
Statistics of Income Division's Uses of Administrative Business Tax Records: An Overview

Jeri Mulrow, Susan Hinkins, and Jonathan Shook, Internal Revenue Service

The Statistics of Income (SOI) Division is a statistical group within the Internal Revenue Service (IRS). The SOI Division takes annual samples of tax returns, cleans and checks the data, and provides population and subpopulation estimates of tax, income, and other financial items of interest to economists and policy makers. In this paper, we discuss only one of the SOI samples, the annual sample of corporate tax returns.

Before describing some specific aspects of the SOI environment, an overall picture of our current state can be described using the perspective of a dynamic system. The dynamic system model described in Kauffman's *The Origins of Order* is used to describe biological processes: the evolution of species, the development of proteins, etc. But it can also be loosely adapted as an analogy for the development or evolution of a sample design and information system.

Greatly simplified, this model begins with a space of possible characteristics. In our example, this would include all the sample design options, data collection methodologies, and estimation techniques that we can choose from. For each point in this space, a measure of its "fitness" is defined. In the case of the sample design and estimation, "fitness" could be a general measure of "usability," encompassing all the customers' needs: accessibility, efficiency, accuracy, cost, etc. This results in a "fitness landscape" over the space of possible options. This is usually a "rugged fitness landscape" as it has more than one local maximum; i.e. there is more than one set of viable combinations.

Naturally such a system will migrate towards the peak, the best choice. If it were a static world, the best design and product would be achieved and no further adjustments would be necessary.

But it is not a static world; the landscape changes over time. Properties of the population change; the users' needs change; budgets change. The computer capabilities and relative costs have certainly improved over the last 20 years and are still changing. There are advances in statistical techniques, in many cases directly related to the increased capacity for fast, inexpensive computing.

Therefore the relative fitness of a design will change over time, and it is necessary to constantly evaluate and make adjustments in order to stay at the top of the "fitness landscape." Usually these are incremental changes: increasing or decreasing sample sizes, changing strata definitions, modifying estimation techniques, etc.

Eventually it may happen that the rugged landscape changes so much that incremental changes are not enough. A "long jump across a rugged fitness landscape" is required; one needs to move to an entirely different set of options in order to make needed improvements in the "fitness" of the final product.

The SOI Division could be described as being in the process of making such a long jump, making more than incremental changes, in order to meet the changing needs of our users and taking advantage of changes in technology.

Since the landscape is constantly changing, the jump may not land at the top of the higher fitness peak. But it is necessary to make the jump in order to move in the direction of a significantly better product. Many people at SOI are involved in different aspects of this major shift. But the creative vision and the direction for this long jump are due to Fritz Scheuren.

In this paper, we give some historical background about the SOI environment, describe briefly the current corporate program, and give some examples of the direction we hope to move with a long jump.

■ Information From Administrative Data

The Statistics of Income (SOI) program of the IRS came into being soon after the adoption in 1913 of the Sixteenth Amendment to the Constitution and the subsequent enactment of the first modern income tax law, the Revenue Act of 1916. In that Act, Congress specifically called for the annual publication of statistics by stating that:

The preparation and publication of statistics reasonably available with respect to the operation of the

income-tax law and containing classifications of taxpayers and of income, the amounts allowed as deductions and exemptions, and any other facts deemed pertinent and valuable shall be made annually by the Commissioner of Internal Revenue, with the approval of the Secretary of the Treasury.

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. Consequently, the Statistics of Income Division has published annual tabulations and reports on the operation of the income tax laws for every year since 1916 (Scheuren & Petska, 1992).

The following sections describe the type of data available from corporate tax returns and some considerations for making this information more useable.

IRS Administrative Data

The SOI Division has access to two basic sets of administrative data within the IRS. The first and underlying basis of the IRS administrative data is a vast volume of documents supplied by taxpayers and IRS agents regarding tax liabilities and payments. In 1916, this was a large, but manageable, data set to work with directly. The early SOI reports were population summaries. Now the corporate population is over four million. The majority of this information is still being supplied on paper documents, which fill warehouses. Fortunately the use of electronic data transfer is increasing.

