the business activities associated with these profits also began with Income Year 1917. They continued annually thereafter through Income Year 1939.

For 1917 and each year for the next 16 years, the sole proprietorship data were published in the single reports that contained SOI data from all tax returns. Then, from 1934 through 1955 (with exceptions for most alternate years after

Figure I
**Partnership Returns: Number in Population, SOI Total Sample and SOI Certainty Sample, Income Years 1956-1986**



1941 when there were little or no sole proprietorship data), the statistics were published in the SOI report that focused on individual income tax returns [50, 51].

Statistics that excluded farming were compiled for 1956, but they were not published [52]. *Statistics of Income—Business Income Tax Returns,* the new series that began with 1957, concentrated on unincorporated businesses—sole proprietorships and partnerships [53]. This series was discontinued after Income Year 1976, and was replaced by separate reports for each of these two types of business [54].

For budgetary reasons, these reports are no longer published. The last sole proprietorship report was for Income Year 1981. However, sole proprietorship statistics have continued to be produced, although in a much reduced form and with data on farming no longer tabulated on a regular basis. After 1981, just one or two tables on sole proprietorships have been published each year in the Summer issue of the *Statistics of Income Bulletin.*

**Scope of Sole Proprietorship Studies**

The history of the SOI sole proprietorship program is characterized by several breaks in continuity. Even though the series dates from 1917, the industries and businesses covered, as well as the content of the program, have changed over time. These changes were due primarily to changing user needs, changes in the tax law, introduction of the first SIC system in 1938, processing decisions, as well as budgetary considerations. Sole proprietorship studies predate those on partnerships by many years and have been far more varied in terms of their scope and coverage.

Data for most of the earlier years were confined to the number of businesses with net income and an amount for net income, classified by 19 industry groupings. The 1917 data also included business receipts, but the statistics for that year were limited to returns with net income $2,000 or more. For 1917, there were 252,000 such businesses reported in SOI, with net income that totalled $2.7 billion [55]. Loss business were not tabulated for any of the early years.

This emphasis on businesses "with net income" continued for many years. With exceptions for 1921 through 1925 and then for 1928, the earlier statistics by industry were confined not only to businesses with net income, but to returns that showed a net income from all sources of at least $5,000 [56]. Most individuals were not subject to the income tax and of those who were, many were thus excluded from the sole proprietorship statistics. Business coverage was, therefore, extremely limited. As a result, the industry data published were probably of limited use for most economic analyses. Nevertheless, for the early years of the series they were the only financial data published by the U.S. Government about sole proprietorship businesses, just as they are now.

Starting with 1933, totals by industry were published for businesses reporting a loss. The data also began to be classified by size of business net income or deficit. In addition to profits, data were presented for business receipts and wages paid. Meanwhile, the industry detail gradually increased so that by 1939, information was provided for more than 100 groups.

Throughout the 1930's until 1939, the statistics continued to be limited to businesses reported on returns with net income of $5,000 or more. The expansion in coverage for 1939 to include returns with net income under $5,000 was thus a major improvement. Mainly as a result of this change, the number of sole proprietorships shown in the statistics increased from almost 130,000 for 1938 to nearly 1.1 million for 1939 [57]. The corresponding net income (less deficit) rose from $0.8 billion to nearly $2.5 billion [58]. In addition to business receipts and net income or deficit, data for the cost of goods sold, purchases of merchandise bought for sale, and total business deductions were also included in the 1939 statistics; and another size distribution was added, based on business receipts. This classifier continued to be used until recent times.

During the 1940's and early 1950's, detailed business data were not produced as frequently, fewer industry groupings were used, and no industry detail was provided for loss businesses. On the other hand, there were major improvements in coverage. When the tax code was broadened during the early 1940's to include most of the population (and therefore most of the sole proprietors), the resulting increase in the businesses reported on individual income tax returns was clearly reflected in the statistics. Sole proprietorships grew in number from 2.0 million for 1940 to 5.7 million for 1945 [59]. A further refinement in coverage resulted from enactment of the social security self-employment tax in 1951. When appended to the income tax filing requirements based on gross income, all proprietors with net earnings of $400 or more from self employment had to file individual income tax returns that also included information about their business or professional activities. The filing requirement for self-employment tax purposes increased the stability of the sole proprietorship coverage; while the income tax filing thresholds were to change several times over the years that followed, the low filing threshold for self-employment tax remained the same.

