

INFORMATION FROM TAX RETURNS: THE STATISTICS OF INCOME PROGRAM OF THE INTERNAL REVENUE SERVICE -- PART II *

Tom Petska and Fritz Scheuren
Internal Revenue Service

This article is the second of two reviewing the projects and operations of the Statistics of Income (SOI) program of the Internal Revenue Service; the first part was published in the April 1992 issue of *Business Economics*. (Editors' Note: Part I has not been reprinted here as most of the content of that article is already contained in the other Petska-Scheuren article reprinted on pages 3-11 of this volume.) Although SOI is a relatively small and obscure statistical agency, SOI data are, nonetheless, a part of the bedrock of the U. S. statistical system and central to the understanding of the economy as a whole. Here we conclude with a description of short-term modernization of work processes and an overview of longer-term innovations geared toward shaping the future.

MODERNIZATION OF WORK PROCESSES

The 1980's were a time of major change in SOI. As a relatively small part of IRS, statistical work has sometimes been given less attention and a lower priority than needed. Despite growing requirements, the SOI budget has fluctuated in this period; currently, it resides below inflation-adjusted levels of a decade ago.

In 1980 the SOI program consisted of 26 projects; now, in 1992, this number has more than doubled to nearly 60. This two-fold growth in programs was accompanied by a parallel four-fold increase in the amount of data extracted from the various tax and information returns. The growing SOI workload was absorbed in the early 1980's by efficiency gains and

sample size reductions. By 1984, however, conventional means of increasing efficiency became (self) limiting. Gradually, other ways of achieving productivity gains were adopted, including a total quality organization (TQO) strategy.

Technological Innovations

SOI's statistical processing has historically been separate from the mainline processing of tax returns for administrative purposes. SOI operations begin by sampling from tax or information returns in the basic tax administrative (or master file) system, which offers a sampling frame that enables efficient sample designs to be used. After returns are sampled, data elements already captured for administrative purposes are used as a starting point in statistical processing. These data are augmented with other items from tax returns, tested for consistency, and identifiable errors or inconsistencies are resolved.

Until a few years ago, the basic SOI information processing was conducted in a "batch-mode" manner, in which several sites often had roles in processing data for each tax return. A system such as this denied "ownership" and accountability for field processors and was not conducive to maintaining high levels of quality.

To address these processing problems, SOI built its own network of mini-computers solely for statistical processing. This new system, which has just been fully implemented, uses an on-line, real-time approach, so that all data processing is completed in a one-pass operation. In addition to reducing handling costs and removing overlapping responsibilities, accountability and ownership improve, since one person is now responsible for assuring the validity of all data processing for any tax return.

The SOI mini-computer network electronically links all SOI national office and field operations so that data can be efficiently transferred between locations.

*Part II of this article was originally published in *Business Economics*, July 1992, pp. 66-69.

This capability enables "experts," wherever located, to better monitor processing and to accelerate efforts to attain higher standards of quality. For example, an economist who specializes in corporate tax law in Washington, DC can periodically examine the information being added to an SOI file in the field and can point out any systematic problems for early resolution.

Many other recent innovations in statistical techniques have been introduced or enhanced by the mini-computer network. SOI has an ongoing interest in developing improved methodologies. Towards this goal, research conducted each year on these methodological improvements appears in papers given at professional meetings, such as the annual meetings of the American Statistical Association. Routinely, these papers are compiled and published in an SOI methodology series [1].

Restructuring Management Practices

Long-term initiatives have been underway in SOI to improve "quality" management in SOI. During the 1980's, conventional quality control techniques for detecting errors were slowly replaced by quality improvement techniques designed primarily to prevent errors. Management practices in both the public and private sector have been studied, and SOI initiated the first phase of a total quality organization (TQO) deployment plan in 1991. In addition to initiating a benchmarking effort, the strategies adopted in pursuit of TQO goals are: a sharper focussing on customer needs; continuing incremental improvements in processing; creating a participatory management structure to foster change; and setting tangible and measurable intermediate objectives in all parts of our operations: from capturing data, to writing reports -- even answering the telephones. Goals chosen are a blend of traditional customer concerns and new ideas "stolen" from the best practices found elsewhere.

As a result of these initiatives, the relationships between SOI and our customers and suppliers are changing. Customers and suppliers have been brought into SOI work processes in a largely informal network of cross-functional relationships. Efforts have also been made to get customers involved in long-term planning. Redesign of the SOI individual program was the result of such a partnership, in which the

long-term needs of Treasury's Office of Tax Analysis were built into the reformulated plan.

Similar partnerships have been made with the SOI field suppliers. One such effort involved shrinking the supplier base for the individual and corporate programs so that a greater level of expertise and human capital could be developed. Beginning in 1992, the processing of these two largest studies was pared down from ten to five service centers. Other initiatives have focussed on the design and deployment of improved quality systems aimed at prevention or early detection with a goal of better addressing root causes.

