

INDIVIDUAL STATISTICS OF INCOME: ADVANCING THE CLOSEOUT DATE

Jim Dumais and Ray Shadid, Internal Revenue Service

This paper reports on the results of research done in the IRS Statistics Division exploring various alternatives for streamlining processing of and providing earlier estimates from the Statistics of Income (SOI) sample of individual income tax returns. Organizationally, this paper is divided into 5 parts. Section 1 provides background on the current SOI processing system. In section 2 each of the proposed changes is discussed. The methodology employed is described in section 3. Results and recommendations, in section 4, are followed, in the fifth section, by an outline of future plans.

1. BACKGROUND

As part of the Statistics of Income program, individual income tax returns filed (Forms 1040 and 1040A and related forms and schedules) are sampled to produce aggregate estimates of taxpayers' income, exemptions, deductions, credits and tax. These estimates are published in an annual Internal Revenue Service report in the Statistics of Income series [1].

Under the current processing system, sample designation for a given program begins with the first week, or cycle, of the processing or calendar year (usually in January) and proceeds through the following December. After the returns for a given program are sampled, they are edited; consistency and validity checking are performed; any transcription errors detected are resolved and a "clean" file is produced. Weight factors are calculated and applied; finally, tabulations are produced and the annual report, Statistics of Income--Individual Income Tax Returns [e.g., 9], is developed and issued.

In addition to the basic SOI program, the Treasury Department's Office of Tax Analysis (OTA) requires estimates of income and tax liability from Forms 1040 and 1040A, filed during the year, by late November of that year. In order to provide these estimates, the IRS Statistics Division has traditionally created a preliminary (or advance data) SOI file using all sample returns processed at the ten IRS Service Centers through the first week of October. From this file of early sampled returns, "advance data" estimates are provided to OTA [e.g., 5]. Traditionally, additional tabulations have also been produced from this file and the report, Preliminary Statistics of Income--Individual Income Tax Returns was issued [e.g., 7]. The preliminary reports have recently been replaced by the quarterly Statistics of Income Bulletin [e.g., 8].

As a result of budget constraints and requests for earlier release of SOI data [1], new concepts in SOI design and processing are being explored. Three specific issues or concepts are discussed here: advancing by two weeks (to mid-September) the sampling and processing

cut-off date for the preliminary SOI file; changing processing at the Service Center level to make sample counts more nearly equal to designation counts for the advance data cut-off; and a proposal for radically different treatment of prior-year returns in the SOI files.

The primary data base used for this research and testing was the Internal Revenue Service Individual Tax Model file for Tax Year 1978 [6]. The tax model is a micro-data file comprised of an abbreviated version of each of the sample return records included in the 1978 SOI file that was used to produce the complete report for 1978. The tax year 1978 sample of 157,518 return records was weighted (by IRS District and sample code) to an estimated population of 89,771,551 Forms 1040 and 1040A returns filed during calendar year 1979.

In order to evaluate the results of testing the proposed modifications, this paper presents a comparison of a full simulation of the 1978 advance data tabulations (incorporating all of the proposed changes) with the 1978 complete SOI estimates and with the actual 1978 advance data tabulations transmitted to the Office of Tax Analysis. Results of the simulation will be explored in detail following a discussion of each of the proposed changes to the current SOI design and processing system.

2. PROPOSED CHANGES

Earlier Advance Data Cut-off--Accelerating the preliminary SOI sampling and processing cut-off by two weeks is the first issue to be explored. The obvious criticisms of this proposal are: (A) that estimates will be based on about 1,500 fewer sample returns, and (B) that the returns not sampled tend to differ from earlier-filed returns.

Returns filed in late September and early October (as well as later-filed returns) exhibit different characteristics than those filed earlier: income amounts (positive or negative) tend to be larger. In Table A we highlight the average adjusted gross income (AGI) of \$14,457 on early-filed returns and \$22,306 on returns filed in late September (cycles 38 and 39), to illustrate this point. Returns also tend to be more complex the later they are filed. The level of complexity of various return categories can be implied from the data presented in Table 1. Late-filed returns exhibit higher relative incidences of filing on Form 1040, having itemized deductions or being classified as business returns than do earlier-filed returns.

In support of this proposal, it should be noted that the early cut-off will result in earlier release of SOI data. Also, adjustments are possible for the bias that would otherwise be introduced by simply cutting off earlier. [1]

The explanation of the methodology which appears in section 3, Simulation of 1978 Advance Data and Final Estimates, includes a discussion of the measures taken to test this proposed change as well as those that follow.

Improving Sample Counts at Advance Data

Cut-Off.--The reasons for the discrepancies between designation counts and actual sample counts at advance data cut-off can be summarized into two major categories: (A) the inability to associate the edit sheet with its return document for abstraction of additional data in time to meet the processing deadline for the early file, and (B) unresolved errors from Service Center level consistency and validity testing not corrected in time to meet the early deadline.

The category in Table A labelled "Returns Missing from Advance Data" presents a summary of the 518,157 such cases (weighted estimate) identified in the 1978 file. The distribution of these returns by size of income is comparable to that of late-filed returns and indicates that, although fewer in number, these cases are adequate substitutes for some of the sample returns excluded from advance data due to the earlier cut-off.

Over recent years, the system of transcription of data from the tax return to computer tape during the processing of returns for revenue purposes has expanded to the point where almost all the data items necessary to produce the advance data tabulations are available to the SOI program from the revenue processing computer system. In addition, the quality level for the aggregated totals of a number of the available items (such as the major sources of income, AGI, and tax) is comparable to the SOI quality level for those items.

