Statistics of Income: A By-Product of the U.S. Tax System

By Robert A. Wilson*

U.S. Government statistics are the product of a decentralized statistical system that involves over 70 Federal Government organizations, one of which is the Statistics of Income Division of the Internal Revenue Service (IRS) in the Department of the Treasury [1]. Another characteristic of the U.S. system is that statistics are often a by-product of an administrative function and are based on an administrative record. In the case of the Statistics of Income Division, the economic, financial, and tax statistics it produces are a by-product of tax returns that are processed in administering the tax laws. The report series in which the data are released is called Statistics of Income (SOI).

This paper reviews the relationships between processing for tax administration and processing for statistics through about 1985. It begins with a description of some of the SOI programs and their uses. It then reviews IRS and SOI processing and their limitations and some of the improvements in SOI processing now under consideration. These improvements will help the Statistics of Income Division to operate more efficiently and effectively in meeting the needs of its major users and to adjust to the continuing climate of reduced budgets for statistics.

THE STATISTICS OF INCOME PROGRAM

The SOI series came into being after the 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. This 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 tax statute since that time. It is currently contained in the Internal Revenue Code of 1986, which is the basis for the present tax law [2].

SOI data from the very beginning (1913) have been used extensively for tax research and for estimating revenue, especially by officials in the Office of the Secretary of the Treasury. At the start, the reports were almost entirely designed to meet these needs. With the growth of research groups both within and outside the Federal Government and with the increased needs of the 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 of data with which to fulfill 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 for a wide variety of Government statistics had become more and more apparent. Each of these events made the tax return a valuable source of economic as well as tax information. While the tax definitions of data items presented some obstacles, the obstacles were far outweighed by the likelihood that taxpayers' response to the tax return tended to be more accurate than their response to special surveys. Moreover, with experience, users learned how to adjust the tax data for these definitions in order to meet their own particular needs. Meanwhile, SOI processing methods contributed by making some of the adjustments for the major users in the course of "editing" the tax return data for the statistics [3].

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 include more general purpose statistics that would assist economists and financial analysts [4].

The main emphasis of the annual statistics has always been individual and corporation income tax return data. Other studies based on other types of returns for which data have been tabulated either annually or periodically are partnerships, estates and gifts, fiduciaries, farmers' cooperatives, private foundations and other tax-exempt organizations, and employee plans. Schedules attached to some of the returns became the subject of separate SOI programs. The sole proprietorship schedules attached to individual income tax returns were a relatively early source of statistics which, when taken together with data from partnership returns, helped shed light on unincorporated business activity.

Another development in the growth of SOI has been the increasing tendency for new provisions of the tax law to require separate reports to Congress by Treasury's Office of Tax Analysis (OTA). These reports have required statistics on such topics as individuals with high income who are nontaxable, capital gains taxation, international boycott

* Coordination and Publications Staff. Reprinted (with minor revisions) from the *Multi-National Tax Modelling Proceedings*, Revenue Canada Taxation, September 1985.

participation, taxation of corporate income from U.S. possessions, and income of citizens working abroad.

Today, information obtained from the SOI program is used extensively throughout the Federal Government for a variety of purposes. Besides OTA and the congressional Joint Committee on Taxation, there is a third major Federal user of SOI, the Bureau of Economic Analysis in the Department of Commerce. Profits data for corporations in the National Income and Product Accounts are benchmarked to the SOI profits obtained from corporation income tax returns, which are then adjusted for conceptual differences and extrapolated to more current years based on more fragmentary data from other sources [5]. Returns of unincorporated businesses, i.e., 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 reported on individual income tax returns is also used for the National Accounts.

THE STATISTICS OF INCOME DIVISION

The 1916 Act that first called for the publication of Statistics of Income necessitated the creation of a statistical organization within IRS to carry out this mandate. The present successor to that original organization is the Statistics of Income Division in the National Office in Washington, D.C. The Division is part of the Office of Taxpayer Service and Returns Processing under the Deputy Commissioner (Operations) which is charged with the responsibility for processing tax returns.

The Statistics of Income Division is comprised of a staff mostly of statisticians and economists who work closely with users to determine the content of each program and report, to design the statistical samples used, and to develop processing procedures. Complications arise from the fact that the processing is currently decentralized in as many as eleven different locations throughout the country; hence there is a need for a strong coordinating role by the Statistics of Income Division, including adequate quality controls to assure uniform and accurate processing.

The role of the Division has changed over the years. Until recent times, it had the additional responsibility of producing management statistics to assist IRS in its day-to-day operations. However, SOI has always been the Division's major responsibility.

ADMINISTRATIVE VS. SOI PROCESSING: A BRIEF HISTORY

Within IRS, statistical processing of the tax return data has historically been a separate off-line operation, divorced from the mainline processing of tax returns for administrative purposes. There were reasons that dictated this separation, some of which are still applicable:

- SOI is designed to serve tax policymakers in particular and economists in general. Consequently, it is of little interest to tax administrators (the role of IRS is, above all, tax administration) most of whom are attorneys and accountants whose statistical needs, where they exist, are quite different from those of policymakers.

- As a corollary to the first point, SOI was and continues to be a byproduct of the IRS tax administration system. Therefore, SOI and the processing for it have often been given a lower priority by IRS. In recent years the SOI budget has reflected this, with continued declines in the resources set aside for statistics. As a program administered by IRS, these budget declines might ordinarily have been steeper. However, the fact that needs for Treasury Department tax policymaking had to be considered often served to mitigate the size of the declines.

- Most of the SOI programs are based on samples of returns and for many years these samples were manually designated. This sampling was accomplished most effectively only after administrative processing was completed. Moreover, after the sampled returns were selected, they were sent to a central location, the Statistics of Income Division in Washington, for special processing.

The administrative processing which preceded statistical processing is and has been a decentralized operation. Until the 1950's, all of this processing took place in the more than 50 IRS district offices throughout the United States where taxpayers filed their returns. This processing was in large part manual, assisted, beginning in 1948, by punchcard equipment to service the larger district offices. Area service centers, also with punchcard equipment, were created in the mid-1950's to assist these same larger districts. Administrative processing consisted largely of mathematical verification to assure that returns were "in balance," plus certain other basic checks that included a review of the tax computation [6].

