STATISTICS OF INCOME: AN OVERVIEW

Robert A. Wilson and John DiPaolo, Internal Revenue Service

In December 1980, the Statistics Division of the Internal Revenue Service prepared for consideration its first multi-year operating plan, in part to meet the directive of the then Office of Federal Statistical Policy and Standards (OFSPS) and in part to meet the requirements of the first IRS "strategic plan".[1] These new reporting requirements now give users more of an opportunity to review the long-range plans for the Internal Revenue Service Statistics of Income (SOI) program than was provided formerly.

This paper is based on material included in the introduction to the long-range plan and reviews some of the major procedural and methodological strategies being considered for the future. The presentation begins with an introduction to the SOI program as background, an explanation of the general concerns that have been raised about resource needs relative to the program, and a summary of how SOI data are now processed. This is followed by examining several of the processing innovations which will be researched and evaluated for possible implementation during the planning period as a means of increasing productivity.

THE STATISTICS OF INCOME PROGRAM

The Internal Revenue Service, in addition to its primary mission of enforcing compliance with the Federal tax laws, is also charged with the responsibility of publishing statistics on the operation of these tax laws. The data, based on tax returns, are published in a series of reports called Statistics of Income.

This series came into being soon after adoption of the Sixteenth Amendment to the Constitution and the subsequent enactment of the first modern U.S. income tax law, the Revenue Act of 1916. The Act specifically called for the annual publication of statistics. The wording contained in the 1916 Act has been repeated, with practically no change, in each major rewrite of the Internal Revenue Code since that time. It is currently contained in the 1954 Code, which is the basis for the current tax law.

The SOI reports from the very beginning (1916) have been used extensively for tax research and for estimating revenue, especially by officials in the Department of the Treasury. At the start, the reports were geared almost entirely to meeting these needs. With the growth of research groups both within and outside of the Federal Government and with the increased needs of tax planners and revenue estimators, new types of data soon were also required. At the same time, the tax returns were expanded to reflect the growing number of new provisions of the law, thus providing a ready source with which to meet these needs.

By the close of World War II, most of the population was subject to the income tax. At about the same time, the economies of using existing administrative files as the source of data on a wide variety of statistics had become more and more apparent. While the tax definitions of data items presented some obstacles, the obstacles were far outweighed by the likelihood that taxpayers' response tended to be more accurate than their response to special surveys. Moreover, with experience, users learned how to adjust for these definitions to meet their own particular needs.

The upshot of all these developments was an SOI increasingly different in its orientation from the early SOI. Several multi-purpose reports replaced the single tax-oriented report. While tax data continued to be included (all the more so as the tax law expanded both in scope and in complexity), the emphasis changed to more general purpose statistics geared to meeting the needs of economists and financial analysts.

The main emphasis of the annual statistics has always been individual and corporation income tax data. Other subjects based on other types of returns for which data have been tabulated either annually or periodically have been partnerships, estates and gifts, fiduciaries, farmers' cooperatives, foundations and other tax exempt organizations, and employee plans. Schedules attached to some of the returns become the subject of their own SOI reports. The sole proprietorship schedules were a relatively early source of statistics, which together with data from partnership returns, shed light on an important part of the economy not covered anywhere else to any appreciable extent.

Another development in the growth of SOI was the increasing tendency for new revisions to the tax law to require separate reports to Congress by Treasury's Office of Tax Analysis (OTA). These reports required statistics on such topics as individuals with high income who were nontaxable, the operation of the jobs credit provisions, Domestic International Sales Corporations (DISC's), international boycott participation, taxation of corporate income from U.S. possessions, and income of citizens working abroad.

Organizational Relationships

The Statistics Division in Washington is part of the IRS Office of Planning and Research. This office plays a leading role in developing taxpayer compliance studies and quality control systems, conducting new systems feasibility studies, and in identifying administrative problems in adapting to new law changes. The Statistics Division is responsible not only for SOI, but also for supplying IRS long-range workload projections and for conducting special statistical studies for the Service and supplying advice on sample designs for use in helping other organizations in IRS conduct studies of their own.

In connection with SOI, a staff of statisticians and economists works closely with users to determine the content of each program and publication, to design the samples used, and to develop field procedures. Complications arise from the fact that the processing is decentralized in twelve different locations throughout the country (see figure 1); hence there is a need for a strong coordinating role by the Statistics Division, including adequate quality controls to assure uniform and accurate processing.

The SOI program has the following basic character. Returns filed with the ten service centers are processed for administrative purposes to determine the correct tax liability. During processing, the returns are entered on tape for eventual posting to the IRS Master File. It is when the return records are on tape that they are selected for SOI. After the returns are selected, they are subjected to additional editing for SOI by specially trained technicians. The data thus extracted from the sample returns are entered on tape and tested for consistency. Any errors detected are then resolved to produce a final data file which is used to prepare SOI tabulations.

SOI Users

Information obtained from the SOI program is used extensively throughout the Federal Government for a variety of purposes. Besides OTA and the Joint Committee on Taxation, the third major Federal user of SOI is the Bureau of Economic Analysis (BEA) in the Department of Commerce. Data on corporations in the National Income and Products Accounts [2] are benchmarked to the amounts reported on corporation income tax returns which are then adjusted for conceptual differences and extrapolated based on more fragmentary data from other sources. Returns of unincorporated businesses, i.e., for sole proprietorships and partnerships, are also used for the national accounts; they constitute the only complete and reliable source of financial statistics for this segment of the economy. Investment income from individual income tax returns is also used in the national accounts.

In prior years the detailed planning for an SOI year began with user meetings which were held during the spring of the tax year under consideration. These meetings were attended primarily by representatives from OFSPS, OTA, Joint Committee on Taxation, and BEA; some of the other agencies that also participated included the Social Security Administration (SSA), Bureau of the Census, Federal Trade Commission, Department of Agriculture, and Small Business Administration.

The format for these meetings consisted of presenting the users with a marked-up copy of the tax forms or return schedules showing which items were proposed for inclusion in SOI for that year. These proposals were based on the frequency or content of recent prior-year programs that were reflected in previous plans; informal discussions held earlier at lower management and technician levels; known or anticipated law changes for which data would likely be needed; and, of course, the extent of available statistical resources. Often, because of lead-time constraints, only limited changes to the proposed program content were possible.

NEW PROGRAM CONTENT STRATEGIES

The basic assumption used in developing the present multi-year strategic plan was that the demand for statistical data was likely to increase in the 1980's and that resource constraints on Government statistical programs would probably continue. To this end, the Statistics Division recently reevaluated the size of each of the SOI samples and presented a new plan to its major users.

The resulting sample size reductions are to be coupled with improved methods of weighting the data. The introduction of post-stratification in all SOI programs is being examined as a possible means for maintaining reliability in the face of new sample size reductions. These reductions are to be accomplished by basing the estimates on subsamples of the former full sample sizes of the late 1970's; the larger samples will continue to be designated, but their use will be confined, for the most part, to improving the weights for the subsample. The larger samples will also be available for reimbursable projects (see figure 2).

Another strategy under examination is the separation of program content into "core" and "other". The core programs would generally be stable, from year to year, and would consist of the basic elements of each program which change only occasionally, when the law or tax forms change. The rest of a program would continue to vary from year to year to meet the changing needs of tax policymakers.

The core program for individual income tax returns would consist of the various sources of income, personal exemptions and deductions, income tax computation, tax credits, and tax payments. The "other" program could consist of studies of the minimum or maximum tax computation schedules, sales of capital assets by type and computations of various tax credits, to cite some examples. In the case of corporations, the core program might consist of the income statement, balance sheet, income tax computation, tax credits, tax payments, and distributions to stockholders. Thus the "other" category could consist of computations of the investment, foreign tax, targeted jobs and work incentive credits and of the mimimum tax. Anything else could either be a Treasury Special Project, or a reimbursable project under this proposal.

Statistics for the core program would be produced in such a way that the entire computer system would not have to be redesigned to facilitate its processing each year. To be consistent with this, more of the statistical table outlines would also remain the same from year to year. Manual and computer processing would thereby remain constant with resultant economies.

When computer programs could not be simply updated, because of the necessary changes in the SOI program content, increased use of generalized systems would be substituted, thereby still achieving a net saving. Only a limited amount of data from the non-core program would be published and only in summarized form; the extent to which special OTA items are used further, such as in the SOI reports, would be dependent on OTA's needs.

DATA ABSTRACTION FROM RETURNS

For most SOI programs, up until now, Master File data have been used sparingly because of their limitations.[3] Until recently, the primary use made of Master File data for SOI had been in identifying returns for the samples used and for advance or early tabulations to meet special requests.

Beginning with Tax Year 1981 or 1982, manual editing or data abstraction from returns for statistics using a specialized abstract sheet will become economically obsolete for many programs. Instead, return data for the SOI sample will be obtained from the Master File system. When possible, adjustments to overcome shortcomings in the Master File data will be introduced through computerized routines. This method will be gradually extended to all SOI programs.

