MEMBERSHIP IN A LINKED PANEL OF INDIVIDUAL TAX RETURNS: REVIEW AND RESULTS

Susan C. Hostetter, Internal Revenue Service
Statistics of Income Division R:S:P, PO Box 2608, Washington, DC 20013-2608

KEY WORDS: Capital gains, measuring change, linking units

In 1985 tax policy economists at the U.S. Treasury and Congress's Joint Committee on Taxation requested Statistics of Income (SOI) Division in IRS to design and produce a panel of individual income tax returns for which detailed data on capital gains transactions would be collected each year. Why did they request a panel, and why the emphasis on capital gains data? For several years Treasury staff had used a small panel of returns with limited income and tax information for modelling and analysis. As a result, they realized that, for some policy questions, it was more than important -- almost necessary -- to review year-to-year change at the individual taxpayer level. This need was particularly true for policy analysis relating to capital gains (Holik et al, 1989).

Not only did these two major SOI customers request the new capital gains panel, they reviewed the design, the processing methods, and later heralded the success of the data. Particularly for Treasury staff, the level of involvement conveyed the importance of the new 1985 Sales of Capital Assets Panel.

The focus of this paper is the review and perfection of panel unit links within and across years of the survey. Why is the linking and careful review so important? Because the accuracy of measuring change at the individual level (the reason for having a panel) is dependent upon correctly identifying and linking individuals to the correct panel unit and linking the panel units correctly across years.

The importance of the panel and accuracy of review and perfection are heightened, in this case, by the subject area -- capital gains. No area of tax law has changed so much; nor has any caused so much controversy as capital gains. To add to the controversy, wealthy taxpayers are primarily responsible for the volume of capital gains, and tax law, politics, and news copy are all more controversial when discussing "the rich."

This paper will cover:

- Historical perspective
- The 1985 Sales of Capital Assets Study
- Identification and tracking individuals
- Development of review criteria
- Expected results and future plans

HISTORICAL PERSPECTIVE

Capital Gains Taxes -- Continually Changing

Capital Gains are gains from the sale of assets, which, for tax purposes, include all property held for personal use or investment. Examples of such assets today are personal residences, furniture, automobiles, stocks and bonds.

Taxation of capital gains has varied greatly and caused considerable controversy since the first U.S. income tax in 1913. In fact, there have been more changes to capital gains tax laws than to any other single area of income tax law. 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 type of income. During the 1920's the concept of different tax rates for different holding periods (the length of time you own the asset) was initiated, but only for gains taxed under the alternative tax rate. Assets held for a longer period of time were taxed at a lower rate than those held short-term.

From 1942 to 1969 capital gains on long-term assets (held over 6 months) were taxed at half the rate for ordinary income, with a 25 percent maximum rate. Beginning in 1970 gains over $50,000 were taxed at the higher rate of 35 percent. This, plus the inclusion of the minimum tax, increased the maximum effective tax for capital gains to about 40 percent. Holding periods were extended to 9 months and then to 1 year in 1977 and 1978, respectively.

The Revenue Act of 1973 reduced maximum

rates on long-term capital gains to 40 percent of rates on ordinary income. This change reflected a belief, developed during the 1970's, that tax rates on capital income were too high and would discourage growth in capital formation and investment. Some economists predicted that reduced tax rates on capital gains would actually increase tax receipts by stimulating capital asset transactions. Many economists were still echoing these views in the late 1980's, but others disagreed (Congressional Budget Office, 1991).

By 1979 the highest capital gains tax rate was 28 percent -- 40 percent of the 70 percent maximum rate on ordinary income. This is interesting, given today's experience, where we also have a maximum 28 percent tax rate for capital gains, but it has a different effect, since it is 85 percent of the maximum rate on ordinary income.

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 long-term holding period for capital assets from 1 year to 6 months -- back to what it was prior to 1977. Finally, the Tax Reform Act of 1986 completely changed the rules. It stripped all exceptions for taxing capital gains, and treated them as ordinary income, regardless of holding period. Taxation of capital gains had come full circle. This has not, however, stopped the debate, the objections, the suggestions for change, and particularly not the intensity of the discussion surrounding taxing capital gains.

Previous Capital Assets Studies

Considering the extensive and frequent change in tax law for capital gains, the consistent disagreement concerning the effects of tax provisions on taxpayer behavior and tax revenues, and the intensity surrounding the disagreements, there have been few studies covering capital asset transactions by holding period and/or asset type. Although IRS conducted studies for Tax Years 1936, 1958, 1962, 1973, 1977, 1981, and finally, the study on which this paper is based, 1985, the primary users of these studies -- Treasury's Office of Tax Analysis and the Joint Committee on Taxation in Congress -- continue to request improvements for classifying transactions by type and for capturing purchase and sales price and holding periods accurately.

