

Paul Rose and Linda Taylor, Internal Revenue Service

The Statistics of Income (SOI) Division of the Internal Revenue Service has been extracting data for many years from samples of income tax returns filed by various legal forms of organization, in particular, corporations, partnerships and sole proprietorships. The Office of Tax Analysis (OTA) of the U.S. Treasury Department is a major user of these data. OTA analyzes SOI data in order to evaluate the impact of changes in legislation on the deductions, tax credits, net profit and tax liability of businesses. In connection with this effort, OTA has a need to determine the growth of employment and payroll by size of business resulting from the implementation of tax-related legislation such as the Jobs Credit Program, which offers tax credit incentives to employers of newly-hired disadvantaged persons. Since the income tax return does not contain information on employment, IRS has turned to the use of employment tax returns in order to satisfy this need for data.

This present paper describes the plans and the work currently underway at IRS to develop a system for adding employment and payroll to its basic Statistics of Income files. The first part of the paper provides an overview of this system. The second part describes the SOI sample files and the files of employment tax returns that will be used. Following is a description of the procedures to be used in linking these files and plans for tabulation of data. Finally, the long-range plans for the development of linked files and data tabulation are briefly described.

#### OVERVIEW

The Statistics of Income samples of corporation, partnership and sole proprietorship returns are a rich source of data on business activity. They contain detailed income statement data for all three types of business organization and balance sheet data for partnerships and corporations. The statistics are classified by industry at the 2, 3, and sometimes 4-digit levels of the Standard Industrial Classification System [1].

Two of the three frequently used measures of the size of a business, receipts and assets, are contained in these files. They do not contain the third measure, employment, because this is not reported on business income tax returns. However, employment and also payroll are reported on the employment tax returns filed by these same businesses. In order to enhance the usefulness of the SOI files to researchers, the SOI Division has undertaken studies to add employment and payroll from the employment tax return to the files of business income tax returns. The income tax return files to be used in these studies are the Statistics of Income samples of corporations,

partnerships and sole proprietorships for tax year 1979. The figures for employment and payroll that will be added to these files are reported by the taxpayer on the Employer's Quarterly Federal Tax Return, Form 941 series, and Employer's Annual Return for Agricultural Employees, Form 943.

#### SOURCES OF THE DATA

The Statistics of Income samples that will be used in these studies and the source of data from employment tax returns are described below.

##### SOI Corporation Sample

About 80,000 corporation returns were selected from a population of approximately 2.6 million returns (Form 1120 series) filed with accounting periods ended from July 1979 through June 1980. Calendar year returns account for over 38 percent in number and over 79 percent in total assets. The sample was a stratified probability sample, selected at rates proportional to size, as measured by the higher of total assets or net income/deficit for broad industrial classifications. In addition, all life insurance companies, and most resident foreign corporations engaged in a trade or business in the United States were included.

The sample was designed to include all corporations with \$10 million or more in total assets, except for corporations in the financial industries where a minimum of \$25 million in total assets was required to assure selection.

Delinquent returns covering accounting periods which ended before July 1979 were accepted into the sample for all corporations except those of the larger returns. These delinquent returns, in theory, offset the returns with accounting periods ended July 1979 through June 1980 which were not timely filed and therefore were missing from the sample. Delinquent returns from the larger corporations (where both delinquent and current returns were present) were not included to avoid overstatement of assets [2].

##### SOI Partnership Sample

About 50,000 partnership returns were selected from a population of approximately 1.4 million returns (Forms 1065) filed during 1980. Over 96 percent of these returns had accounting periods for the calendar year ending in December 1979. The sample was a stratified probability sample selected at rates proportional to size, as measured by the higher of gross receipts or total income/deficit, and total assets. Separate sampling rates were designed for real estate operators apart from other partnerships. The sample was designed to

include all partnerships in which there were \$5 million or more in gross receipts, total income or deficit or total assets.

The incidence of prior year delinquent returns is also present in the partnership sample but to a greater degree than in the corporation sample because no prior year delinquent returns of partnerships were excluded [3].