Every five years, a more comprehensive manual statistical edit, often involving many more items than are available from the Master File system, might take place for the SOI unincorporated business programs, possibly using an abstract sheet. This special editing would coincide with the Agricultural and Economic Censuses planned for 1982, 1987, etc. Special requests for data may be accommodated in a like manner. For example, the Department of Agriculture has expressed interest in obtaining tax return statistics on farming activities in addition to information that would normally be provided as part of SOI for use in connection with the Agricultural Census.

Since the cost to Agriculture of obtaining the required information through conventional survey methods is prohibitive, it may be possible in the future to increase the farm portion of the SOI sample to obtain this information for them on a reimbursable basis. In the interim years, changes in program requirements would be kept to a minimum so that all programming and manual instructions may be held constant to the maximum extent. This would facilitate meeting completion dates for major functions in each program, thereby speeding delivery time of the final product to users while conserving both professional and clerical resources.

For those SOI items which are not key-entered to the Master File tapes during revenue processing, an abbreviated abstract sheet may be required. The size of the sheet, however, will be kept to a minimum, providing perhaps for only those items that are to be manually abstracted. Under this approach, data from the Master File system would be transferred directly to an SOI tape for later consolidation with the manually-edited items.

Current thinking is to base some SOI programs, namely individuals, sole proprietorships, partnerships, and fiduciaries, almost entirely on Master File information. These data may be augmented each year, to a limited extent, by additional data that are manually edited for statistical purposes and that are not available through the Master File, although how this might be done is still being explored. For the annual individual income tax return statistics program, the number of Master File items available will be far more numerous and comprehensive than for the unincorporated business and fiduciary programs. For corporations, the relatively few data elements for SOI that are transcribed for revenue processing are currently under study in order to determine the extent to which they can be utilized for SOI; their use may be possible at least for smaller corporations in the SOI sample.

The current explorations will also determine whether there are some relatively inexpensive changes that can be introduced into the administrative processing system which would facilitate statistical use of Master File data. These might include the processing of limited additional data elements now not required for administrative processing.

To the extent such steps can be accommodated at this earlier stage in return processing, added costs at later stages, i.e., during statistical processing, may be avoided. Items still not used in administrative processing, or for which adjustments during administrative processing are inconsistent with their use for statistics, may be obtained as in the past by manually abstracting the data in an off-line statistical processing operation. In some cases, this may be facilitated by use of specially designed, smaller, samples for this purpose; presently a general-purpose sample is used for all statistics from a given return form.

COMPUTERIZED EDITING, ERROR DETECTION AND CORRECTION

Integration of the two sets of data, from the Master File system and from the statistical processing system, will be facilitated by a computerized error resolution system which would increase the role of the computer either in editing certain data which were manually edited in the past or in estimating data missing from the returns as filed. To the extent that this can be accomplished, in part with the aid of prior-year "perfected" statistical data for the same taxpayers, a more economical substitute for former procedures may be achieved.

For some programs, more of the computerized testing of each record for internal consistency testing and error resolution associated with this testing will take place concurrently with editing to shorten the feedback cycle to editors, verifiers, and data transcribers and to enable the correction of errors while the tax return is still available.

Much of the return editing will be computerized as part of this operation, thus replacing to a varying extent, the former manual operation. While past studies point to significant problems in any extensive use of Master File data without some form of statistical verifi- cation, the plan now under development calls for flushing out discrepancies, insofar as possible, using the computer to identify returns with computations "out of balance" or with other problems. Only the returns that fail this preliminary screening would be manually edited. This approach assumes some redefinitions of data items now manually edited because certain adjustments now made in manual editing might not be identifiable by computer. The extent of these redefinitions will depend on the SOI program under consideration.

At the same time, an automated approach is contemplated that will deal with schedules and items missing from the return. For example, a significant number of partnership returns are filed with balance sheet or other data missing; research is therefore needed to develop a methodology for the imputation of this missing information.[4] The Statistics Division is actively seeking outside funding for this purpose. For other returns, identification of missing schedules and items early in processing will permit followup to obtain various missing data in time to prevent delays later on in processing.[5]

The new methodology would contribute to a lower cost of controlling overall data quality because of the reduced error rates following the initial institution of more timely feedback of error conditions to the originators. Longitudinal characteristics of the sample would be used to advantage in consistency testing. Selected ratios based on tax return data would also be computed for comparison to the prior-year's ratios. For the business and corporation programs, industry codes would be systematically compared to prior-year codes to detect gross errors. Many of the errors would then be corrected by computer, while errors of a more complex nature would be read out for resolution by professional subject-matter staff members in the Statistics Division.

Finally, the IRS is currently engaged in a study to evaluate a new overall system for handling key entry and error resolution. The present system involves many hours of complicated separation of printed registers, and the association of registers with the related returns or other input documents. Error resolution clerks must then manually correct the register which is then batched and controlled for key entry. The use of on-line systems are now under study. These would utilize direct access to documents in error through a terminal that is connected to a minicomputer, permitting the corrections to be made without intermediate processing. We look for this approach to have an important long-run beneficial impact on the SOI program.

The success or new approaches to or substitutes for the present statistical editing process and of the expanded use of Master File data will be largely dependent on the adequacy of a quality control system. Presently, the quality control system that is used in statistical processing is concerned mainly with the effectiveness of the data abstracting or editing operation. Its major limitation is lack of timeliness for corrective action. The "system of the '80's" will check, not only on the manual editing (for those programs for which manual statistical editing is still applicable), but also on the processing at each subsequent stage, so that it will be possible to identify on a more timely basis the exact stage at which changes to the "original" data are made for any given return. The appropriateness of the changes made can then be more adequately assessed. As a byproduct, additional measures of nonsampling error will become available.

Industry Coding

Currently, most of the business and corporation tax returns are industry coded by the taxpayer using the numbered groupings that appear in the return form instructions and that are based, for the most part, on the Standard Industrial Classification. (For sole proprietorship schedules, the IRS attempts to code the return based on the taxpayer's description in the absence of a perceived need for a self-coding requirement.) An independent statistical coding operation is now included for returns selected for the SOI samples and involves, in general, consistency of the reported code with other information from the return itself (including the source of the receipts shown on the return and the business' narrative description of its principal business industrial activity and product) or from reference books. It is estimated based on the results of this independent coding that up to one-third of the self-coded entries may be in error. Therefore, the taxpayer-reported codes which are transcribed in revenue processing are not acceptable for most statistical purposes. On the other hand, economies may be realized if perfected codes can be obtained elsewhere in Government, either annually or periodically. These codes could be used each year in place of those reported by the taxpayer. To accomplish this, legal and practical problems would first need to be overcome. The former involve confidentiality rules affecting IRS and other agencies; the latter involves differences in the statistical reporting unit among agencies which could limit the appropriateness of any interagency use of a given code for a given business. [6]

The longitudinal aspects of the basic business samples might permit increased utilization of the SOI industry code from the prior year.[7] The SOI industry code previously obtained would be used; then, if the taxpayer's self-reported present and prior-year's code were the same, the prior-year SOI code would be used again without further research. On the other hand, if there were a difference in the taxpayer's industry code from year to year, the return would be examined to determine if there appeared to have been a real change in business activity. Among

other things, this type of two-year comparison would result in more stable estimates of industry from one year to the next at less cost.[8]

EXPANDING THE SOI DATA BASE

If SOI is to serve tax policymakers in a more responsive manner and on broader issues, it will be necessary to build a data base from as many sources as possible. With this in mind, the Division is now establishing exchange agreements with other agencies with regard to information furnished to them by the Internal Revenue Service under provisions of Internal Revenue Code section 6103, as amended by the Tax Reform Act of 1976 (which limits access to return records to specified governmental agencies for specified purposes). The new agreements will provide that the IRS, on request, will be entitled to receive back a copy of the information furnished which will also include any perfection, modifications, or enhancements, or the addition of any other information prepared by the other agency for inclusion in, or for use with, the IRS-supplied data (to the extent possible, given the confidentiality rules of the other agencies).

The larger data base made possible by the inclusion of data from other agencies would make the Division more responsive to the research needs of other activities within the IRS and within the Treasury as a whole. Combined uses of SOI and the IRS Taxpayer Compliance Measurement Program are contemplated, for example.[9] As another illustration, working with SSA and the National Cancer Institute, Statistics Division would be able to provide mortality and morbidity data within demographic subgroups by an individual's occupation and industry.

Considerable research is, of course, necessary to develop or perfect methods of overcoming the many known difficulties that would be encountered in trying to expand the data base. For example, techniques would have to be developed for linking employer, taxpaying entity, establishment, pension plan, payroll entity, and employee. Such linkages would encompass all types of employers, including corporations, sole proprietorships, and partnerships.

Long-range plans might require the addition of an individual taxpayer's sex and age to the Master File system, along with an occupation code. Age and sex could be obtained from SSA files. Inclusion of age would permit a study of the relationships between income and age, and measurement of income differences between individuals with income from different kinds of retirement plans and individuals with no income from formal retirement plans. The existing SOI sample design results in an oversampling of individuals at the peak of their income-producing years. Including age in the Master File would permit stratification of the SOI sample to yield better measures of income for both younger and older taxpayers.