The complete income statement was first tabulated for 1945, by industry division. Only business receipts and profits were then presented separately for more detailed industry groupings. A similar pattern of presentation was resumed starting with 1959, in the *Business Tax Returns* series, and also continues to the present (although currently farm data are excluded). Data for 1986 (the most recent year available) show that there were 12.2 million nonfarm proprietorships distributed over more than 200 industries.

Their net income (less deficit) was $90.4 billion [60]. In comparison, for 1959, there were 5.8 million nonfarm proprietorships, and their net income (less deficit) totalled $18.9 billion [61].

During the 1960's and 1970's, SOI sole proprietorship statistics also explored a variety of special subjects in efforts to meet as many data needs as practicable. Included were depreciation and inventory practices; the cost of depreciable property; information about the investment and jobs tax credits claimed by proprietors; and, in general, the nonbusiness and total income characteristics of proprietors. For several years the income and deductions unique to farming operations were shown and geographic data were featured. As with the partnership statistics, most of these special subjects were not repeated on any regular basis and none of them are included in the current statistics.

### Principal users of the data

As in the case of partnerships, the Bureau of Economic Analysis is the principal requester of nonfarm sole proprietorship statistics, for use in the National Income and Product Accounts. The Bureau of Economic Analysis was instrumental in reviving an annual sole proprietorship program starting with 1956 and in expanding both its item and industry coverage. For more recent years, the Department of Commerce has also funded most of the program. Prior to 1956, Commerce used SOI sole proprietorship data for years when they were available and filled in the gaps for the national accounts by using statistics for small corporations as a proxy for extrapolation, as it did for partnerships. Sole proprietorship statistics are also used extensively by the Office of Tax Analysis and the Joint Committee on Taxation. Each utilizes the data in its continuing evaluation of the effects of the tax law on small businesses and their owners.

The Department of Agriculture is the principal user of the statistics on farm proprietorships. For years when SOI data on farming are compiled, it uses them as a check on its own farm income and expense data derived from surveys. Data that show farmers' principal source of receipts by type of farm commodity and on farm deductions are of particular interest.

As with corporations and partnerships, the Office of Tax Analysis and the Joint Committee are authorized to receive the microdata files containing information for each sole proprietorship return in the SOI sample. These files also include data for items not shown in the published statistics. The Bureau of Economic Analysis does not have access to the data files. However, the detailed tabulations it receives are generally sufficient to meet its needs.

Public use files were created for Income Years 1980 and 1983. For 1982 only, a file was created that contains information about farm businesses only. In order to preserve taxpayer confidentiality, all three files exclude business returns with AGI of $200,000 or more and the data included have been "edited" [62]. In addition, steps have been taken so that the sole proprietorship data cannot be linked with other individual tax returns information, such as that contained in the Individual Tax Model file [63]. There are no public use files for more recent years, however, special tabulations can be produced on a reimbursable basis.

### Population coverage

*Number of Businesses.*—In general, the sole proprietorship studies provide data on all unincorporated, single-owner businesses, whether conducted on a full-time or part-time basis, so long as the owner (the sole proprietor) meets the income tax or social security self-employment tax filing requirements. Beyond this though, the population shown in SOI is somewhat imprecise, especially over a period of years. This imprecision in counting the number of businesses is quite apart from the exclusion from the earlier statistics of businesses reporting a loss and those reported on returns with total net income under $5,000.

Often, classification for the statistics depends on how the business income is reported and whether a business schedule is attached as part of the tax return. Also, until 1937, the statistics included the relatively small number of sole proprietorship businesses reported by fiduciary agents of estates and trusts. Starting with 1937, the data have been limited to businesses owned by individuals. More notable is the imprecision caused by changes made in the treatment of multiple businesses with the same owner. These changes also complicate comparisons with the number of business "establishments" published by the Bureau of Labor Statistics and the Bureau of the Census.

