

ANALYSIS OF SOI CRITICAL CASE PROCEDURE

Nina Shumofsky and Paula Karvounis

Internal Revenue Service (R:S:C), P. O. Box 2608, Washington, DC 20013-2608

In July 1991, the Statistics of Income (SOI) Division of the Internal Revenue Service, published their annual data compilation entitled the *Corporation Source Book of Statistics of Income* based on 1988 tax data. The volume provides detailed industry data on corporation income tax returns. Data are shown for assets, liabilities, income, deductions, tax liability, and other financial items by industry groups and by size of total assets. In the process of gathering the sample data, editing them for consistency and publishing the final statistics, the Statistics of Income staff employ many different research methods to look for inaccuracies. This paper will focus on one method known as critical case research. After defining a critical case, this paper will explain the critical case procedure, demonstrate how it is statistically important to the final data and, finally, discuss improvement initiatives being explored.

DEFINING A CRITICAL CASE

In studying corporate tax data for statistical purposes¹, SOI uses a stratified sample of corporation tax returns. To ensure the highest amount of accuracy in the statistics, certain returns from this sample are labeled super-critical or critical to an industry, based on total asset size. These are returns which are so large that there cannot possibly be accurate statistics on an industry without their inclusion. For this reason SOI has to account for these returns.

The following specifications determine which returns are labeled super-critical or critical. A return is labeled super-critical when:

$$\frac{\text{total assets of the return}}{\text{total assets for its specific industry}} \geq 2\%$$

If this ratio is less than 2%, but meets the following industry-specific criteria, the critical label is applied:

- Form 1120S (Small Business Corporation) with total assets over \$50 million;
- Form 1120PC (Property & Casualty Company) or 1120 REIT (Real Estate Investment Trust), with total assets over \$100 million;
- Form 1120L (U.S. Life Insurance Company) or 1120 RIC (Regulated Investment Company) with total assets over \$500 million;

- Form 1120 or 1120F (Foreign Corporation), with total assets over \$500 million and classified as a bank or savings and loan; and
- All other Forms 1120 or 1120F with total assets over \$100 million.

TRACKING CRITICAL CASE RETURNS

The lists of companies designated as super-critical or critical are compared from one year's SOI file to the next. Then research is undertaken to determine why those returns are missing from the current-year file. This research is undertaken in the ten field service centers and the National Office in Washington, DC.

The critical case research process begins in the service centers. A major resource used by the service centers in locating critical cases is a database called the Integrated Data Retrieval System (IDRS). This system provides employees in the service centers with instantaneous access to the most current taxpayer information. The most significant information provided by the IDRS for critical case analysis is the "merge related transaction codes." These codes indicate whenever activities such as corporation mergers, acquisitions or status changes are recorded on the master file. This information can explain why a return is not included in the sample and if the reason is legitimate. Even when information is not available to the analysts, they may still be able to determine if a company has a filing requirement. A company may no longer be required to file a return if they are included as a subsidiary of a consolidated company, have declared bankruptcy or do not meet other filer requirements.

The critical case research process goes through three cycles between the service centers and the National Office over a seven month period, during which time information is continually updated. After service center researchers have learned all that is currently available to them regarding the returns listed on the critical case list, they forward this list to the National Office. Researchers in Washington, DC then verify and confirm this information by examining microfilm of the tax return. For example, if a return is thought to be filing as a subsidiary of another corporation, National Office staff would review the Affiliations Schedule of the parent company to confirm its filing status. If a company was thought to have liquidated, prior-year microfilm would be examined for a final year return

annotation. When the information is confirmed, that return is removed from the missing critical case list.

When a service center has not been able to get specific information on a missing return by the time the list has been sent to Washington, DC, the National Office staff then turns to other research tools not available at the service centers. The National Office maintains a microfilm database which catalogues returns on microfilm by name and employer identification number. As the system is cross-referenced by these two key identifiers, it can also be used for critical case research. Querying the system by employer identification number, for example, it can easily be determined if the company has undergone a name change. Similarly, an employer identification number change can be found by querying the system by name. In addition, reference volumes such as *Mergers & Acquisitions*, *Moody's*, and the *Directory of Corporate Affiliations* contain information regarding recent corporate status changes. The National Office staff also has access to prior-year microfilm of the tax returns, which may indicate a corporation's future plans.

After National Office research is completed, those companies which have not been removed from the critical case list are deemed missing returns. There are two categories of missing returns: those already selected for the sample and those not selected for the sample. If a return has been selected but is not available to SOI, it is considered missing for SOI purposes. These returns are usually delayed in standard IRS processing, such as correspondence, research or examination. When these missing returns later become available, the National Office staff may then add them to the final SOI sample.