For the administration of tax collecting, the IRS also maintains more accessible, computerized data bases. A nationwide master file system, which has an account for every active taxpayer required to file U.S. income taxes, including a limited amount of information from every tax return filed with the IRS. The data primarily relate to tax consequences: taxes paid and taxes due, amendments to the tax return, refund requests, and information about audits. However, there is also some demographic data and descriptive information such as a change in filing status, which may include a merger or acquisition, a filing of bankruptcy or re-consolidation, address changes, and some information on parent companies and their subsidiary filers. Some of the information is more easily acces-

sible than others, and the SOI Division has not had access to or used much of the administrative data that is available. Just recently, efforts are being made to collect some of this additional information on the master file system based on requests from the Treasury Department.

The master file keeps accounts open for five years after the last activity. Therefore, there are varying amounts of information available for different taxpayers and companies. Companies which have had a lot of activity over the years may have accounts that have records dating back to the inception of the master file system. Others may only have the last five years worth of data or less depending on the date of incorporation. Thus, some longitudinality exists, but the structure is not complete or easily defined.

The master file is maintained basically for keeping track of taxes owed and taxes paid. It is not designed for providing the details, available from the tax return, that are of interest to economists and policy makers. For example, it does not contain balance sheet detail, all the tax credit items, or schedule components. Also, although the information on the master file is generally reliable, certain variables are less reliable than others. Data directly related to tax consequence are extremely reliable, such as, taxes paid. On the other hand, data not related to tax consequences may not be quite as good. Industry code is an example. It is a self-reported item which is on the master file but is not checked for accuracy or consistency from year to year.

The primary users of IRS tax data within the Federal government are the Office of Tax Analysis in the Department of the Treasury, the Bureau of Economic Analysis in the Department of the Commerce, and the Joint Committees on Taxation in the U.S. Congress. The master file system, although it has information on every taxpayer, does not have enough detail for the extensive economic modeling performed by these primary users. In order to provide the detail needed for their work, the SOI Division collects additional information from a sample of approximately 85,000 tax returns annually.

Over the nearly 80 years of SOI data collection, the population has changed dramatically, and with it, the

amount and types of data being collected. The following sections briefly describe two basic aspects of the sample data:

- ❑ What information is collected from each corporate return, and
- ❑ How many and what type of corporations are included.

Qualities of the Corporate Tax Return

In the early years, only a modest amount of economic and tax data was available on SOI's files. For example, the 1917 corporation file consisted of approximately 30 data items covering the population of corporation tax return filers. At that time, the collected data were predominantly used by Treasury Department officials to estimate revenue and to conduct research on tax policy, which are still two of the primary uses of the data.

By contrast, the modern day SOI corporation program collects a wealth of financial, economic, and tax data. For example, the 1991 corporate program collects approximately 1,800 data items from a sample of eight of the Forms 1120 Corporate income tax return types and 16 of the supporting forms and schedules. In addition, the SOI foreign studies program collects and disseminates data from three more specialized forms attached to the corporate return having to deal with foreign-owned corporations.

Depending on the specific Form 1120 corporate form type, the corporate tax return generally includes detailed schedules for the income statement, balance sheet, tax computation, cost of goods, dividends, reconciliation of book income to tax return income, and an analysis of retained earnings. Specialized corporation returns such as the Form 1120L, Life Insurance, and 1120-PC, Property and Casualty, also include business-specific schedules. For example, the Form 1120L provides schedules for such items as policyholder dividends, increase/decrease in reserves, and policy interest. Similarly, the Form 1120-PC provides schedules to compute premiums earned and losses incurred.

Beyond the data captured from the internal schedules of a Form 1120, many of the attached forms, used to calculate bottom line totals, are also collected. For example,

most of the data from the Schedule D, Capital Gains and Losses, used to calculate the amount of capital gains reported on the income statement, are collected. Similarly, most of the detailed data available on the Form 4626, Alternative Minimum Tax--Corporations, the Form 4562, Depreciation and Amortization, and the Form 3800, General Business Credit, are collected by SOI. See Department of the Treasury, Publication 16 (Pub. 16) for a complete listing of the forms, schedules, and number of data items abstracted for the 1991 program.