#### SOI Sole Proprietorship Sample

About 92,300 sole proprietorships were included in a sample from a population of 11.5 million sole proprietorships. The population of individual tax return filers was divided into two parts (1) nonbusiness and farm proprietors without other businesses, and (2) other proprietors. These two parts were, in turn, stratified by sampling classes defined by state groupings and by either adjusted gross income (AGI) or the largest of certain income items, whichever was larger, and business receipts (farm receipts for farm proprietors without other businesses and business plus farm receipts for other proprietors). The sampling rates were proportional to the size, as measured by the three criteria above for defining the sampling class within each state group. The rates ranged from .05 percent to 100 percent. The sample was designed to select all of the following - (1) farm proprietors with at least \$200,000 in AGI and no tax, (2) proprietors with AGI or an income item of \$500,000 and over, (3) farm proprietors with \$5 million or more in gross farm receipts, and (4) other proprietors with \$5 million or more in gross receipts. Farm investors or landlords not materially participating in the business from which they received a rent were included in the sample. Businesses operated by estates or trusts were not included. While a return was selected for the sample based on the AGI and the highest income item reported on Form 1040 (for the individual) tax return, the business receipts were determined by the amounts reported on each business schedule (Schedule C for non-farm, Schedule F for farm) filed with the tax return [4].

#### Employment Tax Returns

Employment tax returns, Form 941 and 943 Series, provide for the reporting by employers of withheld income taxes and FICA (Social Security) taxes. They are filed by employers of all types, including partnerships, corporations, and sole proprietors. The Form 941 is filed by non-farm employers and covers a calendar quarter. The Form 943 is an annual return used to report agricultural employment and covers four calendar quarters. In both cases, the taxpayer is required to report the number of employees on the payroll for the week including March 12 of each calendar year.

Nonfarm employers report their payroll as "Total wages and tips subject to withholding, plus other compensation" on Form 941. Agricultural employers report their payroll as "Taxable cash wages paid during the year" on Form 943. These are the figures which will be

added to the SOI files.

Employment tax returns are filed by the employer/taxpayer at the ten IRS Service Centers. As with other returns, various checks are made on the validity of items reported on these returns to ensure that the tax data are accurate. With respect to the perfection of employment and total payroll, however, the checks are limited. This is because the withholding rates as a percentage of payroll vary with each employee, making it impossible to relate, with certainty, total withheld taxes to total payroll at the reporting unit level. Because of this, the IRS files of employment tax returns contain some missing and inaccurate data on employment and payrolls. For this reason, it was decided to look for an alternative source for these data.

The Bureau of the Census receives Form 941/943 tape files from IRS on a regular basis [5]. The amounts of payroll and employment on these files are subject to additional validation and imputation for missing and incorrect data. Because of the extent to which Census perfects these two figures, it was decided that the employment tax return data to be used in the IRS link studies would be that which Census produces. Therefore arrangements were made to obtain the employment tax return data from the Census Bureau for all Forms 941 and 943 filed during calendar years 1978, 1979 and 1980.

#### LINKING PROCEDURES

The Form 941/943 files will be linked to each of the three SOI files using the Employer Identification Number (EIN). The accuracy of the EIN's in each of these files is therefore a major factor affecting the ability to link the files. In addition, there are two other features of the files which complicate the linkage of the records and the tabulation of data from linked records. The first relates to the reporting periods of the tax return records. The employment tax returns relate to a calendar quarter or calendar year while the income tax returns frequently relate to a fiscal year period. As mentioned above, over 38 percent of the returns in the corporation SOI sample were filed for a calendar year accounting period. The remaining 62 percent were filed for a fiscal year period. For these returns, the employment tax return data will have to be adjusted to reflect the differences in accounting periods between the two types of tax returns.

The other complicating factor relates to the noncomparability of reporting units. The income tax return reporting unit is a company or composite of companies. Unlike income tax returns, there is no requirement that the employment tax return cover employment at the company level. Therefore, an employment tax return may relate to employment in an establishment (part of a company) or subsidiary, or a group of establishments or subsidiaries. This difference in levels of reporting will be especially critical in the corporation link where it will be necessary to search the Form 941 files for employment and payroll related to subsidiary corporations which are included in

the corporation income tax return.

The accounting period distribution of the returns in the partnership file covers tax years 1976-1980, but returns with accounting periods ending in 1976 and 1977 represent less than 4 percent of all returns in the file, with the vast majority having calendar year accounting periods ending in 1978, 1979 and 1980. In addition, according to the experience of the Census Bureau, partnerships for the most part are single-unit companies [6]. Therefore, the possibility of a particular partnership filing one income tax return and several employment tax returns, (which include only one establishment or group of establishments), does not exist to the same extent as it does in the corporate area.

The partnerships (Forms 1065) were selected as the first SOI file to be matched to the employment tax files because (1) there is less disparity in coverage between income tax and employment tax returns, compared to corporations, and (2) the EIN's which are the identification variables for linkage are always present on partnership returns in contrast to sole proprietorship returns where many EINs are missing.

