

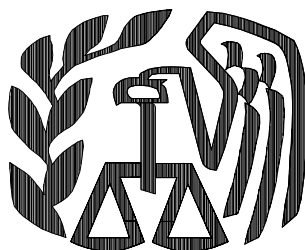
---

---

**STATISTICAL OPERATIONS AND STUDIES  
IN THE SOI PROGRAM OF THE IRS**

*by*

*Tom Petska, James Hobbs, and Fritz Scheuren*



*Statistics of Income Division*

*Internal Revenue Service*

*January 2001*

---

---

# STATISTICAL OPERATIONS AND STUDIES IN THE SOI PROGRAM OF THE IRS

by  
*Tom Petska, James Hobbs, and Fritz Scheuren* (\*)

## Origins of SOI

The modern U.S. income tax was enacted in 1913 with the passage of the sixteenth amendment to the U.S. Constitution. Subsequently, the Revenue Act of 1916 required the annual publication of statistics, establishing a role for the Statistics of Income (SOI) function. Despite many revisions to the tax law, the original requirement of that Act continues today. Specifically, the current Internal Revenue Code states that "The Secretary (of the Treasury) shall prepare and publish not less than annually statistics reasonably available with respect to the operations of the internal revenue laws..."[1].

The mission of the SOI program is to collect and process data so that they become meaningful information and to disseminate this information to its customers and users. The SOI Division conducts the Internal Revenue Service's studies on the operations of the tax laws with respect to individuals, corporations, partnerships, sole proprietorships, estates, nonprofit organizations, and trusts, as well as specialized studies covering both inbound and outbound international activities.

The annual budget of the SOI program is currently about \$30 million. While this amount represents a very small portion of the total IRS resources, SOI is among the leading Federal statistical organizations. SOI's budget covers designing the various projects; overseeing and participating in the statistical processing necessary to accomplish them; and tabulating, reviewing, describing, documenting, and publishing or otherwise releasing the results, all of which are done in Washington, DC. Most of the statistical processing takes place in IRS service centers located throughout the United States.

## Customers of SOI Data

Since the mandate for the SOI program is a responsibility required of the Secretary of the Treasury, it is not surprising that SOI's primary customer is the Office of Tax Analysis (OTA) in the Office of the Secretary of the Treasury. Another primary customer is OTA's legislative counterpart, the Congressional Joint Committee on Taxation (JCT). OTA and JCT use the microdata files produced by SOI as their primary source of information for analysis of the functioning of the tax system. In both agencies, microsimulation modeling is employed using SOI data as the primary database for tax policy analysis and tax revenue projections. The SOI data are also sometimes matched with other data to build comprehensive databases that can be used in estimating the overall impact of tax law changes and their effects on tax collections.

---

(\*) This is the latest update in a series of papers documenting the operations and studies of the Statistics of Income (SOI) Division of the IRS. Earlier papers on which this is based include: (1) Petska, Tom, Scheuren, Fritz, and Wilson, Bob, "Information from Tax Returns: The Statistics of Income Program of the Internal Revenue Service, Parts 1 and 2," *Business Economics*, April and July 1992; (2) Scheuren, Fritz and Petska, Tom, "Turning Administrative Systems into Information Systems," *Journal of Official Statistics*, Volume 9, Number 1, 1993; (3) Petska, Tom and Scheuren, Fritz, "Turning Administrative Systems into Information Systems: The Statistics of Income Program of the IRS," *1992 Proceedings of the National Tax Association*, 1993; and (4) "Statistics on Federal Taxation: The Tax Statistics Program of the IRS," presented at the Mid-Year Meeting of the American Tax Association, February 1994. In addition, James Hobbs updated and expanded the latter and placed it on the SOI Internet site in 1999.

Although the bulk of SOI's resources are focused on the statistical needs of OTA and JCT, SOI has many more customers. The Department of Commerce's Bureau of Economic Analysis (BEA) is a significant user of SOI data for estimating components in the National Income and Product Accounts related to individuals, corporations, partnerships, and sole proprietorships. The Census Bureau, also in the Department of Commerce, is another significant data user; however, its needs (unlike those of OTA, JCT, and BEA) are primarily met by SOI using IRS master file "population" data, rather than the lower-volume, but content-rich, SOI samples.

Many other Federal agencies are also users of SOI data, including the Federal Reserve Board, the General Accounting Office, the Social Security Administration, and the Health Care Financing Administration. Tax data are rich, and sometimes the only, source of information on the financial activities of the study populations. As with BEA and Census, many of these data needs are often fulfilled as a result of reimbursable contracts, since the resources of SOI are heavily committed to meeting the informational needs of OTA and JCT.

Outside of the Federal Government, SOI data are used by a broad array of tax practitioners, policy researchers, demographers, economic analysts, consultants, business associations, State and local Governments, universities, public libraries, and the media, as well as the public at large. In addition, other areas of IRS use SOI data for their internal operations.

### **Access to SOI Data**

Tax returns are protected by law from public scrutiny, and strict procedures govern the handling of returns and computer files containing such information. SOI's primary customers (OTA and JCT) are authorized to receive detailed tax return (microdata) files, so computer files of tax return information are regularly provided to them. However, most other users of SOI data can only have access to summary tabulations. The purposes and details describing such access are specified in Section 6103 of the Internal Revenue Code [2].

**Distribution of SOI Products and Services** -- Statistics of Income information is made publicly available through both printed publications and electronic media. The *Statistics of Income (SOI) Bulletin* is published quarterly, with each issue containing four to eight articles and data releases of recently completed studies, as well as historical tables covering a variety of subject matter, from Treasury Department tax collections to taxpayer assistance and tax return projections [3]. SOI produces separate annual "complete reports" on individual and corporation income tax returns, which contain more comprehensive data than what are published in the *Bulletin* [4,5].

The *Corporation Source Book* is also published annually, presenting detailed income statement, balance sheet, and tax data by industry and asset size [6]. Periodically, SOI produces special compendiums of research and analysis, covering topics such as nonprofit organizations, estate taxation and personal wealth, and international business activities [7,8,9]. Research articles documenting technological and methodological changes in SOI programs and other related statistical uses of administrative records are also published in a series of reports [10].