Much of this detailed corporate tax data, whether viewed in aggregate or in microdata form, is absolutely essential to assess the efficacy of current tax law and estimate the effects of proposed changes in the law. Additionally, economic and tax items such as corporate profits, total tax liability after credits, cash distributions, total depreciation, interest received, and interest paid are used extensively by the Commerce Department's Bureau of Economic Analysis in its National Income and Product Accounts.

Beyond economic and tax data, a limited amount of organizational information is available for collection. For example, information on consolidated/non-consolidated, initial or final filing, 50 percent or more foreign ownership, and number of shareholders are routinely collected. However, parent/subsidiary relationships and merger/reorganization activities are currently not well captured by SOI, although this information is of interest. Tax law requires the parent corporation to complete Form 851, Affiliations Schedule, identifying its affiliated corporations included in the consolidated return. However, large consolidated returns frequently include hundreds of subsidiaries, making the data capture particularly onerous. Consequently, Form 851 data are not currently collected by SOI. Reorganization, merger and acquisition data are frequently provided on attachments describing the taxpayer actions and the entities involved. As mentioned, SOI does capture the taxpayer-reported indication as to initial, final, or merging status of the filer. However, data are not currently collected from taxpayer-provided attachments as to the entities involved.

One of the most critical data items in the SOI corporate file is the industry code assigned to a given corporation return. The industrial classification used by SOI conforms to the Enterprise Standard Industrial Classification (ESIC), which classifies companies, rather than

individual establishments. It follows closely the detailed Standard Industrial Manual (SIC) which is designed as a means of classifying establishments. Both, the ESIC and SIC are publications issued by the Office of Management and Budget. Some departures from the ESIC system were made by SOI for financial industries in order to reflect particular provisions of the Internal Revenue Code. See Department of the Treasury's Publication 647 for an example of the relationships between the SOI industry code and the SIC and ESIC codes.

Although SOI's classification is designed to apply to the company rather than the establishment, its application to corporation income tax return statistics has limitations. A return is classified by industry based on the activity accounting for the largest percentage of total receipts. This means, large corporations with diversified activities are included in only one industry, even though many of their business operations are unrelated to that industry. Consequently, statistics for an industry may be understated by amounts reported by corporations whose principal activity lies elsewhere, or overstated by amounts reported by corporations who have substantial business operations in other industries.

A final consideration concerning corporate tax return data is its comparability (or lack thereof) to publicly available sources. In general, the largest caveat must be that accounting requirements for tax purposes frequently differ significantly from Generally Accepted Accounting Principles (GAAP) employed for financial statement purposes. Differences in statutory and book accounting for items like depreciation, installment sales, and amortization of goodwill can lead to significant differences in reported profits. Additionally, for tax purposes, companies may file a consolidated return but in filing their financial statements, they may report their results separately. These data differences make any effort to replace missing tax data with publicly-available financial data, a somewhat dubious procedure. The SOI Division makes no attempt to reconcile these differences and collects only tax data.

Data Collection: From Census to Sample

Along with the increased complexity of the data, in terms of the number of different forms and the amount of information per return, the total number of returns has also increased dramatically since 1916. Up through the early 1950's the SOI statistics were based on a census of

all returns filed with IRS. Since then the statistics have been based on samples of returns. The change from a census to a sample might be considered a "long jump" rather than an incremental change.

The sample size has stayed relatively constant or decreased in size since 1951. The overall sampling rate has decreased from nearly 40% in 1951 to the current level of just over 2%. The sample size is approximately 85,000 returns out of a population of nearly 4 million active corporate returns.

The dramatic decrease in the relative size of the sample over time has caused the structure of the sample design to become more complex and to be increasingly dominated by the larger corporation tax returns. The population of corporation returns is highly skewed, with a relatively few large corporations accounting for a large percentage of the total money amounts. For example, in 1990 the largest 0.5% of the corporations contained 89% of Interest Income, 82% of Long Term Capital Gains, and 98% of the Foreign Tax Credit. Obviously these largest returns must be in the sample in order to make reasonable estimates of populations totals. Of the 85,000 sampled 1990 corporations, approximately 20,000 were these largest corporations which are selected with certainty.