As previously stated, files containing data from all Form 941/943 returns filed for calendar years 1978, 1979 and 1980 were obtained from the Census Bureau. Each record for a particular employer in these files contains the following: EIN; a record code indicating whether the record is from a non-farm (Form 941) or a farm (Form 943) employer; employment for the weeks including March 12, 1978, 1979 and 1980; and Total Compensation (payroll) for each calendar quarter of 1978, 1979 and 1980. The EIN on the partnership file (base file) will be used as the sole matching variable to select records from the Form 941/943 file (transactions file).

Although the exact match will be on EIN only, an attempt was made to evaluate whether the use of one variable is sufficient to minimize spurious matches and failure to match because of errors in that variable. The EIN on the Form 941 file was verified or supplied during processing by IRS and Census operations made no change to this number. Therefore, the EIN on the transaction file will be considered correct. The EIN on the base file, i.e., the partnership SOI file, was transcribed from the SOI returns and therefore was subject to error, despite the fact that transcription was verified. Although the SOI record was consistency tested, the EIN was not subject to testing for validity or accuracy and, in addition, there is the possibility that a correct EIN may have been inadvertently changed while correcting another item on the partnership SOI record.

In order to evaluate the quality of the EIN's in the SOI file, a small probability sample of 139 partnership tax returns filed for 1978 was used to test taxpayer entries as well as IRS coding [7]. In this study, the EIN's on the partnership sample returns were compared to those in the Social Security Administration's (SSA) records for the same partnerships. SSA's records were searched using the name and address shown on the partnership tax return.

The results were that of the 139 returns examined, 129 were matched exactly with SSA records. The remaining 10 were not found in the SSA records, but none of these returns reported payroll. The implication of these findings is that SOI partnership records generally reflect the correct EIN at least up to transcription. This does not mean that the SOI partnership file has a correct EIN for each record, but the need for a confirmatory variable is not as great as it would be in the absence of this information.

In any event, the Form 941 does not have another variable which would necessarily agree with a similar variable on the Form 1065 that could be used as a confirmatory variable. If necessary, payroll which appears on both records could be investigated. However, there is reason to believe that the amount of payroll in the SOI record will differ from that in the Form 941 file, even if the reporting unit is the same. This is because of the differences in accounting periods already mentioned and because of the way payroll expenses are reported on the partnership return. Although there is a line on the Form 1065 for reporting "salaries and wages" and "cost of labor" (the components of payroll in the SOI file) the expenses reported here are frequently not complete. To the extent that taxpayers include some payroll amounts with other line items such as "rent," "repairs" and "other deductions," the SOI payroll figures will be understated when compared with those on the Form 941. To the extent that the taxpayer includes the costs of "casual labor" or other payments to individuals who are not employees, the SOI payroll figures will be higher.

In order to use payroll as a confirmatory variable, some expected differences between Form 941 payroll and Form 1065 payroll would have to be computed and matching records would have to be rejected or accepted as correctly matched on the basis of the distribution of this expected difference. Although it will not take this form, there will be some verification of the matching which takes into account the amount of payroll reported on the SOI return.

As a first step in the linking process, the records in the partnership file were sequenced by EIN and accounting period. This was done in order to link the files and also to determine the number of returns in the partnership file with duplicate EIN's. Returns with duplicate EIN's occurred for various reasons. In some cases, returns were filed during 1980 for both the current year and a prior year. Such returns will have the same EIN, but a different accounting period. In a few cases, the partnership return is filed by an agent of the partnership, usually a lawyer or accountant, representing the partnership, and the EIN of the representative is used rather than the EIN of the partnership. If the agent files multiple Forms 1065 for different partnerships, returns with duplicate EIN's will result. In other cases, particularly for partnerships engaged in oil and gas exploration and underwriting syndicates, some partners, who may be partnerships themselves, are designated to file the tax return. Some of these may be in-

## FUTURE PLANS

involved in many ventures and if they use their own EIN, rather than the EIN of the syndicate, multiple returns with the same EIN will be filed. These can usually be identified by the industry code of the SOI return. In all cases, those records with duplicate EIN's were flagged and will be researched and processed separately during subsequent matching operations.

After the initial match of the Form 1065 and Form 941 files, a sample of matched cases and unmatched SOI returns with payroll will be further researched to determine whether or not a "true" match occurred and also reasons for nonmatching, respectively. Using the IRS Business Master File and/or the retrieval of the tax return itself, it will be determined whether or not the EIN on the SOI record is correct. If it is correct, the existence of a Form 941 filing requirement will be researched. Nonmatched records with correct EIN's and Form 941 filing requirements will indicate late or non-filing of Forms 941. For records with incorrect EIN's, an attempt will be made to obtain correct EIN's and reinput these records to a match against the Form 941/943 files. Spurious matches will be corrected, and non-matched income tax returns without employment indicators on their records will be purged from the files which are input to tabulations. Then, tabulations will be prepared from both the matched and the unmatched files. This will be done in order to analyze the joint distributions of business receipts, assets, SOI payroll, industry, and state so that an imputation strategy can be devised.