The IRS World Wide Web site provides users an easy option for accessing SOI data. At present, 40,000 files are downloaded monthly from the Tax\_Stats portion of this site. While SOI manages Tax\_Stats in order to make tax-related data on individuals, corporations, and other entities available to the public, we are also the conduit for releasing other IRS information, including the Internal Revenue Service *Data Book* (containing tax collections and other tax administration data), tax return projections, and microdata records of exempt organizations [11].

Electronic media products available from SOI also include magnetic tapes, CD-ROM's, diskettes, and files sent via e-mail. These products include the Individual Public-Use Microdata File (for which taxpayer identifiers have been removed); Exempt Organizations and Private Foundations Microdata Files (whose returns are open to the public); the Corporation Source Book; individual income tax return data shown by State, county, or ZIP code; and individual migration data shown on either a State or county basis. SOI also has a Statistical Information Services (SIS) office to facilitate the dissemination of SOI data.

### **Statistical Operations**

This section highlights statistical operations and procedures utilized in the development and implementation of many of the SOI statistical studies.

**Sample Design and Selection** -- U.S. tax returns are filed and administratively processed at one of ten IRS regional sites, called "service centers." Once processed, IRS compiles selected information from most return forms into a computerized "master file" system, which is the informational backbone of the agency. Most SOI operations begin by sampling returns from the master file system; the master file offers a sampling frame that enables use of sophisticated and efficient sample designs.

Statistics compiled for the SOI studies are generally based on stratified probability samples of tax or information returns. As returns are processed into the master file system, they are assigned to sampling classes (strata), based on criteria such as size of income or assets (or other measures of economic size), industrial activity, accounting period, or the presence of certain supplemental forms or schedules.

Each taxpayer, whether an individual or a business, has a unique number--- the social security number (SSN) for individuals or the employer identification number (EIN) for businesses. These unique taxpayer identification numbers (TIN's) are used as the seed for a pseudo-random number (using an algorithm, this is a transform of the TIN) which, along with the sampling strata, determines whether a given return is to be selected for the SOI sample [12]. The probability of a return being designated for the SOI sample depends on the sampling rate prescribed for its sample class or stratum and may range from a fraction of 1 percent to 100 percent.

The samples are selected from each stratum over the appropriate filing periods. Thus, sample selection can continue for a given study for several calendar years because of the prevalence of fiscal (non-calendar) year reporting. Because sampling must take place before the population size is known precisely, the rates of sample selection within each stratum are fixed. This means, in practice, that both the population and the sample size can differ from that planned. However, this does not compromise the validity of the estimates.

**Data Capture Techniques** -- After sampling, the relatively few data items pulled electronically from the master file system are substantially augmented with additional items key-entered from hardcopies of taxpayers' returns. Statistical abstracting can take as little as a few minutes for a simple return, to as long as several days for a large corporate return.

SOI has built a network of mid-range servers in selected IRS service centers that are dedicated to SOI statistical processing. "Hub" sites are located in Ogden (Utah) and Covington (Kentucky), with other processing centers located in Atlanta (Georgia), Austin (Texas), and Kansas City (Missouri). The processing system uses on-line transaction processing, so that all data capture operations are completed in a single pass. One editor is responsible for ensuring the validity of all data processing for a given return.

Due to substantial penalties for misreporting, the income and expenditure data reported on tax returns have proven to be more reliable than comparable survey data. Even so, SOI employees go to great lengths to protect against nonsampling errors, such as those due to taxpayer reporting variations or inconsistencies, or data processing errors. In order that final statistics be consistent and reliable, SOI economists develop extensive on-line tests and error resolution procedures that are applied to each sampled return. The tests and correction procedures are based on the structure of the tax laws and forms, generally accepted accounting principles, and the improbability of various data combinations.

Editors in service centers and SOI economists "statistically edit" data items in order to make each sampled return internally consistent. Missing data problems arise, albeit infrequently (under 1 percent of the time). Missing items can be obtained through direct contact with taxpayers, or be estimated through imputations based on other return data, prior-year data for the same taxpayer, or same-year data from a "statistically similar" return. SOI economists serve as subject matter experts to the editors in answering their questions, and also resolve errors in the more difficult cases.

Subsamples of returns are independently reprocessed and analyzed for a quality evaluation. Additionally, in order to provide high quality statistics, economists conduct on-line review trips to the processing centers and review quality logs written by editors.

**Weighting and Estimation** -- As noted above, the probability with which a return is selected for inclusion in an SOI sample depends on the sampling rate prescribed for the stratum in which it is classified. Weights are computed by dividing the population count of returns filed for a given stratum by the count of sample returns for that same stratum. "Weights" are used to adjust for the various sampling rates used --- the lower the rate, the larger the weight.

For some studies, it is possible to improve the estimates by subdividing the original sampling classes into "post-strata," based on additional criteria or refinements of those used in the original stratification. Weights are then computed for these post-strata using additional population counts.

The data on each return in a stratum are multiplied by the weight assigned for the given stratum. To produce the tabulated estimates, as shown in the *SOI Bulletins* and other publications, weighted data are summed to produce statistical totals.

Of over 200 million tax returns processed each year for administrative purposes, only about half a million are sampled for the various SOI programs. However, since sampling rates generally increase with increases in the size of financial amounts (such as income or assets), the returns in the samples are, on average, disproportionately larger and more complex than those in the master files, which include the population of returns.

Thus, in comparison to IRS administrative processing, which captures 100 percent of the tax returns but with limited item content, SOI programs collectively represent a smaller volume, but with a proportionately higher fraction of complex returns and with much greater item content.

The particular sample used in a study is only one of a large number of possible random samples that could have been selected using the same sample design. Estimates derived from the different samples usually vary. The standard error of the estimate is a measure of the variation among the estimates from all possible samples and is used to measure the precision with which an estimate from a particular sample approximates the average result of the possible samples. The

sample estimate and an estimate of its standard error permit the construction of interval estimates with prescribed confidence that this interval includes the actual population value. In SOI publications, the ratio of the standard error to the estimate itself is presented in percentage form. This ratio is called the coefficient of variation (CV).

## **SOI Programs**

For most of SOI's 80-plus-year history, the main emphasis has been focused on individual and corporation income tax information. However, growth has occurred in the nature and number of studies now conducted by SOI. Brief overviews are provided below.