An essentially fixed sample size, an increasing number of returns in the population, and the need to sample all "large" returns, result in dramatic reductions in the sampling rates for the smaller size classes. If SOI's users were only interested in estimates of annual, total amounts, this might not be a concern. But the users have other needs, including interests in subpopulations of smaller corporations. These needs must currently be met with the remaining 65,000 in the sample. Therefore, the sample design has become increasingly more complex over time. The sample design is currently a highly stratified probability design with 48 strata defined by tax form type and size of corporation. A description of the 1990 sample of corporate returns can be found in Publication 16.

Returns are initially selected for the SOI sample from the IRS master file system. They are obtained for statistical processing, where the data are put into an electronic format, cleaned, and checked for inconsistencies. In recent years, SOI has made tremendous production gains

in this area due to a modernization of the computer systems used to collect the data. Instead of weeks to obtain a clean record it, now may take less than an hour, depending on the size of the return.

Due to substantial penalties for misreporting, the detailed income and expenditure data on tax returns are generally regarded as more reliable than similar survey data. Even so, SOI goes to great lengths to protect against nonsampling errors, such as those due to taxpayer or data entry errors. Extensive on-line tests for consistency and reliability are made based on the structure of the tax law and the improbability of various data combinations. The SOI Division also has a rigorous quality review program.

Missing data are not much of a problem. Typically, less than one percent of the data are missing or inconsistent. Missing items can sometimes be obtained through telephone or written follow-ups, but more often imputation procedures are employed. For item imputation, prior year information is used, and variations on hot deck and nearest neighbor methodologies have been used with current data.

Longitudinal Data

Because of the interest in longitudinal studies and in improving estimates of change for various economic variables, SOI has designed the sampling selection process so that there is an overlap of sampled returns from one year to the next. To do this, SOI began using the Taxpayer Identification Number (TIN), an individual's Social Security Number (SSN) or a corporation's Employer Identification Number (EIN), as a basis for sample selection in 1968. This procedure allows for overlap of companies in year-to-year samples, while retaining randomness within a given year. It is not a panel per se, since SOI does not guarantee that any particular company will be in the file over several years. However, longitudinal data are available for many of the corporations in the sample, particularly the larger ones. The overlap in samples decreases over time, and the smaller a corporation is, the less likely SOI is to have information for it several years in a row.

The procedure using the Taxpayer's Identification Number has changed over the years. In the corporate program, from 1968 to 1978, random digits were se-

lected in specific positions of the EIN. The corporate sample was actually a cluster sample under this type of sampling procedure. However, there was a concern that the assignment of EINs to corporations might cause an appreciable intra-cluster correlation, since EIN's have some built-in structure to them.

Beginning in tax year 1978, SOI decided to use a transformation of the EIN rather than the EIN itself. This method was first proposed and studied at the Bureau of the Census by B.J. Tepping. The general formula for computing the transform is: $Y = c * X \pmod{p}$ where Y is the transformed number and equals the remainder when $c * X$ is divided by p ; X is the EIN; p is a large prime number; and c is a constant which is relatively prime to the number of subsets the population is partitioned into.

This transformation accomplishes two important purposes: (1) the transformed number is pseudo-random, and (2) the transform, corresponding to a given EIN, is always the same. The companies in the sample from a given stratum will be a random sample. If p and c remain constant over the years then (1) the sample is self-adjusting for births and deaths and (2) there will be a large overlap in the sampled returns from year to year. For example, beginning with the 65,469 unique, non-Forms 1120S returns in the 1987 sample, 36,002 are also present in the following three years' samples. To properly use these longitudinal data, however, one must take into account the sample design and the selection mechanism. Basically, the larger corporations and corporations that increase in size over time will be over-represented, and corporations that decrease will be under-represented.

This sample selection mechanism also provides a subset representing a simple random sample from the 1987 population, that should be selected in every following sample. This subset is defined by taking corporations with EIN's that would be selected at the lowest sampling rate. It can be used to estimate 'birth and death' rates for some classes of corporations. However, it is a small sample containing, 6,165 corporations with only 3,623 corporations with data for all 4 years.

■ Looking to the Future

Historically, SOI, like many other government statistical agencies, has maintained a strong descriptive or enu-

merative focus rather than an analytic or inferential one (Norwood, 1989.)