REDUCING SOI PUBLICATIONS

The SOI reports for the 1980's will be streamlined in that they will emphasize the presentations that change but little each year. The more dynamic presentations highlighting data on detailed computations from the tax return may be presented only in short summary tables. Besides the basic SOI reports, vehicles for releasing statistics could be news releases or special supplemental SOI reports, such as those already used to shed light on the foreign tax credit and on sales of capital assets, for example.

The 1980's are expected to witness a continuation of the trend already well underway, namely, direct employment by SOI users of the microdata records on computer tape. While disclosure rules effectively limit the extent to which this can now occur, it is expected that public use files containing microdata in a form not inconsistent with the current IRS disclosure provisions will be developed in the next few years and that their use will no longer be restricted to Treasury and to those other users now already authorized under the law to receive these data. Much more research needs to be done in this area, and much better documentation on the content of the SOI tape files as they already stand will be required, too. This initial investment can be expected to be costly in time and resources.

Other, perhaps short-run, solutions to more timely release of the SOI complete report statistics will include elimination of the preliminary reports long associated with the major SOI programs. For many years now, about half of the preliminary reports have been based on early cutoffs of the samples. However, for corporations, in order to produce meaningful estimates based on an early cutoff, an elaborate system had to be developed in order to estimate data for returns of many of the larger corporations. Elimination of the processing steps unique to the release of preliminary data, such as in the case of corporations, can lead to concentrated efforts, resource-wise, to develop a single system for each program in order to perfect data for the complete reports on a timelier basis.[10]

This curtailment will present a void, however. A publication vehicle was recently developed in the SOI Bulletin; the Bulletin is a quarterly report, that began with the summer issue which was released in July 1981. In the future, this report will include an advance release of selected tables from forthcoming SOI complete reports, as a partial substitute for the former preliminary SOI reports. The Bulletin will also include, among other subjects, tabular summaries of early data based on the Master File system. These Master File data are now produced routinely each month based on individual income tax returns for use by IRS, OTA and the Joint Committee on Taxation. More fragmentary data from the Master File are available annually for corporations and tax-exempt organizations which may also be included in the Bulletin.

CONCLUDING COMMENTS

Streamlining the SOI programs is not confined to cutting the size of samples, programs, and publications. Methodological and processing changes have to keep pace or even lead the way. The proposals to introduce concurrent computerized consistency testing of the data while SOI returns are still accessible, and to make more use of data for other years for the same taxpayer in perfecting return data for the current year, have already been mentioned. Other innovations, now well along in development, include use of generalized systems and of electronic composition as a substitute for typesetting tables to be published. Neither of these steps is a true innovation; rather, each is an example of steps that would have been introduced earlier, had resources been available with which to conduct the needed research. In fact, most statistical agencies have long since made use of them in their own programs.

A Generalized Tabulating System (GTS), initially developed by the Census Bureau, is now already in use in developing the tables for some SOI projects. Attention will now need to be focused on developing a generalized system applicable to "front-end" processing of the return data themselves, including the consistency testing and any automatic error resolution. Complete tape-to-tape electronic composition is soon to be phased in for use in all SOI reports.

Savings realized from economies due to reduced samples and more efficient methods of data processing will enable the Statistics Division to meet the needs for more statistical data expected in the '80's, and to release the regular SOI reports and studies on a more timely basis. They should also enable the Division to devote increased resources to new areas of research and to satisfy the needs of its major users.

ACKNOWLEDGMENTS

The authors wish to thank Ross Summers and Ralph Bristol for reviewing the manuscript copy of this paper, Wendy Alvey and Beth Kilss for their help in presenting this paper at the Annual American Statistical Association meetings, and Terry Smith for preparing the illustrations. Thanks are also due to Denise Herbert who typed the several drafts of this paper and to Bettye Jamerson who edited the manuscript.

NOTES AND REFERENCES

[1] The OFSPS requirements were stated in the Statistical Reporter, May 1980; the Internal Revenue Service requirements were defined together with the results in the report entitled Strategic Plan for the IRS, December 1980.

[2] See the Current Business Statistics published monthly in the Survey of Current Business, Bureau of Economic Analysis, U.S. Department of Commerce,

[3] Data are "perfected" for administrative processing only to the extent they have a direct bearing on the ultimate computation and verification of tax. However, not all of the procedures are consistent with statistical needs.

[4] Internal Revenue Service follows up through correspondence with the taxpayer on only selected schedules found missing during administrative processing of the returns.

[5] Presently, missing schedules and incomplete data are identified only at the time of the final consistency testing which occurs after data abstracting is complete. This contributes to processing delays.

[6] Report on Statistical Uses of Administrative Records, Statistical Policy Working Paper 6, Subcommittee on Statistical Uses of Administrative Records, Federal Committee on Statistical Methodology, Office of Federal Statistical Policy and Standards, U.S. Department of Commerce, December 1980.

[7] Longitudinal designs which include the same sample returns in the sample each year are utilized to maintain the reliability of estimates of year to year changes.

[8] While the resultant increase in the stability of the industry estimates would facilitate certain kinds of year-to-year comparisons, it could also mask the effect of bonafide changes in industrial activity in a given year. This would also occur if industry codes for given businesses were reassessed only periodically e.g., once every five years.

[9] The IRS Taxpayer Compliance Measurement Program (TCMP) compiles statistics on the results of comprehensive audits of taxpayers based on representative samples of various classes or types of income tax returns in order to estimate the total potential effects of audit. TCMP results might thus be used to "update" SOI, which is based on unaudited data.

[10] Left unresolved for purposes of this paper is the means by which the Statistics Division will be able to provide corporation data on an expedite basis to the Department of Commerce for use in benchmarking the national accounts in July of each year. Formerly, this need has been met by emphasizing the same early cutoff of the SOI sample used for the preliminary SOI statistics. With elimination of the preliminary statistics, the timing of the cutoff may be revised to a later date for the SOI complete statistics. This may prove incompatible with Commerce needs.

FIGURE 1.--INTERNAL REVENUE SERVICES - GEOGRAPHIC LOCATIONS

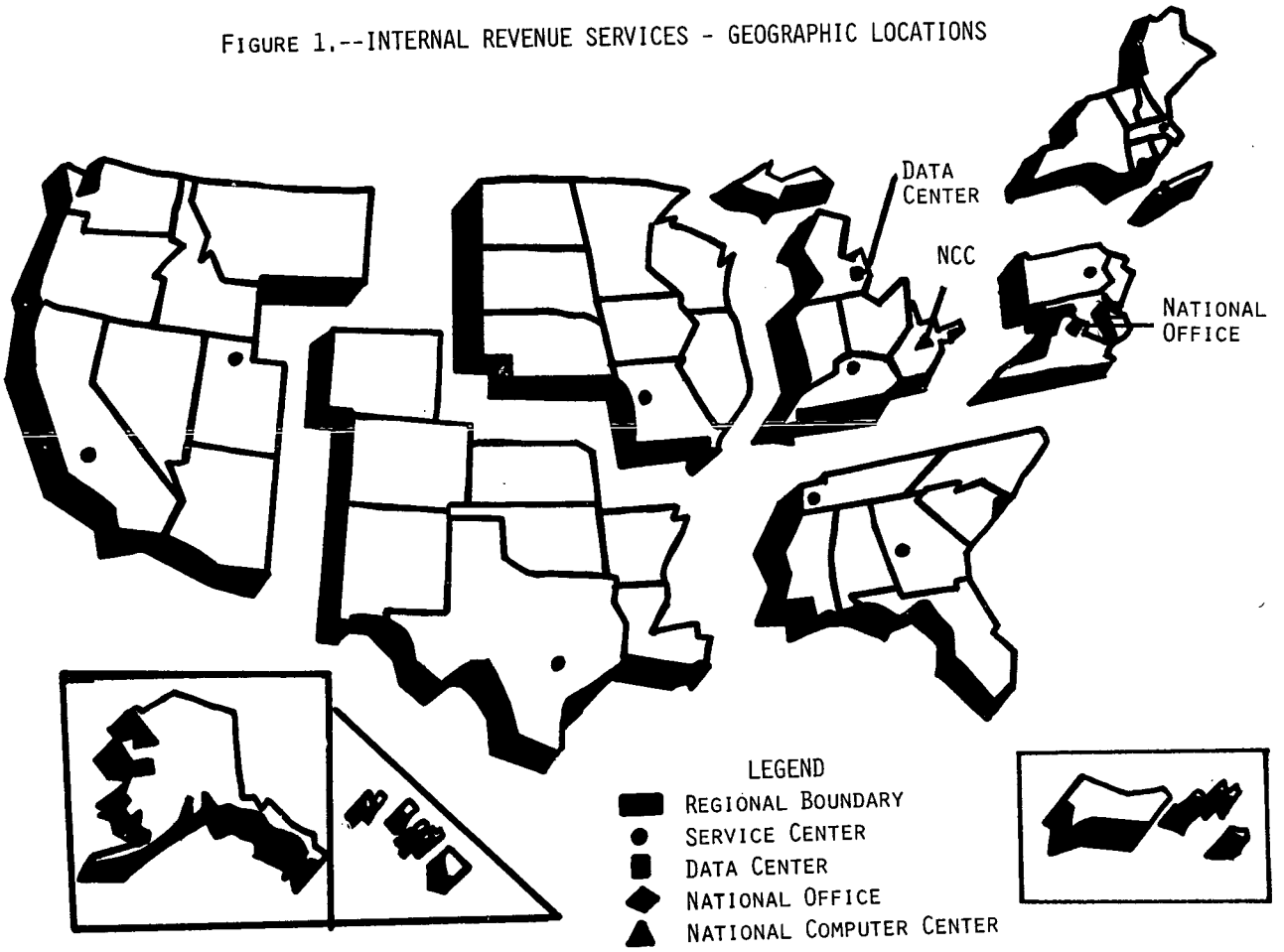

Figure 2.--Number of Returns Included in Statistics of Income Samples, by Tax Year