**Individual Taxation Studies** -- Income and tax statistics from individual income tax returns have been published annually since 1916. The content of the program is largely determined by OTA, for use in tax policy research and in revenue estimation. Included in the program are data on sources of income (e.g., salaries and wages, interest, and capital gains), exemptions, deductions, taxable income, income tax, tax credits, and tax payments. Classifications of the data are by tax status, size of adjusted gross income, marital status, and type of tax computation.

The sample design for SOI Individual Studies is a stratified probability sample, in which the population of tax returns is classified into strata and a sample is randomly selected independently from each stratum. Strata are defined by: (1) size of total gross positive or negative income; (2) nontaxable returns with large amounts of income; (3) large amounts of combined business and farm receipts; (4) presence of special tax forms or schedules; and (5) the potential usefulness of the return for tax policy modeling.

The sample also includes a large panel of individuals embedded within the annual cross-sectional samples, in order to measure multi-year economic events such as capital asset transactions. Because "family economic units" are a desirable focus of tax analysis, dependents' returns are included in the samples and combined with parents' returns to form "tax families."

In addition to the annual complete report, the following other studies and reports are produced in the general area of individual taxation:

- **Early Tax Estimates:** This study is a subset of returns from the basic annual SOI individual sample. The data are derived from a sample of all individual income tax returns received at IRS service centers from January through April. The sample is selected in a random manner, based on the primary Social Security Numbers reported on the tax returns. About 20,000 returns are used as the basis for the statistics.
- **Sole Proprietorship Study:** Annual data are collected from Schedule C for nonfarm proprietors. Schedule C is an attachment to the individual income tax return. This study is further described in the following section covering the business taxation programs.
- **Sales of Capital Assets Study:** This annual study provides detailed data on the sales of capital assets reported on capital gains (or losses) schedules of individual income tax returns, as well as gains (or losses) on personal or depreciable business property reported elsewhere on the returns. The study includes returns from a panel that covers a period of years and now from the cross-sectional files [13].
- **Family Cross-Section and Panel Files:** In the primary SOI individual file, taxpayers (including dependents) are selected for the sample largely based on the size of their own

incomes. An alternative method for sampling dependent taxpayers, however, is used in the Family Cross-Section Panel. In these files, dependent returns are deleted and replaced by the actual dependent returns from the non-dependent returns in the file. This allows data from the dependent returns to be added to the parent's return(s) in order to create "tax family" units.

- **Taxpayer Usage Study:** Data from this study are released weekly during the filing season, beginning in early February and going through April. The data include general characteristics of the individual taxpayer population (e.g., use of paid tax return preparers); and the reporting of certain forms, schedules, or items (especially those that are new for the year). The sample size for this study is relatively small, usually between 10,000 and 20,000 returns. TPUS data are presented in an article for the *SOI Bulletin* and placed onto the Tax\_Stats Internet site.
- **High-Income Tax Returns Study:** This study grew from a requirement in the Tax Reform Act of 1976 to analyze the taxation of high-income taxpayers, generally defined as those with incomes of \$200,000 or more. Results of the study are presented in the *SOI Bulletin*. Part of the analysis is devoted to showing the major reasons high-income returns become nontaxable [14].
- **Individual Income Tax Rates and Tax Shares Data:** These data cover both average and marginal tax rates, as well as tax shares, by percentiles, under alternative income concepts. Annually, an article for the *SOI Bulletin* is produced showing this information [15].
- **Migration and Geographic Data:** In a joint project with the Census Bureau, the population of returns on the Individual Master File is used in conjunction with Census geographic codes to produce annual migration flow data (based on year-to-year changes in the addresses reported on returns). This file is also used to produce county and State income data by year. Individual income data by ZIP Code data are also occasionally produced.
- **Americans Living Abroad Study:** This periodic study on the foreign earned income and foreign tax credits of individuals is further described in a following section covering the international studies.
- **Public-Use File:** This file is produced annually from the individual tax returns, containing record-by-record detailed information (i.e., microdata), but with identifiable taxpayer information omitted in order to make the files available for public dissemination. The file is available on a reimbursable cost basis.

**Business Taxation Studies** -- Most business activity in the United States is conducted by corporations, partnerships, and sole proprietorships. Annual SOI studies are conducted in each of these areas. Corporation income is generally taxed directly at the business level, then again at the shareholder level, at the applicable rates on dividend income. However, certain provisions in the Internal Revenue Code lessen this effect. For example, S corporations that are closely held by small numbers of shareholders are not taxed directly. Rather, their income is generally subject to tax only at the owner level, much like partnerships. Partners and S corporation shareholders report their allocated shares of income and expenses on their own tax returns. The profits of sole proprietorships are taxed only at the personal (i.e., owner) level. The net income or loss from proprietorships is added to personal income from all other sources and taxed at the applicable individual income tax rates.

- **Corporations:** The SOI corporate data are the only publicly-available source of financial information on all corporations, since other sources include only large or publicly-held corporations or those in certain regulated industries. The data are estimates based on a stratified random sample of about 130,000 pre-audited returns, selected from a population approaching 5 million returns. Returns of the largest corporations are generally included in the sample each year, subject to consolidations and mergers.

The corporation program is rich in item content, containing complete income statement, balance sheet, and tax computation information. The data are classified by industries, accounting periods, and sizes of assets, receipts, and income taxes after credits.

Separate information on selected subjects is also available from the corporate program, including data for S corporations. S corporations have 75 or fewer shareholders, all of which are individuals, estates, or trusts. The income of S corporations is generally subject to tax only at the shareholder level. S corporations account for nearly one-half of the total returns filed by corporations. They also account for about 40,000 sampled returns. Restrictions on the number of shareholders and, consequently, on the ability to raise capital means that, on average, S corporations are considerably smaller than other corporations.

The S corporation program is run independently of the basic corporate program in many respects (e.g., a separate *SOI Bulletin* article is published on S corporations), although S corporation data are also included in the statistics for all corporations [16].

There are a number of smaller, specialized studies covering international activities which are either subparts of the basic SOI corporation program (e.g., foreign controlled domestic corporations and foreign corporations with U.S. business operations) or offshoots of the program (e.g., controlled foreign corporations, corporation foreign tax credits, transactions of foreign owned corporations, and U.S. possessions corporations). These studies are described in a following section covering the international studies.