Changes in the "fitness landscape" have resulted in the need for SOI to consider major changes in the process and the product. More information and more accessible data are required in the following ways:

- More timely data
- Simpler data sets
- More available micro-data (masked to maintain confidentiality)
- More information.

In the following sections, we present some examples of major limitations of the data base. We outline the accessibility of the data and we include our current and future plans to provide better products for our users, by expanding our statistical, graphical, and processing tools.

Timeliness

Because the SOI corporate data are used to analyze the effects of current tax policy, to estimate effects of proposed policy changes, and to measure and analyze the U.S. economy, timeliness is important for the efficacy of these data. Unfortunately, timeliness is a problem with the administrative data based on tax returns.

For example, the 1991 corporate tax returns will generally be filed between June 1991 and March 1993. Most of the largest or most complex returns will be filed after September of 1992. Sample selection continues through June of 1993. After sample selection, there is still the process of retrieving selected tax returns, abstracting the data, checking, and cleaning the data file. In rare cases, the returns may not be available to SOI at all, because another branch of the IRS is using them or the corporation has been granted a special exception by the IRS due to special circumstances.

The data base for the 1991 corporate activity would typically not be complete until October 1993, and the tabulations would not be published until December 1993 or January 1994. The data collection process is being changed to dramatically shorten the time to enter and check the data, and provide more information with fewer errors. This should move the production of the final data

base and the final estimates up by four to six months over the existing system.

But this is not enough. The users need estimates on demand, or a continuum of data over time. This is our ultimate goal, which requires major changes in our sampling perspective and in estimation techniques. In particular, increased use of model-based estimates will be required. As a starting point, an incremental change was made beginning with the 1990 program; SOI is providing advance data estimates and an advance data file before the sample is complete. For the 1991 data, advance data were available by May 10, 1993.

Because the "late" returns are not like the "early" returns, the properties of the late returns need to be modeled. The first models used have been simple ratio adjustments, based on prior year results. But in the future we are considering estimating the propensity to be in the "early" sample versus "late," to be used to weight the advance sample.

Since the distribution of many of the economic and tax variables is extremely skewed, a very, very small number of the corporations account for a large percentage of the total amount, making these entities extremely influential to the data base and the tabulations. If even a few of these largest corporations are missing, the resulting tabulations, and possibly conclusions, may change dramatically. Some of these critical corporations will not be in an advance sample. Since these largest corporations are so influential and, at least for certain variables, so unstable from year to year, modeling these records for the advance data is not a reasonable option (Hinkins & Mulrow, 1992). Therefore, a small survey has been added to the administrative data base. For these critical corporations, if at the time of the advance data, these tax returns have not been filed or have not been available for statistical processing, a short questionnaire is sent directly to the corporation requesting information on approximately 20 tax items. In this way, at least some of the current information for these corporations is obtained and used in the file.

The move from providing only one final product to producing a continuum of data files and estimates, with associated measures of reliability or confidence, is a major component of our "long jump." It requires more than

incremental changes; it requires a shift in perspective and in the tools needed.

Simpler Data Sets

The fact that the sample design is fairly complex is both a strength and a weakness of the data. There are many advantages to the complex sample design. It generally accommodates the many needs of the primary users within a world of fixed and sometimes shrinking resources. It allows our primary users to make more accurate estimates and projections.

To correctly use the microdata resulting from such a complex sample requires a detailed understanding of the design, how it changes over time, and the mechanism for sample selection. For a less sophisticated or more casual user, i.e., one who is not intimately involved with the sample design and the data structure over the years, it is very difficult to correctly take advantage of this information. It is even likely that such a user will draw incorrect conclusions from the data because properties of the sample design were ignored. For example, as mentioned earlier, if the sample overlap from year to year is used to estimate change, the estimates will be biased unless the correct adjustments are made for the sample design and selection mechanism.

We cannot simplify the sample design, but we are considering methods of providing users with alternative data files that could be tailored for their specific needs and be equivalent to a simple random sample. This would make the information in the corporate sample more accessible.