Taxpayers report capital gains, or losses, to IRS when they sell an asset. This sale is called a transaction, and each transaction is reported separately on the Schedule D of Form 1040. Capturing these transactions data accurately is difficult and expensive. Some taxpayers have several thousand transactions, and many have considerable numbers of them. Partly because their own records are incomplete, and partly through error, taxpayer reporting of these data shows pervasive errors and missing data. So, accurate purchase and sale price and/or dates were, and still are, inconsistent at best. The type of asset sold is also difficult to determine, and methods for classifying were less than systematic in early surveys.

THE SOCA STUDY AND PANEL

Goals for the 1985 SOCA Panel

In 1986 Treasury and the Joint Committee requested a new study beginning with 1985. Survey designers from SOI and these organizations collaborated to design a comprehensive study that would meet the needs of tax policy modelers into the 1990's. The important goals identified for the new study by our primary users were to:

- Maintain, with some added sophistication, a representative sample selection methodology that would give reliable estimates of taxpayer behavior regarding the sales of capital assets.
- Change the processing standards and methodology to capture and test data defining holding periods more systematically and accurately.
- Develop a new, systematic methodology for classifying transaction types.
- Collect the capital assets transactions data annually for a panel of returns (a subset of the initial study). The panel should support national estimates of change for key assets.
- Design a data processing and data entry system that would enhance the accuracy of the review and data entry process. The system should be designed to retain taxpayer-reported values when data were corrected.

The designers of the 1985 SOCA Study placed major emphasis on developing the panel, so that

policy analysts would have an effective tool for measuring change from year-to-year at the individual taxpayer level. And, because in this capital gains panel we looked at individual transactions, treatment of gains and losses were automatically separated. Analysis at this new return-level and transaction-level insured that a large change in one area would not obscure that in another. Compare this to previous methods where only one year's data were available at a time and gains were measured by subtracting losses from gains if gains were larger, and losses were measured by subtracting gains from losses if losses were larger. This netting effect loses much of the specific activity and the magnitude of both gains and losses. (For example, if a taxpayer has $5 million in gains and $4.5 million in losses, the netting effect would show $500 thousand in gains.) For further discussion of collaborative sample design efforts see Hostetter and O'Conor, 1991.

The Base-Year, 1985 Sample

The 1985 Sales of Capital Assets (SOCA) Study was conducted using a representative subsample of returns selected from the 1985 Statistics of Income

Table I.--Sample Size for SOCA Studies

Study/Population	Number of Returns
1985 Population	101,836,347
1985 SOI Ind. Program	121,480
1985 SOCA Study	56,649
1985 SOCA Panel	12,980

Individual/Sole Proprietorship Program; the SOCA Panel was selected as a representative sub-sample of the 1985 SOCA Study (Internal Revenue Service, 1986). Capital gains transactions were reported on over 60 percent of the 1985 returns. Table I shows the relationship of these studies.

For both the 1985 SOCA Study and all years of the SOCA Panel, editors captured the purchase price and date and the sales price and date on all returns for which capital gains transactions were reported. Generally, data reported on Schedules D include many omissions and inaccuracies. Purchase price and/or date is frequently omitted; purchase and sale dates are reversed; the description of the asset is missing, illegible, or so vague as to be uncodable; and taxpayers, generally, have a myriad of other omissions and errors. Sometimes there are notes on returns that more fully describe some of these characteristics. There was considerable discussion among the survey designers about how to systematically treat these issues. Users felt strongly that they did not want to lose information, even erroneous information, through the editing process.

Each transaction was coded as to the asset type, such as corporate stock, put and call options, residential rental property, etc. The choice and treatment of asset types was strongly influenced by Treasury staff. Efforts to improve asset classification included: reducing the number of asset types from 30 to 20 and organizing types in clusters; having discrete codes for "missing" data versus "unable to code" data; conducting all review at a single site; providing intensive initial training with follow-up training and continued National Office oversight; and standardizing and tracking the resolution of unique coding issues.

The SOCA Study was the first on-line editing system for SOI, developed to improve the data-entry and editing quality of the 1985 SOCA data. The other hurdle the new, on-line system was expected to overcome was timeliness. One of the ongoing problems/challenges for capital gains studies is that the asset coding and editing are labor- and time-intensive. This is the major reason the study has always been processed outside the schedule of the basic Individual Program, which has a rigid production schedule. For a periodic and smaller study, such as the SOCA it is hard to generate the continued level of attention necessary to meet deadlines, such as you can for a mainline annual program. Thus, many of the earlier SOCA Studies were a lengthy process, which, in turn, inconvenienced our customers. The on-line system; new, well-defined asset types; new, structured editing guidelines, backed up with detailed manuals; and extensive training provided to support all of these features finally contributed to a successful, and more

timely 1985-based study.