In the early years of SOI and currently, all businesses reported by a taxpayer were combined and classified based on the predominant business activity so that, in effect, the industry statistics are actually for sole proprietors (the business owners) rather than for sole proprietorships (the businesses themselves). This simplifies data processing and the discrepancy created is not large. In fact, the relationship between the number of businesses and the number of owners is close, about 1.1 to 1 [64].

In the intervening years, efforts were made to improve the counting of multiple businesses for SOI, so that only multiple businesses with the same business activity (and the same spousal owner, in the case of joint returns) were combined for the statistics. On the other hand, if the income and expenses of an identical business were equally divided between husband and wife on a joint return and separate business schedules filed for each spouse, they were combined for SOI. Moving toward more recent times, businesses were counted however they were reported, although for awhile they were limited to a maximum of three per owner.

The sole proprietorship population is often thought to equal the establishment population, especially for smaller-size firms. However, if ownership of a business establishment changes during a year, the establishment is counted more than once for SOI because each owner files his or her own business schedule to report for different parts of the year. Some establishments can consist of more than one business, e.g., the restaurant facilities in a bowling alley or the chair in a barber shop that is leased out by the owner of the shop. Similarly, in the sense that an establishment has a fixed place of business, some businesses may not be establishments at all, e.g., self-employed taxicab drivers.

*Sampling of returns.*—Sampling of individual income tax returns, including those with business schedules, started with 1925. For many years, only returns with total net income from all sources under $5,000 were sampled at less than 100 percent [65]. The sampling classes were determined exclusively with reference to the return processing categories used for tax administration. An early SOI report notes that, with the exception of 1928 when the sample of returns with net income under $5,000 was increased so that certain returns including those with business schedules were sampled at 100 percent, the SOI samples were inadequate for purposes of producing sole proprietorship statistics for all businesses [66]. Limitations due to sampling were avoided by limiting the sole proprietorship data for most of the early years to businesses reported on returns with net income of $5,000 or more [67].

For 1943, many new sampling classes were added and other improvements introduced in selecting returns from the various processing categories used for tax administration. Sampling of returns with net income under $20,000 at less than 100 percent had become a necessity because of the sharp increase in returns filed after the income tax was extended to most of the population. This dollar cutoff for the certainty sample was gradually increased as more sophisticated designs were implemented and the size of the overall samples declined substantially.

From 1945 through 1955, the so-called "basic" SOI sample, now based on AGI rather than net income, was augmented for the smaller returns with business schedules in order to produce sole proprietorship statistics by industry [68]. The additional returns were not used for the rest of the individual income tax return statistics, however. Since 1956, the so-called business returns in the sample have been selected at rates that are generally higher than those used for other individual income tax returns. In addition, the same samples have generally been used for all of the individual income tax return statistics, including those for sole proprietorships. A major change for 1966 was the first use of the social security number as the basis of sample selection. Previously, the sample had been selected based on document or taxpayer account numbers.

Beginning with 1968, sample designation was computerized, based on the tax return records processed for the IRS Master File. (Previously, the samples had been manually designated from the IRS tax return categories.) This enabled further improvements to be made in the sample designs which no longer had to be based on the IRS categories. The revised criteria were based on combinations of the larger of AGI and the largest specific source of income, and of business receipts reported by farm and nonfarm proprietors. Until 1973, farm proprietorships were sampled at the same rates as nonfarm proprietorships. Since then, with an exception for 1982, they have been sampled at the same lower rates used for nonbusiness returns.

For 1982, the sample was redesigned and is currently based on the larger of total positive income or total negative income (i.e., total loss) per return, as well as on the size of farm and nonfarm receipts of sole proprietorships. For 1982 only, farm returns were sampled separately at rates somewhat higher than nonbusiness returns in order to accommodate a reimbursable request by the Department of Agriculture for farm industry statistics. Currently, there are over 30 sampling classes.