Disclosure and Public Access

Finding ways to obtain wider public access, while protecting taxpayer's confidential information, is considered extremely important to SOI. Several outside users have expressed extreme interest in tax microdata for running economic models. However, tax return data are protected by law from public scrutiny, and strict procedures govern the handling of returns and computer tape files containing such information. Even after specific identifiers, such as name, address, and EIN are removed, the remaining tax return data may still be confidential. It may be possible to identify firms by linking taxpayer information with publicly available data; access to microdata such as Standard and Poor's COMPUSTAT and Dun and

Bradstreet data compounds the problem of releasing business tax data (Spruill, 1983). It is difficult to address disclosure concerns while preserving data utility (Greenberg, 1990).

Currently, SOI has a public-use file for individual tax data, but nothing comparable for the business tax data. However, a "blurred" microdata file for specific use by the Texas Legislative Budget Board has been developed by KPMG Peat Marwick in cooperation with SOI. The microdata file is actually a composite of information from 11 different data sources, including the SOI corporate sample. Although many disclosure concerns have been addressed in the blurring procedure, SOI is still investigating the data further. The question of data utility still remains. Univariate statistics have been preserved throughout the process, but no analysis of the underlying correlation structure has been undertaken. Thus, the blurring procedure is a good start towards producing a corporate microdata file, but still has a ways to go. The SOI Division will use this as a foundation for future projects and research in the area of developing a business public-use file.

More Information

For SOI, the focus has been to strictly gather information available on the tax return only. Not much emphasis has been put on trying to obtain information from other sources or to expand the data beyond the initial narrow bounds. For example, the SOI data are collected from pre-audited returns and no attempt is made to update the data with any further information the IRS may receive after the initial filing of the return. The data are just a snapshot in time. For some users, this is a limitation of the data.

Another limitation for our users is that the population unit of interest does not always correspond to the corporate unit represented by the tax return. Statistics Canada's work on a Business Register (Armstrong, 1990) shows both the great usefulness and importance of documenting the relationships between corporate units and also the difficulty of keeping track of these relationships.

This is a problem inherent in the tax return data available. If a tax return corresponds to more than one unit of interest, it is not necessarily possible to extract the information corresponding to the parts. It would be useful to

at least know something about these relationships and how they change over time. That is, the user would like to know that the 1989 tax return corresponds to corporate unit X and that the 1990 tax return corresponds to corporate units X and Y.

The plans for improving information of this type involve several processes. The corner stone of this process is being able to explicitly define the corporate units of interest and the rules for defining relationships between them. Once we have at least a simple set of definitions, the more straight forward part is to see what information is available for tracking these relationships. Finally, a system must be devised for using the available data to automatically track and make this information accessible.

We don't have the cornerstone yet, but we are working on it. The definition of the corporate units of interest (establishments, companies, corporations, etc.) depends on our users. The priority with which we work on this also depends to some extent on our users. We are currently discussing with economists just what kinds of definitions and relationships are of interest.

A Changing IRS Environment

Some fairly dramatic changes in the tax collection environment appear to be coming. Not surprisingly, some will aid us and some will make our task more difficult.

For example, there is a move to an increased use of electronic filing. This has begun in the Individual tax filing (Forms 1040's) and will expand to the corporate area sometime in the future. The advantages of electronic filing, both to the taxpayer and to IRS, are many and obvious. For SOI's collection of the administrative data, the advantages could be great. There should be more data available, sooner, with fewer errors. The information can be available to more than one IRS user, so there should no longer be a problem of SOI not being able to get the information because another part of IRS has the return.

But this system cannot eliminate all errors, and with electronic filing the taxpayer's original data are no longer available as a source to fall back on. If inconsistent information is found on a schedule it may be because the taxpayer made an error or because there was an error in transmission. Currently we can go back to the original tax return to check the information. In the future, this

may no longer be an option. Also, it will require more care and control so that data are not "electronically lost."

There may also be a movement afoot to reduce the amount of data that the taxpayer must supply. Instead of supplying all the information to substantiate the tax calculation, the taxpayer may only be required to initially send in the major items of concern. The taxpayer could be required to send in additional information after the initial information was reviewed. Obviously, this would seriously affect the type of administrative data available and necessitate SOI using a very different approach for collecting a sample of complete administrative data.

■ Summary and Acknowledgments

There have been significant changes in our "landscape," both in terms of our users' needs and in terms of available technology, making a "long jump" necessary. A common difficulty in such a situation is focusing all of the separate parts into a complete picture; seeing the need to make a major shift. The SOI Division is starting to see this need and lay the ground work for such a change.