SOCA Panel Characteristics

A Panel subsample of 13,000 returns was selected from the 56,649 returns in the 1985 periodic SOCA Study. All social security numbers (SSN's) reported on the 12,980 original Panel returns are included annually on a list file used to select returns for the panel. IRS did not require taxpayers to report dependent SSN's until Tax Year 1987, so the SOCA Panel is limited to primary and secondary SSN's. Any return reporting these SSN's as the primary or secondary taxpayer will be selected and included in regular Individual Program processing. Following that processing the returns with capital gains transactions are sent to the Cincinnati Service Center for SOCA processing -- the coding of assets and transcription and editing of asset types, prices, and holding periods.

One new twist was added for the 1985 periodic SOCA Study that is quite unusual for SOI. Because the study was begun late in the processing year, some returns that we knew had capital gains transactions could not be physically obtained for transcription and editing. These returns were treated as refusals and nonresponse adjustments were made for their absence. As for the base-year study, the subsample Panel could be weighted to produce estimates of the population.

PANEL MEMBERSHIP LINKS: REVIEW AND CORRECTION

Importance of Accurate Panel Links

The primary reason for developing longitudinal data is to measure change at the individual unit level as opposed to measuring net change of aggregate data. Panels are expensive to design, to select, and particularly to maintain. However, tax modelling experts are convinced of the value of these data for policy analysis. To measure change accurately you need very accurate unit identification throughout all years. Furthermore, those units must be linked accurately across years, because the links are the basis for measuring change. Accurate weighting adjustments necessary to reflect changes in sample unit makeup are also essential for accurate measurement of change. Since maintaining the panel membership file is one of the largest expenses of using panel data -- and certainly an important one -- SOI developed a review process to assess the accuracy of the panel units.

Decision Model Based on Previous Review

In 1992 SOI completed the review of 331,000 returns, covering three years of panel returns from a large (90,000 units) panel of individual returns. (Unlike the SOCA Panel, this basic Individual Program Panel was initiated for Tax Year 1987 and included the SSN's, and therefore panel membership, for taxpayer dependents.) The returns were mechanically screened, using 15 sets of rigid screening conditions, to classify about 180,000 returns as "error-free." The remaining 150,000 returns were manually reviewed, based on an additional 18 sets of screening conditions. Using this experience, SOI staff examined a small sample of SOCA returns to determine the most valuable conditions for their review. Because the SOCA Panel is much smaller and covers only primary and secondary taxpayers, the review criteria could be less complex than for the larger panel. We found much of the error and manual review centered around the dependent SSN's on the large panel. (For additional background on the design of the basic Individual Program Panel and its clean-up process see Hostetter, 1992.)

Review Criteria

Using what was learned from the 1992 review of the basic Individual Study, a SOCA review was undertaken. The following basic definitions are important to discussing the review and matching process:

- **Panel Unit** - The originally selected return, and the individuals defined by the primary, or primary and secondary, SSN reported on it.
- **PSSN AND SSSN** - The primary SSN and the secondary SSN.
- **MFS** - A return with a married filing separate filing status. In such cases, for each Tax Year, there should be two returns with each reporting the other's SSN as the secondary.
- **Name Control** - The first four letters of the last name reported on the tax return. These are transcribed from the return for the primary taxpayer. SOI also has the name control

associated with the SSN when it was assigned by the Social Security Administration (SSA).

The review process included all returns filed with Panel SSN's for Tax Years 1985 through 1990. Panel units were linked across all six years. This link was preliminary, subject to correction based on the review. If a potential error or reweighting condition was identified, data for all years were displayed for review. If a Panel unit was included for review in any group, it was reviewed for all potential error or reweighting conditions. Following is a list of conditions identified for review, only some of which were actually errors. The review conditions also include changes in the scope of the Panel unit that, in many cases, require an adjustment to weighting:

- The primary and secondary SSN's on a return represent different original Panel units. **Individuals from two different returns in the base year are now filing jointly on the same return representing a merging of sample units (e.g., two members from different panel units married).** Requires weight adjustment.
- Any return with a MFS filing status. **There are few of these returns and they need to have the matching return from the spouse to comprise a full reporting unit. Since individuals filing MFS returns frequently change filing status in subsequent years, we felt these returns would benefit from a thorough review for sample unit continuity.** Potential for reweighting.
- A Panel SSN appears on two returns. **This is a potential duplicate or error.** Potential for deleting a return.
- The name control for the name reported on the return is different from the name control that SSA associates with the same SSN. **Probable reporting or transcription error in the SSN.** Probably will either correct the SSN or delete the return.