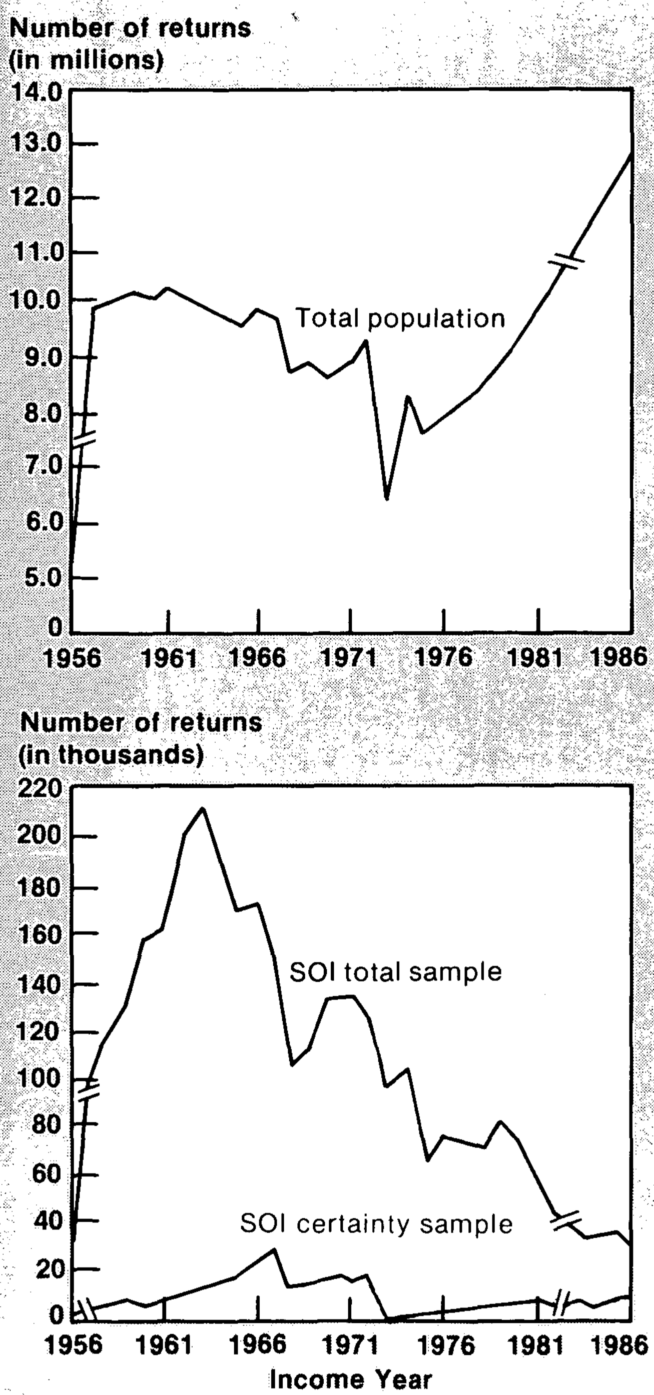
Figure J is a summary for 1956 through 1986 of the total number of individual income tax returns classified as business returns, the total number in the SOI samples, and the number in the samples that were included at the 100-percent rate, i.e., in the certainty sample. For Income Years 1957 through 1980 the data in Figure J are based on the number of individual income tax returns with Schedule C (nonfarm proprietorships) or Schedule F (farm proprietorships). For 1956 and in general since 1981, the data have been based on returns with Schedule C (to reflect the exclusion of farm proprietorships from the statistics).

## Future Plans

In response to the increased attention focussed on "tax shelters" and taxpayers' participation in activities designed to yield "passive" losses, information relating to these two areas is being included in the sole proprietorship studies, beginning with Income Year 1987 [69]. Through a series of questions on the return form, the taxpayer is asked to indicate whether the business is a registered tax shelter and whether he or she materially participated in the operation of the business. Responses to these questions are being coded during statistical processing and statistics will be developed based on them.

Figure J

**Sole Proprietorship Returns: Number in Population, SOI Total Sample, and SOI Certainty Sample, Income Years 1956-1986**




In response to requests from users, information regarding other aspects of taxpayers' businesses is being gathered and will be tabulated. To assist the Bureau of the Census, the number of months the business was in operation during the year and information indicating whether or not the business was operating at the end of the year will be coded and included in the statistical file. In addition, data will also indicate whether expenses for a home office were deducted.

The "three martini lunch," a popular topic among tax reformers, also is being addressed in the sole proprietorship program. Using the expanded reporting requirements in the travel and entertainment area resulting from the Tax Reform Act of 1986, SOI data beginning with Income Year 1987 will contain the total amounts spent, as well as the limited amounts actually deducted.

