Corporation Supercritical Cases: How Do Imputed Returns on the Corporate File Compare to the Actual Returns?

Lucy Davitian, Internal Revenue Service

Statistics of Income (SOI) corporation "supercritical" cases are certain large corporations that SOI samples at the 100-percent rate. These supercritical cases account for 58 percent of the total assets of the corporation study while comprising only .03 percent of the total corporation returns; thus, their absence from the Corporation Study would affect the final statistics. Any unavailable returns must therefore be added to the file to protect the validity of the SOI Corporation Study. One method of adding those missing data is to collect the information through surveys sent directly to the corporations. Data collected are then used to create alternate records in the file through various imputation routines. These alternate records are later replaced with the actual return when that information is secured. This paper will give a brief overview of critical cases and the survey process, compare the data in the alternate records to that of the actual returns, evaluate the accuracy of the imputation routines, and make subsequent recommendations for changes to improve data quality where necessary.

▶ Background on Critical Cases

The critical case list for each program year is created based on the critical cases in the last two program years of the corporation study.[1] In general, there are three levels of critical case classifications: the top level, or supercritical cases, which are the largest corporations; critical cases that comprise 5 percent or more of the total assets of the industry they are classified in; and all other critical cases. The classifications are made based on three different criteria: type of return filed, industry classification, and corporation total assets.

During SOI's corporation Advance Data processing (beginning after the critical case list creation in December and running through April), all supercritical cases that are unavailable for statistical processing are searched for. Clerks at the IRS submission processing centers in Ogden and Cincinnati search for information on these critical cases. If the clerks cannot secure these returns, they provide information to assist National Office (N.O.) analysts with additional research. N.O. analysts then use this information to verify mergers between companies or other reasons why the return may be unavailable for SOI's processing.

Companies that are found to have no tax liability for the tax year, are liquidated or bankrupt, have changed Employer Identification Numbers (EIN's), or merged into other companies are suppressed from the study file and will not appear on future critical case lists. Between program years 1997 and 2002, an average of 85 supercritical cases were suppressed (see Table 1), thus reducing the number of critical cases that are researched or included in subsequent studies.

Table 1.--Number of Suppressed Critical Cases

Program Year	Total Super Criticals	Number Suppressed
1997	1,006	55
1998	1,160	70
1999	1,416	93
2000	1,622	95
2001	1,584	109
2002	1,595	85

However, if there is no evidence to conclude that a return does not have a filing requirement for the current tax year, and the returns are not located during this advance data period, alternate records, also called added records, are created as a substitute for the unavailable returns. There are four classifications of added records based on the type of information SOI has available to process the corporation return. The most ideal added record is one that uses data from both the IRS Business Master File (BMF)[2] and a survey sent to the corporation since it contains the most current information on the corporation return. The next level of preference is the use of BMF information only. Then, there are added records created using only survey information. Lastly,

records created based only on prior-year information are included when no other current information is sufficient to create the added record. For the purposes of this paper, only the added records created from survey information will be discussed and analyzed.

▶ Filling in for Missing Information: Overview of the Survey Process

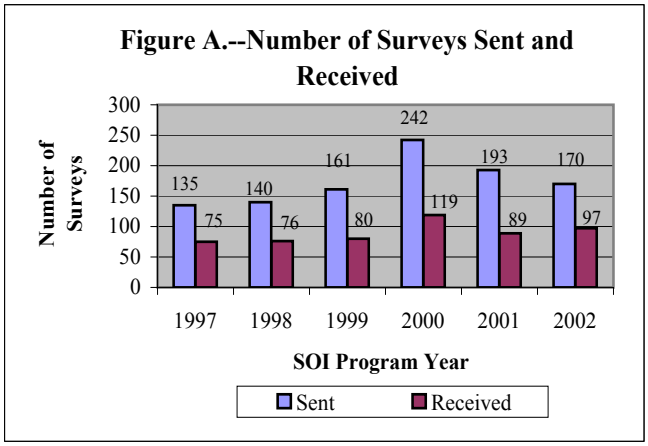
The surveys that are sent to missing corporations initially go through an approval process (renewed every 5 years) through the Office of Management and Budget (OMB). The approval process considers taxpayer burden in filling out and returning the survey, as well as other factors to ensure it meets established OMB guidelines. Once approved for distribution, the survey is sent with an accompanying memorandum signed by the Director of the Statistics of Income Division that states the nature of the survey and informs the corporations that the survey is voluntary. It also notes that the information collected is for statistical use only and not the result of any ongoing or forthcoming examination of the corporation's income tax return. The survey lists approximately sixteen data items from the corporation's tax return relevant to the SOI program year, and asks that the data be returned within 3 weeks of receipt.

