

THE 1985 SALES OF CAPITAL ASSETS STUDY

Dan Holik, Susan Hostetter, and John Labate, Internal Revenue Service*

The Statistics of Income (SOI) Division of the Internal Revenue Service has just completed the 1985 Sales of Capital Assets (SOCA) Study using two new and innovative approaches -- a cooperative venture in directing the study and SOI's first on-line, real-time editing system. Data on Sales of Capital Assets are reported on Schedule D of the Individual income tax Forms 1040 series. They indicate the profit/loss from sales of stocks, bonds, commodity contracts, livestock, land, real property (both personal and business), and other capital assets. The data are used by Congress and the Treasury Department's Office of Tax Analysis for modeling the impact of changes in the tax laws for capital gains.

Due to the intense discussion of capital assets taxation raised by the Tax Reform Act of 1986, the SOCA studies have attracted much attention. This paper represents the first public release of the 1985 data. The objectives of this paper are to present processing methodology and data derived from the SOCA study. First, some historical perspective for this study is provided. Next, the new data processing approaches are described. Then, some of the results will be examined, with a focus on issues of concern to our primary users. Finally, some plans for the future will be proposed.

HISTORICAL PERSPECTIVE

Taxation of capital gains has varied greatly and caused considerable controversy since the income tax was enacted in 1913. Until 1922 capital gains were taxed as ordinary income, and treatment of capital losses varied from not deductible at all, to deductible against capital gains only, to deductible in full from any income. The concept of different tax rates for different holding periods was initiated in the mid-1920's, but only in regard to the alternative tax rate. From 1942 to 1969 capital gains on long-term assets (held over 6 months) were taxed at half the ordinary rates, with a 25 percent maximum tax. Beginning in 1970 gains over \$50,000 were taxed at a greater rate of 35 percent. This plus the inclusion of the minimum tax increased the maximum effective tax to about 40 percent. The holding period for long-term transactions was extended to 9 months in 1977, and to 1 year in 1978. Throughout these years, it should also be noted that tax provisions regarding capital losses offsetting income changed frequently.

Effect of Tax Rates on Capital Growth

During the 1970's there was a perception that tax rates on capital income were high enough to discourage growth in capital formation and investment, and that reducing rates on capital income would stimulate capital asset transactions. Some economists predicted that reduced rates would actually increase tax receipts.

Following this logic, the Revenue Act of 1978 reduced maximum rates on long-term capital gains to 40 percent of rates on ordinary income. As a result, for 1979 the highest capital gains tax rate was 28 percent (40 percent of the 70 percent maximum rate on ordinary income).

In 1982, as a result of the 1981 Economic Recovery Tax Act, the top rate on ordinary income was reduced from 70 to 50 percent, effectively reducing the maximum rate on capital gains to 20 percent. The Tax Reform Act of 1984 reduced the holding period for long-term capital assets from 1 year to 6 months. Finally, the Tax Reform Act of 1986 completely changed the rules -- stripping all exceptions for taxing capital gains, treating them as ordinary income regardless of holding period.

Previous Capital Assets Studies

Considering the extensive and frequent change in tax law for capital gains and the consistent disagreement concerning the effects of tax provisions on taxpayer behavior and tax revenues, there have been few studies covering capital asset transactions by holding period and/or asset type. Previous Statistics of Income studies were conducted for Tax Years 1936, 1958, 1962, 1973, 1977, 1981, and finally, the study on which this paper is based -- 1985. Designers of these studies have had considerable difficulty both in capturing, testing, and reporting holding period information and in classifying transactions by types of assets.

Cooperative Approach to Data Problems

In 1986 the Office of Tax Analysis and the Joint Committee On Taxation requested that the 1985 study be conducted with a new, intensified, and cooperative approach to solve what seemed to be unsolvable data problems. SOI accepted this challenge, and worked closely with Treasury and Joint Committee staff to assure the success of the 1985 Capital Gains Study. The next sections of this paper describe the sample design, processing methodology, and asset categories used for the 1985 study.

METHODOLOGICAL IMPROVEMENTS

Sample Selection

The 1985 Sales of Capital Assets Study was conducted using a representative subsample of returns selected for the 1985 Statistics of Income Individual/Sole Proprietorship Program. The complete SOI individual sample for 1985 was comprised of 121,480 returns, selected from a population of 101,836,347 returns.[1] The subsample for the 1985 SOCA Study was comprised of 56,649 returns--46.6 percent of the complete individual sample.