Since all sampled returns, including those with errors detected or those that require editing for special studies, have sufficient data available on tape to produce the early estimates, all returns designated for the early SOI cut-off will be transmitted to the Detroit Data Center for extensive consistency testing, error resolution, and posting to the advance data SOI file.

Prior-Year Returns.--A prior-year return is defined as one filed for an income year earlier than that for which the majority of the tax returns are being filed. Most tax year 1978 returns were filed during 1979. Thus, returns filed in 1979 for tax years 1977 or earlier were classified as prior-year returns. We estimate from the 1978 complete SOI file that there were 1,045,897 prior-year returns filed in 1979 (1.2 percent of the total). Table A includes a brief distributional analysis of the prior-year returns included in the 1978 complete SOI file and Table 1 includes a characteristics analysis of these same returns (as well as other categories of returns).

Prior-year returns present two problems to the

SOI program. In the first place, prior-year returns require exception processing and testing because they relate to prior-years' tax laws. In the second place, prior-year returns are being tabulated with records for a tax period to which, it can be conceptually argued, they do not necessarily belong.[2,4]

The rationale for including prior-year returns in the current SOI year was that they were an acceptable substitute for current year returns yet to be filed. This made sense so long as inflation rates were low, and relatively few (or minor) year-to-year changes in tax law occurred.

Analysis now indicates that prior-year returns, as a group, tend to differ significantly from other returns from the tax year for which they were filed, and to differ from current year returns processed during the same filing year. In comparing prior-year returns with other returns from the tax year for which they were filed, we found that prior-year returns tend to have higher incomes and to be more complex. The second observation, that prior-year returns have a lower overall income level than current-year returns, may be attributable primarily to the effects of inflation.

TABLE A.-- AVERAGE ADJUSTED GROSS INCOME FOR SELECTED CATEGORIES OF RETURNS, BY SIZE OF AGI

CATEGORY	SIZE OF ADJUSTED GROSS INCOME			
	TOTAL	DEFICIT	\$1 under \$2,000,000	\$2,000,000 or more
All returns, total	14,520	-15,431	14,665	3,844,367
Processing cycle:				
1 through 37	14,457	-12,711	14,583	3,843,069
38 through 39	22,306	-42,931	23,707	3,449,222
40 or later	19,578	-68,143	22,005	3,954,600
Prior year returns:				
Total	11,659	-18,858	13,157	6,659,455
Processing cycle:				
1 through 37	11,583	-17,354	13,145	5,849,125
38 through 39	16,966	-28,010	18,604	7,290,000
40 or later	11,712	-29,083	12,923	9,585,500
Returns missing from Advance Data	17,279	-24,775	17,596	4,015,909

In terms of the concept of SOI as a vehicle for analyzing and evaluating the operation of the tax laws in a given tax year, it would seem beneficial to isolate prior-year returns by the tax period for which they were filed. Once isolated, these returns could be consistency and validity tested with a simplified battery of tests designed for that specific tax year only. Once tested, the prior-year returns could be reassociated with the other returns filed for the same tax period. The resulting "tax year" SOI file should be a conceptually stronger data base from which to analyze the operation of our tax system, in that assumptions made about prior-year returns will have been eliminated.

However, there will be a considerable time lag in producing this "tax year" file, because the majority of prior-year returns are filed either one or two years late. Until these returns are filed, it will be impossible to build an accurate representation of the "ever-filed" population for a given tax year.

3. SIMULATION OF 1978 ADVANCE DATA AND FINAL ESTIMATES

Methodology.--A simulation of the 1978 SOI file, as it would have been, was created as a vehicle for evaluating the results of incorporating the three proposed changes discussed above. The simulation was also used as a preliminary step in evaluating the use of an early cut-off file to produce the complete SOI report for a given tax year.

In creating the simulation file, all sample returns on the 1978 SOI tape file with a tax year prior to 1978 or with a return processing cycle code greater than 37 (i.e., filed later than the third week of September) were assigned a weight factor of zero. This step excluded prior-year returns from the simulation, and included returns that would have been designated prior to the proposed cut-off for the simulation but processed after this date. These latter returns would not have been included had we followed the current processing method.

The second stage of the simulation, developing weight factors for the remaining returns, required: first, producing sample counts by sample code (stratum) within IRS districts; then, computing simple ratio weight factors, by dividing the sample count into the population. This paralleled the original 1978 sample weighting technique.

In order to maintain comparability between SOI and simulation estimates, the simulation file (which excludes prior-year returns) was weighted to represent the entire processing year population for 1978 (which includes prior-year returns). Columns 7 through 9 of Table 2 present a distribution, by size of AGI, of the simulation file after initial application of the simple ratio weight factors. Two obvious deficiencies exist at this stage: the deficit class and the \$2,000,000 or more AGI class.

To overcome the deficiencies evident in the deficit and very high income classes, we assumed the institution of a special control system that would take over after the early cut-off, and continue until some specified time in the processing year, insuring that all returns designated in these two classes are included in the final sample. For purposes of this simulation, we assumed that the "specified time" was the end of the processing year. For all returns falling in these two classes, the original 1978 complete SOI weight factor was transferred to the simulation record. Columns 10 through 12 of Table 2 present the simulation results after this adjustment.

A final refinement was made to the simulation weights to adjust for the absence of prior year returns in the simulation sample. In order to accomplish this, we first developed a basis for adjustment by applying the ratio of aggregate AGI between 1977 and 1978 to the AGI amount (on a record by record basis) on prior-year returns in the 1978 complete SOI file. Deficit prior-year returns were not adjusted. Columns 1 through 3 of Table 2 present the distribution, by size of adjusted gross income, of this adjustment to the 1978 complete SOI file. This should be an approximate representation of the frequency distribution of an "ever-filed" population for 1978.