Credit and praise should be given to those inside the Division who have recognized the need and are willing to meet the challenge of making a "long jump." Thanks should be given to those outside the division, our primary users, for supporting and encouraging our efforts in these times of change.

■ References

- Ahmed, Y., Scheuren, F. (1991). The U.S. Statistics of Income Program: Issues and Challenges, *Bulletin of the International Statistical Institute*, No. 48.
- Armstrong, G. (1990). An Overview of the Generation of Statistical Entities. Working paper, Statistics Canada, Ottawa, Canada.
- Dalenius, T. (1988). *Controlling Invasion of Privacy in Surveys*. Stockholm, Sweden: Statistics Sweden.
- Fellegi, I. (1987). Some Mathematical Statistical Research Problems in Statistics Canada. A paper presented at the Annual Meeting of the Statistics Society of Canada. Quebec, Canada.

- Greenberg, B. (1990). Disclosure Avoidance Research at the Census Bureau, Paper presented at the 1990 Annual Research Conference, Arlington, VA.
- Harte, J.M. (1986). Some Mathematical and Statistical Aspects of the Transformed Taxpayer Identification Number: A Sample Selection Tool Used at IRS, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 603-608.
- Hinkins, S., Scheuren, F. (1986). Hot Deck Imputation Procedures Applied to a Double Sampling Design, *Survey Methodology*, 12, 181-196.
- Hinkins, S., Scheuren, F. (1989). Evaluating Sample Design Modifications: Balancing Multiple Objectives, *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, 654-658.
- Hinkins, S., Mulrow, J., Collins, R. (1990). Design and Use of an Imbedded Panel in the SOI Corporate Sample, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 760-765.
- Hinkins, S., Mulrow, J. (1992). Preliminary Estimates from the 1990 SOI Corporate Sample, *Statistics of Income: Turning Administrative Systems into Information Systems*, 115-118.
- Jones, H., McMahon, P., (1984). Sampling Corporation Income Tax Returns for Statistics of Income, 1951 to Present, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 437-442.
- Kauffman, S.A. (1993). *The Origins of Order*. Oxford University Press.
- Mulrow, J. (1990). Description of the Sample and Limitations of the Data, *Statistics of Income 1987, Corporation Income Tax Returns*, Publ. 16, 9-15.
- Mulrow, J., Woodburn, L. (1990). An Investigation of Stratification Errors, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 754-759.
- Norwood, J.L. (1989). The Influence of Statistics on Public Policy. *Proceedings of the Symposium on Statistics in Science, Industry and Public Policy*, 30-42.
- Petska, T., Scheuren, F., Wilson, B. (1992). The Statistics Corner: The Statistics of Income Program of the Internal Revenue Service, *Business Economics*, 56-60.
- Petska, T., Scheuren, F. (1992). The Statistics Corner: Information From Tax Returns: The Statistics of Income Program of the Internal Revenue Service: Part 2, *Business Economics*, 66-69.
- Scheuren, F., Petska, T. (1993). Turning Administrative Systems into Information Systems, *Journal of Official Statistics*, Vol. 8, No. 4, 109-120.
- Spruill, N. (1983). The Confidentiality and Analytic Usefulness of Masked Business Microdata, *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 602-607.
- Sunter, A.B. (1987). Implicit Longitudinal Sampling from Administrative Files: A Useful Technique. *Journal of Official Statistics*, 2, 2.
- Tepping, B. J. (1969). Memo to Mr. J. F. Daly, dated January 15, 1969, U.S. Department of Commerce, Bureau of the Census.
- Triplett, J. (1991). The Federal Statistics System's Response to Emerging Data Needs. *Journal of Economic and Social Measurement*, 17, 3-4, 155-177.
- Westat, Inc. (1974). Results of a Study to Improve Sampling Efficiency of Statistics of Corporation Income. Bethesda, MD (unpublished technical report).
- U. S. Department of Treasury (1992). *Statistics of Income - 1990 Corporation Income Tax Returns*, Publication 16, Internal Revenue Service.
- U.S. Department of Treasury (1991). *A General Description of the Corporation Source Book*, Publication 647, Internal Revenue Service. ■