SOCA returns, including the supplementary data picked up from the individual income tax returns (in this case transactions and holding period detail), were weighted to estimate National totals by dividing national population frequen-

cies for each sample stratum by the number of SOCA returns in the stratum. To assess whether the SOCA subsample was representative, a comparison chart was prepared showing the weighted data for the complete 1985 SOI sample and for the 1985 SOCA sample. The comparison chart presented the number of returns and adjusted gross income (AGI) by AGI class and showed that the distributions of returns and income across AGI classes are similar, when both samples are weighted.[1]

Processing the Data

Returns selected for Statistics of Income samples routinely are subjected to additional data capture and editing after the regular revenue processing has been completed. More line items are picked up and totals are balanced to correct taxpayer reporting errors. Generally, SOI editors make corrections/changes to resolve data differences which occur.

This extra handling is necessary to include those items which are important for statistical research but may not be needed for administrative purposes. Furthermore, returns for the Capital Assets Study and other special studies often undergo additional processing, which focuses on the specific detail available and the needs of the users.

SOI processing has long been basically a manual process, with trained "editors" extracting data from the returns onto transcription forms, which are then keyed into the computer. Several "loops" are usually required before manual processing of a return is completed. Needless to say, this two-step process has great potential for introducing delay and additional error.

The processing system designed to capture the SOCA data was the result of intensive discussions between SOI personnel and the project sponsors, the Office of Tax Analysis (OTA) and the Joint Committee on Taxation (JCT). Previous SOCA studies had raised concerns relative to the problems of taxpayer reporting on the complex SOCA schedules, difficulties of "error resolu-

tion" of taxpayer data, and the accuracy of data capture.[2] For the 1985 Study, OTA expressed a need to see the uncorrected taxpayer data for most fields.

SOI's First On-Line Editing System

The 1985 SOCA study was designed as SOI's first major on-line real time editing system, using a new computer system and relational data base software. The editing system was planned to thoroughly test the data as it was transcribed directly from the returns, starting from the smallest conceptual unit of information. In addition, quality review and management reporting systems were integrated into the database design. These features of the system were menu driven.

An editor initiated transcription of a SOCA return by entering identifying information into the computer. After a search of the database, additional return information from a control file was displayed to the editor, in order to ensure that the proper return had been selected for edit. The data entry screens were designed to present the SOCA tax forms in a systematic order.

The order of data editing was set up so that information entered on the supporting schedules could be used to test the data entered on the main SOCA schedules, Form 4797 and Schedule D. Editors were required to check for the presence or absence of each schedule as they moved through the data entry screens. The system had provisions made for capture of multiple occurrences of schedules (for example, more than one Form 6252), as well as allowing for multiple taxpayer transactions (e.g., several short-term sales of corporate stock). In areas where there was a particular concern over the accuracy of the data capture, editors were prompted by the system to double check their work. One example of such checking was in the application of asset categories. When an editor entered an asset code, the system displayed the name of the asset and asked the editor to verify his or her entry. An example of a Schedule D editing screen is provided in Figure 1.

Figure 1.--Sample Editing Screen for Schedule D

SSN 999999991 TXPD 8612		SCHEDULE D SALE OF CAPITAL ASSETS				Schedule D Present Y SHORT TERM	
Part I, Short-Term Part Present		Is This The Correct Asset Code - Enter The First 20 Characters					
Starting SEQ #							
SEQ # (A)(B)	(C)	(D)	(E)	(F)	(G)	(TC)	
1 2	XX/XX/XX XX/XX/XX	2345	1234		0	1111 2	
2 3	XX/XX/XX XX/XX/XX	1234	345		0	889 2	
3 5	XX/XX/XX XX/XX/XX	7865	4544		0	3321 2	
Part I(a) Present							
		3. Form 2119, Lines 6 or 12.....3.					597546
		4. Form 6252, Lines 22 or 30.....4.					2345
		5. Partnerships, S. Corp., Fig. 5					1254 4362
		6. Loss Carryover after 1969.. 6.					3456
		7. Add Lines 2a thru 6 Col. f/g 7.					4690 609744
		8. Combine Columns (f) and (g) Line 7.....8.					605054
Char Mode: Replace Page 1		Count: 1					

Verification and Testing

Data entry was also verified by real-time consistency testing of the data entered in each form. That is, the SOCA system performed the same mathematical operations that the taxpayer was instructed to do when completing the tax forms. If data entered by an editor failed a mathematical test, the SOCA system prompted the editor to check the entry. In the event of a data transcription error, the editor was instructed to re-key the data. In a more complex situation involving a taxpayer reporting error, special computation or other data discrepancy, the editors were instructed to consult with a supervisor. Some common types of taxpayer reporting problems are:

- Reporting transactions on the wrong forms--for instance, sales of principal residences or installment sales which should not be shown on Schedule D;
- Incomplete or missing acquisition or sales dates for transactions; and
- Gains or losses that do not balance to the reported cost subtracted from the sales price.