Ratios were then computed, on an income class by income class basis, between the expected number of returns (Column 1) and the simulation number of returns (Column 10). The ratios thus developed were applied to the data in columns 10 through 12 and the results presented in columns 1 through 4. For income less than \$30,000, the classes used for computing the ratios were much broader than those presented in Table 2. The computation of ratios based on broad classes and applied to narrow classes accounts for the minor discrepancy in the number of returns between columns 1 and 4 for classes below the \$30,000 income level. These broader classes (than those presented in Table 2 for income levels less than \$30,000) were used for computing ratios because the ranges correspond to similar tabulations available for 1979 and future years. When we do simulations for years later than 1978, these tabulations will become the basis for this type of adjustment.

These ratios were applied to the existing simulation weights on a record by record basis to generate the final simulation weights. These final simulation weights were used to produce all the "simulation" estimates that appear labelled as "simulation after all adjustments". A brief explanation of the weighting technique employed in generating the 1978 advance data estimates is presented in note [12].

Two unmeasurable differences exist between the original 1978 advance data estimates and those generated through any 1978 simulation run. Although few in number, duplicate returns in the 1978 SOI file were deleted at final SOI closeout, not at preliminary cut-off. Thus any duplicate returns in the 1978 preliminary file had been deleted from the complete 1978 SOI sample file before we began creating the simulation file. Also, as part of normal consistency testing of SOI returns at the Data Center, information listings of returns with unusual, unexpected, or out-of-range items are produced. These sampled returns are located (whenever possible) and reviewed by statisticians in the Statistics Division. Most corrections or changes posted to the SOI file as a result of this review were not available at preliminary cut-off, but were made to the final SOI file which was the starting point for this research file. Simulations for future years, 1980 and on, will measure these differences.

4. SIMULATION RESULTS AND RECOMMENDATIONS

Results.--A careful comparison of columns 3 and 9 in Table 2 would lead one to conclude that a straightforward simulation of 1978 advance data that incorporates the three basic changes discussed earlier (with no subsequent refinements) accurately reproduces the 1978 complete SOI estimates for adjusted gross income except in the deficit and very high income classes. The discrepancy in the deficit class should be expected because the average deficit on returns filed after the cut-off is more than 5 times larger than the average deficit on returns filed before the cut-off. The differences encountered in the \$2,000,000 or more class also appear to be the result of excluding late filed returns.

In terms of producing the advance data tabulations from an early SOI file that incorporates the three changes explained above, a special control and handling system must be instituted. This system would begin at the early cut-off and would maintain strict controls on deficit and very high income returns, insuring that any sample returns designated in these two categories after the cut-off date would be included in the early cut-off SOI file. For advance data, this system could provide an additional six weeks worth of these cases (this is four weeks longer than under the current processing system). This system, for advance data, would end early in November in time to develop weight factors for the advance data (or preliminary) file.

In attempting to simulate the complete SOI for 1978, we carried the idea of a special control system a little further. We assumed this type of system would continue to the end of the processing year. The results of including this control system through the end of the processing year, as well as incorporating the ratio refinement (to adjust for the exclusion of prior year returns), resulted in simulation estimates that were within the range described by coefficient of variation (at the 68% confidence level) for all but the three items listed below.

In reviewing the simulation estimates we were quite concerned with the levels of Business Net Profit, Business Net Loss and Net Capital Gain. A distribution of returns with these items by size of the item and by returns included in and excluded from the simulation indicated that an early cut-off sample was not representative of these categories of returns. It appears from this information that late filed returns with large amounts for any one of these three items should be included in our special handling and control system.

On balance, it appears from the results of the various simulation runs produced to date that it will be possible to modify the SOI processing system, conserve resources, produce earlier estimates, and only marginally (if at all) affect the reliability of the SOI figures.

Recommendations.--Our recommendation for constructing an early cut-off advance data sample would incorporate the three basic proposals outlined earlier. In addition, a special control system would be instituted to include deficit, very high income, and large "special item" (business, capital gain or other) returns designated within six weeks after the cut-off in the advance data SOI file. [3]

In constructing an early cut-off complete SOI file, we recommend continuing the "special control" system through early December. In addition to this, any sample returns processed error-free through the Service Centers between the mid-September and the early December cut-off dates should also be included in the final SOI file. The inclusion of these additional sample units in the final SOI will reduce the sampling variability of the estimates made from that sample. On the other hand, this procedure could introduce an element of bias into the sample if the error-free records are not representative of all returns processed during that period of time. The proposed 1980 simulation will analyze this problem.

5. FUTURE PLANS

The research and testing of an early cut-off for preliminary or advance data SOI estimates is really the first step in a longer-range plan to produce the complete SOI report for a given tax year from an earlier cut-off SOI file than is currently being used. The benefits of an early cut-off for SOI publication purposes are two-fold. Resources are conserved and data is available for release much earlier.

One of the proposals we are giving very serious thought to calls for closing out the basic sample file, from which the complete SOI report will be produced, after the third week of September (as was done in this 1978 simulation). Sample designation and data transcription will continue through the end of the processing year. Error free returns sampled after the cut-off date, as well as any returns subjected to special handling (deficits, very high incomes, etc.), will be included in the publication version of the SOI file for any given tax year.

The early cut-off advance data recommendation will be simulated (exactly as specified) using the 1980 SOI File. The early cut-off complete SOI File will also be simulated (again, exactly as specified) using the 1980 SOI File. The 1980 file is the first one available containing all the indicators necessary to isolate each specific category of return. The 1980 file will also contain the necessary information to measure the effects of duplicate returns and post-processing improvements mentioned in the methodology section of this report.