Program	Tax year						
	1979	1980	1981	1982	1983	1984	1985
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Individuals, total[1]	204.0	168.0	132.4	127.4	122.4	117.4	112.4
Nonbusiness	121.2	96.0	76.8	73.9	71.0	68.1	65.2
Business	82.8	72.0	55.6	53.5	51.4	49.3	47.2
Partnerships	50.0	40.0	35.0	35.0	35.0	35.0	35.0
Corporations:							
Sample, transaction tape	108.0	104.0	200.0	200.0	200.0	200.0	200.0
Subsample, total	77.6	90.0	95.0	95.0	95.0	95.0	95.0

[1]The size of the statistical sample for tax years beyond 1981 may be increased if unit processing costs can be reduced through revised methods.

INDIVIDUAL STATISTICS OF INCOME: ADVANCING THE CLOSEOUT DATE

Jim Dumais and Ray Shadid, Internal Revenue Service

This paper reports on the results of research done in the IRS Statistics Division exploring various alternatives for streamlining processing of and providing earlier estimates from the Statistics of Income (SOI) sample of individual income tax returns. Organizationally, this paper is divided into 5 parts. Section 1 provides background on the current SOI processing system. In section 2 each of the proposed changes is discussed. The methodology employed is described in section 3. Results and recommendations, in section 4, are followed, in the fifth section, by an outline of future plans.

1. BACKGROUND

As part of the Statistics of Income program, individual income tax returns filed (Forms 1040 and 1040A and related forms and schedules) are sampled to produce aggregate estimates of taxpayers' income, exemptions, deductions, credits and tax. These estimates are published in an annual Internal Revenue Service report in the Statisics of Income series [1].

Under the current processing system, sample designation for a given program begins with the first week, or cycle, of the processing or calendar year (usually in January) and proceeds through the following December. After the returns for a given program are sampled, they are edited; consistency and validity checking are performed; any transcription errors detected are resolved and a "clean" file is produced. Weight factors are calculated and applied; finally, tabulations are produced and the annual report, Statistics of Income--Individual Income Tax Returns [e.g., 9], is developed and issued.

In addition to the basic SOI program, the Treasury Department's Office of Tax Analysis (OTA) requires estimates of income and tax liability from Forms 1040 and 1040A, filed during the year, by late November of that year. In order to provide these estimates, the IRS Statistics Division has traditionally created a preliminary (or advance data) SOI file using all sample returns processed at the ten IRS Service Centers through the first week of October. From this file of early sampled returns, "advance data" estimates are provided to OTA [e.g., 5]. Traditionally, additional tabulations have also been produced from this file and the report, Preliminary Statistics of Income--Individual Income Tax Returns was issued [e.g., 7]. The preliminary reports have recently been replaced by the quarterly Statistics of Income Bulletin [e.g.,8].

As a result of budget constraints and requests for earlier release of SOI data [1], new concepts in SOI design and processing are being explored. Three specific issues or concepts are discussed here: advancing by two weeks (to mid-September) the sampling and processing cut-off date for the preliminary SOI file; changing processing at the Service Center level to make sample counts more nearly equal to designation counts for the advance data cut-off; and a proposal for radically different treatment of prior-year returns in the SOI files.

The primary data base used for this research and testing was the Internal Revenue Service Individual Tax Model file for Tax Year 1978 [6]. The tax model is a micro-data file comprised of an abbreviated version of each of the sample return records included in the 1978 SOI file that was used to produce the complete report for 1978. The tax year 1978 sample of 157,518 return records was weighted (by IRS District and sample code) to an estimated population of 89,771,551 Forms 1040 and 1040A returns filed during calendar year 1979.

In order to evaluate the results of testing the proposed modifications, this paper presents a comparison of a full simulation of the 1978 advance data tabulations (incorporating all of the proposed changes) with the 1978 complete SOI estimates and with the actual 1978 advance data tabulations transmitted to the Office of Tax Analysis. Results of the simulation will be explored in detail following a discussion of each of the proposed changes to the current SOI design and processing system.

2. PROPOSED CHANGES

Earlier Advance Data Cut-off.--Accelerating the preliminary SOI sampling and processing cut-off by two weeks is the first issue to be explored. The obvious criticisms of this proposal are: (A) that estimates will be based on about 1,500 fewer sample returns, and (B) that the returns not sampled tend to differ from earlier-filed returns.

Returns filed in late September and early October (as well as later-filed returns) exhibit different characteristics than those filed earlier: income amounts (positive or negative) tend to be larger. In Table A we highlight the average adjusted gross income (AGI) of $14,457 on early-filed returns and $22,306 on returns filed in late September (cycles 38 and 39), to illustrate this point. Returns also tend to be more complex the later they are filed. The level of complexity of various return categories can be implied from the data presented in Table 1. Late-filed returns exhibit higher relative incidences of filing on Form 1040, having itemized deductions or being classified as business returns than do earlier-filed returns.

In support of this proposal, it should be noted that the early cut-off will result in earlier release of SOI data. Also, adjustments are possible for the bias that would otherwise be introduced by simply cutting off earlier. [11]

The explanation of the methodology which appears in section 3, Simulation of 1978 Advance Data and Final Estimates, includes a discussion of the measures taken to test this proposed change as well as those that follow.

Improving Sample Counts at Advance Data Cut-Off.--The reasons for the discrepancies between designation counts and actual sample counts at advance data cut-off can be summarized into two major categories: (A) the inability to associate the edit sheet with its return document for abstraction of additional data in time to meet the processing deadline for the early file, and (B) unresolved errors from Service Center level consistency and validity testing not corrected in time to meet the early deadline.

The category in Table A labelled "Returns Missing from Advance Data" presents a summary of the 518,157 such cases (weighted estimate) identified in the 1978 file. The distribution of these returns by size of income is comparable to that of late-filed returns and indicates that, although fewer in number, these cases are adequate substitutes for some of the sample returns excluded from advance data due to the earlier cut-off.

Over recent years, the system of transcription of data from the tax return to computer tape during the processing of returns for revenue purposes has expanded to the point where almost all the data items necessary to produce the advance data tabulations are available to the SOI program from the revenue processing computer system. In addition, the quality level for the aggregated totals of a number of the available items (such as the major sources of income, AGI, and tax) is comparable to the SOI quality level for those items.

Since all sampled returns, including those with errors detected or those that require editing for special studies, have sufficient data available on tape to produce the early estimates, all returns designated for the early SOI cut-off will be transmitted to the Detroit Data Center for extensive consistency testing, error resolution, and posting to the advance data SOI file.

Prior-Year Returns.--A prior-year return is defined as one filed for an income year earlier than that for which the majority of the tax returns are being filed. Most tax year 1978 returns were filed during 1979. Thus, returns filed in 1979 for tax years 1977 or earlier were classified as prior-year returns. We estimate from the 1978 complete SOI file that there were 1,045,897 prior-year returns filed in 1979 (1.2 percent of the total). Table A includes a brief distributional analysis of the prior-year returns included in the 1978 complete SOI file and Table 1 includes a characteristics analysis of these same returns (as well as other categories of returns).

Prior-year returns present two problems to the SOI program. In the first place, prior-year returns require exception processing and testing because they relate to prior-years' tax laws. In the second place, prior-year returns are being tabulated with records for a tax period to which, it can be conceptually argued, they do not necessarily belong.[2,4]

The rationale for including prior-year returns in the current SOI year was that they were an acceptable substitute for current year returns yet to be filed. This made sense so long as inflation rates were low, and relatively few (or minor) year-to-year changes in tax law occurred.

Analysis now indicates that prior-year returns, as a group, tend to differ significantly from other returns from the tax year for which they were filed, and to differ from current year returns processed during the same filing year. In comparing prior-year returns with other returns from the tax year for which they were filed, we found that prior-year returns tend to have higher incomes and to be more complex. The second observation, that prior-year returns have a lower overall income level than current-year returns, may be attributable primarily to the effects of inflation.