In order to produce final tabulations, the payroll data from the employment tax return will undergo further processing. The total annual payroll for the three calendar years will be calculated. In addition, in order to present a payroll figure that is comparable with the other entries on the SOI return, a payroll figure will be computed on the basis of the accounting period covered by the SOI return.

Once the linked file is produced, it will be an important source of data. Although the file will contain income tax return information for only one year, it will contain employment and payroll for three calendar years, thus providing levels and changes over time in employment and payroll by industry. It will also provide income statement and balance sheet data which can be classified by size of employment. This is significant because the time series data which are currently produced use receipts and assets as measures of business size and are subject to fluctuations resulting from price changes in an inflationary economy. The use of employment size precludes the necessity of using constant dollars to analyze growth, stagnation or decline. Supplementing the SOI partnership aggregate data with employment and payroll will also be a step toward fulfilling the Congressional mandate (P.L. 96-302 Title IV) that requires the Small Business Administration to build an external data base to evaluate public policy and economic trends affecting small business [8].

While work on the partnership link continues, research is underway on linking the employment tax return records to the SOI corporation and sole proprietorship samples described earlier. Although current plans call for publication of tables produced from the linked files, the first effort in each of these areas must be considered primarily a pilot or research project. It is anticipated that a system will eventually be developed to produce a linked file for corporations annually and in the other areas, possibly every five years.

A more complete version of this paper, which includes additional detail on all phases of the system including linking methodology and data tabulation, is available by writing to: Paul Rose, D:R:S:C, Internal Revenue Service, Statistics of Income Division, Corporation Statistics Branch, Washington, DC 20224.

## ACKNOWLEDGMENTS

The authors wish to acknowledge the contribution of Michael Howland, who assisted in the design of the data processing system used in the link studies. We also wish to thank Michelle Cobb and Nancy Robinson who assisted in the preparation and typing of this manuscript. Our appreciation also goes to the Small Business Administration and Dave Hirschberg, of that agency, for their continuous interest and support.

## NOTES AND REFERENCES

- [1] Standard Industrial Classification Manual, Enterprise Standard Industrial Classification Manual, Office of Management and Budget, Statistical Policy Division, Washington, DC, 1972.
- [2] For additional information, see also Statistics of Income, 1978-1979 Corporation Income Tax Returns, Internal Revenue Service, Washington, DC, 1982.
- [3] For additional information, see also Statistics of Income, 1979 Partnership Returns, Internal Revenue Service, Washington, DC, 1982.
- [4] For additional information, see also Statistics of Income, 1979 Sole Proprietorship Returns, Internal Revenue Service, Washington, DC, 1982.
- [5] The chief purpose in obtaining current employment and payroll data is to update Census' Industrial Directory and consequently any publications produced from this, e.g. County Business Patterns. See also:  
Farrell, Michael G. and Sullivan, John M. "Industrial Directory - Status and Direction," 1981 American Statistical

Association Proceedings, Section on Survey  
Research Methods.

- [6] Based on unpublished data from the Industrial Directory of the Census, May 1982.
- [7] Comparison of EIN and Business Activity Coding on 139 IRS Partnership Returns with Administrative Records at Social Security Administration. Unpublished paper.
- [8] For additional information on the development of the Small Business Data Base, see:

Kirchhoff, Bruce A. and Hirschberg, David A. "Small Business Data Base: Progress and Potential," 1981 American Statistical Association Proceedings, Section on Survey Research Methods, and

Hirschberg, David A. and Phillips, Bruce D. "Using Financial Statement Data to Evaluate the Status of Small Business," 1982 American Statistical Association Proceedings, Section on Survey Research Methods.

## PART E. DISCUSSION

## (Sources, Notes and References)

Selection Number (Page)	Title and Author(s)	Source*	Notes and References
2.13 (193)	Discussion, PETER DAVIS  Discussion (Excerpt), NEIL BARCLAY	<u>Statistics of In- come and Related Administrative Record Research, 1981</u>	Davis' comments refer to papers presented by Wilson and DiPaolo (2.01); Dumais and Shadid (2.03); and Powell and Stubbs (2.08). Barclay's comments relate to two of these same selections (2.01 and 2.03), and also to a paper presented by Durkin and Schwartz (2.05).

\*A full citation of this source is:

U.S. Department of the Treasury, Internal Revenue Service. Statistics of Income and Related Administrative Record Research, 1981.