The publication SOI file will become the basis for the IRS individual tax model and a version of this file will be provided to the National Archives for distribution as a public use data

base. As with the current system, OTA will have access to this final SOI file for generating their tax model. The time frame for the availability of this final SOI file to OTA and to the National Archives will be considerably earlier than under the present system.

Even though prior year returns have been excluded from this simulation and would be excluded from future SOI publications, they will still be designated as part of the sample, isolated, and maintained separately. At some point in time, these prior year returns will be associated with tax year SOI file in which they belong and basic tabulations will be produced. This updated file will be made available as a public use data base.

In order to successfully produce the complete SOI report from an early cut-off file, a number of issues (potential problem areas) must be explored and resolved. Some of the more critical issues are: imputing missing data items resulting from an early cut-off, adjusting for late filed returns and improving population estimation techniques, see [12]. These issues will be explored in a simulation of an early cut-off 1980 SOI File, and the findings presented in a subsequent, related report.

Instead of the simple ratio estimation weighting technique now being used, a raking ratio estimation technique might better adjust for some of the skewing tendencies exhibited by the late-filers. Raking is a procedure for iteratively ratioing sample data to known (outside) marginal totals [10]. The raking ratio method will be tested against the proposed 1980 simulation file and the results also presented in a later report.

The future simulations mentioned in this section and in the methodology section should provide a more realistic test of our proposals than did the 1978 simulation. Because of the unavailability of information from the intermediate stages of 1978 processing, we simulated a "best case" situation. For coming simulations, we expect to reproduce the exact conditions we plan to implement.

It should be kept in mind that a fail-safe system is implied by the continuation of sample designation and data transcription after the early cut-off. If it becomes apparent that the reliability of the SOI data will be compromised beyond acceptable limits by using an early cut-off for the published SOI report, it will always be possible to produce the necessary tabulations from a complete SOI file comparable to that used for previous SOI years.

One final point must be made. The trend in the filing pattern over recent years indicates that returns are being filed later and later each year [8]. This would imply that, in the long run, it will become increasingly difficult to justify an early cut-off SOI sample. The early cut-off proposal is a reasonable short run strategy for meeting the commitment to earlier release of SOI estimates. A better long run

strategy would be to standardize and streamline the processing that occurs between sample designation and the publication of estimates. If such standardization can be achieved, it could allow for a cut-off that is even later than under the current system.

ACKNOWLEDGEMENTS

The authors would like to take this occasion to acknowledge the helpful comments and suggestions of Ralph Bristol and Thomas Vasquez of the Treasury Department's Office of Tax Analysis, Neil Barclay of Revenue Canada Taxation and Peter Davis of the Senate Budget Committee. In addition, we would like to thank Mary Haigler and Dawn Nester for their help in typing the various drafts and tables for this presentation.

NOTES AND REFERENCES

- [1] Blacksin, Jack and Raymond Plowden, Statistics of Income for Individuals: A Historical Perspective, 1981 Proceedings: American Statistical Association, Section on Statistical Use of Administrative Records, 1981.
- [2] Deming, W. Edwards, Review of Quality Standards of the Procedures Used by the Internal Revenue Service to Produce Statistics of Income From Individual Returns with Special Emphasis on the Sampling Procedures, 1963.
- [3] Hirsch, Werner Z., Ways in Which the Internal Revenue Service can Improve its Service to the Treasury and Other Users, 1961.
- [4] Houthakker, H.S., The Future of the Statistics of Income Program: A Preliminary Report to the Office of Tax Analysis, U.S. Treasury Department, 1966.
- [5] Individual Income Tax Returns for 1978 - Advance Data, 1979 (unpublished).
- [6] Internal Revenue Service, Individual Tax Model File for 1978, Order Number 374-109(a), National Archives and Records Service, 1980.
- [7] Internal Revenue Service, Preliminary Statistics of Income--1978, Individual Income Tax Returns, Publication 198, U.S. Government Printing Office, 1980.
- [8] Internal Revenue Service, Statistics of Income Bulletin, Publication 1136, U.S. Government Printing Office, 1981. The Summer 1981 issue contains information that would have appeared in Preliminary SOI report [7]. The Fall 1981 issue includes articles on return filing patterns.
- [9] Internal Revenue Service, Statistics of Income--1978, Individual Income Tax Returns, Publication 79, U.S. Government Printing Office, 1981.
- [10] Scheuren, F. et. al., Studies from Interagency Data Linkages, Report No. 10, 1981.
- [11] Vanik, Charles A., Letter of November 1 to Commissioner of IRS transmitting report of GAO study of timeliness of Statistics of Income data. 1p., and attachments, 1976.

[12] For advance data (or preliminary) estimates of a given SOI year, Y, the sample is cut-off at some point before the sampling time frame is complete. The population continues to be counted and the sample continues to be selected beyond the cut-off date until the time period (frame) is satisfied; however, sample returns selected after the cut-off are not included in the advance data estimates. The portion of the population to be counted (and selected from) after the cut-off in year Y is

unknown, but is estimated from the previous year's (Y-1) known population counts for the similar time period, then added to the current year (Y) known population counts. This method of estimation is applied by sample stratum within districts. Therefore, the advance data weight factors are based on a known sample count and an estimated population. Over the years, this method has proven to be reliable in estimating the full year SOI population counts.