Once a survey is returned, SOI processes the data to create an added record, also called a short-edit, in the file until the actual return can be processed. The survey data items are manually typed in, and the program then uses these numbers to calculate the remainder of the current-year amounts (those not included in the survey).[3] It does so by using current and prior-year amounts to create ratios that are used to help fill in for the missing data. The returns are then processed through the normal edit function used on all corporate returns to ensure that the total amounts balance and no additional errors are present. Returns created through this short-edit process are then given a weight and included in the study file.

After the close of the Advance Data file and throughout the remainder of the program year (for the 2002 program, file closeout was November 2004), these short-edits (and all types of added records) are replaced once the actual returns are available for SOI processing.

▶ Survey Statistics

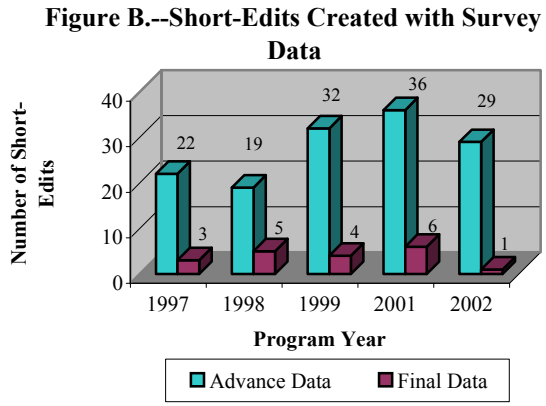
Since 1997, an average of 173 surveys have been sent each year to corporations, with average response rates of 51 percent (see Figure A). Over the course of the program years analyzed, many attempts were made to try to increase the response rates. For the 2000 program year, however, there was a higher number of unavailable returns. This was due to the IRS processing center realignments, which resulted in SOI's processing of corporate returns being scaled down from four centers to two. This also created some confusion and resulted in many corporate tax departments still mailing their returns to the same centers as in prior years. This caused a need for the returns to be shipped from these centers to the newly realigned ones. The changes in these processes and the delays they caused directly affected SOI's ability to process the returns for the Advance Data. For the 2001 study, to try to avoid a possible repeat of the prior year, the surveys were mailed earlier. Unfortunately, since many of the corporations were filing extensions, we did not receive as many surveys back until after the extension period was over. Also, in the wake of the September 11 attacks, longer extension periods were granted to corporations that were directly affected by the attacks, and many of these companies were either no longer in business or had portions of their businesses that were dissolved. Since some of the tax departments of these corporations were in New York City, the addresses that the surveys would normally be sent to were no longer valid. This directly attributed to the decline in the number of surveys sent, as well as the number of survey responses. In addition to these challenges with the earlier mailing, we observed the need to call more corporations to obtain the data; they had either misplaced the initial survey or were too busy at the time to fill it out within the 3-week timeframe mentioned in the memo. With that in mind, for the 2002 program, we mailed the surveys a few weeks later than we had for the 2001 study and noticed better response rates and fewer followup calls being necessary to secure the survey data, though, given the circumstances for the prior year files, we will need to evaluate this method further.



Each year, there is also an attempt to try to increase the number of survey responses and decrease the use of prior-year data. However, despite our efforts, there are still many instances of nonresponse. One reason is that the surveys are voluntary; many corporations do not return the data or do so weeks or months after the specified timeframe. Even though the survey states it has nothing to do with an ongoing or forthcoming investigation of the return, many corporate tax departments are hesitant to submit data that might catch someone's attention--especially if they do not have to. In such nonresponse cases, we attempt to contact the company's tax department directly to see if we can obtain the information we need. This usually causes the corporation to question the need for filling out a survey when it has already filed a return. We explain why the survey is necessary, and that the Statistics of Income Division, while under the IRS, is a statistical organization that uses the data for statistical purposes only and obtains the tax data after the other IRS processing functions. Another reason the survey may not be returned is due to various filing extensions that many corporations file. Depending on the date of the closeout of the Advance Data file, the company might not have enough time to provide the data needed.

