

CHOOSING THE APPROPRIATE INCOME CLASSIFIER FOR ECONOMIC TAX MODELING

Susan Hostetter, John L. Czajka, Allen L. Schirm, and Karen O'Connor
Internal Revenue Service and Mathematica Policy Research, Inc.
S. Hostetter, Internal Revenue Service (R:S:P), PO Box 2608, Washington, DC 20013-2608

KEY WORDS: Individual Income Tax Returns, Tax Policy Analysis, Sample Redesign

The word "INCOME" has many definitions, most of which are controlled by the source of the income information and/or the use of the information. This paper will describe income derived from individual income tax returns and used in the development and administration of tax policy. In 1987, the Statistics of Income (SOI) Division of the Internal Revenue Service (IRS) began a major redesign of its annual sample of individual income tax returns. This resulted in three major changes to the old design: the incorporation of a longitudinal panel with 1987 as the base year, addition of sample units based on the tax family, and development of a new strata design for the annual cross-sectional sample.

All three phases of the redesign were based on needs expressed by the Treasury Department's Office of Tax Analysis (OTA) and, thru OTA, the Congressional Joint Committee on Taxation (JCT) so that they could more accurately model the effects of tax policy recommendations. Incorporating a longitudinal panel met the need to measure the effect of tax policy on individual taxpayer behavior over time, as opposed to measurement based on aggregate change or differences within a cross-section. Developing a tax family unit met OTA's and JCT's needs to model the effect of tax law changes on family or economic units. The strata redesign addressed the twin goals of strengthening the sample of income components which are reasonably thought to be the subject of tax policy interest and of obtaining better coverage for certain demographic groups which are of policy interest.

The first two phases of the redesign have been implemented [1] and results will be included in Tax Year 1988 data. This paper will focus on the third -- the strata redesign. It will cover (1) some income concepts affecting the strata redesign; (2) a description of the new design; (3) the goals and development process used; (4) an evaluation of how well the new design met the goals; and (5) finally, some effects of the redesign on future IRS plans.

HISTORY OF INCOME CONCEPTS

IRS began producing income statistics as early as 1916, using "net income" as its classifier until 1944, when the concept of adjusted gross income (AGI) was first introduced. Net income was defined as positive sources of income less negative amounts (as provided for under the tax law for each particular year). As most of us know, AGI, too, has changed definition frequently over the years -- most dynamically and recently, as a result of the Tax Reform Act of 1986.[2] IRS still produces most of its

statistics based on AGI, although since Tax Year 1985 some tables have also been produced using a broader income classifier that uses all income reported on income tax returns consistently since 1979.[3]

Treasury and the Joint Committee, our primary users, already employ income concepts different from AGI to evaluate tax policy. Of course, in selecting such classifiers, they attempt to overcome the limitations of AGI for their own purposes. In fact, it was OTA that was the primary driving force behind the Individual Program Redesign -- and especially behind the strata redesign. A brief discussion of the income concept they currently use is provided below.

Since the release of Tax Reform for Fairness, Simplicity and Economic Growth in 1984, OTA has used "family economic income" (FEI) as its measure of income in distributional analyses of tax proposals.[4] Family economic income attempts to approximate economists' ideas of income as consumption plus change in real net worth. The classifier is derived by adding imputed data from several non-IRS sources to the Statistics of Income data to develop family units with greatly expanded income. It is independent of prevailing tax law and, therefore, a stronger standard for measuring the consequence of a particular tax proposal or a comprehensive tax overhaul, such as the Tax Reform Act of 1986. OTA's goals in developing FEI are also the basis of the SOI Individual Program Redesign.

The JCT has not embraced Treasury's family economic income concept; instead, they have used a slightly more conservative approach to income, called "expanded income." Although "expanded income" includes some imputed data, such as amounts for employer contributions, it continues to use the return as the unit of measure; and the JCT staff is quick to agree that it, like AGI, is not the ideal income classifier for analyzing tax policy.