Error Bypass Methods

As a rule, in SOCA editing data were not "error resolved" using the traditional SOI method, which forces totals to balance to detail items. Instead, for SOCA a special consistency test "bypass" function and bypass acknowledge system (BAKS) were developed, which allowed the editors to leave the data as reported by the taxpayer. A careful attempt was made to transcribe the taxpayer's entries as accurately as possible. If a consistency test failed during data transcription, the bypass system wrote an audit record to the SOCA database, with complete information about the field bypassed, the edited field and computed values, and a comment field for the editor to indicate the reason for bypass.

The BAKS system was designed, at Treasury's request, to eliminate distortions caused when the taxpayer's original entries were deleted or recomputed during data processing. It should also be noted that some methods of "error resolution" can be applied to the raw data in the file at any time after original transcription. The bypass system audit records maintain explanations of the nature of the field error on the file, allowing a researcher to understand the problem caused by the taxpayer's entry, without having to refer to the return.

The primary advantage of the BAKS system is that it permits users of the data to choose which pieces of information they need for analysis. To investigate taxpayer behavior, for example, a user may wish to select the "raw" SOCA data. If data records need to balance for econometric analysis, the "computed" data may be selected, by matching the SOCA bypass records to the appropriate SOCA schedules. The bypass records also allow data users to investigate "outliers" or unusual cases, by examining the comments the editors have placed with the data. The edited data were also subjected to an "edit verification" program. This batch program re-consistency-tested the SOCA data relationships that "carried over" from form to form --

for example, data transferring from Form 6252 to Schedule D -- so as to assure that a taxpayer's data are correctly represented throughout the SOCA record.

Asset Classification

Asset transactions are reported by taxpayers on different schedules, depending on the type of asset being sold, method of sale, whether the asset is classified as an ordinary or capital asset, and whether the taxpayer chooses to treat sales of assets under certain provisions of the Internal Revenue Code.[3] Forms and schedules used by taxpayers to report asset sales and the transactions reported on them are listed below in Figure 2.

Figure 2.--Forms Used for Capital Gains Asset Transactions

Form or Schedule:	Used to Report:
Form 2119	Sales of principal residence Installments sales Sales of "Sec.1231" and other capital assets and also sales of "ordinary assets"
Form 6252	
Form 4797	
Schedule D	All other asset transactions and summarizes capital gains or losses for entry on Form 1040

The classification of asset transactions reported by taxpayers is crucial, since policy alternatives may be based upon the types of transactions reported by taxpayers. Twenty asset categories were designated for the 1985 SOCA study. The categories were designed to be non-overlapping and as precise as possible. The categories are listed in Figure 3.

Figure 3.--SOCA Asset Categories

1. Corporate Stock
2. U.S. Government Obligations
3. State and Local Government Obligations
4. Other Bonds, Notes and Debentures
5. Put and Call Options
6. Commodity and Other Futures Contracts
7. Tax Exempt Mutual Bond Funds and Trusts
8. Capital Gains and Losses from Partnerships, Fiduciaries and S-Corporations
9. Capital Gains and Losses from Regulated Investment Companies and Mutual Funds (includes Capital Gain Distributions)
10. Livestock
11. Timber
12. Involuntary Conversions
13. Residential Rental Property
14. Depreciable Business Personal Property
15. Depreciable Business Real Property
16. Land Other than Farmland
17. Farmland
18. Principal Residence
19. Other Assets
20. Unidentifiable Transactions

Asset coding of the SOCA returns was handled in a two-stage process, in order to ensure quality of data entry. Asset codes were marked onto the photocopies of the returns and reviewed prior to on-line data entry. As described above, the editing system also prompted the editor to verify the asset code after it was entered.