The response rates mentioned above also do not consider those corporations that were sent surveys but did not respond because the corporation filed as a subsidiary of another; there are times that our initial research either does not provide all the information about the corporation or it does so after we have already mailed out the survey. In addition, given the time it takes between when the survey is mailed and returned to SOI, the return may have been selected for processing during subsequent selection cycles and edited before imputation of the survey data is necessary. In such cases, we make no attempt to contact the corporation in nonresponse cases and if the taxpayer calls to ask about the survey, we inform them that the survey information is no longer needed.

Between SOI Program Years 1997 and 2002, of the surveys received, an average of 28 (about 30 percent of all added records) were used in the Advance Data file (see Figure B).[4] By the end of the Final Data closeout, only an average of 4 remained in the file (19 percent of all added records), the others having been replaced with the actual returns.



▶ Comparisons of Survey Data to Edited Returns

During Advance Data, the short-edit records accounted for 0.6 percent of the total assets for all corporations in the study file, nearly $288.7 billion. In addition, all added records comprised 2.7 percent of total assets, or $1.4 trillion. While the percentages themselves are small, we can see that the missing data could potentially grossly underestimate the total assets in the overall file as well as all the other data items that are collected. To further examine the impact of these variances and see which schedules and forms needed further review, a sample of 50 returns were used to evaluate the trends within the data.[5] Fields with discrepancies between the added record and actual return were reviewed using a number of different criteria.

Data were first researched by comparing the added record to the actual return for the year studied to view the overall trends within the data. This was then broken into two categories--data that were collected directly from the taxpayer survey, and data that were imputed using the prior-year ratio amount.

Table 2 shows that data items created directly from the information provided by the taxpayer on the survey exhibited little to no change between the added record and the actual return. These small variances may be attributed to differences in taxpayer reporting on the survey and the actual return filed or minor differences in SOI processing of these data items.

Data items for the fields created using the ratio calculations, as exhibited in Table 3, however, showed a much different picture. The largest percent changes were concentrated in the dividends schedule. Using 2002 as an example, for this schedule, dividends from domestic corporations on the added records were $148.3 million compared to $0.06 million on the actual returns. This is due to SOI's processing for statistical information purposes where dividend distributions among member corporations electing to file a consolidated return were eliminated from the statistics as part of the consolidated reporting of tax accounts.[6] The data item, "dividends received deduction," also exhibited similar changes between the added records and actual returns, decreasing from $129.9 million to $0.04 million on the actual returns filed. This schedule will need additional review to compensate for these large differences so that amounts imputed on this schedule will more closely match those following SOI's processing of the actual return.

The remaining majority of data items with variances were scattered throughout all parts of the return, and most did not show significant changes between the actual and imputed returns. Many changes, like those on the balance sheet and income and deduction statement of the returns were more susceptible to variances in general. Since the imputations are based on the current-year totals and prior-year data, highly variable data fields like "cash" and "accounts payable" on the balance sheet and "deduction for bad debts" on the deduction statement were susceptible to higher variances from one year to the next. These imputations were not made based on corporation behavior, and, as such, large accounts payable or receivables, etc. in one year can have an impact (which subsequently disappear once the actual return is filed) on the imputed data items on the added records.

In addition to the above criteria, return types were also evaluated to observe whether a particular return type was susceptible to larger variances. It was observed that, while the type of return filed may contribute to the overall number of variances (especially for larger, more complicated returns), it is not a good indicator of whether or not a data item will change from year to year nor is it a good predictor of trends within the data.

Lastly, companies in the file as added records over multiple years were evaluated to see if they showed distinct trends for the data variation from year to year, and also to see if any one company was driving the changes. For these evaluations, the corporations showed no distinct trends beyond what was observed for the overall sample, other than showing that the same data items changed from year to year.