In contrast, statistical processing was conducted in Washington and was limited to returns selected for the SOI samples. Compared to the administrative processing of these returns, it was a lengthier and more complicated procedure. Many more lines on the return forms were utilized and some of the totals reported on these lines had to be edited for the statistics using amounts identified through examination of supporting forms and taxpayer schedules that were included as part of the return.

The advent of automatic data processing (ADP) for tax administration purposes in the early 1960's had a profound

effect on IRS processing and statistical processing was directly affected. Statistical abstracting, editing, and transcription, which after 1955 had begun to be decentralized from Washington to the pre-ADP area service centers, were completely decentralized during the 1960's to the ten ADP regional service centers where returns were now filed by taxpayers and then processed there for administrative purposes under the new system.

However, with a view to relieving regional service centers of all processing not directly related to the administrative processing of returns, a data center (now called a computing center) was later established in Detroit, Michigan. Among its duties was SOI processing, especially the computer programming—to test and tabulate the data. The Data Center also manually resolved the errors uncovered by these tests and, in addition, for many years was responsible for the initial manual abstracting, editing and transcription of data from returns for the corporation and certain other SOI programs.

While statistical processing continued to be an off-line process, under the new ADP system processing was no longer under the control of the Statistics of Income Division. The role of the Division in response to these changes evolved into one of planning, coordinating and overseeing a field operation. This was in addition to its continuing role of meeting with users to identify their data needs and to publish the SOI reports. As a result of the transition, the Statistics of Income Division ceased to have its own manual and computer processing operation and no longer had access to the computer which it had previously shared with the Bureau of the Census. Instead, it became a "paying customer" for the processing services that were provided by other IRS organizational units whose principal functions were return processing in general. The Division developed the data "requirements" (including standards and goals for completion) and, at least in principle, these other organizations determined the best way to carry out and meet these requirements. The loss of control over its own processing operation and the administrative problems that developed under the new system have continued to this day to present challenges to timely, accurate, efficient and economical processing for statistics.

CURRENT IRS ADMINISTRATIVE PROCESSING

The concept of a centralized Master File containing a computerized account for each taxpayer was adopted by the Service in the mid-1950's. Then, in 1959, the concept of regional centers located in each of the Internal Revenue Regions was adopted. These centers were designed as a means of centralizing the processing that had previously taken place in the district offices and area centers, and of introducing computer processing of tax return data to replace a large part of the previously-manual operation. The first regional center was opened in 1961.

Using individual income tax returns as the example, Figures A and B summarize the major steps (also described below) in the administrative processing system as it has now evolved [7]:

- Tax returns are received at the service centers throughout the country where processing begins. Tax examiners in the service centers first check the returns received to be sure they are signed and meet the IRS criteria for a completed return. Returns are sorted by type and accompanying checks compared with the amounts reported. A quick search is made for unallowable deductions and obvious errors and the returns are coded for computer processing.

- Information and tax data comprising nearly all of the information reported on the main part of the return and much of what is included on selected supporting schedules are transcribed for computer processing. This is accomplished by means of direct data entry onto magnetic tape, using key station terminals.

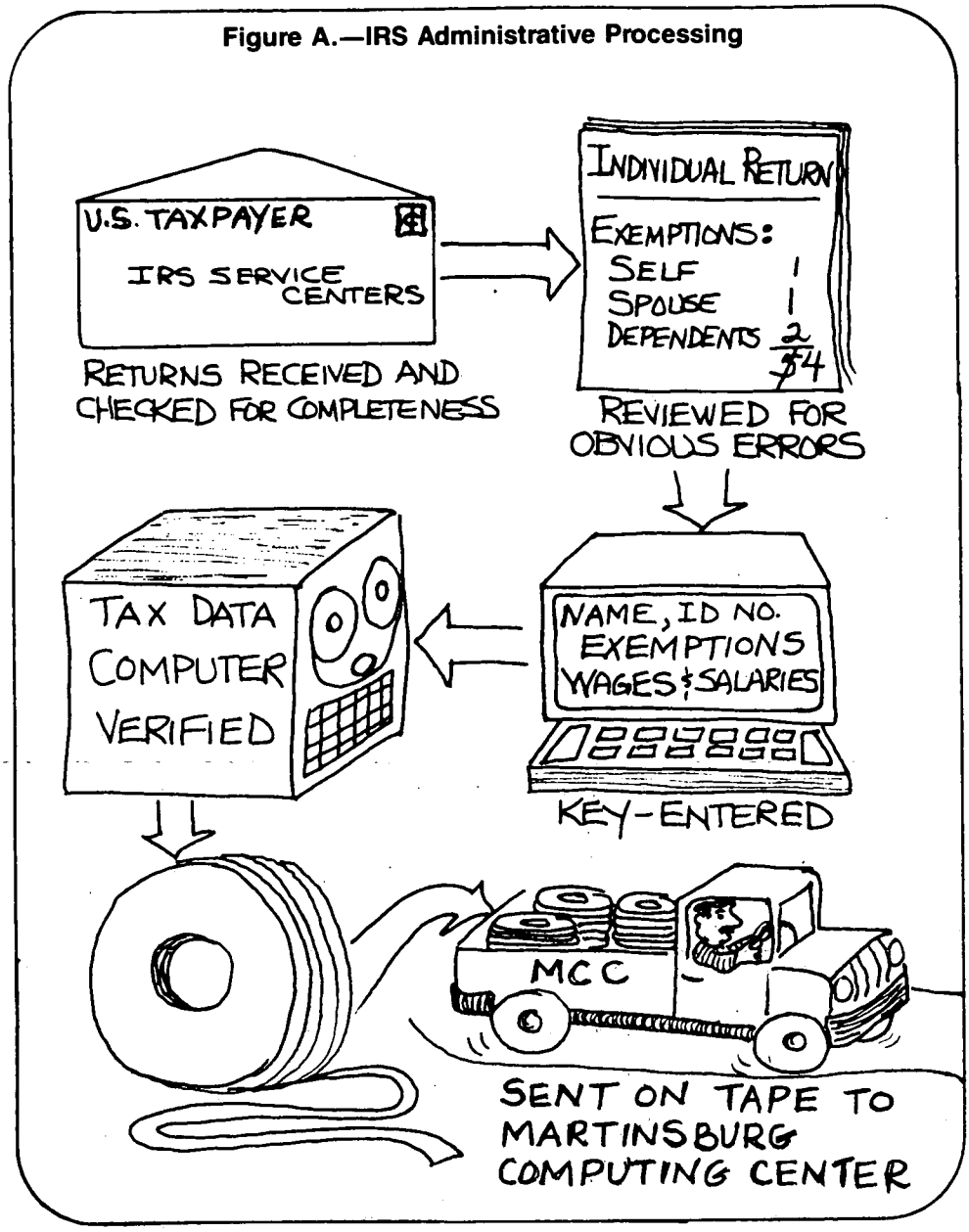
- After verifying the return count per "block," the computer verifies the tax computation used or, if appropriate, computes the tax for the taxpayer [8]. A number of consistency and validity checks are made to the information transcribed in order to identify certain mathematical errors made by taxpayers and mistakes made in the actual data capture process. (However, unless errors have a direct bearing on the tax reported, they may not be corrected. The implications of this on SOI are discussed below.)