There were some other conditions that we considered reviewing for the SOCA Panel, such as marital status changes or nonpanel members on the return of a panel member through joint filing. However, neither of these conditions alone represents an error; in fact, they represent expected demographic changes. Because of our experience in reviewing the large Individual Panel, we felt confident in limiting our review to the four conditions described above. Our previous analysis of the much larger panel indicated that real errors related to these conditions would be identified by nonmatching name controls or duplicate SSN's in almost all cases.

FINDINGS

Because of changes in behavior of individuals selected in the original SOCA Panel, we tend to see an increase in panel units over the years. The following table shows the total number of units in the SOCA Panel from 1985 through 1990 and the number of these units that have capital gains transactions.

Table II.--Panel Size and Returns with Transactions

Year	SOCA Panel Size	Returns with Transactions
1985	12,980	8071
1986	13,126	8910
1987	13,316	8912
1988	13,588	8952
1989	13,513	8912
1990	13,519	8786

Although work was initiated on this project, the actual production work has not developed on schedule, and final results are not available yet. However, using the results of the three-year study of the basic Individual Program Panel, we are able to predict approximately the number of returns in the SOCA Panel that will need manual review. We are even able to estimate approximate coding and correction incidence for some review conditions. But, these estimates cover a three-year study period, and we can expect a slightly higher incidence of error within a panel unit over a six-

year review period. Taxpayers have six years, or chances, in which to make an error, rather than three.

Table III shows potential error patterns in SOCA Panel returns based on Tax Year 1989 characteristics.

Table III.--1989 SOCA Panel Characteristics

Characteristic	Number of Returns
Two Units, Same Return	**6**
MFS Returns, Total	**533**
MFS Matched Returns*	**(less) 337**
MFS Returns to Review	**196**
Two Returns, Same SSN	**419**
Name Control Nonmatch	
Joint Returns	**1011**
All Other Returns	**35**
TOTAL 1989 REVIEW	**1667**

* MFS matched returns are a pair of correctly filed married filing separate returns where each taxpayer reports the other person's SSN as the secondary SSN. These pairs would not require review for correction or reweighting.

Our projection for the number of panel units requiring review is actually somewhat higher than the review rate shown in Table III for 1989 -- about 2000 returns. Although these error/review numbers represent one year of data, we have certain expectations concerning the effect these error conditions will have on the quantity of overall review. For example, if there are only six combined panel units (where SSN's of two different panel units appear on the same return) by 1989, we can assume that there will not be many more for the next year either. Most likely, the married filing separate returns needing review will represent some with simple error that can be fixed, and many that show inaccurate reporting methods by taxpayers. These are frequently repeated year-after-year and will be reviewed for all years together.

Duplicates (two returns with the same SSN) are frequently specific to a year, but may also represent repetitive error, caused by copying the previous year's return or unusual errors in filing patterns. When the name control from the return fails to match the name control SSA associates with the SSN, there is almost always an error in the primary SSN on the return. Frequently, on joint returns, we will be able to identify the correct SSN for the primary taxpayer in the second year because the secondary return causes the correct unit to be selected. About 80 percent of returns in this panel are joint returns.

THE FUTURE SOCA PANEL

Plans for the SOCA Panel improvement over the next year or more include the following:

- Review hard copy output for errors and additional coding; code and enter corrections to database; develop and produce tabular output
- Present final results of the SOCA Panel perfection and linkage at the 1994 Winter ASA meetings
- Study methods, confer with Treasury and the Joint Committee, and provide Panel weights for all years -- more than one weight per year to meet multiple needs
- Study Panel replenishment
- Develop a public-use cross-sectional file for the 1985 Study

ACKNOWLEDGMENTS AND REFERENCES

The author is greatful to Wendy Alvey and Beth Kilss for their assistance in the preparation of graphic materials and editorial advice.

Congressional Budget Office, 1991, **Capital Gains Taxes in the Short Run, 1991.**

Holik, D.; Hostetter, S.; and Labate, J. (1989), The 1985 Sales of Capital Assets Study, **American Statistical Association 1989 Proceedings of the Section on Survey Research Methods.**

Hostetter, S. (1992) Exploring Nonsampling Error in a Longitudinal Survey of Individual Taxpayers, **Symposium 92: Design and Analysis of Longitudinal Surveys,** Ottawa, Ontario, Canada.

Hostetter, S. and O'Conor, K. (1991), Satisfying the Need of Income Policy Modelers While Preserving the Reliability of Descriptive Statistics, **Statistics of Income, Turning Administrative Systems into Information Systems: 1991-1992,** Internal Revenue Service.

Internal Revenue Service, **Statistics of Income...1986 Individual Income Tax Returns,** Pub. 1304.