SOI corporate data have been published annually, beginning with tax returns for 1916. In addition to the annual corporation complete report, corporate data are also published in the *SOI Bulletin* and in the detailed *Source Book* [6]. The 500-plus pages of the *Source Book* show nearly 100 data items for 12 asset size classes, for over 150 industrial classifications used by SOI. There are separate statistics for S corporations and for those returns that reported amounts of net income.

In addition to the needs by OTA and JCT related to tax policy analysis, SOI corporate data are used by BEA to estimate corporate profits for the National Income and Product Accounts.

- **Partnerships --** Partnerships serve as conduits between businesses and their owners, who are, in this case, partners. The partnership entity is not taxed directly; each partnership files an annual information return, which includes an income statement, balance sheet (in most cases), a schedule of allocations or distributions made to each partner, and other supporting schedules. Partners report their allocated shares of income and expenses on their own tax returns. Partners are predominantly individuals, but may also include corporations, tax-exempt organizations, nominees, other partnerships, or any other legal entity. Partners are classified as either general (i.e., those who assume liability for the partnership's debts and losses) or limited (i.e., those whose liability in the partnership does not exceed their investment).



The annual SOI partnership program is important to BEA in estimating the National Income and Product Accounts, since it is the only source of data on these businesses. Partnership data are also published annually in the *SOI Bulletin* [17]. Data are classified chiefly by industry group, but also by type of partnerships and partners. The program includes data for limited liability companies, which are hybrid business entities that combine the corporate characteristics of limited liability for all owners (members) with the pass-through tax treatment of partnerships. Real estate rental income is featured in the partnership program.

- **Sole Proprietorships** -- Information about nonfarm sole proprietorship business activities is reported on Schedule C of the individual income tax return (i.e., Form 1040). Profits from these activities are combined with income from other sources in order to compute individual "adjusted gross income." From this perspective, a proprietorship acts essentially as a conduit through which the income of the business is passed through to the business owner.

Schedule C data, produced and published annually in the *SOI Bulletin*, are classified by industry, and show business receipts, deductions, and net income [18]. Data on nonfarm sole proprietorships provide information on the unincorporated business sector for the National Income and Product Accounts. The tax return is the only annual source of financial information covering these businesses.

**International Studies** -- Many of the SOI international studies are conducted on an annual basis, while others are periodically undertaken once every two or more years. Each of the studies can be placed into one of two broadly defined areas: foreign investment and activity abroad by U.S. corporations or other "persons" and investment and activity in the United States by foreign "persons."

The following studies cover aspects of foreign investment and activity abroad by U.S. persons (commonly referred to as "outbound" activities):

- **Controlled Foreign Corporations:** U.S. corporations can establish a business presence in foreign countries either by using foreign branch operations or through separate subsidiary companies incorporated under the laws of the foreign countries (e.g., Controlled Foreign Corporations). CFC's have several advantages over branch operations, one of which is the potential ability to defer U.S. taxation on part of the CFC earnings. The studies collect data on the assets, liabilities, income, deductions, earnings and profits, foreign taxes, and transactions of CFC's with related parties. Data from the studies are classified by country and industry and presented in an article for the *SOI Bulletin* [19]. For the tax years between the full-scale studies, SOI conducts smaller studies covering the foreign corporations controlled by nearly 200 U.S. parent corporations pre-selected by OTA as a panel. While data files for these smaller studies are provided to OTA, they are not published.
- **Corporation Foreign Tax Credits:** Although the United States taxes the worldwide income of U.S. persons, foreign-source income is often taxed as well by the country where the income is earned. To mitigate the potential impact of double taxation of foreign-source income, U.S. persons can credit their foreign taxes paid against their U.S. tax liabilities, subject to limitations. This annual study provides data on foreign income, certain foreign taxes paid, and foreign tax credit reported on corporation income tax returns, classified by industry and country. Data are also presented in an article for the *SOI Bulletin* by the type of income earned, such as passive income, financial services income, and shipping income [20].

- **Americans Living Abroad:** Like corporations, individuals may claim a foreign tax credit against their U.S. income taxes for foreign taxes paid on certain foreign-source income. In addition, certain individuals may claim an exclusion for foreign earned income and an exclusion or deduction for certain foreign housing expenses. This study covers foreign income, taxes paid, and foreign tax credit shown on individual income tax returns. Data are classified by size of adjusted gross income, country, and occupation.
- **Foreign Sales Corporations (FSC's):** These corporations were established and provided certain tax benefits relating to their U.S. exports. An FSC is incorporated in a qualifying foreign country or U.S. possession (except Puerto Rico), and is usually controlled by a U.S. parent corporation. A portion of the FSC's income attributable to the sale or lease of "export property" and the performance of export services outside of the United States is exempt from U.S. income taxation. Balance sheet, income statement, and export-related data are classified by major product or service, size of assets of the principal shareholder, size of foreign trading gross receipts, country of incorporation, and intercompany pricing method [21]. FSC's have recently been declared an illegal export subsidy, so alternative means are being explored.
- **Interest-Charge Domestic International Sales Corporations:** IC-DISC's are small domestic corporations formed to export U.S. products; they receive limited tax incentives to encourage smaller businesses to export U.S. products. The income of an IC-DISC is generally exempt from taxation. However, the shareholders of an IC-DISC, usually U.S. parent corporations, are subject to an interest charge on the tax that would have been paid on the deferred income, and a tax on actual and deemed distributions from the IC-DISC. Balance sheet, income statement, and export-related data are classified by major product or service, size of assets of the majority corporate shareholder, size of export gross receipts, and intercompany pricing method.
- **U.S. Possessions Corporations:** A credit against U.S. income tax for the tax otherwise payable on possessions-source income is allowed to U.S. corporations that locate in U.S. possessions and meet certain requirements (e.g., 80 percent or more of gross income must be from sources within the U.S. possession). The majority of possessions corporations are located in Puerto Rico. This study provides data on the balance sheets, income statements, taxes, credits, and payments for employee services of possessions corporations [22]. The results of this study help researchers examine the effect of the credit on unemployment in the U.S. possessions.
- **International Boycotts:** This study provides data on the business operations of U.S. persons (mostly corporations) in boycotting countries, as well as the requests and agreements to participate in international boycotts not sanctioned by the U.S. Government. The purpose of the boycott provisions is to discourage U.S. persons from participating in certain boycotts, such as the Arab League of Nations boycott of Israel. All U.S. persons who participate in unsanctioned boycotts are denied certain tax benefits related to the boycott income. Data are classified by type of boycott request, boycotting country, and method of computing the loss of tax benefits [23].
- **Foreign Trusts:** This study provides data on foreign trusts that have U.S. persons as grantors, transferors, or beneficiaries. Data include the country where the trust was created and the value of the transfer to the trust.