DESCRIPTION OF THE NEW INCOME STRATA

The income concepts discussed above describe how the major users of SOI Individual Program data treated the data after they were collected and processed. In this section, the focus is on defining or controlling what returns will be selected in the Individual Program sample.

Income sources for the Statistics of Income sample are limited to information supplied on the Individual Tax Form 1040 and supporting schedules that have been processed by IRS's tax collection system. After returns are received and processed in each of the 10 IRS service centers, each center transmits data for returns completed each week to IRS's computing center in Martinsburg, West Virginia. These data are

further processed, and if the social security number for the primary taxpayer is valid, the data are "posted to the Individual Master File." It is from this file, processed on a weekly basis, that the SOI sample of returns is selected. Fortunately, the revenue processing data from individual returns that are included on the Master File and its supporting Returns Transaction File are extensive. Items from a combination of these files are used to select SOI returns.

Since 1982, IRS' sample of Individual Income Tax Returns has been stratified on the larger of total income or total loss amounts and the size of business plus farm receipts. In addition, the strata were based on the presence or absence of a Form 2555, Foreign Earned Income; a Form 1116, Computation of Foreign Tax Credit; a Schedule C, Profit or (Loss) from Business or Profession; and a Schedule F, Farm Income and Expenses. Sample selection rates could vary based on the inclusion of specific form types. Twenty income or loss variables were used to derive the total income and loss amounts.

Our principal customers at Treasury, however, felt there was room for improvement in the (1) conceptual definition of the income stratifier;

(2) treatment of negative income returns; and (3) selection of returns with certain demographic characteristics. In particular, they wanted to maximize the selection of "interesting" returns for policy analysis purposes. Figures 1-6, below, describe the new selection methodology IRS will use to select returns.

Figure 1 shows the sources of positive income and expense items from these files that are used as selection criteria in the new sample design. Figure 2 shows the same information for negative sources of income.

When redesigning the income strata for sample selection, two exception strata which are not based on the income concept were included in the design to meet specific goals. Less than 1,000 returns will be selected in these two strata. One of these groups is the High Income Nontaxable returns (defined as returns with Adjusted Gross Income in excess of \$200,000 and no income tax), which -- because of a mandatory study -- will be selected with certainty, as in the current sample design.

The other exception stratum is for sole proprietorships or farms with receipts of \$50 million or more. They will also be selected at the 100 percent rate, so as to reduce year-to-year variations in the Bureau of Economic Analysis' aggregate estimates of proprietors' income due to the movement in and out of the sample of a few very large sole proprietorships.

Figure 1.--Sources of Positive Income and Expense Items for Sample Selection.[5]

| <u>Income or Expense Item</u> | <u>1989 IRS Form Where Information Reported</u> |
|-------------------------------------------------------|-------------------------------------------------|
| Strictly Positive Items | |
| Wages, salaries, and tips | (Form 1040, line 7) |
| Taxable interest income | (Form 1040, line 8a) |
| Tax-exempt interest | (Form 1040, line 8b) |
| Dividends received | (Form 1040, line 9) |
| Alimony received | (Form 1040, line 11) |
| Capital gain distributions | (Form 1040, line 14) |
| Total pensions and annuities (taxable and nontaxable) | (Form 1040, line 17a) |
| Taxable IRA distributions | (Form 1040, line 16b) |
| Unemployment compensation | (Form 1040, line 20) |
| Social security benefits (taxable and nontaxable) | (Form 1040, 21a) |
| Potentially Off-setting Items: | |
| Gain Items | |
| Short-term capital gain | (Sched. D, line 7g) |
| Long-term capital gain | (Sched. D, line 16g) |
| Rents received | (Sched. E, line 4) |
| Royalties received | (Sched. E, line 5) |
| Partnership and S Corporation income | (Sched. E, line 32) |
| Estate and trust income | (Sched. E, line 37) |
| Business Income | |
| Schedule C gross profit (positive) | (Sched. C, line 5) |
| Schedule F gross income (positive) | (Sched. F, line 11) |
| Adjustments to Income | |
| Gain from sale or exchange of your home | (Sched. D, line 10) |
| Other gains (from Form 4797) | (Form 1040, line 15) |
| Other income, net gain | (Form 1040, line 22) |
| Farm rental income (positive) | (Sched. E, line 28) |