The other category of missing returns are those not selected for the sample. Unlike the selected returns, these returns are not being used in the service center, but are missing for other reasons. Generally, the returns are missing because the company filed a late return, merged with another company, liquidated, is in unresolved tax return status, or changed in asset size. Occasionally a return goes through IRS processing but total assets are miskeyed, and it fails to meet the SOI sampling criteria. When a return is found to be miskeyed or mis-sampled, it is also added to the file.

ADJUSTING FOR MISSING RETURNS

Returns which are deemed missing are analyzed to assure that the sample is compensated for them. A tax return can be:

- Imputed (dummied), using a variation of

prior-year data.

- Added, if the return is located through research.
- Not included, under the assumption that it is filing as a subsidiary of another company.

The dummied procedure involves imputing returns from the prior-year file for use in the current-year file. A return is only dummied if there is every reason to believe that the company still meets the filing requirement. Also, in order to reduce the amount of dummied records and, thus, less accurate data, only super-critical returns are dummied. In order to impute records for returns which have not been selected, a match will be made by employer identification number against the prior-year file. In creating records for selected returns, current-year corporate file information is used to impute missing data. If, after a return is dummied, information is found to support the fact that the company merged, liquidated, or should not be filing in this tax season, then the return is rejected from the file.

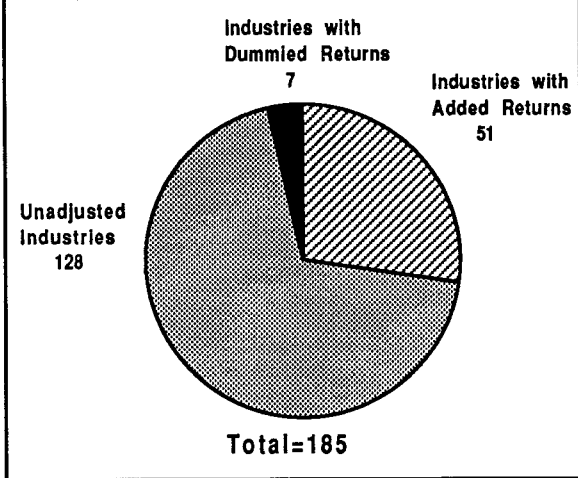
A second procedure used to adjust the file for missing returns is the added return process. If a return is located after the review process is completed at the service centers, the return is added to the file at the National Office. The procedure of adding a return to the sample depends on the type of the return, selected or non-selected, and how soon it is obtained before the end of the review process.

A third option is not adding or dummied a return, under the assumption that the company no longer meets the filing requirement. This decision is based on information obtained from the IDRS and other research tools discussed previously. It is important not to include these companies as that would result in inflation of the data.

EVALUATIVE RESEARCH

In order to determine the statistical importance of the critical case procedure, each industry that had an added and/or dummied return was examined to show the change in the data when the procedure was modified.² Of the 185 industries evaluated, there were 51 industries with added returns and 7 industries with dummied returns,³ as shown in Figure 1. Three alternative methods of compiling the data⁴ were examined. The added and/or dummied returns were removed from the data to show what the data would look like if only the added return process were used, if only the dummied process were employed, and finally if neither method were used. To represent the large number of items SOI abstracts from each tax return, the statistics of four

Figure 1: Breakdown of Industries



items have been selected to examine the impact of the process: total assets, net income, total receipts and total deductions.⁵

Figure 2 presents the data of Industry 2040 (Grain Mill Products) with and without critical case adjustments. This industry contains both added and dummied returns. The first bar shows the 1988 data using all current research procedures; total assets are 38.4 billion dollars. The second bar shows the modified data when only the added process is employed; total assets now are only 33.3 billion