The two codes at the end of the list, "Other Assets" (19) and "Unidentifiable" (20) warrant special discussion. Previous asset coding experiences had indicated a need to closely examine assets that editors were not able to place in a distinct category of property. For the 1985 SOCA study, whenever an editor entered an asset code of 19 or 20, the editing system prompted the editor to enter the asset description the taxpayer had written on the form or schedule, so that reclassification of assets was possible, in case of editor error.

While asset code 19 was used to record transactions such as disposition of bad debts, patents, and collectibles, asset code 20 was especially designated for descriptions of assets that were incomplete, unreadable, or blank. The typed descriptions of these assets were reviewed during editing, in order to further minimize editor error.

Over 400,000 transactions (unweighted) were transcribed for the 1985 SOCA project. Unidentifiable transactions comprised less than two percent of the total number of transactions reported.

Quality Review

After the SOCA data were keyed, further quality review was performed to ensure accuracy of the file. This was done for each editor by dumping a random sample of returns edited from the database. An independent quality reviewer, then, checked the prints of the edited data against photocopies of the actual tax returns.

This process yielded two benefits:

- First, questions or problems that were identified were referred to the unit supervisor, and frequently were identified to the SOI staff who developed the consistency tests for the field editors.
- Second, although editors were very sensitive to this systematic check of their work, we know from results when similar techniques were used in revenue processing units, that such checks improved the standards and quality of a process.

During the initial asset coding there was a 100 percent review of all coding. In addition, editors were required by the system to give 100 percent review of the asset code entered. This extensive asset coding review was designed to support and emphasize the difficulty of the task and the need for great accuracy.

The error bypass system, described above, also included a hierarchical review procedure, where all bypassed records were automatically produced in BAKS output for senior staff review. This BAKS review triggered questions for the system designers, as well.

Results of New Methodologies

As a result of the innovations described above, the 1985 SOCA File provides newer, faster and more complete data than has been possible in the past. The quality controls built into the system also ensure that the resulting file is of better quality, as well as permitting more versatility for policy researchers.

Data Synopsis

Early results of the data analysis reveal that for 1985 there were 39.9 million capital assets transactions, an increase of 39 percent over the 28.6 million transactions reported for 1981. As in 1981 and earlier studies, corporate stock transactions represented by far the largest asset category, with 60 percent of transactions in 1985 and 48 percent in 1981 falling into this group.[4]

The 1985 study shows \$171 billion in capital gains and \$29 billion in losses. This contrasts with \$152 billion in capital gains and \$29 billion in losses for 1981. The new study approaches and the emphasis on data capture apparently realized considerable benefit in reducing unclassified assets, decreasing the number of unclassified transactions from 13.5 percent of all assets for 1981 to less than two percent for 1985.

Data Presentation

Because of space considerations, data presented here are limited to one summary table which shows transactions, gains, and losses by asset type (see Table 1). A more extensive presentation by form type, asset type, and holding period is included in the latest volume of the IRS Methodology Reports series; Statistics of Income and Related Administrative Record Research: 1988-1989. [5]

FUTURE SALES OF CAPITAL ASSETS STUDIES

Although the news media and economic researchers have given more projections and advice regarding capital gains than about any other part of the tax law, what solid information has backed up such advice? In most cases only obsolete or incomplete information has been available. The summary data provided here and the detailed tables in Statistics of Income and Related Administrative Record Research: 1988-1989 represent the first major release of detailed data of capital assets transactions and holding periods since the Tax Year 1981 data were released.[5] The importance of more recent data is amplified by changes in investment behavior, in general, and by changes in the way many taxpayers respond to revisions or potential revisions in the tax law.

The 1985 Sales of Capital Assets study represents the start of a multi-year study of taxpayer behavior. Present plans include continuation of transcription of capital assets data for a panel of taxpayers into the 1990's. The Office of Tax Analysis of the Treasury selected approximately 13,000 returns for inclusion in the SOCA panel, and selections for this panel are made each processing year at a

Table 1.--Long Term and Short Term Capital Gains Transactions, Gains, and Losses by Tax Form and Asset Type

[Money amounts in millions of dollars]