▶ Conclusion and Plans for Future Research

Critical cases are an integral part of the corporation study and, in some cases, necessary for the statistical validity of the file. This is why studying the alternate records is imperative to ensuring a complete and accurate program file. Reviewing the short-edit records showed the need for further analysis of these returns. While the variances in general are not unreasonably large, there are still some very large changes noticed within the data that could potentially have an impact on the overall corporation file.

The dividends schedule, in particular, is an area that will require further examination for future program years. For the time being, this may involve the manual editing and review of this field by the analyst in charge of the critical case program until additional line items may be added through the OMB authorization process. Once the process is in place for adding the necessary data items, adjustments can be made to the program where necessary to account for the data on this schedule and further improve the data quality.

Table 2.--Selected Items, Tax Years 1997- 2002: Corporation Super Critical Case Short-Edits Compared to the Actual Returns.*

[All figures are averages based on samples-money amounts are in thousands of dollars.]

	Program Years														
	2002			2001			1999			1998			1997		
	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change
Total assets	12,612,889	12,611,857	0.0%	9,893,770	9,893,770	0.0%	32,208,388	32,208,181	0.0%	9,658,437	10,258,437	-6.2%	6,471,497	6,471,497	0.0%
Total receipts	2,750,783	3,019,442	-9.8%	1,921,713	2,082,870	-8.4%	6,799,734	6,370,017	6.3%	3,935,096	4,015,147	-2.0%	2,359,071	2,424,352	-2.8%
Interest	148,084	148,084	0.0%	188,489	188,494	0.0%	595,166	595,152	0.0%	96,599	96,599	0.0%	849,567	843,248	0.7%
Interest on Government Obligations	3,093	3,093	0.0%	1,199	1,199	0.0%	4,215	4,215	0.0%	49,181	48,724	0.9%	28,119	28,119	0.0%
Net gain, noncapital assets	7,980	7,980	0.0%	2,099	2,098	0.0%	0	0	0.0%	6,639	6,639	0.0%	0	0	0.0%
Total deductions	2,555,159	2,579,181	-0.9%	1,671,554	1,663,048	0.5%	5,501,018	5,531,349	-0.6%	3,590,440	3,579,160	0.3%	106,264	77,316	27.2%
Cost of goods sold	1,381,860	1,433,028	-3.7%	630,873	673,015	-6.7%	2,295,827	2,179,638	5.1%	1,153,308	1,162,667	-0.8%	2,079,082	2,076,564	0.1%
Interest paid	110,277	110,277	0.0%	116,161	116,161	0.0%	370,584	370,584	0.0%	120,066	120,066	0.0%	43,978	51,626	-17.4%
Depreciation	76,343	76,343	0.0%	54,241	53,829	0.8%	176,781	178,019	-0.7%	111,781	112,430	-0.6%	36,049	14,009	61.1%
Net loss, noncapital assets	-3,353	-3,353	0.0%	-5,362	-5,362	0.0%	-7582	-7582	0.0%	-718	-718	0.0%	34,716	32,945	5.1%
Income subject to tax	145,777	145,777	0.0%	185,620	185,620	0.0%	586,081	586,081	0.0%	288,296	288,303	0.0%	3,290	40,085	-1118.3%
Total income tax after credits	42,116	42,026	0.2%	51,398	52,054	-1.3%	164,606	162,619	1.2%	84,146	83,572	0.7%	40,255	43,026	-6.9%

* There were no short-edit returns added for the Tax Year 2000 program. Data items shown here were items requested as part of the taxpayer survey. Averages were used in the table to protect taxpayer confidentiality.

Table 3.--Selected Imputed Items, Tax Years 1997- 2002: Corporation Super Critical Case Short-Edits Compared to the Actual Returns.*

[All figures are averages based on samples-money amounts are in thousands of dollars.]