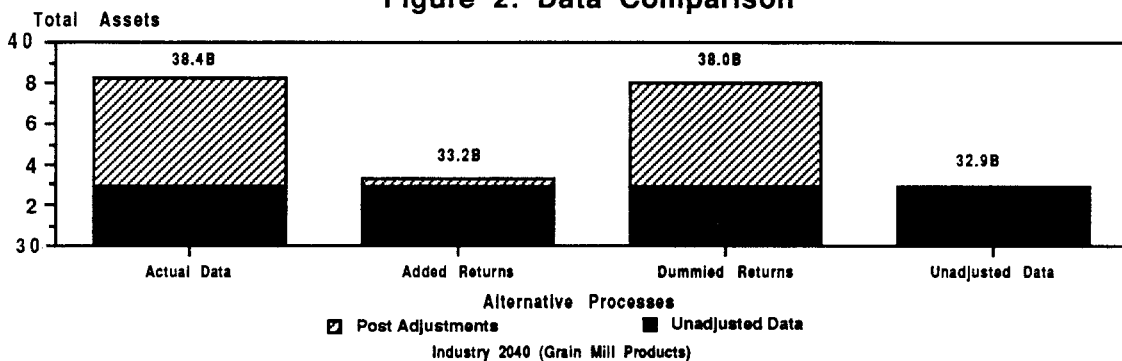
dollars--a difference of 5 billion dollars. Bar 3 presents the data including only the dummied returns; total assets are 38.0 billion dollars. Finally, Bar 4 depicts the altered data when no critical case research is done. Total assets are now only 32.9 billion dollars. Using this industry alone, it is apparent that the adjustment process vastly increases the estimates.

In Figure 2, it appears that the dummied process results in greater additions to the data; however, Figure 3, a comparison of dummied and added effects, illustrates the more global effects that the added returns have. Since the added returns involve many industries, their effect on the final data is greater. Conversely, the dummied returns have larger effects in their specific industries because only super-critical returns are dummied. As the graph shows, a total of \$183 billion dollars was added to the file through the added return process, while only \$11 billion dollars was added through the dummied procedure. The data differences in the analyzed 57 industries show the importance of including added and dummied records, thus producing more accurate data.

IMPROVEMENT INITIATIVES

Currently, there are many improvement initiatives being researched regarding the critical case procedure. It is apparent that SOI's critical case research results in a significant number of returns

Figure 2: Data Comparison

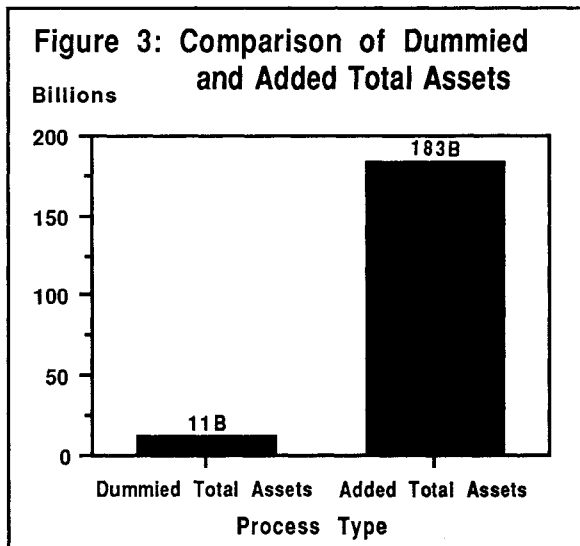


DATA ANALYSIS OF CRITICAL CASE METHODS

(money amounts in thousands)

	Data from added returns	% of data from added returns	Data from dummied returns*	% of data from dummied returns	1988 SOI data	Data using added returns only	Data using dummied returns only	Data with no critical case research
2040: Grain Mill Products								
Total Assets	361,099	0.94%	5,155,113	13.43%	38,395,214	33,240,101	38,034,115	32,879,002
Net Income	43,511	1.11%	249,025	6.38%	3,903,318	3,654,293	3,859,807	3,610,782
Total Receipts	443,729	0.69%	8,716,503	13.63%	63,928,488	55,211,965	63,484,759	54,768,256
Total Deductions	400,218	0.66%	8,384,751	13.90%	60,328,009	51,943,258	59,927,791	51,543,040

* Data in this column have been masked due to risk of disclosure.



added to the sample that would have otherwise been missed. In addition, the dummied process compensates for missing data of super-critical corporations by using the prior-year information. Although the current system improves the accuracy of SOI data, it is very labor intensive and time consuming. Hence, certain enhancements could be made to create an even better procedure.

1989 INITIATIVES

Research has been undertaken relating to the criteria employed to designate particular returns as critical to an industry. Since SOI resources are utilized to research each entry on the critical case list, it is important to analyze the number of cases labeled critical and the effect that the research has on the overall data. In reviewing the data from the 1988 file, some industries had nearly 50 percent of their returns designated as critical. This situation has arisen because the critical case criteria have remained static over the last several years, while corporate assets have risen over time. Although the criteria were appropriate when adopted, they have become outdated and no longer effective in determining which returns are the most important. It should be noted that, unlike the critical case criteria, the super-critical criteria are percentage-based, and thus, self-adjusting over time.