- **Entity Classification Elections:** For Federal tax purposes, both domestic and foreign entities can elect to be treated as a corporation, a partnership, or to be disregarded as an entity separate from its owner. Because an entity can elect to change its classification once every 60 months, SOI continuously maintains a database of elections, which is based on the population of election forms.

The following studies cover aspects of investment and activity in the United States by foreign persons (commonly referred to as "inbound" activities):

- **Foreign-Controlled Domestic Corporations:** This study covers domestic corporations whose voting stock is 50 percent or more owned by a single foreign "person" (usually a corporation). Data include balance sheets, income statements, and tax-related items, classified by country of residence of the foreign owner, as well as by industry, size, and age of the corporations [24].
- **Foreign-Owned Corporations:** This study focuses on domestic corporations with total receipts of \$500 million or more and with at least 25 percent ownership by a single foreign "person." Data are collected on the transactions (e.g., sales and purchases of stock in trade, and loans) between these corporations and their foreign related persons. Data are classified by country and industry.
- **Foreign Corporations with U.S. Business Operations:** Some foreign corporations engage in business activity in the United States through U.S. branches (as opposed to the U.S. subsidiaries discussed above). These foreign corporations are subject to U.S. tax on income "effectively connected" with the conduct of a U.S. trade or business in a manner similar to that used to tax domestic corporations. This study tabulates income statement and tax items classified by industry and country.
- **Minimum Effectively Connected Net Investment Income (MECNII):** This study collects investment income and asset data from the population of about 5,000 domestic life and property / casualty insurance companies. The data are used to determine a percentage that is later applied in the taxation of foreign insurance companies operating in the United States. The MECNII study is conducted annually.
- **Foreign Recipients of U.S. Income:** U.S. source income of nonresident alien individuals, foreign corporations not engaged in a U.S. business, or other foreign recipients, is generally subject to a 30-percent gross withholding tax, unless such income is exempt from U.S. taxation, or unless a lower rate is established by an income tax treaty. Income subject to the withholding includes dividends, interest, rents, royalties, compensation for personal services, and retirement payments. Income and tax withholding data are available by recipient's country of residence, type of income, type of recipient (individuals, corporations, fiduciaries, etc.), and income-size classes [25].
- **U.S. Partnership Income of Foreign Partners:** U.S. partnerships are required to withhold income tax on that portion of the partnership's taxable net income classified as effectively connected with the conduct of a U.S. trade or business and deemed allocable to foreign partners. Data covering U.S. partnership payments to foreign partners, and the related tax withheld, are classified by country and recipient type.

- **Sales of U.S. Real Property Interests by Foreign Persons:** A buyer or other transferee of a U.S. real property interest purchased from a foreign person must withhold a 10-percent tax from the amount paid to the foreign seller. The tax is transmitted to the IRS, along with information about the transaction. This study includes data on the amounts realized from these transfers, U.S. tax withheld, and the countries of the foreign persons.
- **Nonresident Alien Estates:** The United States imposes an estate tax on the portion of a nonresident alien's estate that is located within the United States, after certain allowable deductions and credits. Data from this study include types of U.S. property, sizes of U.S. estates, and countries of domicile of the decedents [26].

**Tax-Exempt Organizations and Obligations Studies --** SOI conducts separate studies based on information returns filed by charities and other tax-exempt organizations, private foundations (and charitable trusts), exempt organizations with taxable "unrelated business income," and for tax-exempt bond issues. As is true of most SOI studies, the main users of these studies are OTA and JCT. However, the studies also generate interest from certain special data users, such as the Independent Sector, a nonprofit coalition of corporations, foundations, and voluntary organizations whose mission is to encourage giving, volunteering, and not-for-profit initiatives.

- **Charities and Other Tax-Exempt Organizations:** This study covers nonprofit charitable organizations exempt from Federal income tax under Internal Revenue Code section 501(c)(3), except for private foundations (discussed below), whose purposes are religious, charitable, scientific, literary, or educational, or to foster national or international amateur sports competition, or to prevent cruelty to children or animals, or to test for public safety and who derive their funds from potentially tax deductible contributions from the general public.

Also included in this study are organizations that are tax-exempt under Code sections 501(c)(4) through (9). These organizations include civic leagues and social welfare organizations; labor, agricultural, and horticultural organizations; business leagues, chambers of commerce, and real estate boards; social and recreational clubs; fraternal beneficiary societies and associations; and voluntary employees' beneficiary associations. Unlike the nonprofit charitable organizations exempt under Code section 501(c)(3), most of these organizations are not eligible to receive tax-deductible contributions [27].

- **Private Foundations and Charitable Trusts:** A private foundation is a nonprofit, tax-exempt corporation, association, or trust with a narrow source of funds, which supports social, educational, scientific, charitable, religious, or other programs dedicated to improving the general welfare of society. The primary difference between private foundations and other charitable organizations tax-exempt under section 501(c)(3) is that foundations usually receive their funds from a narrow base of support such as an individual, family, or corporation, while other nonprofit charitable organizations derive their funds from a larger base of support.

This study also includes "nonexempt" charitable trusts, defined in the Tax Code as organizations that have exclusively charitable interests, for which donors are allowed to claim income tax deductions for charitable contributions. Nonexempt charitable trusts that are not publicly supported are subject to the same excise tax provisions as private foundations and are required to file the same information return (Form 990-PF) [28].