- After a block of return records meets the standards for acceptability, the service center transmits it on computer tape to the national computing center in Martinsburg, West Virginia (MCC), for central account posting and "settlement." The function of MCC is to post taxpayer transactions to the so-called Master File. In the process, MCC performs several functions. It analyzes the data from the service centers by making comparisons with the prior year and generates refund and tax deficiency notices; it identifies returns for possible audit examination; and it creates special listings for the centers, one of which identifies the returns for inclusion in the SOI samples.

SOI USE OF THE TAX ADMINISTRATION SYSTEM

SOI began to piggyback directly on the tax administration system soon after the Master File for individuals was developed. The Master File offered a vastly improved sampling frame. Moreover, it could be accessed by com-

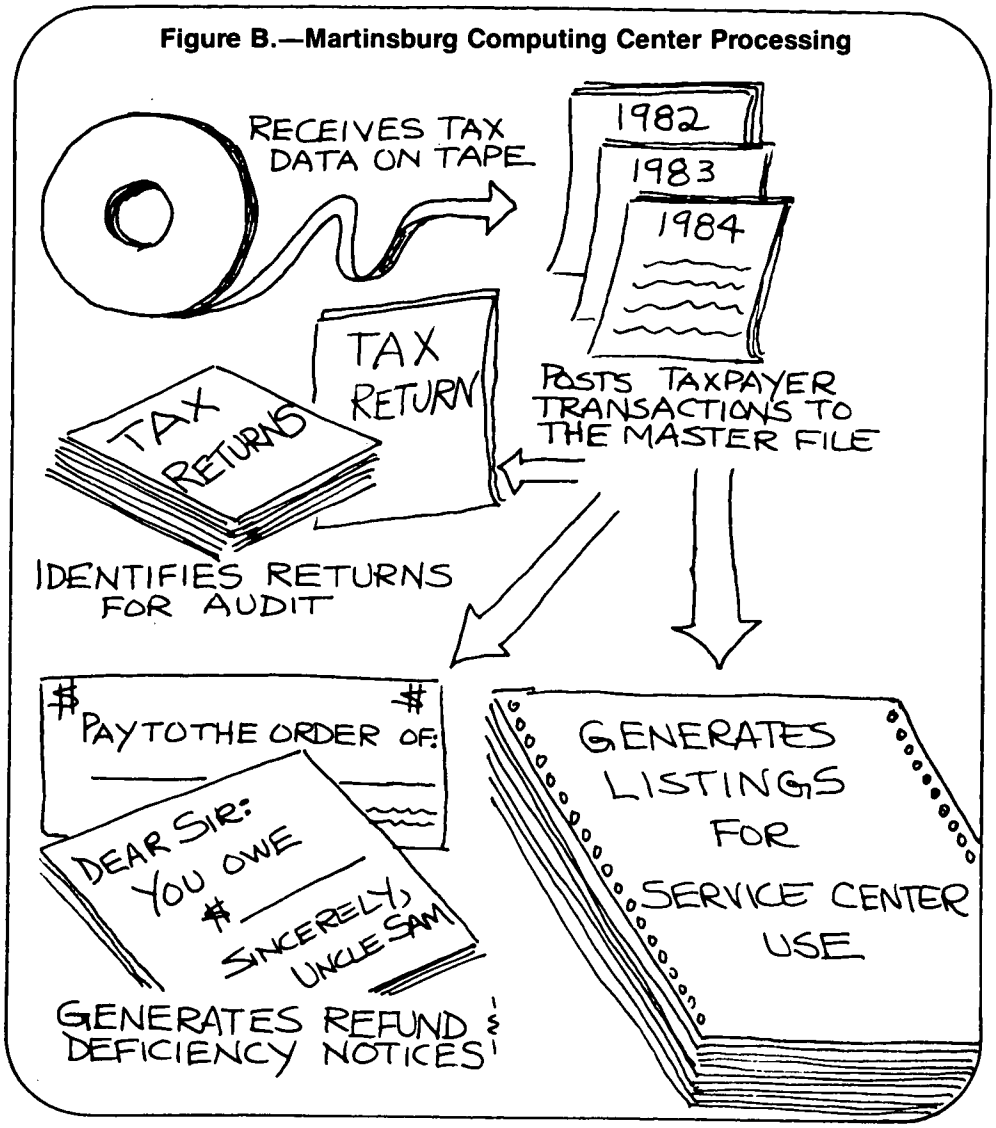
Figure A.—IRS Administrative Processing



puter, enabling more efficient, sophisticated and effective sample designs to be used than under manual sampling. Since, in addition, these samples were usually smaller in size, economies as well as improved data often resulted. Sampling by computer was then gradually extended to the other major SOI programs. Within a few years, nearly all of the SOI samples were designated by computer, using information from the Master File system.