Topics which have been included in past studies will be continued for future SOI sole proprietorship studies. Information regarding sex of proprietor, accounting and inventory practices, and "at risk" investments in the business, has been requested by various users for their on-going research endeavors and the items supporting this research will continue to be included in the sole proprietorship program [70]. In addition, data on farming may be included from time to time, depending on the availability of funds from the Department of Agriculture.

**NOTES AND REFERENCES**

[1] See Skelly, Daniel F., and Hobbs, James R., "Statistics of Income Studies of International Income and Taxes," *Statistics of Income Bulletin*, Fall 1986, Volume 6, Number 2; Skelly, Daniel F., and Kozielec, John A., "Statistics of Income Domestic Special Studies," *Statistics of Income Bulletin,* Fall 1987, Volume 7, Number 2; and Coleman, Michael J., "Statistics of Income Studies of Individual Income and Taxes," *Statistics of Income Bulletin,* Winter 1987-1988, Volume 7, Number 3.

[2] See Skelly, Daniel F., and Kozielec, John A., "Statistics of Income Domestic Special Studies," *op. cit.*, Scheuren, Fritz J., *Statistics of Income—1963, Farmers' Cooperative Income Tax Returns,* U.S. Department of Treasury, Internal Revenue Service, 1966, and Estep, Gary J., "Fiduciary Income Tax Returns, 1982," *Statistics of Income Bulletin*, Spring 1985, Volume 4, Number 4.

[3] Unlike the corporation and partnership statistics, those for sole proprietorships currently exclude farm proprietorships. For 1982, the most recent year that *Statistics of Income* data are available on farming, there were nearly 2.7 million farm proprietorships. See

Wolfe, Raymond, "Sole Proprietorship Returns, 1982," *Statistics of Income Bulletin*, Summer 1984, Volume 4, Number 1. See also the discussion on sole proprietorships later on in this article.

[4] Much of the material in this and following sections is taken from the 50-year historical summary contained in *Statistics of Income—1965, Corporation Income Tax Returns,* U.S. Department of the Treasury, Internal Revenue Service, 1968. Prior to enactment of the income tax in 1913, there was a so-called excise tax on the income of corporations (1909–1913).

[5] For the first 18 years, i.e., Income Years 1916 through 1933, only one *Statistics of Income* report was prepared each year. Beginning with the 1934 statistics, a separate report for corporations was instituted. From time to time, supplemental SOI reports for corporations have been issued, particularly on the foreign tax credit and on Controlled Foreign Corporations.

The first separate report was *Statistics of Income for 1934, Part 2,* U.S. Treasury Department, Bureau of Internal Revenue, Washington, DC, 1937. The most recent report is *Statistics of Income—1985, Corporation Income Tax Returns*, U.S. Department of the Treasury, Internal Revenue Service.

[6] The return types included for Income Year 1985, which is the most recent year for which statistics are available, are:

Form 1120 (U.S. Corporation),
Form 1120-A (U.S. Short-Form Corporation),
Form 1120S (U.S. S Corporation),
Form 1120L (U.S. Life Insurance Company),
Form 1120M (U.S. Mutual Insurance Company),
Form 1120F (U.S. Returns of Foreign Corporation),
Form 1120 IC-DISC (Interest-Charge Domestic International Sales Corporation), and
Form 1120-FSC (Foreign Sales Corporation).

[7] These items and classifications consisted of State in which the return was filed, industrial activity, "gross income," total deductions, net income (or deficit), and income tax.

[8] In regard to current assets and liabilities, previously only selected accounts on the tax return balance sheet could be identified as current or long-term.

[9] For a more complete explanation of "income subject to tax," see footnote 8, Table 13, in the Selected Historical Data section of this issue of the *Statistics of Income Bulletin*.

[10] For a discussion of this income tax definitional concept (in the context of individual income tax returns), see the analysis contained in Lerman, Allen H., "High Income Returns for 1984," *Statistics of Income Bulletin,* Spring 1987, Volume 6, Number 4.

[11] See Carson, Chris R., "Corporate Foreign Tax Credit, 1982: A Geographic Focus," *Statistics of Income Bulletin*, Fall 1986, Volume 6, Number 2, and Barlow, Mary, "Foreign Tax Credit by Industry, 1982," *Statistics of Income Bulletin*, Spring 1986, Volume 5, Number 4.