As a result of this study, the criteria for the 1989 critical cases have been changed. The asset size criteria in the finance and utility industries have been raised to reduce the number of returns which are labeled critical. It has been estimated that the change will reduce the critical case list from 6,751 returns to 5,046 returns, a reduction of 1,705 returns. Figure 4 illustrates the breakdown of this reduction by industry.

FIGURE 4.-- Critical Cases Eliminated under New Criteria

<u>Industry</u>	<u>Number of Returns</u>
Electric Services	32
Gas Production and Distribution	38
Combination Utility Services	15
Water Supply/Sanitary Services	16
Business Credit Institutions	31
Other Credit Agencies	73
Security Brokers, Dealers, Etc.	68
Commodity Contracts Brokers	23
Mutual Insurance	93
Other Insurance Companies	154
Real Estate Investment Trusts	54
Mutual Savings Bank	35
Bank Holding Companies	93
Savings and Loan Associations	151
Life Insurance	29
<u>Regulated Investment Companies</u>	<u>800</u>
Total Eliminated	1,705

The improved criteria are more industry-specific than those currently used, which only divide returns into financial and non-financial industries. This more specific categorization results in increased accuracy, because it eliminates the grouping of industries with differing characteristics. Thus, SOI can allocate more time to the industries where critical case research is necessary and save time by not researching industries where the accuracy will not be affected.

1990 INITIATIVES

A second initiative being investigated regards the 1990 SOI Program. New obstacles will be encountered as SOI converts its gathering of statistics from a manual to a computerized on-line system. This new system will enable SOI to obtain the corporate data seven months earlier, thus making the information available to the users on a more timely basis. The earlier date, however, will make even less data available for critical case review, due to circumstances such as taxpayer filing extensions. SOI procedures for obtaining data regarding companies whose returns are unavailable will become an even greater issue. The issue is already sensitive, as SOI has an overall commitment to publishing data as reported on tax returns, but at the same time, must realize that the information would be incomplete without certain compensations.

Thus, new alternatives for estimating corporate tax data and other sources of information are currently being researched. First, information could be gathered from sources other than tax returns, such

as annual reports, *Moody's* reference books, SEC 10-K reports,⁶ or direct inquiries to companies. Second, the current dummying method could be greatly improved upon by developing methods to try to predict current-year data rather than merely substituting prior-year data. These methods would include applying industry-wide trends and time series analysis to prior-year data to make it more accurate. Finally, rather than waiting until the return completes the standard IRS processing cycle, SOI may attempt to intercept the return as it goes through that process to obtain it earlier. After the return is located in the service center, a copy of the return, or at least pertinent parts of it, could be sent for SOI use.⁷ This process would enable SOI to include the maximum number of returns in the file when meeting the earlier deadline for sending preliminary information to the users.

SYNOPSIS

This paper has outlined and critiqued an important SOI procedure of compensating for missing data. With the major changes associated with the development of the computerized on-line system, this is an excellent time to explore new options. Criteria and time-scheduling changes are already underway and different ideas for the dummying process are being discussed. The purpose of this paper is to explain the critical case procedure and show the steps which are used to ensure the highest amount of accuracy in SOI data.

ACKNOWLEDGMENTS

Thanks are due to the following people in the Internal Revenue Service, who reviewed this paper and provided helpful comments: Fritz Scheuren, Karen Cys, Ken Szeflinski, and members of the Corporation Returns Analysis Staff. Additional thanks are also due to Beth Kilss and Wendy Alvey.

NOTES AND REFERENCES

- [1] It should be noted that this discussion confines itself to the selection of the sample for statistical research and not for audit or revenue processing purposes.
- [2] A compilation of the industries researched and analyzed can be found in the full report: Shumofsky, Nina and Karvounis, Paula "Analysis of SOI Critical Case Procedures," *Statistics of Income and Related Administrative Record Research: 1991*, Internal Revenue Service (Forthcoming).
- [3] It should be noted that only super-critical returns were subjected to the dummying process.
- [4] The data from this evaluation study were taken from Internal Revenue Services' Statistics of Income 1988 Corporation Data File.
- [5] A complete description of these items can be found in *Statistics of Income - 1988, Corporation Income Tax Returns*, Internal Revenue Service Publication 16, November 1991.
- [6] SEC 10-K reports are annual reports that all U.S. corporations must file with the Securities and Exchange Commission. Since 1974, corporations must release this report, which generally contains more information than an annual report, to any interested stockholders (information from *Banking Terminology*, American Bankers Association, 1985).
- [7] An Advanced Data Study is currently being undertaken in the SOI Corporation Branch to determine how to get return information earlier.