Yet, for many years, little use was made of the actual data captured from the returns for tax administration purposes [9]. As a result, there appeared to be some duplication between the administrative and SOI systems because, for returns in the SOI sample, many of the same data items were processed twice, once for tax administration and once for statistics. However, a number of items were processed differently under the two systems. On the one hand, all data used for SOI were subject to rigorous testing and to statistical editing when necessary. On the other hand, because of the sheer volume of returns processed by IRS some reporting and processing errors were accepted under the administrative system. (The 193.2 million documents processed by the Service in Fiscal Year 1987 were about twice the number processed in Fiscal Year 1962, at the inception of ADP.) To conserve resources, such shortcuts in administrative processing had to be taken whenever possible. Because the two systems served such different purposes, it was not without reason that the SOI Division

Figure B.—Martinsburg Computing Center Processing



was reluctant to use administrative data.

Traditionally the emphasis in administrative processing has been on production. To meet production standards the expectation is that the reporting and processing errors that initially are allowed to "pass" will be caught later on if significant, through computer review. However, since this computer review generally addresses only data items necessary for tax administration (generally the return entries that have a direct bearing on the tax or on the payments reported or refund claimed), errors in other return items that are important for statistics often remain in the system. Although service centers often instituted some sort of verification system of their own design when resources were available, it has only been recently that IRS has begun to institute more of a balance between quantity and quality.

Despite these shortcomings in the adequacy of some of the administrative data for statistical use, efforts were made starting in the mid-1970's to utilize them for the SOI individual income tax return program, but only to a limited extent. Administrative data items that came from the same entries on the return as those required for SOI were manually verified or edited at the same time that additional data needed for the statistics, but not available from the administrative system, were abstracted and edited for SOI. (The principal economy achieved through this process was in not having to retranscribe the administrative data that were accepted for the statistics.) Then, toward the end of processing all of the transcribed data comprising a return record were brought together and tested for internal consistency at the Detroit computing center. Inconsistencies were resolved, either manually or by computer, on the basis of logical relationships among return items inasmuch as the returns were no longer accessible for statistical purposes at this stage.

The SOI system has since been modified. Formerly,

clerical tax examiners verified the administrative data for SOI and corrected or edited them as necessary. Now, all data items from the individual income tax return are "accepted" for the statistics as reported, unless the return fails a computerized check for acceptability which is made in the service centers. Return records failing this check are flagged for the specific item questionable and the return entry to be used for the statistics corrected or subjected to SOI editing while the return is still available.

This same computer-assisted editing process also identifies returns for which additional data need to be obtained for SOI only. This identification is based on the presence of entries in the administrative system for which the taxpayer has to file supplemental or computational information, often in a supporting schedule that is the source of the data needed for SOI. This means that, at least in the case of individual income tax returns, most of the computer checking for consistency is decentralized from the Detroit computing center to the service centers which are where the returns are processed, both for administrative and statistical purposes. It also means that some of the additional editing of the return totals that was formerly accomplished by recourse to supplemental data in supporting schedules may no longer be possible because the need for certain adjustments formerly made in the manual edit process cannot always be determined by computer [10].

For the sole proprietorship and partnership programs, a more comprehensive manual statistical edit, often involving many more items than are available from the Master File system, especially for farm data, is now being considered for certain future years. This special editing may coincide with the Agricultural and Economic Censuses for 1992, 1997, and so on. Special requests for data may be accommodated in a like manner. At the present time, the corporation program is the only major SOI program in which administrative data have not yet been used.

At least in the case of individuals, little appears to have been lost by the inclusion of administrative data, with no noticeable breaks in the historical time series apparent. This change in the approach to return editing would probably have been necessitated in any case by outside events; with the decline in statistical budgets throughout the Federal Government in recent years, more economical and efficient methods of obtaining data have become a necessity. At the same time, the drive to reduce respondent reporting burdens has also been made applicable to tax returns. As a result, some of the return form schedules used to facilitate SOI editing may no longer be required. Changes in the mode of taxpayer filing can also be expected to make changes in statistical processing, particularly in the data editing. This will result, for example, from the filing by computer of certain tax returns by paid tax preparers.

PRESENT STATUS OF SOI PROCESSING

How does SOI fit into the current IRS processing scheme? Figures C and D pick up where Figure B leaves off in showing the connection between the administrative and SOI systems. Figure E shows this relationship in more detail.