[12] Data classified by State were produced through 1951 and again for 1953. For 1955–1969, in order to emphasize that the data were intended primarily for use by tax administrators, the geographic data were labelled as for Internal Revenue Districts and Regions, rather than States. The geographic statistics needed most for tax administration continue to be tabulated from other Internal Revenue sources and are summarized in the *Annual Reports of the Commissioner and Chief Counsel, Internal Revenue Service.*

[13] A general description is contained in *Source Book of Statistics of Income,* Publication 1053, including ordering information, is available from the Director, Statistics of Income Division TR:S, Internal Revenue Service, 1111 Constitution Avenue, NW, Washington, DC 20224.

[14] U.S. Department of Treasury, Internal Revenue Service, "Number of Returns to be Filed," *Projections,* Document 6149 (Rev. 12–87).

[15] The last study on the growth and "age" of corporations was contained in *Statistics of Income—1959-60, Corporation Income Tax Returns.* This report also compares the 1959 data with the results of related studies published in *Statistics of Income* reports for the mid-1940's.

[16] Much of the material in this discussion of sampling was obtained from Jones, Homer W., and McMahon, Paul B., "Sampling Corporation Income Tax Returns for Statistics of Income, 1951 to Present," *Statistics of Income and Related Administrative Record Research: 1984*, U.S. Department of the Treasury, Internal Revenue Service, 1984.

[17] For a discussion of how this identifying number is used, see Harte, James M., "Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS," *Statistics of Income and Related Administrative Record Research: 1986–1987,* U.S. Department of the Treasury, Internal Revenue Service. See also, Jones,

Homer W., and McMahon, Paul B., "Sampling Corporation Income Tax Returns for Statistics of Income, 1951 to Present," *op. cit.*

[18] See Leszcz, Michael R., Oh, H. Lock, and Scheuren, Fritz J., "Modified Raking Estimation in the Corporate SOI Program," *1983 Proceedings of the American Statistical Association, Section on Survey Research Methods,* and Harte, James M., "Post-Stratification Approaches in the Corporation Program," *1982 Proceedings of the American Statistical Association, Section on Survey Research Methods.*

[19] This description of statistical editing was taken from Wilson, Robert A., "Statistics of Income: A Byproduct of the U.S. Tax System," *Multi-National Tax Modelling Symposium Proceedings,* Revenue Canada Taxation, 1985.

[20] For Income Years starting with 1963, the Enterprise Standard Industrial Classification (ESIC), also issued by the Office of Management and Budget (OMB), was used for the corporation Statistics of Income Program. This system was specifically intended as a means of classifying "businesses" rather than "establishments" and was based on groupings of appropriate Standard Industrial Classification (SIC) codes. OMB discontinued the ESIC system when the SIC system was revised in 1987, so that Federal statistical agencies responsible for enterprise data now use their own classification systems. For a brief discussion of "establishments" versus "businesses," see the section on SOI sole proprietorship statistics later on in this article.

[21] The 1972 Standard Industrial Classification (as revised in 1977) and the accompanying 1974 Enterprise Standard Industrial Classification, both issued by the Office of Management and Budget, are presently used as the basis of Statistics of Income industry classifications for corporations, partnerships, and sole proprietorships. However, research is currently underway to determine how best to adapt the revised industry groupings contained in the recently released 1987 Standard Industrial Classification to Statistics of Income. Decisions will be based on knowledge of the economy's changing industry mix, the statistical needs of principal users of the data, the legal forms of organization in which an industry is likely to be best represented, the likelihood that the statistics will be reliable given the present and anticipated size of the SOI samples, and the limited number of industry titles and codes for which space can be devoted in the tax return instructions used for taxpayer self-coding.

[22] See Clark, Allison, "Corporation Income Tax Returns for 1985: An Initial Look," in this issue. (Some data for 1985 were also included earlier, in the historical tables contained in the Winter 1987-1988 issue of the *Statistics of Income Bulletin.*)

[23] For a description of how the Bureau of Economic Analysis uses SOI data for corporations, see U.S. Department of Commerce, Bureau of Economic Analysis, *Corporate Profits: Profits Before Tax, Profits Tax Liability, and Dividends,* Methodology Paper Series MP-2, May 1985.