- **Charitable Remainder Trusts:** Charitable remainder trusts are a form of "nonexempt" trusts, having both charitable and noncharitable beneficiaries. These trusts are exempt from income taxation; however, distributions made to noncharitable beneficiaries are taxable. SOI restarted studies of these trusts beginning with Income 1998 [29].
- **Exempt Organizations' Unrelated Business Income:** Some nonprofit organizations earn supplementary income, which is considered unrelated to their charitable purposes. The unrelated business income tax is levied on this income. Unrelated business income (UBI) is defined as income from a trade or business that is regularly carried on by an exempt organization and is not substantially related to carrying out the exempt purpose or function for which the organization has tax-exempt status. The purpose of the UBI tax is to restrict the potential for unfair competition between nonprofit organizations and taxable for-profit businesses that provide similar services [30].
- **Tax-Exempt Bond Issues:** State and local Governments, or their authorized agents, are able to issue bonds for which the interest paid to the holders is exempt from Federal income taxation. This tax exemption is a subsidy for both Government projects and private-purpose projects, by making it possible to obtain funding at lower interest rates. The information collected on both "public purpose bonds" and "private activity bonds" includes the types of property financed and the face amounts of the bonds, classified by State. SOI processes the populations of returns filed for these studies [31,32].

**Estate, Gift, and Personal Wealth Studies --** SOI conducts separate studies on Estate and Gift Taxation, Personal Wealth estimation derived from Estate returns, and supported the work of the Federal reserve Board in conducting the Survey of Consumer Finances.

- **Estate Taxes:** Federal estate tax is a tax on the transfer of assets from a decedent's estate to its beneficiaries that is levied on the estate. The purpose of the tax is to raise revenue and to offset concentrations of wealth. The SOI study includes information on the sizes of estates, composition of assets, deductions, and taxes, as well as information on the age, sex, and marital status of the decedents. Statistics are available for both filing year and year-of-death [33].
- **Gift Taxes:** The Federal gift tax return is filed for transfers, or gifts, of property that are completed during a donor's life. The gift tax applies to all such transfers above the \$10,000 annual exclusion. The first SOI study will provide retrospective panel data for filers of 1998 gift tax returns. SOI will be able to examine patterns of giving, usage of the annual exclusion and unified credit, as well as patterns in the usage of the marital and charitable deductions.
- **The Personal Wealth Study** is a companion to the Estate Study. It provides estimates of personal wealth of the top living wealthholders that are generated from estate tax return data using the "estate multiplier" technique, in conjunction with both filing-year and year-of-death estate databases. The inverse of the mortality rates is used to weight decedents' returns to the size of the living top wealthholders. For the study, general mortality rates are adjusted to reflect that wealthy people have mortality rates that are lower than those of the population as a whole [34].
- **The Survey of Consumer Finances:** SOI staff have supported the work of the Federal Reserve Board in conducting the periodic Survey of Consumer Finance. Under very

controlled conditions, individual income tax return data have been used to help identify high wealth households to improve the coverage in the survey.

### **SOI Computer Systems**

SOI personnel use a wide range of data processing platforms, operating systems, programming languages, and commercial off-the-shelf software packages. Mainframe computers, mid-range computers, and microcomputers are utilized by the SOI staff. UNIX, MVS, and Windows NT are the operating systems residing on these computers. Currently, connectivity to information and resources on these computers is achieved through the SOI local area network (LAN) and the SOI distributed processing system (DPS). Internet and Intranet access is also available.

**Local Area Network** -- Each SOI staff member has a microcomputer residing on his or her desktop. The microcomputers are used for client-based statistical studies, for communicating with mid-range and mainframe computers available to SOI, and for accessing relevant Internet or Intranet sites. All SOI microcomputers have high resolution, color graphics display capability. Some microcomputers are configured with peripherals such as scanners, plotters, tape backup drives, label printers, laser jet printers, and CD-ROM drives.

The typical user workstation is a Pentium-class or better microcomputer operating under Windows NT 4.0. Routinely, SOI personnel use the Microsoft Office 97 Professional Edition products which include Word (for word processing), Excel (spreadsheets), Access (database development), and PowerPoint (creating presentations). Additional software packages used on the microcomputers are Internet Explorer (Internet/Intranet access), Pagemaker (desktop publishing), Lotus Notes (GroupWare and e-mail), Reflections (PC to UNIX / mainframe communication), and the Adobe Acrobat Reader.

All SOI personnel in Washington have available to them Internet and Intranet access for use in their day-to-day work. Access is accomplished through Microsoft's Internet Explorer software via the Treasury Communications Systems (TCS) wide area network (WAN). The microcomputers are connected to a local area network (LAN) located in SOI's computer facility at 500 North Capitol Street, NW, Washington, DC. The Ethernet based network uses the Microsoft Windows NT operating system. The LAN serves approximately 200 microcomputers and 20 high-speed laser printers located throughout the Division's offices. The laser printers are capable of color and duplex printing, collating, stapling, envelope feeding, and contain multiple paper trays. SOI provides technical support of computer hardware and software to all SOI personnel.

**Distributed Processing System** -- The SOI DPS is used by SOI personnel to develop and process moderate to large size SOI projects. The DPS is a nationwide multi-tier system that utilizes fiber circuits and the TCS (with point-to-point encryption) as a communications backbone and the TCP/IP protocol. Mid-range server/computer hosts are located at SOI's offices in Washington, DC, as well as in Ogden, Utah, and Covington, Kentucky. Connectivity to IRS mainframes located in Martinsburg, West Virginia and Detroit, Michigan is accomplished through DPS. Selected SOI employees have access to the Service's Information Data Retrieval System (IDRS), in which specific taxpayer records can be extracted.

### **Organizational Structure**

SOI is a structured organization comprised of four Branches, each responsible for different aspects of its overall mission. Three of the Branches are primarily responsible for statistical projects covering specific subjects --- Individuals (including sole proprietorships), Corporations

(including partnerships), and Special Studies (including international, tax-exempts, and estates). The Statistical Computing Branch provides support to the other three Branches with statistical expertise and computer know-how. It also assists other IRS offices with their statistical needs.

Each of the three subject-matter Branches is comprised of four Sections. Two Sections are staffed with economists, one is staffed with computer specialists, and the fourth is responsible for research (in the case of Individuals and Corporations) or for publications and information dissemination (in the case of Special Studies).

The Statistical Computing Branch is also composed of four Sections. Two of these Sections provide computer support through staffs of computer specialists, while the other two Sections provide statistical support through staffs of mathematical statisticians.