Limitations of Master File Data

Not all information on the Form 1040 and its supporting schedules is available on the

Figure 2.--Sources of Negative Income and Expense Items for Sample Selection.[5]

| <u>Income or Expense Item</u> | <u>1989 IRS Form Where Information Reported</u> |
|--------------------------------------------|-------------------------------------------------|
| Potentially Off-setting Items: | |
| Loss Items | |
| Short-term capital loss | (Sched. D, line 7f) |
| Long-term capital loss | (Sched. D, line 16f) |
| Deductible rents and royalties loss | (Sched. E, line 26) |
| Partnership and S Corporation loss allowed | (Sched. E, line 33) |
| Estate and trust deduction or loss allowed | (Sched. E, line 38) |
| Alimony paid | (Form 1040, line 29) |
| Moving expenses | (Sched. A, line 19) |
| Business Losses | |
| Schedule C gross profit (negative) | (Sched. C, line 5) |
| Schedule F gross income (negative) | (Sched. F, line 11) |
| Net Income Items (negative) | |
| Other losses (from Form 4797) | (Form 1040, line 15) |
| Other income, net loss | (Form 1040, line 22) |
| Farm rental loss (negative) | (Sched. E, line 28) |
| Deductions | |
| Schedule C expenses | (Sched. C, line 29) |
| Farm expenses | (Sched. F, line 35) |

Individual Master File when SOI returns are selected. For example, it would seem logical to consider unreimbursed employee business expenses as a business deduction; however, this field (Sched. A, line 20) is not included as a separate item on the Master File. Rather, it is combined with other miscellaneous deductions.

Another example is that a capital loss carryover from a prior year is included in the capital loss amounts we use. We would choose to limit the selection amount to current-year income and losses -- and would, therefore, choose to reduce the loss by the prior-year carryover -- but we are unable to identify the prior-year carryover amount at the time of selection.

Differential Factors for Income Items

An additional feature, built into the new design, is that any income or loss item can be assigned a multiplicative weighting factor that is greater or less than one, allowing us to give a specific item more or less importance in calculating the sum of items that make up the income classifier. This will permit us to respond to Treasury's request for "interesting" returns. We have not chosen to use this feature for the initial Tax Year 1990 production, but we anticipate that such flexibility will be valuable in the future.

For example, the greatest single change caused by the Tax Reform Act of 1986 was to the treatment of capital gains.[6] Since then many proposals for changing the tax law have centered around changing the capital gains treatment again. Therefore, if Treasury and the Joint Committee need additional returns that include capital gains in order to make reliable estimates, we could apply a factor of, say, 1.5 to such income or loss. Presumably, this would result in the selection of many more returns with such income in our sample.

Determining the Selection Strata

Under the new design, 24 strata were developed, based on positive and negative income ranges. For each return all positive income amounts listed will be summed, and all negative amounts will be summed separately. A return will be classified as a "net positive return" or a "net negative return," according to which of these two sums has the larger absolute value. The selection amount will be the sum of all positive income **only** for "net positive returns" and, likewise, the sum of all negative income **only** for "net negative returns."

The selection amount determines the strata to which the return is assigned, and, of course, the selection rate. Figure 3 shows the income strata and probabilities of selection (expressed as a percentage) for the new strata design.

Description of the Strata Boundaries. -- Note that negative income strata almost mirror the positive income strata, both in income boundaries and in selection rates. In all cases, the selection rate for a negative income stratum is at least as great as the rate for the positive strata with the same absolute value of income.