[24] See Wilson, Oliver, and Smith, William J., "Access to Tax Records for Statistical Purposes," *Statistics of Income and Related Administrative Record Research: 1983,* U.S. Department of the Treasury, Internal Revenue Service.

[25] See Spruill, Nancy L., "The Confidentiality and Analytic Usefulness of Masked Business Microdata," *1983 Proceedings of the American Statistical Association, Section on Survey Research Methods.*

[26] Much of the material in this section was taken from Chapter 4, "Other Corporation Branch Studies," *Basic Operating Plan, Proposed Multi-year Operating Plan, FY 1986-92,* Volume I, Statistics of Income Division, Internal Revenue Service, unpublished report.

[27] Material in this section is abstracted from a draft paper by Nick Greenia, Corporation Statistics Branch, Statistics of Income Division.

[28] See footnote 21.

[29] See footnote 22.

[30] Data for 1956 for sole proprietorships and partnerships excluded agriculture, forestry, and fishing and were contained in a release entitled *Business Indicators—1956-57, Sole Proprietorships, Partnerships, Corporations,* U.S. Department of the Treasury, Internal Revenue Service, April 1958, unpublished.

[31] See Skelly, Daniel F., and Hobbs, James R., "Statistics of Income Studies of International Income and Taxes," *op. cit.*

[32] For this purpose, the Census Bureau subjects the data to additional testing and imputes for missing items by using prior-year data reported by the same taxpayer and current-year data reported by other taxpayers in the same industry and geographic area.

[33] For additional information on the employment linkage studies, see Moglen, Gail, Day, Charles, and Petska, Tom, "Record Linkage and Imputation Strategies in the 1982 Business Employment and Payroll Studies,"

*Statistics of Income and Related Administrative Records Research: 1986–1987,* U.S. Department of the Treasury, Internal Revenue Service. See also Greenia, Nick, "1979 Sole Proprietorship Employment and Payroll: Processing Methodology," *Record Linkage Techniques—1985*, U.S. Department of the Treasury, Internal Revenue Service.

[34] For the last published article see Greenia, Nick, "Partnership Employment and Payroll," *Statistics of Income Bulletin,* Spring 1984, Volume 3, Number 4.

[35] A limitation of the employment estimate is that employers are required to report the number of employees on the payroll for the week that includes March 12 of each calendar year, which then becomes the basis for the estimate for the year. The dollar payroll estimate, on the other hand, is based on the total amount reported for the entire year.

[36] See Greenia, Nick, "1979 Sole Proprietorship Employment and Payroll: Processing Methodology," *op. cit.*

[37] See Day, Charles, "Imputation of Employment and Payroll as an Alternative to Record Linkage," *1988 Proceedings of the American Statistical Association, Section on Survey Research* Methods (in preparation as of August 1988).

[38] U.S. Department of the Treasury, Bureau of Internal Revenue, *Supplement to Statistics of Income for 1939, Part I, Compiled from Partnership Returns of Income,* 1945.

[39] U.S. Department of the Treasury press releases S-2253 and S-2645, dated February 16, 1950, and April 4, 1951, respectively.

[40] U.S. Department of the Treasury, Internal Revenue Service, *Statistics of Income for 1953, Partnership Returns,* 1957.

[41] This series was originally styled *Statistics of Income—U.S. Business Tax Returns.* Starting with 1965, it was retitled *Statistics of Income—Business Income Tax Returns.* From 1957 through 1972, the series also included summary statistics on corporations.

[42] See U.S. Department of the Treasury, Internal Revenue Service, *Statistics of Income—Partnership Returns,* for these years.

[43] For the most recent article, see Zempel, Alan, "Partnership Returns, 1985," *Statistics of Income Bulletin,* Summer 1987, Volume 7, Number 1.

[44] The first compendium was published as *Statistics of Income—1978-82, Partnership Returns,* U.S. Department of the Treasury, Internal Revenue Service, 1985.