Table 1.-- Number of Returns and Column Percents: by Filing Year, Processing Cycle and Selected Classifications

Selected Classification	1978 Total	Filing Year		Processing Cycle			Simulation After
		Current	Prior	1 Through 37	38 Through 39	40 or later	All Adjustments

Part I.-- Frequencies (in thousands of returns)

Total.....	89,771	88,726	1,046	88,803	229	739	89,771
Joint.....	44,483	43,957	526	43,945	129	408	44,528
Nonjoint.....	45,288	44,768	520	44,858	100	330	45,243
Nonbusiness.....	81,224	80,403	821	80,542	163	519	81,194
Business.....	8,548	8,323	225	8,261	66	220	8,577
Itemized.....	25,756	25,482	274	25,388	96	273	25,751
Other.....	64,015	63,244	772	63,416	133	466	64,020
1040.....	53,824	53,026	798	52,995	204	626	53,786
1040A.....	35,947	35,700	247	35,808	26	113	35,985

Part II.-- Column Percents

Total.....	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Joint.....	49.55	49.54	50.31	49.49	56.36	55.29	49.60
Nonjoint.....	50.45	50.46	49.69	50.51	43.64	44.71	50.40
Nonbusiness.....	90.48	90.62	78.49	90.70	71.02	70.18	90.45
Business.....	9.52	9.38	21.51	9.30	28.98	29.82	9.55
Itemized.....	28.69	28.72	26.21	28.59	42.00	36.89	28.69
Other.....	71.31	71.28	73.79	71.41	58.00	63.11	71.31
1040.....	59.96	59.76	76.35	59.68	88.81	84.71	59.91
1040A.....	40.04	40.24	23.65	40.32	11.21	15.29	40.09

Table 2.-- Number of Returns, Amount and Average AGI by Size of AGI for 1979 Complete SOI (Adjusted), Various Stages of Simulation and 1978 Advance Data
 (Number of returns in thousands, amounts in millions and averages in whole dollars)

Size of Adjusted Gross Income	Simulation												1978 Advance Data as Transmitted to OTA		
	SOI Complete After Ratio Adjustment For Prior Year Returns			Simulation After all Adjustments			Raw Data			Adjustment for Deficit and Very High AGI Only			Number	Amount	Average
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)			
Number	Amount	Average	Number	Amount	Average	Number	Amount	Average	Number	Amount	Average	Number	Amount	Average	
Total.....	89,771	1,303,434	14,520	89,771	1,303,453	14,521	89,773	1,305,728	14,545	89,822	1,303,706	14,515	89,890	1,304,189	14,509
Deficit.....	484	-7,473	-15,431	484	-7,473	-15,431	435	-5,569	-12,796	484	-7,473	-15,431	484	-7,473	-15,431
Break-even.....	8,459	9,477	1,118	8,432	9,452	1,122	8,422	9,476	1,122	8,442	9,476	1,122	9,058	9,330	1,037
\$2,000 under \$4,000.....	9,234	27,405	2,989	9,237	27,430	2,991	9,261	27,702	2,991	9,281	27,702	2,991	9,280	27,697	2,991
\$4,000 under \$6,000.....	8,388	42,236	5,035	8,362	42,111	5,036	8,384	42,220	5,036	8,384	42,220	5,036	8,401	42,292	5,035
\$6,000 under \$8,000.....	8,259	57,530	6,966	8,303	57,558	6,968	8,325	58,008	6,968	8,324	58,008	6,968	8,263	57,579	6,968
\$8,000 under \$10,000.....	6,758	62,000	8,922	6,986	62,388	8,931	6,986	62,528	8,931	6,986	62,528	8,931	6,944	62,146	8,950
\$10,000 under \$12,000.....	6,882	72,434	10,528	6,924	72,434	10,528	6,924	72,434	10,528	6,924	72,434	10,528	6,924	72,434	10,528
\$12,000 under \$14,000.....	5,584	72,634	12,971	5,580	72,374	12,971	5,574	72,300	12,971	5,574	72,300	12,971	5,604	72,691	12,969
\$14,000 under \$16,000.....	5,016	75,198	14,993	5,022	75,302	14,994	5,017	75,224	14,994	5,017	75,224	14,994	5,016	75,212	14,993
\$16,000 under \$18,000.....	4,669	79,388	17,002	4,678	79,333	17,002	4,673	79,452	17,002	4,673	79,452	17,002	4,683	79,611	17,002
\$18,000 under \$20,000.....	4,584	81,361	18,192	4,271	81,323	18,192	4,268	81,040	18,192	4,268	81,040	18,192	4,277	81,234	18,192
\$20,000 under \$25,000.....	8,269	191,104	22,326	8,253	190,732	22,326	8,244	190,728	22,326	8,244	190,728	22,326	8,255	190,989	22,323
\$25,000 under \$30,000.....	5,394	147,023	27,258	5,388	146,877	27,260	5,382	146,728	27,260	5,382	146,728	27,260	5,385	146,794	27,260
\$30,000 under \$35,000.....	6,546	239,253	36,584	6,546	239,304	36,584	6,529	238,658	36,584	6,529	238,658	36,584	6,534	238,846	36,584
\$35,000 under \$40,000.....	1,475	96,499	65,430	1,475	96,447	65,395	1,473	96,319	65,395	1,473	96,319	65,395	1,469	96,005	65,358
\$40,000 under \$50,000.....	286	37,591	131,469	286	37,577	131,421	288	37,867	131,421	288	37,867	131,421	285	37,471	131,405
\$50,000 under \$60,000.....	60	16,737	279,018	60	16,705	278,684	61	17,035	278,684	61	17,035	278,684	60	16,726	278,424
\$60,000 under \$70,000.....	7	4,387	666,074	7	4,386	665,885	7	4,764	665,885	7	4,764	665,885	7	4,572	665,356
\$70,000 under \$80,000.....	1	1,991	1,335,083	1	1,984	1,330,443	2	2,138	1,330,444	2	2,138	1,330,444	2	4,172	1,330,444
\$80,000 under \$100,000.....	1	2,126	3,844,367	1	2,126	3,844,367	1	2,153	3,824,190	1	2,153	3,824,190	1	4,186	2,000,773
\$100,000 or more.....	1	2,126	3,844,367	1	2,126	3,844,367	1	2,153	3,824,190	1	2,153	3,824,190	1	4,186	2,000,773