SHAPING THE FUTURE

Despite recent strides, SOI information systems continue to have chronic weaknesses. Currently, SOI is focussed on several initiatives to address deficiencies in the SOI systems to better meet customers needs. Four examples of these are described, followed by a sketch of some of the plans proposed to address these short-comings. These four are: the need for greater program timeliness, improved consistency of data items over time, better tracking of demographic changes, and improved data access.

Program Timeliness

The fact that users never have enough current information from tax returns is a key weakness of the SOI program. Timeliness of SOI studies has become a primary focus for improvement and one in which some successes have been achieved. All of the major SOI studies have a sampling period that extends for one year (or more) beyond the close of the applicable accounting period, to ensure the inclusion of late filed returns. To illustrate, most tax year 1990 individual taxpayers filed by April 15, 1991, but significant numbers of taxpayers (about 6 million) sought and were granted extensions until August 15. For taxpayers living abroad, further extensions were common.

Significant efforts are being made throughout SOI to complete statistical processing within a minimum time after the close of the return filing period. Figure A provides a summary of 1992 public release dates for major SOI programs. What is not visible in the

**Figure A.--Public Release of
SOI Information**

Program	Release Date (on or before)
1989 Corporations	August 1992
1990 Partnerships	August 1992
1990 Individual:	
Preliminary	February 1992
Final	August 1992
1990 Sole proprietorships	August 1992
Foreign Studies:	
1988 Foreign tax credit (corporation)	January 1992
1988 Foreign corporations related to U.S. corporations (including Controlled Foreign Corporations)	January 1992
Special Studies:	
1988 Tax-exempt organization business income and tax	May 1992
1988 Tax-exempt organizations	April 1992
1989 Private foundations	June 1992
1991 Estate tax returns	October 1992

figure is the extent to which delivery dates have been improving. In the individual SOI program, for example, preliminary information is available 2 months after the end of the 1991 filing period, and final information is available 6 months later. Both of these represent a significant acceleration over earlier schedules.

In addition, internal discussions are underway to provide even more current information by early closure of the sample periods, as has been done for the individual SOI program. For example, an early cutoff has just been reinstated in the corporation program speeding delivery by about 9 months.

Conceptual Consistency of Data Items

The ever-changing concepts of the economic and financial variables, as reflected on tax and information returns, create a multitude of difficulties in trying to examine tax data over time. Unlike other statistical

agencies where the questionnaire content can be customized to meet a statistical need, SOI's "questionnaires" are tax returns, which are developed specifically for tax compliance or other administrative purposes, rather than statistical studies. Not only is there limited flexibility in changing tax return content, but the information provided on returns is often complicated, not uniformly reported, and conceptually inconsistent over time, thereby impeding multi-year economic analysis.

Improving the conceptual clarity and year-to-year consistency of the content on tax and information returns is also a problem that has no easy solution. Since the tax laws have been frequently revised by the U. S. Congress, preserving year-to-year consistency can be challenging. For some data items, such as business depreciation, no attempt is made to make adjustments to the basic source data provided on tax and information returns, as long it is consistent with current provisions of the tax code. In other cases, such as when the overall profitability of partnerships was redefined by a tax form change, efforts have been made to produce consistent data over time.

The lack of direct control of the tax form content is, in addition to the frequent tax law changes, a key source of the general problem of year-to-year consistency. A current initiative is underway to subject tax forms and schedules to study in the Bureau of Labor Statistics' Collection Procedures Research Laboratory. SOI has supported this work since its inception in 1989 and plans a regular flow of IRS tax forms for study and improvement.

Use of more computer-based data editing procedures, including expert systems technologies, is under study as well. Improving the labor-intensive process of data problem resolution by identifying discrepancies and resolving inconsistencies at the individual return level, whenever possible, not only improves consistency but also is cost effective.

Tracking Demographic Changes

The recent redesign of the individual SOI program at the request of Treasury's Office of Tax Analysis has underscored the need to continually improve the longitudinality in SOI studies. Such "unusual" transac-

tions as large capital gains realizations can only be understood by means of a panel database. As a result of this redesign, additional efforts are needed to ensure completeness as well as inter-temporal consistency.

A similar need for greater longitudinality applies to the business sector studies, where tracking business legal changes has become a difficult and costly aspect of file processing. For example, assuring the inclusion of all large corporations in the annual corporation SOI program is a very labor-intensive effort. Often mergers and acquisitions were determined to be the source of what had initially appeared to be missing entities. Recent tax reform legislation has heightened the occurrences of changes of legal type, such as switching from a corporation to a limited partnership. Unfortunately, an automated system to track such changes is under development but does not yet exist.

Data Access

The quarterly *SOI Bulletin* is the primary vehicle for dissemination of SOI information [2]. SOI individual and corporation information are published annually in complete reports as well as in the *SOI Bulletin*. Corporation data are also published in a Source Book, which is available on magnetic tape. Source Books for partnerships and sole proprietorships are prepared too, as are periodic compendiums on tax-exempt organizations, partnerships, and international studies.