[45] Although balance sheet data are compiled annually, they continue to be incomplete. Currently, not all partnerships are required to file a balance sheet; in other cases, balance sheets are not included with the return as originally filed. No attempt is made to impute the missing data for the statistics. However, because partnerships which are exempt from filing balance sheets are small by definition, and those that fail to file them are thought to be generally small, the amount of data "lost" may not be significant.

[46] See footnote 24.

[47] More specifically, partnerships classified in real estate are divided into eight component industries. Only returns classified in the largest of the eight, real estate operators (except developers) and lessors of buildings, are sampled at the special set of rates.

[48] In general, in contrast to active partners or certain other investors, passive partners or investors do not materially participate in the activity of the business throughout the year. The Tax Reform Act of 1986 limited the amount of partnership and certain other losses that passive partners or investors could deduct.

[49] See *Statistics of Income Compiled from the Returns for 1917,* Bureau of Internal Revenue, U.S. Department of the Treasury, 1919.

[50] From 1934 through 1953, the statistics were included in *Statistics of Income—Part 1,* for each year. This report was retitled *Statistics of Income—Individual Income Tax Returns,* starting with 1954.

[51] Only totals were compiled for sole proprietorships for 1940 and, to a lesser extent, for 1944; there were no industry data.

[52] See footnote 30.

[53] See footnote 41.

[54] The sole proprietorship data for 1977 through 1981 were published in *Statistics of Income—Sole Proprietorship Returns,* for these years. Data for 1979 and 1980 were published in a single combined report.

[55] See footnote 49.

[56] In general, "net income" meant gross income from all sources as defined in the tax law, minus personal

deductions. It was before the deduction for taxpayer exemptions and before credits against income, such as those for dependents and "earned income". In contrast. "adjusted gross income" is gross income from all sources as defined in the tax law, minus certain exclusions and statutory adjustments, depending on the year. It is before subtraction of amounts for personal deductions and exemptions for the taxpayer(s) and his or her dependents.

[57] See *Statistics of Income—Part 1*, for these years.

[58] *Ibid.*

[59] *Ibid.*

[60] See Wolfe, Raymond M., "Sole Proprietorship Returns, 1986," *Statistics of Income Bulletin,* Summer 1988, Volume 8, Number 1 (in preparation as of August 1988).

[61] See *Statistics of Income—1959-60, U.S. Business Tax Returns.*

[62] Additional Information about these files can be obtained by writing to the Director, Statistics of Income Division TR:S, Internal Revenue Service, 1111 Constitution Avenue, NW, Washington, DC 20224.

[63] *Ibid.*

[64] For a discussion of this relationship and its effect on the statistics, see Wolfe, Raymond M., "Methodological Changes in the Statistics of Income Sole Proprietorship Program—Dominant Business Processing," Statistics of Income Division, Internal Revenue Service, an unpublished paper. For a copy of this paper, write to the Director, Statistics of Income Division TR:S, Internal Revenue Service, 1111 Constitution Avenue, NW, Washington, DC 20224.

[65] However, the SOI reports indicate that returns with net income under $5,000 were included in the 100-percent sample class if they showed income characteristics similar to those found on returns with net income of $5,000 or more, such as unusually large amounts of income from certain sources.

[66] The inadequacy of the SOI sample for sole proprietorship purposes is mentioned in *Statistics of Income for 1929.*

[67] The statistics for 1939 covered all sole proprietorships, including those reported on returns with net income under $5,000. The SOI report for that year contains no explanation of how this became possible, since the previous sample design was unchanged and the percentage of returns sampled in the less-than-100 percent class had declined in comparison to the immediately preceding years. However, there were changes made in how the sample data for the late 1930's were weighted and this may be part of the explanation. For 1939, the sole proprietorship sample in the under-$5,000 net income class was weighted to an estimated total number of sole proprietorship returns in this class, apparently for the first time. For 1938, sole proprietorship returns appear to have been weighted to the size class total which also included returns for other than sole proprietorships. (Prior to 1938, a different form of "weighting" was used, based on average amounts computed from sample returns in the under-$5,000 net income sample class).

[68] The sample was similarly augmented for 1944, but no industry data were actually tabulated.

[69] See footnote 48.

[70] The "at risk" rules in the tax code limit the amount of losses that most investors can deduct to the amount they have "at risk", i.e., to the amount they have actually invested.