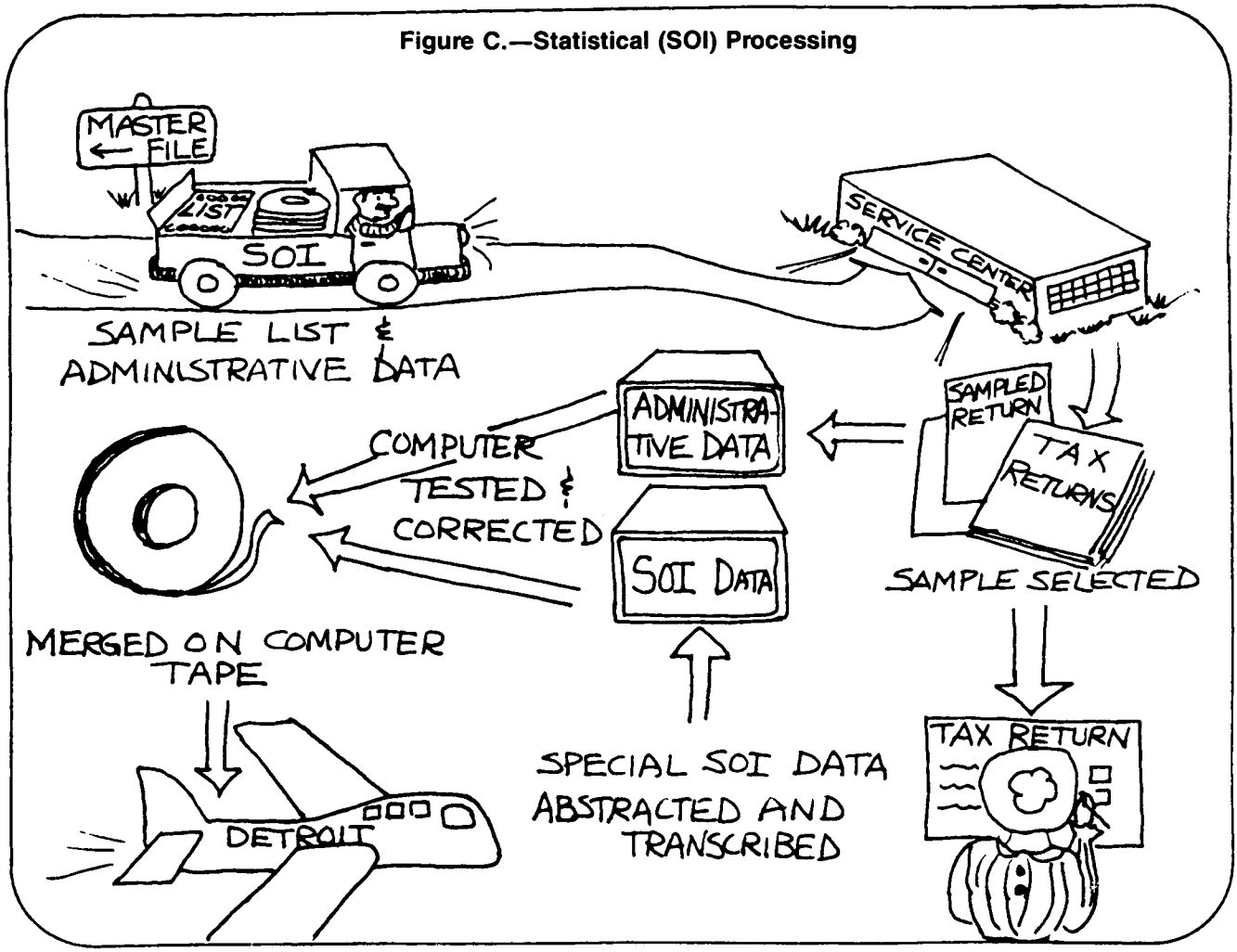
Magnetic tape extracts containing the identification of sampled returns are sent by MCC to each of the service centers where returns for SOI are selected from the files. The sampled returns in some cases may be processed for the statistics in the same service center in which taxpayers file them. Increasingly though, they are shipped to another service center or to the computing center in Detroit for processing. The exact locations chosen are dependent on prior-year experience, e.g., familiarity with the SOI programs and their processing requirements, the efficiency of a center's SOI operation and the quality of its work, and the priority it assigns the program and its ability to meet deadlines.

As already mentioned, SOI programs increasingly utilize data captured during administrative processing, and only the data not available from the administrative processing system are obtained directly from the returns. The latter data are either transcribed directly or are entered onto intermediate "edit sheets" for transcription at the service centers or Detroit computing center. The two sets of data are then merged.

Computer testing is now in two stages. In the first stage, the administrative data are checked to assure that, at least to start with, they can be reconciled with what is reported on the return. It is at this stage that most of the errors left uncorrected during administrative processing are caught. Then, after the complete record (including the additional data obtained solely for SOI) is on tape, it is sent to the Detroit center for the second stage of testing.

It is during this second testing stage that illogical relationships and internal inconsistencies are identified. Misreported or missing information may be imputed by the Detroit center, either manually or by computer (and the Statistics of Income Division sometimes has to make the final decision on how to deal with specific returns). In the case of errors, the output, in the form of hardcopy error registers or information listings, is sent by mail to the originating service center where it is associated with the actual returns. Return information is then used to help correct the SOI records by annotating the error or information registers. This is a "back and forth" process between service centers and the Detroit center until the file is considered error-free. After the second round of corrections is made by the service centers, any remaining errors are corrected at the Detroit center, without recourse to the returns, in order to save time. Having arrived at this point in

Figure C.—Statistical (SOI) Processing



processing, the files are tested for duplicates and other characteristics including those used to evaluate the completeness of the total sample. Weighting factors are then produced.

In addition to tabulations, either for publication or to satisfy special user needs, analytical tables are generated to assist in reviewing the aggregated data. Most of this review takes place in the Statistics of Income Division and takes into account how the data were processed as well as whether or not they are reasonable in light of prior-year data, changes in tax law and data from other statistical series. Disclosure in tables is dealt with mostly by computerized routines. Table programming is now done, not only at the Detroit computing center, but in the Statistics of Income Division or by outside contractors.