Asset Type*	Schedule D and Form 4797				Form 6252	
	Number of Transactions	Net Gain	Number of Transactions	Net Loss	Number of Transactions	Net Gain
Total.....	22,851,460	\$139,188	11,818,689	\$28,645	2,515,400	\$30,283
Corporate Stock.....	14,228,069	65,848	7,025,237	\$14,447	180,032	9,613
U.S. Government Obligations.....	144,544	421	191,965	138	38	**
State & Local Government Obligations.....	762,386	1,529	385,474	612	499	19
Other Bonds, Notes, and Debentures.....	323,901	974	210,831	561	9,041	65
Put & Call Options.....	1,485,490	1,867	1,365,645	2,162	630	26
Commodities & Other						
Futures Contracts.....	132,811	1,453	246,785	1,137	8,099	19
Tax Exempt Municipal Funds.....	288,132	250	176,553	282	320	37
Partnerships, Fiduciaries & S Corporations.....	1,507,271	441	435,599	1,166	54,142	1,272
Mutual Fund Companies.....	723,345	1,254	500,977	490	73	8
Livestock.....	745,552	2,001	220,074	251	12,670	92
Timber.....	57,280	462	12,550	31	5,209	35
Involuntary Conversions.....	51	1	32	**	11	**
Resident Rental Property.....	627,389	13,446	109,793	971	881,319	5,560
Depreciable Business						
Personal Property.....	705,029	2,857	211,726	442	108,913	281
Depreciable Business						
Real Property.....	269,137	10,128	47,251	703	322,481	4,561
Land (Not Farmland).....	372,164	6,943	74,906	481	480,639	4,947
Farmland.....	19,527	726	8,223	308	172,149	1,051
Residence.....	14,458	264	3,964	1	35,682	181
Other Assets.....	463,026	5,960	431,852	3,685	68,752	1,055
Unidentifiable Transactions.....	431,014	4,044	159,252	775	174,702	1,464

* For definition of terms, see [4].

** Less than \$1 million.

centralized IRS computing center, based upon social security number and Tax Year criteria. Currently, data from over 13,000 Tax Year 1987 returns are being processed, using the SOCA data capture system described in this paper.

The continuity of data available in following this capital gains panel through several years is particularly important for studying the results of the Tax Reform Act on capital investment behavior. By far the most dramatic change in the 1986 Tax Law was the elimination of the 60 percent capital gains exclusion. The SOCA panel data for Tax Year 1986 will provide detailed information on the taxpayer behavior (including asset and holding period details) in anticipation of the elimination of the exclusion, and the Tax Year 1987 panel file will show results of capital gains behavior under the new tax law.

Because individual taxpayer decisions from year to year regarding sales of assets are likely to be highly correlated, many issues in capital gains tax policy, and many hypotheses about taxpayer behavior require data analysis over multiple tax year periods. Consequently, plans call for the linkage of the SOCA data files over time.

The SOCA data files may also be linked to additional IRS files, in order to examine other facets of taxpayer behavior and whether there is any systematic correlation to taxpayer capital gains behavior. At the request of the Office of Tax Analysis, SOI has constructed files of Form

W-2, dividend, interest and other information to compare data on assets held by taxpayers with data on assets sold for SOCA study members.

These studies will be most useful for estimating the impact of future tax law changes involving sales of capital assets.

ACKNOWLEDGMENTS

The authors are grateful to Wendy Alvey for her assistance in the preparation of graphic materials and editorial advice. She assisted the authors both for the presentation and in preparing the paper.

NOTES AND REFERENCES

* At the time of this paper, the three authors worked for the Statistics of Income Division of the Internal Revenue Service. Since then, Dan Holik has gone to work for the Joint Committee on Taxation, U.S. Congress, and John Labate is working for Fortune Magazine, both potential users of the SOCA data. Questions about the paper should be directed to Susan Hostetter at the above address.

[1] A complete description of sampling procedures is contained in Statistics of Income...1986 Individual Income Tax Returns, Internal Revenue Service, Pub.1304.

- [2] See, for one example, a discussion of capital assets data quality in How Taxes Affect Economic Behavior, Henry J. Aaron and Joseph A. Pechman, eds., The Brookings Institution, Washington, D.C., 1981, pp.281.
- [3] See "Sales and Other Disposition of Assets," Internal Revenue Service, Pub. 544, for additional information on asset classification.
- [4] For a description of the 1981 Sales of

- Capital Assets study, see "Sales of Capital Assets, 1981 and 1982," by Bobby Clark and David Paris, in the SOI Bulletin, Volume 5, Number 3, Winter 1985-86, Internal Revenue Service.
- [5] Dan Holik, Susan Hostetter, and John Labate (1990). "The 1985 Sales of Capital Assets Study," Statistics of Income and Related Administrative Record Research: 1988-1989, Internal Revenue Service.