TABLE A.-- AVERAGE ADJUSTED GROSS INCOME FOR SELECTED CATEGORIES OF RETURNS, BY SIZE OF AGI

CATEGORY	SIZE OF ADJUSTED GROSS INCOME			
	TOTAL	DEFICIT	$1 under $2,000,000	$2,000,000 or more
All returns, total	14,520	-15,431	14,665	3,844,367
Processing cycle:				
1 through 37	14,457	-12,711	14,583	3,843,069
38 through 39	22,306	-42,931	23,707	3,449,222
40 or later	19,578	-68,143	22,005	3,954,600
Prior year returns:				
Total	11,659	-18,858	13,157	6,659,455
Processing cycle:				
1 through 37	11,583	-17,354	13,145	5,849,125
38 through 39	16,966	-28,010	18,604	7,290,000
40 or later	11,712	-29,083	12,923	9,585,500
Returns missing from Advance Data	17,279	-24,775	17,596	4,015,909

In terms of the concept of SOI as a vehicle for analyzing and evaluating the operation of the tax laws in a given tax year, it would seem beneficial to isolate prior-year returns by the tax period for which they were filed. Once isolated, these returns could be consistency and validity tested with a simplified battery of tests designed for that specific tax year only. Once tested, the prior-year returns could be reassociated with the other returns filed for the same tax period. The resulting "tax year" SOI file should be a conceptually stronger data base from which to analyze the operation of our tax system, in that assumptions made about prior-year returns will have been eliminated.

However, there will be a considerable time lag in producing this "tax year" file, because the majority of prior-year returns are filed either one or two years late. Until these returns are filed, it will be impossible to build an accurate representation of the "ever-filed" population for a given tax year.

3. SIMULATION OF 1978 ADVANCE DATA AND FINAL ESTIMATES

Methodology.--A simulation of the 1978 SOI file, as it would have been, was created as a vehicle for evaluating the results of incorporating the three proposed changes discussed above. The simulation was also used as a preliminary step in evaluating the use of an early cut-off file to produce the complete SOI report for a given tax year.

In creating the simulation file, all sample returns on the 1978 SOI tape file with a tax year prior to 1978 or with a return processing cycle code greater than 37 (i.e., filed later than the third week of September) were assigned a weight factor of zero. This step excluded prior-year returns from the simulation, and included returns that would have been designated prior to the proposed cut-off for the simulation but processed after this date. These latter returns would not have been included had we followed the current processing method.

The second stage of the simulation, developing weight factors for the remaining returns, required: first, producing sample counts by sample code (stratum) within IRS districts; then, computing simple ratio weight factors, by dividing the sample count into the population. This paralleled the original 1978 sample weighting technique.

In order to maintain comparability between SOI and simulation estimates, the simulation file (which excludes prior-year returns) was weighted to represent the entire processing year population for 1978 (which includes prior-year returns). Columns 7 through 9 of Table 2 present a distribution, by size of AGI, of the simulation file after initial application of the simple ratio weight factors. Two obvious deficiencies exist at this stage: the deficit class and the $2,000,000 or more AGI class.

To overcome the deficiencies evident in the deficit and very high income classes, we assumed the institution of a special control system that would take over after the early cut-off, and continue until some specified time in the processing year, insuring that all returns designated in these two classes are included in the final sample. For purposes of this simulation, we assumed that the "specified time" was the end of the processing year. For all returns falling in these two classes, the original 1978 complete SOI weight factor was transferred to the simulation record. Columns 10 through 12 of Table 2 present the simulation results after this adjustment.

A final refinement was made to the simulation weights to adjust for the absence of prior year returns in the simulation sample. In order to accomplish this, we first developed a basis for adjustment by applying the ratio of aggregate AGI between 1977 and 1978 to the AGI amount (on a record by record basis) on prior-year returns in the 1978 complete SOI file. Deficit prior-year returns were not adjusted. Columns 1 through 3 of Table 2 present the distribution, by size of adjusted gross income, of this adjustment to the 1978 complete SOI file. This should be an approximate representation of the frequency distribution of an "ever-filed" population for 1978.

Ratios were then computed, on an income class by income class basis, between the expected number of returns (Column 1) and the simulation number of returns (Column 10). The ratios thus developed were applied to the data in columns 10 through 12 and the results presented in columns 1 through 4. For income less than $30,000, the classes used for computing the ratios were much broader than those presented in Table 2. The computation of ratios based on broad classes and applied to narrow classes accounts for the minor discrepancy in the number of returns between columns 1 and 4 for classes below the $30,000 income level. These broader classes (than those presented in Table 2 for income levels less than $30,000) were used for computing ratios because the ranges correspond to similar tabulations available for 1979 and future years. When we do simulations for years later than 1978, these tabulations will become the basis for this type of adjustment.

These ratios were applied to the existing simulation weights on a record by record basis to generate the final simulation weights. These final simulation weights were used to produce all the "simulation" estimates that appear labelled as "simulation after all adjustments". A brief explanation of the weighting technique employed in generating the 1978 advance data estimates is presented in note [12].

Two unmeasurable differences exist between the original 1978 advance data estimates and those generated through any 1978 simulation run. Although few in number, duplicate returns in the 1978 SOI file were deleted at final SOI closeout, not at preliminary cut-off. Thus any duplicate returns in the 1978 preliminary file had been deleted from the complete 1978 SOI sample file before we began creating the simulation file. Also, as part of normal consistency testing of SOI returns at the Data Center, information listings of returns with unusual, unexpected, or out-of-range items are produced. These sampled returns are located (whenever possible) and reviewed by statisticians in the Statistics Division. Most corrections or changes posted to the SOI file as a result of this review were not available at preliminary cut-off, but were made to the final SOI file which was the starting point for this research file. Simulations for future years, 1980 and on, will measure these differences.

4. SIMULATION RESULTS AND RECOMMENDATIONS

Results.--A careful comparison of columns 3 and 9 in Table 2 would lead one to conclude that a straightforward simulation of 1978 advance data that incorporates the three basic changes discussed earlier (with no subsequent refinements) accurately reproduces the 1978 complete SOI estimates for adjusted gross income except in the deficit and very high income classes. The discrepancy in the deficit class should be expected because the average deficit on returns filed after the cut-off is more than 5 times larger than the average deficit on returns filed before the cut-off. The differences encountered in the $2,000,000 or more class also appear to be the result of excluding late filed returns.

In terms of producing the advance data tabulations from an early SOI file that incorporates the three changes explained above, a special control and handling system must be instituted. This system would begin at the early cut-off and would maintain strict controls on deficit and very high income returns, insuring that any sample returns designated in these two categories after the cut-off date would be included in the early cut-off SOI file. For advance data, this system could provide an additional six weeks worth of these cases (this is four weeks longer than under the current processing system). This system, for advance data, would end early in November in time to develop weight factors for the advance data (or preliminary) file.

In attempting to simulate the complete SOI for 1978, we carried the idea of a special control system a little further. We assumed this type of system would continue to the end of the processing year. The results of including this control system through the end of the processing year, as well as incorporating the ratio refinement (to adjust for the exclusion of prior year returns), resulted in simulation estimates that were within the range described by coefficient of variation (at the 68% confidence level) for all but the three items listed below.

In reviewing the simulation estimates we were quite concerned with the levels of Business Net Profit, Business Net Loss and Net Capital Gain. A distribution of returns with these items by size of the item and by returns included in and excluded from the simulation indicated that an early cut-off sample was not representative of these categories of returns. It appears from this information that late filed returns with large amounts for any one of these three items should be included in our special handling and control system.

On balance, it appears from the results of the various simulation runs produced to date that it will be possible to modify the SOI processing system, conserve resources, produce earlier estimates, and only marginally (if at all) affect the reliability of the SOI figures.

Recommendations.--Our recommendation for constructing an early cut-off advance data sample would incorporate the three basic proposals outlined earlier. In addition, a special control system would be instituted to include deficit, very high income, and large "special item" (business, capital gain or other) returns designated within six weeks after the cut-off in the advance data SOI file. [3]

In constructing an early cut-off complete SOI file, we recommend continuing the "special control" system through early December. In addition to this, any sample returns processed error-free through the Service Centers between the mid-September and the early December cut-off dates should also be included in the final SOI file. The inclusion of these additional sample units in the final SOI will reduce the sampling variability of the estimates made from that sample. On the other hand, this procedure could introduce an element of bias into the sample if the error-free records are not representative of all returns processed during that period of time. The proposed 1980 simulation will analyze this problem.

5. FUTURE PLANS

The research and testing of an early cut-off for preliminary or advance data SOI estimates is really the first step in a longer-range plan to produce the complete SOI report for a given tax year from an earlier cut-off SOI file than is currently being used. The benefits of an early cut-off for SOI publication purposes are two-fold. Resources are conserved and data is available for release much earlier.