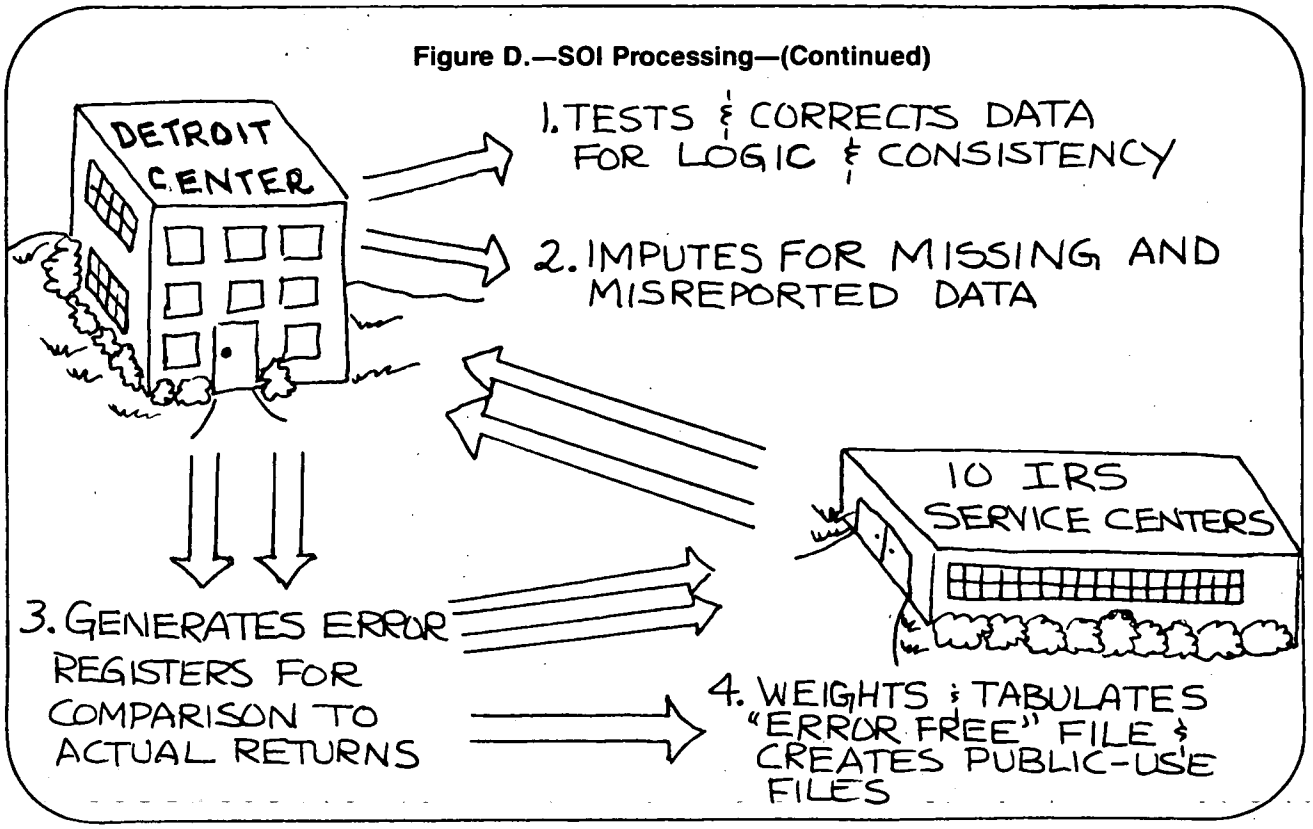
CURRENT PROBLEMS AND FUTURE PLANS

Even as SOI becomes more of a "natural" by-product of the administrative system there are a number of problems that need to be addressed. Some of them are described below [11].

In this period of continuing budget constraints, limited resources pose significant problems for SOI. The present SOI system has developed into one that utilizes whatever resources are available and thus it still relies heavily on manual processing. It does not take advantage of the interactive capabilities of computer systems because they have not been available. To help counter the cost of manual processing, the Statistics of Income Division has been attempting to adopt more sophisticated processing techniques, such as using specialized samples or using longitudinal files to assist in error resolution. However, this cannot always be done, in part because computer programming resources are not always available.

Because processing occurs at so many locations, there is much shipping of documents and data tapes. The monitoring of these shipments is time consuming. Significant delays in processing result from late, missing, or misrouted shipments. Timely project completion as well as the security of tax return information has sometimes been jeopardized by the controlling and shipping process. Moreover, these delays mean that the returns are not always available when needed for IRS compliance activities. Competition for re-

Figure D.—SOI Processing—(Continued)



turns is aggravated by the length of time it takes for error resolution. Service centers must sometimes hold returns until the Detroit computing center has processed two error resolution cycles. The competition sometimes necessitates photocopying. Most frequently, the competition is for the larger returns which are sampled for SOI at the 100-percent rate.

Control of the sample is also complicated by the process used to merge administrative and SOI data. When data obtained through the tax administration system are incorporated into a project, they are included in the same return record that is used for sample control. When this record is matched with the separate record containing the additional data obtained for SOI, problems can arise if there are transcription errors in the return identifiers used to effect the match. When data from the administrative system are not used in a project, sample control is entirely manual. In the absence of a computerized audit trail, there is always the additional problem of returns getting lost at some stage in processing.

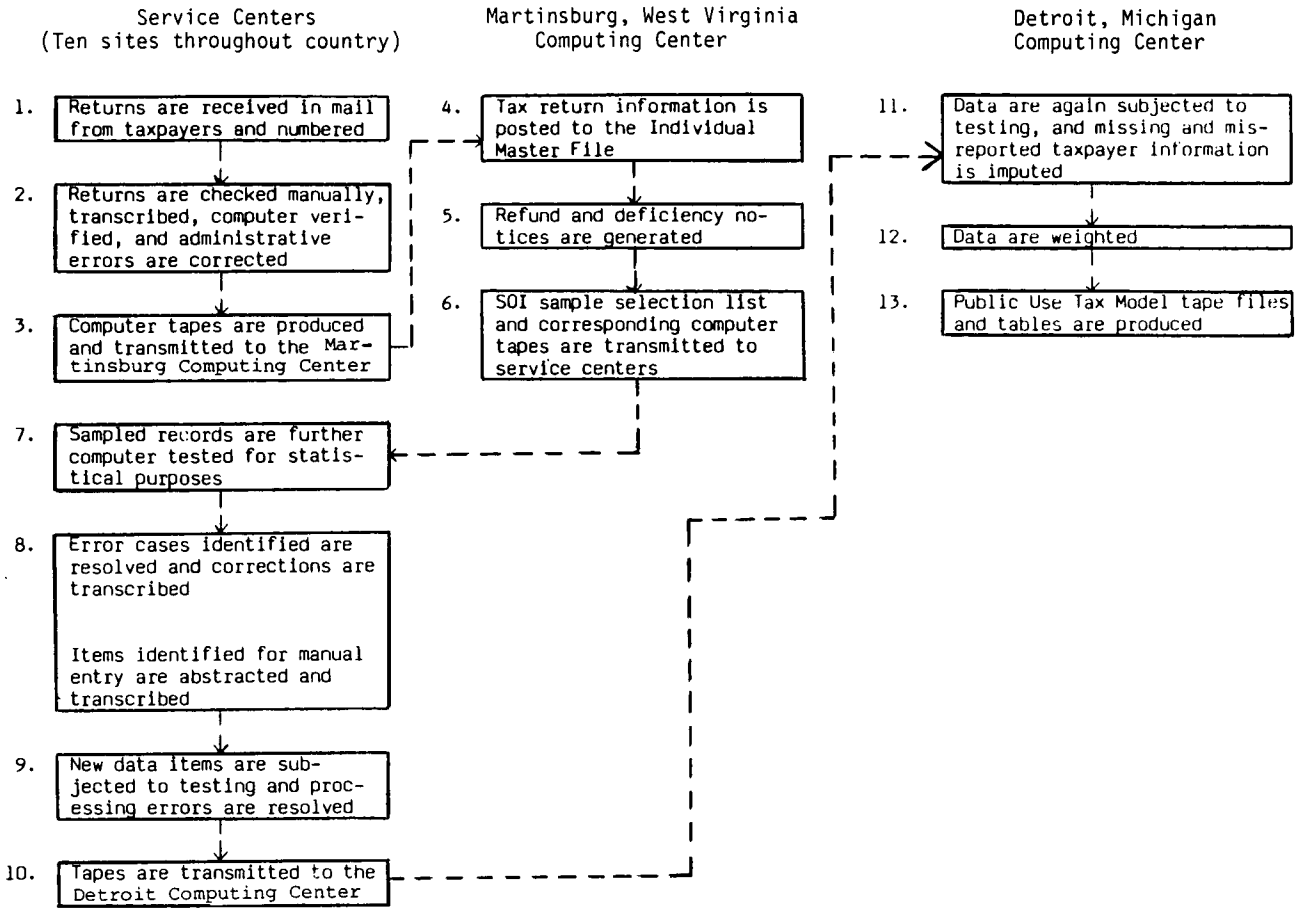
Although the basic SOI process is the one already described, there are many variations from project to project, especially in computer processing. There are two major projects—corporations and individuals—for which processing is complex and for which large computer files are necessary. The numerous smaller projects are forced to compete with these two major ones for available mainframe computer resources. As a result, many of the smaller projects have had to be assigned a lower priority and, therefore, are not always produced on as timely a basis.