There are some differences between the

Figure 3.- SOI Individual Program Income Stratification.

| Stratum | Income Range | Selection Rate |
|---------|--------------------------------|----------------|
| 1 | -\$10,000,000 or less | 100% |
| 2 | -\$ 5,000,000 to -\$10,000,000 | 100 |
| 3 | -\$ 2,000,000 to -\$ 5,000,000 | 50 |
| 4 | -\$ 1,000,000 to -\$ 2,000,000 | 16 |
| 5 | -\$ 500,000 to -\$ 1,000,000 | 4 |
| 6 | -\$ 250,000 to -\$ 500,000 | 1 |
| 7 | -\$ 120,000 to -\$ 250,000 | .4 |
| 8 | -\$ 60,000 to -\$ 120,000 | .25 |
| 9 | \$ 0 to -\$ 60,000 | .1 |
| 10 | \$ 0 to \$ 30,000 | .02 |
| 11 | \$ 0 to \$ 30,000 | .03 |
| 12 | \$ 0 to \$ 30,000 | .08 |
| 13 | \$ 30,000 to \$ 60,000 | .035 |
| 14 | \$ 30,000 to \$ 60,000 | .1 |
| 15 | \$ 60,000 to \$ 120,000 | .08 |
| 16 | \$ 60,000 to \$ 120,000 | .15 |
| 17 | \$ 120,000 to \$ 250,000 | .25 |
| 18 | \$ 120,000 to \$ 250,000 | .4 |
| 19 | \$ 250,000 to \$ 500,000 | 1 |
| 20 | \$ 500,000 to \$ 1,000,000 | 4 |
| 21 | \$ 1,000,000 to \$ 2,000,000 | 16 |
| 22 | \$ 2,000,000 to \$ 5,000,000 | 50 |
| 23 | \$ 5,000,000 to \$10,000,000 | 100 |
| 24 | \$10,000,000 and over | 100 |

positive and negative strata. There were comparatively few returns in negative strata from \$0 to -\$120,000, and Treasury considered all of these "interesting." The selection rates for equivalent positive strata were too low to yield sufficient returns. We used two strategies to resolve the problem. First, we collapsed the income breaks for negative strata to one stratum from \$0 to -\$60,000 instead of the two strata used in this range for positive income returns. And, second, we increased the selection rate for the negative stratum from -\$60,000 to -\$120,000 to .25 percent instead of the .15 percent used for the top positive stratum in this income group.

Identifying Interesting Returns. -- Having increased the homogeneity of the sample strata by separating returns with positive and negative net income, we found that we still needed to do more for low and moderate positive income strata. In these strata (strata 9-18) there are many of the usual people who have wages and some interest or dividends, but not much else. However, in these strata there are also some individuals with moderate-to-complex returns, and there are not enough of them to obtain an adequate sample for tax policy analysis using selection rates based on income alone.

This crucial issue was addressed by developing a second stratifying variable and nesting it in the range of the major income stratifying variable. For the four lowest positive income strata (covering from \$0 to \$250,000), we designed multiple strata with the same income boundaries and different selection rates. Basically, we defined and limited the selection of "uninteresting" returns as a means of selecting more "interesting" returns.

Sub-strata defined to cover less interesting returns were established for each of the positive income strata under \$250,000, and two

such sub-strata for the \$0 to \$30,000 stratum were identified. Below are descriptions of each of the less interesting sub-strata making up the nested stratification. Figure 4 describes the less interesting sub-stratum for low income returns -- basically, Form 1040EZ returns.

