Description of Tax-Exempt Organization Sampling

by Tamara L. Rib

This article describes sample design and sample selection, data capture and cleaning, the method of estimation, and the sampling variability of estimates for the various exempt organization (EO) studies conducted by the Statistics of Income (SOI) Division of the Internal Revenue Service.

Exempt Organization Study Domains

The SOI Division selects annual study samples of many types of tax-exempt organizations. Tax-exempt organizations that engage in charitable, educational, religious, and other nonprofit activities usually report financial and other required information on Form 990, *Return of Organization Exempt from Income Tax* (or the "short" Form 990-EZ), and Form 990-PF, *Return of Private Foundation or Section 4947(a)(1) Nonexempt Charitable Trust Treated as a Private Foundation.* Some of these nonprofit organizations must also file a Form 990-T, *Exempt Organization Business Income Tax Return* to report income from activities that are considered unrelated to their tax-exempt missions. This type of income is taxable; income from related activities is not. For Reporting Year 1998, SOI began selecting an annual study sample of Form 5227, *Split-Interest Trust Information Return*. This form is used to report financial activities of split-interest trusts.

For each SOI study year of the Form 990 series (990/EZ, 990-PF, 990-T), the sample frame consists of all returns "posted" to the IRS Business Master File (BMF) over a period of 2 years. Returns from the Form 990-series are filed over a period of 12 months. The requirement for when to file depends upon an organization's established annual accounting period; however, extensions of time to file can add several months to the required filing date. It is for this reason that Form 990-series returns are sampled over a 2-year period.

For the Form 5227, the sample frame consists of all returns posted to the IRS BMF over a period of only 1 year. All Forms 5227 are required to file by April of the year following the IRS reporting year, which is a calendar year for all filers. Even with extensions of time to file, a 1-year sampling period is sufficient.

Amended returns are not included in the EO studies because the original returns had already been subjected to sampling. Returns that change because of IRS examinations are also excluded.

Sample Design and Selection

The sample design is a stratified probability sample. For each study, stratification is carried out as follows:

1. The Form 990/990-EZ returns are stratified by organizations exempt under Internal Revenue Code (IRC) section 501(c)(3) or the IRC section 501(c)(4)-(9) group, and size of assets based on book value. The sample excludes those returns with gross receipts less than $25,001, the filing threshold. For Reporting Year 1996 only, organizations exempt under IRC section 501(c)(4)-(9) were excluded because of budgetary restrictions.
2. The Form 990-PF returns are stratified by organizations defined as a private foundation or a "nonexempt charitable trust" treated as a private foundation, and size of assets. For Reporting Year 1994 and later years, size of assets is based on fair market value. Prior to Reporting Year 1994, size of assets was based on book value.
3. The Form 990-T returns are stratified by size of gross unrelated business income (GUBI). Prior to Tax Year 1992, net income was used as the stratifier. The sample includes only those returns that had GUBI greater than or equal to $1,000, the filing threshold. Beginning with Tax Year 1993, joint strata samples from the Form 990-T and the Form 990 have been selected. The sample design is based on size of GUBI and the strata from the Form 990 when both a Form 990 and 990-T with the same Employer Identification Number (EIN) have been filed. An integrated match is done between the Form 990-T returns and the Form 990 returns. Sample strata are labeled with sample codes from both forms. This results in numerous sample strata, which are later consolidated into a smaller number of sample groups for weighting purposes. For Tax Year 1996 and later years, the Form 990-T sample only matched to the Form 990 IRC subsection (c)(3)s. After Tax Year 1995, Form 990 subsection (c)(4)-(9)'s were excluded.

Tamara L. Rib is a mathematical statistician for the Exempt Organization Studies for the Statistics of Income Division of the Internal Revenue Service

4. The Form 5227 returns are stratified by organizations defined as charitable remainder annuity trusts, charitable remainder unitrusts, lead trusts, or pooled income funds, and size of assets. For Reporting Year 1998 and 1999, only charitable remainder annuity trusts and charitable remainder unitrusts were sampled. Also, for those reporting years, the size of assets was based on fair market value. Review of this data proved that better estimates were achieved by using size of assets based on book value. For Reporting Year 2000 and later years, size of assets are based on book value only, unless book value assets are less than 10 million dollars and fair market value assets are greater than 50 million dollars.

Neyman allocation is used to assign the designated sample to the stratum. A Bernoulli sample is selected independently from each stratum with rates ranging from less than one percent for the small asset or income classes to 100 percent for the organizations with very large assets or income.

Each return posted to the BMF and belonging to the sample frame is assigned to a stratum and then subjected to sampling. The BMF-SOI "transform" method is based on the EIN assigned to each exempt organization. A unique random number, called the transformed taxpayer identification number (TTIN), is calculated for each EIN and then compared to a range of numbers based on a return's selection rate. If the number falls within that range, the return is selected. The formula that generates the TTIN does not change from year to year. Therefore, once an exempt organization is selected for the sample, it will be selected into the sample every subsequent year providing the organization's assets fall in a stratum with the same or higher rate.

Data Capture and Cleaning

When an exempt organization return is filed, selected data items are computer entered for administrative purposes to determine correct reporting information. All data are then transmitted to the IRS Business Master File (BMF) system. This database serves as the SOI sampling frame. Data are updated on a weekly basis to the BMF; therefore, the SOI sample is selected as the returns are being processed.

EO data are extracted directly from the tax or information return. The SOI data-capture process allows the data to be entered into a database where consistency tests and error resolution tests are run. Problems that arise include returns that are out of scope, sampled in error, or misclassified. The misclassified returns have SOI-generated sample codes different from the sample codes assigned during BMF processing. Tests also check for returns with data inconsistencies, such as items not adding to the correct totals.

Estimation

Each selected return is weighted to reflect the entire population. This weight is computed by dividing the population count of filed returns in a given stratum by the number of sample returns for that same stratum, after adjusting for rejected returns, missing returns, and outliers. A missing return may occur if some other IRS function retrieves the return, and it cannot be located by SOI after selection. The weights are used to produce aggregate estimates for items of interest such as assets, liabilities, revenue, and expenses.

Sampling Variability

Any sample selected for the EO studies is one of any number of samples that could have been selected using the same sample design. The estimates calculated from the different samples would vary. The standard error (SE) of the estimate is a measure of the variation among the estimates of all possible samples. The SE is usually expressed as a percent of the value being estimated. This ratio is called the coefficient of variation (CV), which is the SE divided by the estimate. The use of CV's is explained in the Appendix to this volume, "SOI Sampling Methodology and Data Limitations."

Source: IRS; Compendium of Studies of Tax-Exempt Organizations, 1989-1998; Volume 3, Publication 1416; August 2002