The Statistics of Income Division is taking steps to deal with these and other problems that will improve SOI data and how they are produced. A mini/micro computer system in the service centers and in the Statistics of Income Division is being installed for processing smaller SOI projects [12]. This system will enable the Division to have more of a voice in setting priorities for the statistical operation in the service centers. It will also increase its ability to monitor the costs and timeliness of its products more effectively. The increased SOI processing at the service centers expected as a result will free up program development resources and computer time at the Detroit center for use on the two larger projects. It should also permit the release of SOI returns to compliance activities sooner.

Most of the savings expected through this system will be achieved by centralizing transcription and by combining data editing and testing into a single operation. As a result, the need for edit sheets and error registers will be eliminated. Since a sample return will be controlled only once, at the point it is processed, controlling costs will also be reduced.

Shipping of paper documents will also be curtailed, thus reducing time delays. While most of the processing is now

Figure E.—SOI Data Source and Processing at Various Locations (Individual Income Tax Returns)



being decentralized from the Detroit computing center to the service centers, not all of the service centers will be involved. When a center chosen is other than the one in which the returns were filed, the distance over which documents have to be shipped will be one of the factors taken into consideration in making the selection. With shipping reduced to an optimal level, security problems caused by loss of tax return data will also be reduced.

Under the new system, it will become possible for the first time to accumulate totals from the SOI returns processed and to screen the returns with particular characteristics at any point in time. This will make it easier to detect problem areas noticeable only from aggregated data at an early enough stage so that remedial action can be taken before the complete tabulations are run.

Sample monitoring should be improved and it will become possible to determine the number of missing returns at any time during processing. The earlier that missing returns can be identified, the better the chances are that they can be located before the sample has to be closed out.

The modern design of SOI processing systems for virtually all programs will incorporate some form of administrative data usage. With the recent redesign of the corporation SOI program to include use of administrative data for the first time, all major SOI programs will share the advantages of data abstracting by electronic means. This improvement from manual abstracting to electronic retrieval has changed the nature of SOI processing, which was once considered too complex to be connected with administrative processing, but which will now use this system as the starting point. With the marriage of SOI and administrative data, manual processing operations which previously were the costliest to perform will be dramatically reduced. Use of administrative data will reduce the amount of manual editing done, as well as the need for much of the transcribing and verifying of data.

In the case of corporations, the savings may not be as

dramatic as in other SOI programs because, unlike individuals, manual editing is expected to continue to be a major factor, especially for the larger returns where taxpayer reporting is more complicated and less standardized. While the administrative data will be checked by computer in any case before they are accepted (by the process already described for individuals) and changed when they are inconsistent with SOI definitions, in the future this may be done only for subsamples (in the case of the smaller corporation returns) so that imputation factors can be developed for the rest of the returns of similar size. A related cost-saving innovation may be the use of improved imputation techniques for erroneous or missing data [13].

In regard to quality, steps are being taken to ensure that every statistical study has a formal quality assurance program which provides for a review at each stage in processing. Because of the observed lapses in the quality of data during administrative processing, the Statistics of Income Division has also taken the initiative, in conjunction with other areas of the Service, to undertake an independent review of the quality of the administrative data transcribed from individual income tax returns. If the Division's recommendations are adopted, both IRS and SOI will stand to benefit.

Many other recent initiatives for streamlining SOI processing are ongoing or in developmental stages. A goal is to integrate SOI more fully into the administrative processing stream as that system is revised so that statistics can become more of an on-line IRS operation.

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. One approach being explored is through use of statistical matching with information from other sources. Another is to establish exchange agreements with other Federal agencies with regard to information furnished to them by the IRS under special provisions of the tax code (to the extent possible, given these agencies' own confidentiality rules).

If these efforts are successful, the new agreements would provide that IRS be able to receive back a copy of the information it furnishes which would also include any corrections, modifications, or enhancements; or the addition of any other information prepared by the other agency for inclusion in, or for use with, the IRS data. Even without the assistance of other agencies, much can be done to expand the SOI data base through linkages of records within IRS. For example, efforts are currently underway to relate estate tax returns to the returns of heirs for a specified number of years.

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 employees with the employer, taxpaying entity, "establishment," pension plan and payroll entity, all of which may have different definitions. Such linkages would encompass all types of employers—corporations, sole proprietorships and partnerships. A start in this direction has been the linkages made between business income tax returns and related employment tax returns, between partnerships and partners, and between certain small corporations and stockholders. In another research area, "panels" of data, representing returns of the same taxpayers over several-years time, are also being developed. These will enable the revenue estimators to make more accurate adjustments for the effects of "carrybacks" [14].

CONCLUSION

The 1980's has been a period of major changes in SOI processing methods. The emphasis during this time has been on finding ways to reduce costs and speed up processing.