Tax returns are protected by law from public scrutiny, and strict procedures govern the handling of returns and computer tape files containing tax return information. Even after specific identifiers (name, address, social security number) are removed, the tax return is still considered to be confidential. Unlike most users of SOI information outside of IRS, Treasury's Office of Tax Analysis and the Congressional Joint Committee on Taxation are authorized to receive detailed tax return (microdata) files.[3]. Periodic deliverables to these organizations thus include computer tape files of tax return information, as well as tabulated (hardcopy) output. The Bureau of Economic Analysis (BEA), another primary customer, only has access to identifiable return information under certain restrictive circumstances. As a result, BEA's primary use is in the form of tabulations, not

microdata; in this respect, BEA is like most other users of SOI data.

Making more tax microdata publicly available to researchers, while protecting taxpayers' confidential information, is considered extremely important by SOI, since it allows researchers to simulate, at the individual observation level, responses to policy changes. Such "microsimulation" has become the primary tool for tax policy and distributional analyses [4].

Public-use microdata files of individual tax data have been produced regularly since 1960. For their time, these files constituted a major break with the past policies of not releasing information to private researchers relating to specific tax returns. In recent years, because of the proliferation of private computer files of micro-data, major investments have been needed in better statistical disclosure protection steps to properly safeguard the confidentiality of personal information on these files so they could continue to be released [5].

Expanding the number of public-use files, particularly for business tax returns, has been studied in SOI for some time. Measures have been proposed and utilized to make more public-use files available, while protecting taxpayer confidentiality [6]. A research project to study the feasibility of public-use business files was undertaken in the early 1980's. While this work demonstrated the feasibility of creating public-use data for unincorporated businesses, the cost of producing such data was considered to be prohibitively expensive to potential customers [7, 8].

AFTERWORD

This two-part paper has looked at some of the ways the Statistics of Income Division is striving to better meet the needs of its many customers. Despite the sometimes daunting challenges in developing administrative data for statistical purposes, it should be clear that a number of successful efforts are underway to address current deficiencies, and, thereby, to make information from the tax system more useful for policymakers, tax administrators, business planners, and the public at large.

Toward the goal of addressing the needs of users of SOI information, comments and suggestions as to how

the SOI program could be improved are always encouraged. For more information on programs and services, contact our Statistical Information Services office on (202) 874-0410, or FAX (202) 874-0922, or write the Director, Statistics of Income Division, P. O. Box 2608, Washington, DC 20013-2608. To help our users become better acquainted with our work, a computerized bulletin board is being developed, which will enable SOI data users to better keep in touch with current data releases and other program events.

FOOTNOTES

- [1] See, for example, *Statistics of Income and Related Administrative Record Research, 1988-1989*, Statistics of Income Division, Internal Revenue Service, December 1990, Publication 1299.
- [2] A list of recent publications is provided at the end of these footnotes.
- [3] The Bureau of the Census is another statistical agency allowed access to tax returns for specifically stated statistical purposes.
- [4] See, for example, *Improving Information for Social Policy Decisions: The Uses of Microsimulation Modeling*, Committee on National Statistics, National Research Council, 1991 (two volumes).
- [5] Various measures are employed to make public-use files available while protecting taxpayer confidentiality, including purging names and other unique identifiers, rounding data items to make it more difficult to identify individuals, and averaging the financial data of "similar" returns. For a discussion of these techniques, see Spruill, Nancy, "Measures of Confidentiality," *1982 Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1983. For a description of an application of these techniques on the individual SOI public-use file, see Strudler, Michael, Oh, H. Lock, and Scheuren, Fritz, "Protection of Taxpayer Confidentiality With Respect to the Tax Model," *1986 Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1987.
- [6] The *Corporation Source Book*, which provides industry and asset size data in more detail than the

annual reports, is also available on magnetic tape. However, these data are at table-level and are thus not microdata. For more information on these files, write the Director, Statistics of Income Division.

- [7] Spruill, Nancy, "The Confidentiality and Analytic Usefulness of Masked Business Microdata," *1983 Proceedings of the American Statistical Association, Section on Survey Research Methods*, 1984.
- [8] Public-use files on sole proprietorships were released for certain years in the early 1980's. See, for example, "General Description of the 1980 Sole Proprietorship (Schedule C) Public-Use File," Statistics of Income Division, Internal Revenue Service, July 1983.

LIST OF RECENT STATISTICS OF INCOME PUBLICATIONS

Some of the more recent Statistics of Income publications include:

SOI Bulletin, Volume 12, Number 3, Statistics of Income Division, Internal Revenue Service, January 1993

1989 Individual Income Tax Returns, Publication 1304, September 1992

1988 Corporation Income Tax Returns, Publication 16, November 1991

1988 Source Book, Corporation Income Tax Returns, Publication 1053, June 1991

Compendium of Studies of International Income and Taxes, 1984-1988, Publication 1216, September 1991

Compendium of Studies of Tax-Exempt Organizations, 1974-1987, Publication 1416, July 1991