One of the proposals we are giving very serious thought to calls for closing out the basic sample file, from which the complete SOI report will be produced, after the third week of September (as was done in this 1978 simulation). Sample designation and data transcription will continue through the end of the processing year. Error free returns sampled after the cut-off date, as well as any returns subjected to special handling (deficits, very high incomes, etc.), will be included in the publication version of the SOI file for any given tax year.

The early cut-off advance data recommendation will be simulated (exactly as specified) using the 1980 SOI File. The early cut-off complete SOI File will also be simulated (again, exactly as specified) using the 1980 SOI File. The 1980 file is the first one available containing all the indicators necessary to isolate each specific category of return. The 1980 file will also contain the necessary information to measure the effects of duplicate returns and post-processing improvements mentioned in the methodology section of this report.

The publication SOI file will become the basis for the IRS individual tax model and a version of this file will be provided to the National Archives for distribution as a public use data

base. As with the current system, OTA will have access to this final SOI file for generating their tax model. The time frame for the availability of this final SOI file to OTA and to the National Archives will be considerably earlier than under the present system.

Even though prior year returns have been excluded from this simulation and would be excluded from future SOI publications, they will still be designated as part of the sample, isolated, and maintained separately. At some point in time, these prior year returns will be associated with tax year SOI file in which they belong and basic tabulations will be produced. This updated file will be made available as a public use data base.

In order to successfully produce the complete SOI report from an early cut-off file, a number of issues (potential problem areas) must be explored and resolved. Some of the more critical issues are: imputing missing data items resulting from an early cut-off, adjusting for late filed returns and improving population estimation techniques, see [12]. These issues will be explored in a simulation of an early cut-off 1980 SOI File, and the findings presented in a subsequent, related report.

Instead of the simple ratio estimation weighting technique now being used, a raking ratio estimation technique might better adjust for some of the skewing tendencies exhibited by the late-filers. Raking is a procedure for iteratively ratioing sample data to known (outside) marginal totals [10]. The raking ratio method will be tested against the proposed 1980 simulation file and the results also presented in a later report.

The future simulations mentioned in this section and in the methodology section should provide a more realistic test of our proposals than did the 1978 simulation. Because of the unavailability of information from the intermediate stages of 1978 processing, we simulated a "best case" situation. For coming simulations, we expect to reproduce the exact conditions we plan to implement.

It should be kept in mind that a fail-safe system is implied by the continuation of sample designation and data transcription after the early cut-off. If it becomes apparent that the reliability of the SOI data will be compromised beyond acceptable limits by using an early cut-off for the published SOI report, it will always be possible to produce the necessary tabulations from a complete SOI file comparable to that used for previous SOI years.

One final point must be made. The trend in the filing pattern over recent years indicates that returns are being filed later and later each year [8]. This would imply that, in the long run, it will become increasingly difficult to justify an early cut-off SOI sample. The early cut-off proposal is a reasonable short run strategy for meeting the commitment to earlier release of SOI estimates. A better long run strategy would be to standardize and streamline the processing that occurs between sample designation and the publication of estimates. If such standardization can be achieved, it could allow for a cut-off that is even later than under the current system.

ACKNOWLEDGEMENTS

The authors would like to take this occasion to acknowledge the helpful comments and suggestions of Ralph Bristol and Thomas Vasquez of the Treasury Department's Office of Tax Analysis, Neil Barclay of Revenue Canada Taxation and Peter Davis of the Senate Budget Committee. In addition, we would like to thank Mary Haigler and Dawn Nester for their help in typing the various drafts and tables for this presentation.

NOTES AND REFERENCES

[1] Blacksin, Jack and Raymond Plowden, Statistics of Income for Individuals: A Historical Perspective, 1981 Proceedings: American Statistical Association, Section on Statistical Use of Administrative Records, 1981.

[2] Deming, W. Edwards, Review of Quality Standards of the Procedures Used by the Internal Revenue Service to Produce Statistics of Income From Individual Returns with Special Emphasis on the Sampling Procedures, 1963.

[3] Hirsch, Werner Z., Ways in Which the Internal Revenue Service can Improve its Service to the Treasury and Other Users, 1961.

[4] Houthakker, H.S., The Future of the Statistics of Income Program: A Preliminary Report to the Office of Tax Analysis, U.S. Treasury Department, 1966.

[5] Individual Income Tax Returns for 1978 - Advance Data, 1979 (unpublished).

[6] Internal Revenue Service, Individual Tax Model File for 1978, Order Number 374-109(a), National Archives and Records Service, 1980.

[7] Internal Revenue Service, Preliminary Statistics of Income--1978, Individual Income Tax Returns, Publication 198, U.S. Government Printing Office, 1980.

[8] Internal Revenue Service, Statistics of Income Bulletin, Publication 1136, U.S. Government Printing Office, 1981. The Summer 1981 issue contains information that would have appeared in Preliminary SOI report [7]. The Fall 1981 issue includes articles on return filing patterns.

[9] Internal Revenue Service, Statistics of Income--1978, Individual Income Tax Returns, Publication 79, U.S. Government Printing Office, 1981.

[10] Scheuren, F. et. al., Studies from Interagency Data Linkages, Report No. 10, 1981.

[11] Vanik, Charles A., Letter of November 1 to Commissioner of IRS transmitting report of GAO study of timeliness of Statistics of Income data. 1p., and attachments, 1976.

[12] For advance data (or preliminary) estimates of a given SOI year, Y, the sample is cut-off at some point before the sampling time frame is complete. The population continues to be counted and the sample continues to be selected beyond the cut-off date until the time period (frame) is satisfied; however, sample returns selected after the cut-off are not included in the advance data estimates. The portion of the population to be counted (and selected from) after the cut-off in year Y is unknown, but is estimated from the previous year's (Y-1) known population counts for the similar time period, then added to the current year (Y) known population counts. This method of estimation is applied by sample stratum within districts. Therefore, the advance data weight factors are based on a known sample count and an estimated population. Over the years, this method has proven to be reliable in estimating the full year SOI population counts.

Table 1.-- Number of Returns and Column Percents: by Filing Year, Processing Cycle and Selected Classifications

Selected Classifications	1978 Total	Filing Year		Processing Cycle			Simulation After All Adjustments
		Current	Prior	1 Through 37	38 Through 39	40 or later	
Part 1.-- Frequencies (in thousands of returns)							
Total	89,771	88,726	1,046	88,803	229	739	89,771
Joint	44,483	43,957	526	43,945	129	408	44,528
Nonjoint	45,288	44,768	520	44,858	100	330	45,243
Nonbusiness	81,224	80,403	821	80,542	163	519	81,194
Business	8,548	8,323	225	8,261	66	220	8,577
Itemized	25,756	25,482	274	25,388	96	273	25,751
Other	64,015	63,244	772	63,416	133	466	64,020
1040	53,824	53,026	798	52,995	204	626	53,786
1040A	35,947	35,700	247	35,808	26	113	35,985
Part II.-- Column Percents							
Total	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Joint	49.55	49.54	50.31	49.49	56.36	55.29	49.60
Nonjoint	50.45	50.46	49.69	50.51	43.64	44.71	50.40
Nonbusiness	90.48	90.62	78.49	90.70	71.02	70.18	90.45
Business	9.52	9.38	21.51	9.30	28.98	29.82	9.55
Itemized	28.69	28.72	26.21	28.59	42.00	36.89	28.69
Other	71.31	71.28	73.79	71.41	58.00	63.11	71.31
1040	59.96	59.76	76.35	59.68	88.81	84.71	59.91
1040A	40.04	40.24	23.65	40.32	11.21	15.29	40.09

Table 2.-- Number of Returns, Amount and Average AGI by Size of AGI for 1978 Complete SOI (Adjusted), Various Stages of Simulation and 1978 Advance Data

(Number of returns in thousands, amounts in millions and averages in whole dollars)