Figure 4.-- Definition of "Less Interesting" Sub-Stratum for Low Income Taxpayers. [7]

Stratum 10
(\$0 to \$30,000 -- less interesting) =

Wages, Salaries, Tips + Tax Exempt Interest + Taxable Interest + Dividends = Positive Income
and
Positive Income = Net Income
and
Taxable Interest < \$400
and
Dividends < \$400
and
Total Itemized Deductions = 0
and
Alternative Minimum Tax = 0
and
Filing Status ≠ Head of Household
and
Taxpayer is Not 65 or Older
and
No Dependent Children are At Home
and
No Dependent Parents are Reported

Although this description is designed primarily to cover Form 1040EZ returns, it also excludes returns with dependents and head-of-household status. Clearly, these are considered more "interesting" characteristics. Figures 5 and 6 describe varying income patterns for taxpayers that were standardized to establish sub-strata and limit selection rates for all income strata to \$250,000.

Figure 5.-- Definition of "Moderately Interesting" Sub-Stratum to \$30,000 and "Less Interesting" Sub-Stratum from \$30,000 to \$60,000

Stratum 11
(\$0 to \$30,000 -- moderately interesting) =
AND
Stratum 13
(\$30,000 to \$60,000 -- less interesting) =

Wages, Salaries, Tips + Unemployment Compensation + Taxable Interest + Tax Exempt Interest + Dividends ≥ 90 percent of Positive Income
or
Schedule C Gross Profit + Taxable Interest + Tax Exempt Interest + Dividends ≥ 90 percent of Positive Income
or
Schedule F Gross Income + Taxable Interest + Tax Exempt Interest + Dividends ≥ 90 percent of Positive Income
or
Pensions and Annuities (Taxable and Nontaxable) + Taxable IRA Distributions + Social Security Benefits (Taxable and Nontaxable) + Taxable Interest + Tax Exempt Interest + Dividends ≥ 90 percent of Positive Income
and
Alternative Minimum Tax = 0
and
Total Negative Income (Less Farm Deductions)/ Total Positive Income ≤ 40%

Figure 6.-- Definition of "Less Interesting" Sub-Strata for Higher Income Nested Strata

Stratum 15
(\$60,000 to \$120,000 -- less interesting) =
AND
Stratum 17
(\$120,000 to \$250,000 -- less interesting) =

Wages, Salaries, Tips + Unemployment Compensation + Taxable Interest + Tax Exempt Interest + Dividends ≥ 75 percent of Positive Income
or
Schedule C Gross Profit + Taxable Interest + Tax Exempt Interest + Dividends ≥ 75 percent of Positive Income
or
Schedule F Gross Income + Taxable Interest + Tax Exempt Interest + Dividends ≥ 75 percent of Positive Income
or
Pensions and Annuities (Taxable and Nontaxable) + Taxable IRA Distributions + Social Security Benefits (Taxable and Nontaxable) + Taxable Interest + Tax Exempt Interest + Dividends ≥ 75 percent of Positive Income
and
Alternative Minimum Tax = 0
and
Total Negative Income (Less Farm Deductions)/ Total Positive Income ≤ 40%

Benefits from Separating Negative and Positive Income.

-- The separation of positive and negative values throughout the design and, therefore, throughout the selection process, is a cornerstone of the design. It accomplished several things.

First, there was no blending of gains and losses, such that a person who had \$1 million in capital gains and \$800,000 in capital losses would have each of these values included separately in the income items. Under the previous design, (s)he would have shown only the net (\$200,000) in capital gains for SOI selection criteria. Under the new design individual taxpayers have larger values on both the gain and loss ledgers.

Second -- and related -- such a separation prevents a taxpayer with large income of any kind from having that positive income offset, such that (s)he would be classified in a low (net) income stratum with a low selection rate, and, hence, maybe not selected.

Third, returns with overall net losses are considered "interesting," and this allows us to control their selection rate to make sure it is high enough for policy analysis needs.

Fourth, by grouping income separately, and not allowing offsetting losses, returns generally retain their basic characteristics (or stratum classification) from year-to-year, except when there is actual economic change.

Fifth, as an outgrowth of retaining their sample characteristics, the number of returns that move between strata -- or in or out of the sample -- will be reduced considerably.