There are obvious ways to streamline a program, such as cutting samples, projects and publications, but these are only part of the solution. Methodological and processing changes must continue to keep pace or even lead the way. The decisions to use administrative data, to adopt computerized testing of data while returns are still present, and to make more use of prior-year data in perfecting data for the current year are examples of the kinds of steps that are really needed. Some of these steps might have been introduced earlier, had the incentive to revise the SOI processing system been present. The budget reductions of recent years have provided that incentive.

While samples may be redesigned and better methods found to estimate totals, partly to help offset the reductions in samples necessitated in recent years, further reductions in the size of samples will now seriously compromise the reliability of the data. Therefore, future savings will have to come from continued efforts to develop more efficient methods of data processing. These efforts will enable the Statistics of Income Division to meet the needs for more statistical data that are expected over the next few years, while releasing the regular SOI reports and studies on a more timely basis. They should also enable the Division to devote resources to new areas of research and to satisfy the needs of its major users, while at the same time helping it adjust to any further budget reductions.

NOTES AND REFERENCES

[1] See *The Federal Statistical System, 1980 to 1985*, Committee on Government Operations, U.S. House of Representatives, 1984, a report prepared for the

Congressional Research Service, Library of Congress, by the Baseline Data Corporation.

[2] See section 6103(a), Internal Revenue Code of 1986.

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

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

While statistical editing is minimal in producing individual income tax return statistics (and, when it is necessary, can often be accomplished through computerized imputations), it is a major factor in producing corporation income tax return statistics because of the complexity of many of the returns, particularly those of the larger corporations which dominate the statistics. The need for editing accounts for a large part of SOI processing costs.

[4] For a more detailed history of SOI, see U.S. Internal Revenue Service, Statistics Division, *Statistics of Income, 50th year, 1916–1965, Historical Summary; Individuals, Corporations, Partnerships*, 1965, unpublished. A more concise history is contained in U.S. Internal Revenue Service, Statistics of Income Division, *Statistics of Income Programs, 70th Anniversary Celebration, 1913–1983*, 1983, unpublished. For a history of the individual income tax return statistical program, see Paris, David and Hilgert, Cecelia, "70th Year of Individual Income and Tax Statistics, 1913–1982," *Statistics of Income Bulletin*, Winter 1983–84, Volume 3, Number 3. Other articles in this 75th anniversary commemorative issue should also be consulted.

[5] The adjustments necessary to conform corporate profits reported in Statistics of Income (SOI) with the coverage and definition of corporate profits in the National Income and Products Accounts (NIPA) are described in detail in U.S. Department of Commerce, Bureau of Economic Analysis, *Corporate Profits: Profits Before Tax, Profits Tax Liability, and Dividends*, 1985, Methodology Paper Series MP–2.

In brief, SOI data become available about 3 years after the income year to which they relate and this determines the approach used in preparing the NIPA estimates. Each July, the existing estimates for a year for which SOI data are newly available are replaced with estimates based on SOI. The major adjustments to SOI corporate profits data can be summarized as follows:

a. An allowance is added for the underreporting of corporate income disclosable by Internal Revenue Service (IRS) audit examination.

b. IRS deductions that are not elements of costs of current production are added back, e.g., depletion, amounts "expensed" currently that are in excess of depreciation for mining exploration and development costs and for "intangible drilling costs"; State and local profits tax accruals; and bad debt allowances in excess of actual losses.

c. Elements of costs of current production that are not IRS current deductions, are subtracted, specifically the costs of trading or issuing corporate securities.

d. Elements of domestic income from current production that are not IRS "income" are added, e.g., income of Federal Reserve Banks and other U.S. sponsored credit agencies; savings and loan association profits (SOI data for this industry are not used because of discontinuities in the treatment of bad debts); and income from credit unions (which are tax-exempt). (Note that SOI profits include tax-exempt interest on State and local Government obligations which is reported on the tax return and which is added to taxable net income for the profit statistics.)

e. Elements of IRS income that are not domestic income from current production are excluded, e.g., gain or loss from most sales of property; intercor-

porate dividends received from domestic corporations; and income or equities in foreign corporations and branches.

[6] U.S. Department of the Treasury, Internal Revenue Service, *Annual Report of the Commissioner of Internal Revenue*, various fiscal years, 1945–1960.

[7] Based on U.S. Internal Revenue Service, Returns Processing and Accounting Division, "Overview of Returns Processing in IRS Service Centers," 1984, unpublished.

[8] Returns in each service center are assembled and numbered in groups of 100 returns within each of the return processing categories used in administrative processing. Each grouping of returns is known as a "block."

[9] Initial plans to integate statistical processing into an enlarged administrative processing system in the so-called "System of the 70's" were never implemented because of budget constraints which prevented the Internal Revenue Service from acquiring the new computer system called for in the plan. Budget restrictions were further compounded by concerns in the U.S. Congress about the privacy of taxpayer information under the new system proposed.

[10] For example, without manually examining the individual income tax return, it is not possible to determine if the taxpayer has reported commission income as an item of miscellaneous income under "other income," instead of in salaries and wages (as the tax return instructions request). Formerly, during manual SOI processing, when commissions were identified by statistical editors on the return lines in support of total "other income", they were transferred to salaries and wages. Because the return entries in support of total "other income" are not transcribed for administrative processing, the most that can now be done under computer-assisted editing is to identify the returns with "other income" for manual review. However, because of the large number of returns with "other income" and the need to reduce costs, only those returns with a large amount reported for "other income" are now reviewed.

[11] For additional information about future plans, see, for example, Statistics of Income Division, U.S. Internal Revenue Service. (1985) *Proposed Multi-year Operating Plan, Statistics of Income Division, FY 1986–92*, Volume I, Basic Operating Plan, 1986, unpublished.

[12] The setting for this new system is described in U.S. Internal Revenue Service, Statistics of Income Division, "Information Systems Requirements Analysis Report: Statistics of Income Distributed Processing System," 1984, unpublished.

[13] For example, see Hinkins, Susan, "Matrix Sampling and the Effects of Using Hot Deck Imputation," *1986 Proceedings of the American Statistical Association, Section on Survey Research Methods.*

[14] The need for adjustments for carrybacks is discussed in the article by Ralph B. Bristol, Jr., "Tax Modelling and the Policy Environment of the 1990's", also contained in this issue of the *Statistics of Income Bulletin.*