Table 3.-- Number of Returns, Amount of AGI and Average AGI, by Size of AGI and by Filing Category

Size of Adjusted Gross Income	All Returns Total	All Returns by Processing Cycle			Prior Year Returns Total	Returns Missing From Advance Data
		1 Through 37	38 Through 39	40 or Later		
Part I.-- Number of Returns						
TOTAL.....	89,771,249	88,802,683	229,466	739,100	1,045,897	518,157
Deficit.....	484,299	458,607	4,238	21,454	46,542	5,951
Breakeven.....	39,765	29,381	4,248	6,136	11,326	1
\$1 under \$2,000.....	8,469,208	8,394,533	21,837	47,838	125,785	46,698
\$2,000 under \$4,000.....	9,234,173	9,162,794	13,358	58,020	115,220	40,865
\$4,000 under \$6,000.....	8,387,955	8,277,372	13,360	97,224	125,177	43,703
\$6,000 under \$8,000.....	8,258,760	8,191,922	9,609	57,229	84,981	49,043
\$8,000 under \$10,000.....	6,925,837	6,874,631	13,667	37,539	76,923	56,386
\$10,000 under \$12,000.....	6,088,694	6,048,040	6,490	34,163	53,674	47,100
\$12,000 under \$14,000.....	5,584,491	5,541,056	15,799	27,637	66,396	32,014
\$14,000 under \$16,000.....	5,015,526	4,973,469	6,596	35,461	51,543	13,897
\$16,000 under \$18,000.....	4,669,441	4,622,746	11,921	34,774	40,756	17,627
\$18,000 under \$20,000.....	4,283,867	4,236,597	12,650	34,620	50,923	18,063
\$20,000 under \$25,000.....	8,559,908	8,472,868	24,553	62,488	74,042	48,682
\$25,000 under \$30,000.....	5,393,740	5,331,662	20,300	41,778	47,026	26,190
\$30,000 under \$50,000.....	6,546,202	6,434,101	29,414	82,687	53,345	50,459
\$50,000 under \$100,000.....	1,474,835	1,413,643	17,172	44,019	18,231	14,894
\$100,000 under \$200,000.....	285,931	270,567	3,257	12,107	5,266	4,974
\$200,000 under \$500,000.....	59,987	55,965	868	3,154	652	1,321
\$500,000 under \$1,000,000.....	6,586	5,940	89	557	61	207
\$1,000,000 under \$2,000,000.....	1,491	1,324	22	145	17	62
\$2,000,000 or more.....	553	465	18	70	11	22
Part II.-- Amount of AGI (in thousands of dollars)						
TOTAL.....	1,303,434,144	1,283,845,448	5,118,434	14,470,264	12,194,500	8,953,258
Deficit.....	-7,473,332	-5,829,444	-181,941	-1,461,947	-877,695	-147,439
Breakeven.....	-	-	-	-	-	-
\$1 under \$2,000.....	9,471,532	9,409,834	18,896	42,802	118,835	51,414
\$2,000 under \$4,000.....	27,605,226	27,389,713	43,205	172,308	335,131	130,570
\$4,000 under \$6,000.....	42,235,731	41,669,940	66,811	498,981	622,496	225,542
\$6,000 under \$8,000.....	57,530,195	57,066,905	63,989	399,302	581,714	335,841
\$8,000 under \$10,000.....	62,000,014	61,537,350	124,330	338,334	703,959	496,165
\$10,000 under \$12,000.....	66,869,810	66,413,961	72,385	383,463	591,878	513,059
\$12,000 under \$14,000.....	72,436,262	71,877,675	199,872	358,715	862,518	419,353
\$14,000 under \$16,000.....	75,198,387	74,565,145	100,196	533,046	769,630	206,630
\$16,000 under \$18,000.....	79,387,934	78,594,834	203,009	590,091	690,160	300,111
\$18,000 under \$20,000.....	81,360,933	80,457,588	246,221	657,124	973,122	344,446
\$20,000 under \$25,000.....	191,104,401	189,162,544	555,803	1,386,054	1,652,512	1,117,823
\$25,000 under \$30,000.....	147,023,113	145,333,832	551,732	1,137,548	1,277,477	706,555
\$30,000 under \$50,000.....	239,353,313	235,137,323	1,111,702	3,104,288	1,968,283	1,900,683
\$50,000 under \$100,000.....	96,498,828	92,411,910	1,121,329	2,965,590	1,179,501	1,005,524
\$100,000 under \$200,000.....	37,591,030	35,564,676	424,865	1,601,489	427,303	657,475
\$200,000 under \$500,000.....	16,737,457	15,578,279	246,471	912,708	179,322	375,162
\$500,000 under \$1,000,000.....	4,386,763	3,953,449	59,340	373,974	42,628	143,186
\$1,000,000 under \$2,000,000.....	1,990,612	1,762,907	28,133	199,572	22,476	82,808
\$2,000,000 or more.....	2,125,935	1,787,027	62,086	276,822	73,254	88,350
Part III.-- Average AGI (in whole dollars)						
TOTAL.....	14,520	14,457	22,306	19,578	11,659	17,279
Deficit.....	-15,431	-12,711	-42,931	-68,143	-18,858	-24,775
Breakeven.....	-	-	-	-	-	-
\$1 under \$2,000.....	1,118	1,120	865	895	945	1,101
\$2,000 under \$4,000.....	2,989	2,989	3,234	2,970	2,909	3,195
\$4,000 under \$6,000.....	5,035	5,034	5,001	5,132	4,973	5,161
\$6,000 under \$8,000.....	6,966	6,966	6,659	6,977	6,845	6,848
\$8,000 under \$10,000.....	8,952	8,951	9,097	9,013	9,151	8,799
\$10,000 under \$12,000.....	10,983	10,981	11,153	11,225	11,027	10,893
\$12,000 under \$14,000.....	12,971	12,972	12,651	12,980	12,990	13,099
\$14,000 under \$16,000.....	14,993	14,993	15,190	15,032	14,932	14,869
\$16,000 under \$18,000.....	17,002	17,002	17,030	16,969	16,934	17,026
\$18,000 under \$20,000.....	18,992	18,991	19,464	18,981	19,110	19,069
\$20,000 under \$25,000.....	22,326	22,326	22,637	22,181	22,319	22,962
\$25,000 under \$30,000.....	27,258	27,259	27,179	27,228	27,165	26,978
\$30,000 under \$50,000.....	36,564	36,545	37,795	37,543	36,897	37,668
\$50,000 under \$100,000.....	65,430	65,371	65,300	67,371	64,698	67,512
\$100,000 under \$200,000.....	131,469	131,445	130,447	132,278	130,834	132,182
\$200,000 under \$500,000.....	279,018	278,358	283,953	289,381	275,034	283,998
\$500,000 under \$1,000,000.....	666,074	665,564	666,742	671,408	698,820	691,720
\$1,000,000 under \$2,000,000.....	1,335,085	1,331,501	1,278,773	1,376,359	1,322,118	1,335,613
\$2,000,000 or more.....	3,844,367	3,843,069	3,449,222	3,954,600	6,659,455	4,015,909