Size of Adjusted Gross Income	SOI Complete After Ratio Adjustment For Prior Year Returns			Simulation									1978 Advance Data as Transmitted to OTA		
				Simulation After all Adjustments			Raw Data			Adjustment for Deficit and Very High AGI Only					
	Number	Amount	Average	Number	Amount	Average	Number	Amount	Average	Number	Amount	Average	Number	Amount	Average
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)
Total	89,771	1,303,434	14,520	89,771	1,303,653	14,521	89,773	1,305,728	14,545	89,822	1,303,796	14,515	89,890	1,304,189	14,509
Deficit	484	-7,473	-15,431	484	-7,473	-15,431	435	-5,569	-12,796	484	-7,473	-15,431			
Breakeven	40	-	-	22	-	-	22	-	-	22	-	-			
$1 under $2,000	8,469	9,472	1,118	8,424	9,452	1,122	8,446	9,476	1,122	8,446	9,476	1,122	9,048	3,230	357
$2,000 under $4,000	9,234	27,605	2,989	9,237	27,630	2,991	9,261	27,702	2,991	9,261	27,702	2,991	9,260	27,697	2,991
$4,000 under $6,000	8,388	42,236	5,035	8,362	42,111	5,036	8,384	42,220	5,036	8,384	42,220	5,036	8,401	42,292	5,035
$6,000 under $8,000	8,259	57,530	6,966	8,303	57,858	6,968	8,325	58,008	6,968	8,324	58,008	6,968	8,263	57,579	6,968
$8,000 under $10,000	6,926	62,000	8,952	6,968	62,366	8,951	6,986	62,528	8,951	6,986	62,528	8,951	6,944	62,146	8,950
$10,000 under $12,000	6,089	66,870	10,983	6,103	67,021	10,981	6,097	66,952	10,981	6,097	66,952	10,981	6,097	66,949	10,982
$12,000 under $14,000	5,584	72,436	12,971	5,580	72,374	12,971	5,574	72,300	12,971	5,574	72,300	12,971	5,604	72,671	12,969
$14,000 under $16,000	5,016	75,198	14,993	5,022	75,302	14,994	5,017	75,224	14,994	5,017	75,225	14,994	5,016	75,212	14,993
$16,000 under $18,000	4,669	79,388	17,002	4,678	79,533	17,002	4,673	79,452	17,002	4,673	79,452	17,002	4,683	79,611	17,002
$18,000 under $20,000	4,284	81,361	18,992	4,271	81,123	18,990	4,268	81,040	18,990	4,268	81,040	18,990	4,277	81,234	18,992
$20,000 under $25,000	8,560	191,104	22,326	8,553	190,952	22,326	8,544	190,758	22,326	8,544	190,758	22,326	8,555	190,969	22,323
$25,000 under $30,000	5,394	147,023	27,258	5,388	146,877	27,260	5,382	146,728	27,260	5,382	146,728	27,260	5,385	146,794	27,260
$30,000 under $50,000	6,546	239,353	36,564	6,546	239,304	36,556	6,529	238,658	36,556	6,529	238,658	36,556	6,534	238,846	36,552
$50,000 under $100,000	1,475	96,499	65,430	1,475	96,447	65,395	1,473	96,319	65,395	1,473	96,319	65,395	1,469	96,005	65,358
$100,000 under $200,000	286	37,591	131,469	286	37,577	131,421	288	37,847	131,421	288	37,487	131,421	285	37,471	131,405
$200,000 under $500,000	60	16,737	279,018	60	16,705	278,484	61	17,035	278,484	61	17,035	278,484	60	16,726	278,424
$500,000 under $1,000,000	7	4,387	666,074	7	4,386	665,885	7	4,764	665,885	7	4,764	665,885	7	4,572	665,336
$1,000,000 under $2,000,000	1	1,991	1,335,085	1	1,984	1,330,443	2	2,138	1,330,444	2	2,133	1,330,444			
$2,000,000 or more	1	2,126	3,844,367	1	2,126	3,844,367	1	2,153	3,824,190	1	2,126	3,844,367	2	4,186	2,000,775

Table 3.-- Number of Returns, Amount of AGI and Average AGI, by Size of AGI and by Filing Category

Size of Adjusted Gross Income	All Returns	All Returns by Processing Cycle			Prior Year	Returns Missing From Advance Data
	Total	1 Through 37	38 Through 39	40 or Later	Returns Total	
		Part I.-- Number of Returns				
TOTAL	89,771,249	88,802,683	229,466	739,100	1,045,897	518,157
Deficit	484,299	458,607	4,238	21,454	46,542	5,951
Breakeven	39,765	29,381	4,248	6,136	11,326	1
$1 under $2,000	8,469,208	8,394,533	21,837	47,838	125,785	46,698
$2,000 under $4,000	9,234,173	9,162,794	13,358	58,020	115,220	40,865
$4,000 under $6,000	8,387,955	8,277,372	13,360	97,224	125,177	43,703
$6,000 under $8,000	8,258,760	8,191,922	9,609	57,229	84,981	49,043
$8,000 under $10,000	6,925,837	6,874,631	13,667	37,539	76,923	56,386
$10,000 under $12,000	6,088,694	6,048,040	6,490	34,163	53,674	47,100
$12,000 under $14,000	5,584,491	5,541,056	15,799	27,637	66,396	32,014
$14,000 under $16,000	5,015,526	4,973,469	6,596	35,461	51,543	13,897
$16,000 under $18,000	4,669,441	4,622,746	11,921	34,774	40,756	17,627
$18,000 under $20,000	4,283,867	4,236,597	12,650	34,620	50,923	18,063
$20,000 under $25,000	8,559,908	8,472,868	24,553	62,488	74,042	48,682
$25,000 under $30,000	5,393,740	5,331,662	20,300	41,778	47,026	26,190
$30,000 under $50,000	6,546,202	6,434,101	29,414	82,687	53,345	50,459
$50,000 under $100,000	1,474,835	1,413,643	17,172	44,019	18,231	14,894
$100,000 under $200,000	285,931	270,567	3,257	12,107	5,266	4,974
$200,000 under $500,000	59,987	55,965	868	3,154	652	1,321
$500,000 under $1,000,000	6,586	5,940	89	557	61	207
$1,000,000 under $2,000,000	1,491	1,324	22	145	17	62
$2,000,000 or more	553	465	18	70	11	22
		Part II.-- Amount of AGI (in thousands of dollars)				
TOTAL	1,303,434,144	1,283,845,448	5,118,434	14,470,264	12,194,500	8,953,258
Deficit	-7,473,332	-5,829,444	-181,941	-1,461,947	-877,695	-147,439
Breakeven	-	-	-	-	-	-
$1 under $2,000	9,471,532	9,409,834	18,896	42,802	118,835	51,414
$2,000 under $4,000	27,605,226	27,389,713	43,205	172,308	335,131	130,570
$4,000 under $6,000	42,235,731	41,669,940	66,811	498,981	622,496	225,542
$6,000 under $8,000	57,530,195	57,066,905	63,989	399,302	581,714	335,841
$8,000 under $10,000	62,000,014	61,537,350	124,330	338,334	703,959	496,165
$10,000 under $12,000	66,869,810	66,413,961	72,385	383,463	591,878	513,059
$12,000 under $14,000	72,436,262	71,877,675	199,872	358,715	862,518	419,353
$14,000 under $16,000	75,198,387	74,565,145	100,196	533,046	769,630	206,630
$16,000 under $18,000	79,387,934	78,594,834	203,009	590,091	690,160	300,111
$18,000 under $20,000	81,360,933	80,457,588	246,221	657,124	973,122	344,446
$20,000 under $25,000	191,104,401	189,162,544	555,803	1,386,054	1,652,512	1,117,823
$25,000 under $30,000	147,023,113	145,333,832	551,732	1,137,548	1,277,477	706,555
$30,000 under $50,000	239,353,313	235,137,323	1,111,702	3,104,288	1,968,283	1,900,683
$50,000 under $100,000	96,498,828	92,411,910	1,121,329	2,965,590	1,179,501	1,005,524
$100,000 under $200,000	37,591,030	35,564,676	424,865	1,601,489	427,303	657,475
$200,000 under $500,000	16,737,457	15,578,279	246,471	912,708	179,322	375,162
$500,000 under $1,000,000	4,386,763	3,953,449	59,340	373,974	42,628	143,186
$1,000,000 under $2,000,000	1,990,612	1,762,907	28,133	199,572	22,476	82,808
$2,000,000 or more	2,125,935	1,787,027	62,086	276,822	73,254	88,350
		Part III.-- Average AGI (in whole dollars)				
TOTAL	14,520	14,457	22,306	19,578	11,659	17,279
Deficit	-15,431	-12,711	-42,931	-68,143	-18,858	-24,775
Breakeven	-	-	-	-	-	-
$1 under $2,000	1,118	1,120	865	895	945	1,101
$2,000 under $4,000	2,989	2,989	3,234	2,970	2,909	3,195
$4,000 under $6,000	5,035	5,034	5,001	5,132	4,973	5,161
$6,000 under $8,000	6,966	6,966	6,659	6,977	6,845	6,848
$8,000 under $10,000	8,952	8,951	9,097	9,013	9,151	8,799
$10,000 under $12,000	10,983	10,981	11,153	11,225	11,027	10,893
$12,000 under $14,000	12,971	12,972	12,651	12,980	12,990	13,099
$14,000 under $16,000	14,993	14,993	15,190	15,032	14,932	14,869
$16,000 under $18,000	17,002	17,002	17,030	16,969	16,934	17,026
$18,000 under $20,000	18,992	18,991	19,464	18,981	19,110	19,069
$20,000 under $25,000	22,326	22,326	22,637	22,181	22,319	22,962
$25,000 under $30,000	27,258	27,259	27,179	27,228	27,165	26,978
$30,000 under $50,000	36,564	36,545	37,795	37,543	36,897	37,668
$50,000 under $100,000	65,430	65,371	65,300	67,371	64,698	67,512
$100,000 under $200,000	131,469	131,445	130,447	132,278	130,834	132,182
$200,000 under $500,000	279,018	278,358	283,953	289,381	275,034	283,998
$500,000 under $1,000,000	666,074	665,564	666,742	671,408	698,820	691,720
$1,000,000 under $2,000,000	1,335,085	1,331,501	1,278,773	1,376,359	1,322,118	1,335,613
$2,000,000 or more	3,844,367	3,843,069	3,449,222	3,954,600	6,659,455	4,015,909