### **Staffing**

SOI is composed of approximately 190 employees located in its headquarters offices in Washington, DC and another 280+ persons at IRS field locations, primarily in Ogden, UT and Cincinnati, OH, but also in the other SOI processing sites. A small core is included in every Service Center, even if no statistical processing is conducted on site, to locate and ship returns to processing sites.

The field staff are approximately 90 percent tax examiners who do the bulk of the statistical editing and controlling and shipping of returns. The other 10 percent are programmers, located in Ogden and Cincinnati. The National Office staffs who direct the studies consist of interdisciplinary project teams of economists, computer specialists, and mathematical statisticians, whose responsibilities are described below.

**Economists** -- Economists are members of staffs, comprised of approximately 6 to 15 people each, within the three "subject matter" branches of SOI: Individuals, Corporations, and Special Studies. Generally, each economist is involved with one or more studies and has a wide variety of responsibilities that may include the following:

- **Planning a Study:** Economists plan a variety of studies that use tax and information returns filed with IRS as the primary data sources. They must continually review economic, statistical, and tax literature related to their studies in order to maintain proficiency as technical experts. Economists work with customers, other SOI staff, and IRS field personnel to set parameters for the studies covering sample designs, item coverage, processing schedules and deliverable dates, manual and computer processing systems, personnel and other resource requirements, and more. Economists meet with customers, primarily from OTA, to determine the objectives of the studies, to review changes to the tax laws, forms, and instructions, and to determine how data are processed in the studies.
- **Conducting a Study:** This category accounts for the largest part of an economist's time during the course of a study, which typically lasts two or more years. It includes the following activities:
  - Economists draft specification documents, in conjunction with computer specialists and mathematical statisticians. They review and approve sampling techniques and designs prepared by mathematical statisticians and draft shipping and controlling procedures for returns, data capture procedures, specifications for computerized programs to consistency and validity test the data, error resolution procedures, and specifications for computerized tabulations of the data.

- Economists coordinate the specifications with programmers located in field offices. Also, prior to production, they perform systems acceptability testing of the consistency and validity tests to ensure that the programming was done correctly. Economists assist in the development of quality review (QR) systems, by determining the sample selection criteria and the items to be checked, and by designing statistical reports to display QR results.
- Economists train tax examiners or "editors" in the processing sites and periodically review field processing activities. Typically, an economist will spend one or two weeks per year at a service center performing these functions. Editors in IRS service centers perform most of the data capture and error resolution functions of studies. Economists perform error resolution of complex data anomalies and ensure inclusion of all "large case returns."
- Economists use statistical software packages to examine the data (e.g., Oracle) and to produce the tabulations that present the results of the studies (e.g., SAS). The tables, along with the microdata files, are transmitted to SOI's customers.
- Economists also analyze the data and write articles covering the results of the studies for publication in the *SOI Bulletin*. Economists produce other publications, such as the individual and corporation "complete reports," the *Corporation Source Book*, and special compendiums covering topics such as nonprofit organizations.
- A few SOI economists work in the Statistical Information Services (SIS) office to respond to public inquiries about SOI, IRS, and other statistical data.

**Computer Specialists** -- Computer specialists are composed of two groups: applications software development specialists and hardware / software technical support specialists. These groups perform different functions within SOI, as described below.

- **Computer Applications Software Development Specialists:** The computer applications specialists work in the operations sections of each of the three subject matter branches of SOI and are primarily involved in the development and implementation of applications systems in support of the various statistical projects conducted by SOI. Working closely with the economists and mathematical statisticians, these computer specialists perform the following activities: systems analysis, systems design, write programs, write processing specifications, document computer systems, code programs, design application schemas, test applications software, monitor application production operations, and maintain applications software.
- **Computer Hardware / Software Technical Support Specialists:** The computer hardware / software specialists work on either the Technical or Distributed Data Processing Teams. They are responsible for the procurement, installation, and maintenance of computer hardware, client and network operating systems software, data communications systems, and commercial off-the-shelf software. Additionally, they manage the LAN and various components of the WAN. They are required to be knowledgeable about systems administration, database administration, data communications, local area networks, and software languages used by SOI. They research new technologies for potential application to SOI's computing environment. In addition, they provide consultation and training on computer-related topics to SOI staff members and oversight of SOI's computer resources located at IRS field locations.



**Mathematical Statisticians** -- Mathematical statisticians compose two sections within SOI: the Mathematical Statistics Section and the Statistical Support Section. In addition to their responsibilities described below, mathematical statisticians are encouraged to prepare technical papers to be presented at professional meetings and published in statistical journals. Specifically, SOI mathematical statisticians have been very involved in presenting papers and findings at the annual meetings of the American Statistical Association.

- **Mathematical Statistics Section:** Within this section, staff members develop sample designs for the SOI studies and monitor the execution of those designs. They write technical specifications for sample designs, editing, imputation, and estimation procedures. Mathematical statisticians are encouraged to develop innovative methods at improving sample designs and estimation procedures. They also provide statistical support to economists in data management, quality, and analysis, plus statistical graphics and inference. Empirical studies concerning sampling and non-sampling errors are also conducted.
- **Statistical Support Section:** Mathematical statisticians within this section serve as resident consultants for non-SOI areas of the IRS on a wide variety of statistical issues. They provide high-level statistical support on systems design and analysis, sample design and weighting, statistical analysis, estimation, programming specifications, quality measures, customer satisfaction surveys, and cognitive research. Their projects currently include the following:
  - **Survey Feedback Action / Employee Satisfaction Project:** This is a project that transforms raw data into numeric indices that gauge IRS-wide attributes.
  - **TeleFile Project:** This project covers an alternative method of filing Federal Forms 1040EZ and 941 with the IRS via a touch-tone telephone, and includes a pilot Federal-State tax return filing program.
  - **Customer Service Quality Programs:** These programs measure the quality of written and on-line correspondence and various tax accounts activity in the IRS service centers.
  - **Customer Service Satisfaction Survey:** This project covers an automated satisfaction survey of taxpayers seeking assistance by calling IRS through our toll-free and Automated Collection Site phone lines.
  - **Remittance Processing Strategies Studies:** These studies are designed to minimize the length of time IRS processes taxpayer remittances.