Table 4.-- 1978 Complet SOI Report, Simulation and Advance Data: Amounts, Differences and Coefficients of Variation for Specified Items

Item	1978 Complete SOI Report (\$000)	Simulation After All Adjustments (\$000)	1978 Advance Data (\$000)	Complete Minus Simulation as a Percent of Complete	Complete Minus Advance Data as a Percent of Complete	Coefficient of Variation for 1978 Complete ^{1/}
Adjusted Gross Income.....	1,302,447,386	1,303,647,457	1,304,188,847	0.09	0.13	0.1
Salaries and wages.....	1,090,291,855	1,092,086,262	1,092,017,073	0.16	0.16	0.2
Business net profit.....	61,413,703	60,957,923	60,741,261	0.74	1.09	0.6
Business net loss.....	7,867,195	7,686,962	7,412,957	2.29	5.77	1.5
Farm net profit.....	11,034,552	11,015,275	10,989,231	0.17	0.41	3.4
Farm net loss.....	7,469,259	7,326,396	7,180,316	1.91	3.87	3.4
Partnership net profit less loss.....	15,044,787	14,849,870	15,407,324	1.30	2.41	3.5
Small Business Corp. net profit less loss.....	2,284,806	2,211,392	2,471,275	3.21	8.16	10.1
Net capital gain.....	26,232,396	25,125,251	24,993,143	4.22	4.72	1.4
Net capital loss.....	3,001,020	2,955,025	2,951,478	1.53	1.65	2.9
Sales of property other than capital assets.....	1,256,902	1,232,696	1,246,753	1.93	0.81	9.2
Total dividends.....	31,671,858	31,677,425	31,634,519	0.02	0.12	1.3
Dividends in adjusted gross income.....	30,206,475	30,208,441	30,169,755	0.01	0.12	1.4
Interest received.....	61,222,522	61,419,009	60,947,334	0.32	0.45	1.0
Pensions and annuities in AGI.....	32,743,819	33,141,546	32,883,949	1.21	0.43	1.9
Rent net income.....	10,983,905	11,032,681	10,878,106	0.44	0.96	2.5
Rent net loss.....	7,844,747	7,592,879	7,618,195	3.21	2.89	2.5
Royalty net income less loss.....	2,559,870	2,599,667	2,573,943	1.55	0.55	5.3
Estate or trust net income less loss.....	3,079,603	2,977,860	2,990,517	3.30	2.89	4.8
State income tax refunds.....	2,368,949	2,361,439	2,363,389	0.32	0.23	1.4
Alimony received.....	1,191,389	1,130,013	1,170,453	5.15	1.76	10.7
Other income less loss.....	-921,836	-948,473	-157,729	2.89	82.89	--
Adjustments.....	22,364,088	22,192,133	22,333,986	0.77	0.13	1.3
Disability income exclusion.....	1,066,206	1,097,204	1,126,745	2.91	5.68	10.2
Payments to an Ind. Retirement Acct.....	2,970,121	2,988,498	2,984,424	0.62	0.48	2.0
Payments to a KEOGH.....	1,994,029	2,018,627	1,990,410	1.23	0.18	1.9
Deduction for expense of living abroad.....	314,468	412,145	340,613	31.06	8.31	8.9
Exemption amount.....	164,900,772	164,692,352	165,008,114	0.13	0.01	--
Taxable income.....	1,062,190,322	1,064,317,958	1,063,308,318	0.20	0.11	--
Income tax before credits.....	203,803,653	204,365,014	204,162,283	0.28	0.18	--
Total credits.....	17,085,591	16,990,499	16,989,359	0.56	0.56	--
New jobs credit.....	1,370,406	1,333,778	1,327,934	2.67	3.10	--
Earned income credit used to offset tax before credits.....	152,934	152,788	139,738	0.10	8.63	--
Residential energy credit.....	576,545	583,060	578,405	1.13	0.32	--
Business energy investment credit.....	219,868	219,590	1,314	0.13	99.40	--
Income tax after credits.....	186,718,062	187,374,515	187,172,925	0.35	0.24	--
Total tax preferences.....	18,381,866	17,451,568	17,284,665	5.06	5.97	--
Minimum tax.....	1,514,475	1,423,157	1,404,261	6.03	7.28	--
Total income tax.....	188,232,537	188,797,672	188,577,186	0.30	0.18	0.2
Self-employment tax.....	4,705,994	4,651,382	4,648,370	1.16	1.22	--
Earned income credit used to offset all other taxes.....	94,197	93,378	85,622	0.87	9.10	--
Total tax liability.....	193,184,849	193,685,577	193,464,593	0.26	0.14	0.2
Total taxpayments.....	202,829,400	203,666,001	203,340,183	0.41	0.25	--
Withholding.....	169,984,010	170,745,020	170,537,019	0.45	0.33	--
Estimated payments.....	29,978,499	30,429,454	30,262,953	1.50	0.95	--
All other taxpayments.....	2,866,890	2,491,526	2,540,211	13.09	11.39	--
Earned income credit, refundable portion.....	801,171	796,200	756,708	0.62	5.55	--
Business energy investment credit, refundable.....	397	353	401	11.08	1.01	--
Tax due at time of filing.....	24,969,333	24,608,375	24,729,595	1.45	0.96	--
Total overpayment.....	35,415,451	35,385,352	35,362,293	0.08	0.15	--
Refund.....	33,034,549	33,082,636	33,042,923	0.15	0.03	--
Credit on 1979 tax.....	2,380,903	2,302,716	2,319,370	3.28	2.58	--