	Program Years														
	2002			2001			1999			1998			1997		
	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change	Actual Returns	Short-Edits	Percent Change
Dividends, domestic corporations	63	148,336	-234850.8%	5	259	-4944.1%	169	365,338	-215979.3%	533	62,015	-11537.8%	4,669	59,650	-1177.6%
Dividends, foreign corporations	8,732	120,749	-1282.9%	13,270	38,221	-188.0%	188,174	822,175	-336.9%	26,670	51,516	-93.2%	1,530	11,407	-645.3%
Statutory special deductions, total	64,071	308,099	-380.9%	66,753	236,416	-254.2%	1,974,982	3,080,787	-56.0%	33,027	131,168	-297.2%	125,040	192,764	-54.2%
Net operating loss deduction	8,519	147,626	-1632.9%	28,274	158,824	-461.7%	45,955	529,579	-1052.4%	28,790	51,244	-78.0%	2,869	28,606	-897.1%
Dividends received deduction	44	121,899	-275721.7%	484	24,740	-5014.3%	1130.708	623310.668	-55025.7%	4,236	79,930	-1786.7%	19,993	62,054	-210.4%

* There were no short-edit returns added for the Tax Year 2000 program. Data items shown here were items imputed based on the taxpayer survey. Averages were used in the table to protect taxpayer confidentiality.

There are also a number of additional ways to evaluate and hopefully improve the imputation process and, thus, the resulting data that are produced. Such evaluations could decrease the time it takes N.O. staff to incorporate missing data, thereby freeing up resources that can be used on other projects.

One option to do so would be to compile ratios created as an average of the last few years of the return, and subsequently use those in conjunction with the amounts supplied by the taxpayer to create the remainder of the current-year amounts. This might decrease the effect of instances where a company has an unusually large amount one year--thus creating an extremely large ratio that is used to calculate the current-year amounts. Another would be to use the trend within the corporation's industry to calculate the ratios. This would allow the ratios to more closely mirror those of the entire industry and possibly decrease the chances of the corporation being an outlier within the industry.

If these comparisons are done for prior-year returns already in the program file, the accuracy of these proposed options could easily be tracked to determine which would be a more accurate way to add the data.

However, all evaluations aside, the ultimate goal in improving data quality is first and foremost to reduce the number of unavailable records during Advance Data. The lower the number of added records, the better the overall file will be during both phases of the Corporation studies.

► Acknowledgment

Thanks go to Patrice Treubert of the Corporation Research Section for her help in creating the SAS data sets that were used in the analysis of the data.

► Endnotes

1 As an example, for the Tax Year 2002 SOI corporation study, which included returns with accounting periods ending July 2002 through June 2003, the critical case list was finalized in December 2003 and was based on the critical cases in the Tax Year 2000 and 2001 corporation studies. If the returns met the critical case criteria for either of the two prior years, they were classified as critical cases for the 2002 study. Previous and subsequent years also incorporate the same principles for inclusion of returns in the sample.

2 All tax data and related information pertaining to individual business income taxpayers are posted to the IRS Business Masterfile (BMF) so that the file reflects a continuously updated and current record of each taxpayer's account. For additional information, please visit: http://www.irs.gov/privacy/article/0,,id=130752,00.html.

3 Items from the balance sheet are calculated differently than the remainder of the tax return. Balance sheet items use total assets to impute remaining data items based on ratios of the industry average.

4 There were no survey records added for the Tax Year 2000 program so that year was not counted in the survey data comparisons.

5 This sample represented 36 percent of all short-edits from Tax Years 1997-2002. Data were selected on a number of factors, mainly, the return type and number of times in the file as an added record. This was done to create a variety of evaluation criteria and ensure that other factors did not influence the data variations. Though the above criterion was used in gathering the sample of returns, the sample was not chosen with the name or size of the corporation as determining factors. The weights for these returns were all the same so that variances were not a result of weighting differences. However, we assumed that the data entered from these returns were free of editor error, that is, the N.O. and field editors entered the amounts in the system correctly for the returns they edited. Since the system is thoroughly tested before program implementation, it is assumed that the program is

also free of error and, therefore, did not contribute to variances in the data.

[6] For tax purposes, dividends reported on these returns represented amounts received from corporations that were outside the tax-defined affiliated group. See also section on Explanation of Terms, Internal Revenue Service, Statistics of Income, *Corporation Income Tax Returns*, annual publications 1997-2002.