#### **Footnotes**

- [1] *U.S. Internal Revenue Code of 1986* (as amended by public laws subsequent to Public Law 99-514, October 22, 1986), Section 6108(a), "Statistical Publications and Studies, Publication or Other Disclosure of Statistics of Income."
- [2] *U.S. Internal Revenue Code of 1986* (as amended by public laws subsequent to Public Law 99-514, October 22, 1986), Section 6103, "Confidentiality and Disclosure of Returns and Return Information."
- [3] Statistics of Income Division, Internal Revenue Service, *Statistics of Income (SOI) Bulletin* Publication 1136.

- [4] Statistics of Income Division, Internal Revenue Service, *Statistics of Income -- 1997, Individual Income Tax Returns*, Publication 1304, 2000.
- [5] Statistics of Income Division, Internal Revenue Service, *Statistics of Income -- 1997, Corporation Income Tax Returns*, Publication 16, 2000.
- [6] Statistics of Income Division, Internal Revenue Service, *Source Book of Statistics of Income - 1997, Corporation Income Tax Returns*, Publication 1053, 2000.
- [7] Statistics of Income Division, Internal Revenue Service, *Statistics of Income -- Compendium of Studies of Tax-Exempt Organizations, Volume 2, 1986-1992*, Publication 1416, 1993.
- [8] Statistics of Income Division, Internal Revenue Service, *Statistics of Income -- Compendium of Federal Estate Tax and Personal Wealth Studies*, Publication 1773, 1994.
- [9] Statistics of Income Division, Internal Revenue Service, *Statistics of Income -- Compendium of Studies of International Income and Taxes, 1984-1988*, Publication 1216, 1991.
- [10] Technical papers are published in the annual methodology reports series, *Statistics of Income -- Turning Administrative Systems Into Information Systems*, James Dalton and Beth Kilss, editors, Internal Revenue Service.
- [11] Internal Revenue Service, *Data Book 1998*, David Jordan and Beth Kilss, managing editors, Publication 55B, 2000.
- [12] Harte, James M., "Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS," *1986 Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1987.
- [13] Auten, Gerald and Wilson, Janette, "Sales of Capital Assets Reported on Individual Income Tax Returns, 1985," *Statistics of Income Bulletin*, Spring 1999, Volume 18, Number 4.
- [14] Parisi, Michael, "High-Income Tax Returns for 1997," *Statistics of Income Bulletin*, Winter 1999-2000, Volume 19, Number 3.
- [15] Cruciano Therese, "Individual Income Tax Rates and Tax Shares, 1996," *Statistics of Income Bulletin*, Spring 1999, Volume 18, Number 4.
- [16] Wittman, Susan and Grant, Robert, "S Corporation Returns, 1997," *Statistics of Income Bulletin*, Spring 2000, Volume 19, Number 4.
- [17] Zempel, Alan, "Partnership Returns, 1998," *Statistics of Income Bulletin*, Fall 2000, Volume 20, Number 2.
- [18] Parisi, Michael and Cruciano, Therese, "Sole Proprietorship Returns, 1998," *Statistics of Income Bulletin*, Summer 2000, Volume 20, Number 1.
- [19] Schwartz, Ruth, "Controlled Foreign Corporations, 1994," *Statistics of Income Bulletin*, Summer 1998, Volume 18, Number 1.

- [20] Ward, Nick, "Corporate Foreign Tax Credit, 1996: An Industry and Geographic Focus," *Statistics of Income Bulletin*, Fall 2000, Volume 20, Number 2.
- [21] Belmonte, Cynthia, "Foreign Sales Corporations, 1996," *Statistics of Income Bulletin*, Spring 2000, Volume 19, Number 4.
- [22] Miller, Randy, "U.S. Possessions Corporations, 1995," *Statistics of Income Bulletin*, Summer 1999, Volume 19, Number 1.
- [23] "International Boycott Reports, 1996," (Data Release), *Statistics of Income Bulletin*, Winter 1998-1999, Volume 18, Number 3.
- [24] Hobbs, James R., "Foreign-Controlled Domestic Corporations, 1996," *Statistics of Income Bulletin*, Fall 2000, Volume 20, Number 2.
- [25] "Foreign Recipients of U.S. Income, 1996 and 1997," (Data Release), *Statistics of Income Bulletin*, Winter 1999-2000, Volume 19, Number 3.
- [26] Eller, Martha Britton, "Federal Estate Tax Returns Filed for Nonresident Aliens, 1995-1996," *Statistics of Income Bulletin*, Winter 1997-1998, Volume 18, Number 3.
- [27] Arnsberger, Paul, "Charities and Other Tax-Exempt Organizations, 1997," *Statistics of Income Bulletin*, Fall 2000, Volume 20, Number 2.
- [28] "Domestic Private Foundations and Charitable Trusts," (Data Release), *Statistics of Income Bulletin*, Fall 2000, Volume 20, Number 2.
- [29] Belvedere, Melissa, "Charitable Remainder Trusts, 1998," *Statistics of Income Bulletin*, Winter 2000-2001, Volume 20, Number 3.
- [30] Riley, Margaret, "Unrelated Business Income of Nonprofit Organizations, 1996," *Statistics of Income Bulletin*, Spring 2000, Volume 19, Number 4.
- [31] Nutter, Sarah E., "Tax-Exempt Private Activity Bonds, 1988-1995," *Statistics of Income Bulletin*, Summer 1999, Volume 19, Number 1.
- [32] Nutter, Sarah E., "Tax-Exempt Governmental Bonds, 1991-1995," *Statistics of Income Bulletin*, Fall 2000, Volume 20, Number 2.
- [33] Johnson, Barry W. and Mikow, Jacob, "Federal Estate Tax Returns, 1995-1997," *Statistics of Income Bulletin*, Summer 1999, Volume 19, Number 1.
- [34] Johnson, Barry W., "Personal Wealth, 1995," *Statistics of Income Bulletin*, Winter 1999-2000, Volume 19, Number 3.

Cross sectional studies are common and include surveys, laboratory experiments and studies to examine the prevalence of a disease. Studies validating instruments and questionnaires are also cross sectional studies. The study of urinary concentration of lead in children described in Chapter 1 and the study of the relationship between height and pulmonary anatomical dead space in Chapter 11 were also cross sectional studies. Sample size. Results measured over time require special care.(7) One of the most common mistakes in statistical analysis is to treat dependent variables as independent. For example, suppose we were looking at treatment of leg ulcers, in which some people had an ulcer on each leg.