^{1/} Coefficient of variation at the 68% confidence level.

Table 5.-- Number of Returns, Amount and Average Net Capital Gain or Loss, Business Net Loss and Business Net Profit: by Size of the Item, for Returns Included in and Excluded From the Complete SOI Simulation (using 1978 SOI weights)

Size of Item	Returns Included in the Simulation			Returns Excluded From the Simulation		
	Number	Amount (\$000)	Average (\$)	Number	Amount (\$000)	Average (\$)
Part I.-- Size of Net Capital Gain or Loss						
Total.....	8,375,948	21,140,808	2,524	335,004	2,092,158	6,245
Loss						
Less than \$3,000.....	397	4,911	12,369	1	1,075	1,074,809
\$2,000 under \$3,000.....	708,474	2,038,077	2,877	35,245	97,187	2,757
\$1,000 under \$2,000.....	330,916	474,998	1,435	25,191	30,976	1,230
\$1 under \$1,000.....	973,837	339,922	349	32,433	13,797	425
Gain						
\$1 under \$1,000.....	3,354,669	991,072	295	89,816	27,388	305
\$1,000 under \$5,000.....	1,977,244	4,753,542	2,404	59,850	153,124	2,558
\$5,000 under \$10,000.....	562,396	4,002,862	7,118	33,279	264,892	7,960
\$10,000 under \$25,000.....	349,691	5,359,407	15,326	45,752	701,840	15,340
\$25,000 under \$50,000.....	74,781	2,562,654	34,269	8,142	276,901	34,009
\$50,000 under \$100,000.....	27,998	1,927,571	68,847	3,094	209,588	67,740
\$100,000 under \$200,000.....	9,548	1,292,869	135,407	1,423	197,103	138,512
\$200,000 under \$500,000.....	4,506	1,355,935	300,918	538	164,257	305,310
\$500,000 or more.....	1,491	1,752,805	1,175,590	240	240,100	1,000,416
Part II.-- Size of Business Net Loss						
Total.....	1,948,949	7,303,319	3,747	84,850	562,885	6,634
\$1 under \$5,000.....	1,624,640	2,216,914	1,365	63,122	105,273	1,668
\$5,000 under \$10,000.....	186,378	1,294,931	6,948	11,299	81,167	7,184
\$10,000 under \$25,000.....	105,489	1,553,626	14,728	7,245	103,153	14,238
\$25,000 under \$50,000.....	21,638	747,496	34,546	1,775	63,655	35,862
\$50,000 under \$100,000.....	6,761	458,017	67,744	885	58,421	66,013
\$100,000 under \$200,000.....	2,807	384,584	137,009	323	42,789	132,475
\$200,000 under \$500,000.....	951	279,896	294,318	141	40,069	284,174
\$500,000 or more.....	285	367,855	1,290,721	60	68,358	1,139,302
Part III.-- Size of Business Net Profit						
Total.....	5,838,743	57,161,392	9,790	321,688	4,245,545	13,198
\$1 under \$5,000.....	3,175,511	5,676,113	1,787	134,435	303,349	2,256
\$5,000 under \$10,000.....	1,026,845	7,461,942	7,267	67,701	498,724	7,367
\$10,000 under \$25,000.....	1,067,952	16,761,782	15,695	77,930	1,228,650	15,766
\$25,000 under \$50,000.....	406,954	14,054,200	34,535	28,354	981,268	34,608
\$50,000 under \$100,000.....	136,982	9,144,197	66,755	10,488	720,272	68,676
\$100,000 under \$200,000.....	21,064	2,703,451	128,345	2,173	286,576	131,881
\$200,000 under \$500,000.....	2,972	821,251	276,330	499	138,414	277,382
\$500,000 or more.....	463	538,455	1,162,971	108	88,292	817,521