Table 4.-- 1978 Complet SOI Report, Simulation and Advance Data: Amounts, Differences and Coefficients of Variation for Specified Items

Item	1978 Complete SOI Report ($000)	Similation After All Adjustments ($000)	1978 Advance Data ($000)	Complete Minus Simulation as a Percent of Complete	Complete Minus Advance Data as a Percent of Complete	Coefficient of Variation for 1978 Complete 1/
Adjusted Gross Income	1,302,447,386	1,303,647,457	1,304,188,847	0.09	0.13	0.1
Salaries and wages	1,090,291,855	1,092,086,262	1,092,017,073	0.16	0.16	0.2
Business net profit	61,413,703	60,957,923	60,741,261	0.74	1.09	0.6
Business net loss	7,867,195	7,686,962	7,412,957	2.29	5.77	1.5
Farm net profit	11,034,552	11,015,275	10,989,231	0.17	0.41	3.4
Farm net loss	7,469,259	7,326,396	7,180,316	1.91	3.87	3.4
Partnership net profit less loss	15,044,787	14,849,870	15,407,324	1.30	2.41	3.5
Small Business Corp. net profit less loss	2,284,806	2,211,392	2,471,275	3.21	8.16	10.1
Net capital gain	26,232,396	25,125,251	24,993,143	4.22	4.72	1.4
Net capital loss	3,001,020	2,955,025	2,951,478	1.53	1.65	2.9
Sales of property other than capital assets	1,256,902	1,232,696	1,246,753	1.93	0.81	9.2
Total dividends	31,671,858	31,677,425	31,634,519	0.02	0.12	1.3
Dividends in adjusted gross income	30,206,475	30,208,441	30,169,755	0.01	0.12	1.4
Interest received	61,222,522	61,419,009	60,947,334	0.32	0.45	1.0
Pensions and annuities, in AGI	32,743,819	33,141,546	32,883,949	1.21	0.43	1.9
Rent net income	10,983,905	11,032,681	10,878,106	0.44	0.96	2.5
Rent net loss	7,844,747	7,592,879	7,618,195	3.21	2.89	2.5
Royalty net income less loss	2,559,870	2,599,667	2,573,943	1.55	0.55	5.3
Estate or trust net income less loss	3,079,603	2,977,860	2,990,517	3.30	2.89	4.8
State income tax refunds	2,368,949	2,361,439	2,363,389	0.32	0.23	1.4
Alimony received	1,191,389	1,130,013	1,170,453	5.15	1.76	10.7
Other income less loss	-921,836	-948,473	-157,729	2.89	82.89	--
Adjustments	22,364,088	22,192,133	22,333,986	0.77	0.13	1.3
Disability income exclusion	1,066,206	1,097,204	1,126,745	2.91	5.68	10.2
Payments to an Ind. Retirement Acct	2,970,121	2,988,498	2,984,424	0.62	0.48	2.0
Payments to a KEOGH	1,994,029	2,018,627	1,990,410	1.23	0.18	1.9
Deduction for expense of living abroad	314,468	412,145	340,613	31.06	8.31	8.9
Exemption amount	164,900,772	164,692,352	165,008,114	0.13	0.01	--
Taxable income	1,062,190,322	1,064,317,958	1,063,308,318	0.20	0.11	--
Income tax before credits	203,803,653	204,365,014	204,162,283	0.28	0.18	--
Total credits	17,085,591	16,990,499	16,989,359	0.56	0.56	--
New jobs credit	1,370,406	1,333,778	1,327,934	2.67	3.10	--
Earned income credit used to offset tax before credits	152,934	152,788	139,738	0.10	8.63	--
Residential energy credit	576,545	583,060	578,405	1.13	0.32	--
Business energy investment credit	219,868	219,590	1,314	0.13	99.40	--
Income tax after credits	186,718,062	187,374,515	187,172,925	0.35	0.24	--
Total tax preferences	18,381,866	17,451,568	17,284,665	5.06	5.97	--
Minimum tax	1,514,475	1,423,157	1,404,261	6.03	7.28	--
Total income tax	188,232,537	188,797,672	188,577,186	0.30	0.18	0.2
Self-employment tax	4,705,994	4,651,382	4,648,370	1.16	1.22	--
Earned income credit used to offset all other taxes	94,197	93,378	85,622	0.87	9.10	--
Total tax liability	193,184,849	193,685,577	193,464,593	0.26	0.14	0.2
Total taxpayments	202,829,400	203,666,001	203,340,183	0.41	0.25	--
Withholding	169,984,010	170,745,020	170,537,019	0.45	0.33	--
Estimated payments	29,978,499	30,429,454	30,262,953	1.50	0.95	--
All other taxpayments	2,866,890	2,491,526	2,540,211	13.09	11.39	--
Earned income credit, refundable portion	801,171	796,200	756,708	0.62	5.55	--
Business energy investment credit, refundable	397	353	401	11.08	1.01	--
Tax due at time of filing	24,969,333	24,608,375	24,729,595	1.45	0.96	--
Total overpayment	35,415,451	35,385,352	35,362,293	0.08	0.15	--
Refund	33,034,549	33,082,636	33,042,923	0.15	0.03	--
Credit on 1979 tax	2,380,903	2,302,716	2,319,370	3.28	2.58	--

1/ Coefficient of variation at the 68% confidence level.

Table 5.-- Number of Returns, Amount and Average Net Capital Gain or Loss, Business Net Loss and Business Net Profit: by Size of the Item, for Returns Included in and Excluded From the Complete SOI Simulation (using 1978 SOI weights)

Size of Item	Returns Included in the Simulation			Returns Excluded From the Simulation		
	Number	Amount ($000)	Average ($)	Number	Amount ($000)	Average ($)
Part I.-- Size of Net Capital Gain or Loss						
Total	8,375,948	21,140,808	2,524	335,004	2,092,158	6,245
Loss						
Less than $3,000	397	4,911	12,369	1	1,075	1,074,809
$2,000 under $3,000	708,474	2,038,077	2,877	35,245	97,187	2,757
$1,000 under $2,000	330,916	474,998	1,435	25,191	30,976	1,230
$1 under $1,000	973,837	339,922	349	32,433	13,797	425
Gain						
$1 under $1,000	3,354,669	991,072	295	89,816	27,388	305
$1,000 under $5,000	1,977,244	4,753,542	2,404	59,850	153,124	2,558
$5,000 under $10,000	562,396	4,002,862	7,118	33,279	264,892	7,960
$10,000 under $25,000	349,691	5,359,407	15,326	45,752	701,840	15,340
$25,000 under $50,000	74,781	2,562,654	34,269	8,142	276,901	34,009
$50,000 under $100,000	27,998	1,927,571	68,847	3,094	209,588	67,740
$100,000 under $200,000	9,548	1,292,869	135,407	1,423	197,103	138,512
$200,000 under $500,000	4,506	1,355,935	300,918	538	164,257	305,310
$500,000 or more	1,491	1,752,805	1,175,590	240	240,100	1,000,416
Part II.-- Size of Business Net Loss						
Total	1,948,949	7,303,319	3,747	84,850	562,885	6,634
$1 under $5,000	1,624,640	2,216,914	1,365	63,122	105,273	1,668
$5,000 under $10,000	186,378	1,294,931	6,948	11,299	81,167	7,184
$10,000 under $25,000	105,489	1,553,626	14,728	7,245	103,153	14,238
$25,000 under $50,000	21,638	747,496	34,546	1,775	63,655	35,862
$50,000 under $100,000	6,761	458,017	67,744	885	58,421	66,013
$100,000 under $200,000	2,807	384,584	137,009	323	42,789	132,475
$200,000 under $500,000	951	279,896	294,318	141	40,069	284,174
$500,000 or more	285	367,855	1,290,721	60	68,358	1,139,302
Part III.-- Size of Business Net Profit						
Total	5,838,743	57,161,392	9,790	321,688	4,245,545	13,198
$1 under $5,000	3,175,511	5,676,113	1,787	134,435	303,349	2,256
$5,000 under $10,000	1,026,845	7,461,942	7,267	67,701	498,724	7,367
$10,000 under $25,000	1,067,952	16,761,782	15,695	77,930	1,228,650	15,766
$25,000 under $50,000	406,954	14,054,200	34,535	28,354	981,268	34,608
$50,000 under $100,000	136,982	9,144,197	66,755	10,488	720,272	68,676
$100,000 under $200,000	21,064	2,703,451	128,345	2,173	286,576	131,881
$200,000 under $500,000	2,972	821,251	276,330	499	138,414	277,382
$500,000 or more	463	538,455	1,162,971	108	88,292	817,521