Finally, the coefficients of variation on most of the income and tax items that IRS considers important to quality and publication standards improved noticeably.

DEVELOPMENT PROCESS

The process used by IRS and Treasury to develop the new strata design was cooperative and dynamic in approach and, in fact, had a major impact on the final product. The design team was a diverse group. OTA, representing itself and its Congressional counterpart, JCT, (SOI's main customers) participated with four members directly on the team. Two representatives of the statistical contractor on the project, Mathematica Policy Research, Inc., played the major research and designing functions for the team. Finally, IRS staff represented both its own interests and those of its other customers -- most notably, the Bureau of Economic Analysis.

The design process was iterative; a participant would suggest a change, Mathematica would model the results of implementing the suggestion, and the group would evaluate the outcomes based on the goals of SOI and its customers. (See Figure 7.) The results then provided the basis for the next round of suggested changes. This data-driven approach proved to be highly successful, even if lengthy.

Demographic Coverage of Sample

Some of the goals/concerns that OTA expressed were satisfied simply by further analysis of the old design. For example, OTA staff had a concern that older taxpayers be sufficiently covered in the sample. We provided tables showing an age distribution for the population and for the SOI sample. In the SOI sample 68 percent of the returns were for taxpayers 40 or older, while only 44 percent of returns in the population fell into this group. The SOI sample also had a slightly higher representation of older taxpayers than the general population. Many other issues were more elusive and required several rounds of data analysis, with the issue changing slightly for each iteration.

Defining Interesting Returns

The definition of interesting returns evolved as options were explored. Throughout, interesting returns have been defined by defining uninteresting returns. Initially, returns were defined as uninteresting if a substantial proportion of the total positive income (75% in strata 14 and 16 and 90% in strata 10 and 12) came from wage and salary or retirement income. This definition put too

Figure 7. -- Major User Goals in Redesigning the SOI Individual Sample

Office of Tax Analysis Goals

- All positive sources of income should be grouped separately from negative sources, using gross income rather than any net amount throughout the process -- MET
- Returns should be sampled in either negative or positive strata, depending on the overall greater absolute value -- MET
- Negative strata income classes should mirror those for positive income strata, both in boundaries and selection rates -- MET
- No return should be subjected to a lower probability of selection because of the presence of an attached schedule or schedules -- MET
- The sample design should be supported by a sample size of about 95,000 returns -- MET
- Unique strata that provide for selecting returns other than by income class such as the High Income Nontaxable returns for a Congressionally mandated study, should be minimized -- DESIGN INCLUDES TWO STRATA OUTSIDE THE INCOME CLASSES (ONE COVERS A MANDATORY STUDY)
- An increased number of "large" returns (generally considered to have positive or negative income of \$500,000 or more) should be included -- MET
- Year-to-year movement of taxpayers (either between strata or in or out of the sample) due to receipt or incursion of one-time income or loss (such as that from sale of a

home) should be limited -- MET

- The sample should include approximately 20,000 returns from the Social Security Administration's Continuous Work History Sample.[8] -- MET
- Coverage should be increased for several demographic characteristics, such as taxpayers who are older, wealthy, poorer, have dependents, have real shifts in their economic situation, have capital gains, have large incomes offset by losses, etc. -- in other words, returns that were "interesting" for tax policy analysis -- MET
- Implement the new strata design for Tax Year 1990 -- PRODUCTION SCHEDULED FOR TAX YEAR 1990

Bureau of Economic Analysis Goals

- Maintain the current level of reliability, by industry, for the following Schedule C items: Gross receipts or sales; Gross income; Net profit or (loss); and Expense deductions for depreciation, Employee benefit programs, and Interest -- ALL KEY ITEMS HAVE IMPROVED COEFFICIENTS OF VARIATION
- Maintain or increase the number of returns with Schedule C data -- INCREASED NUMBER
- Provide stability of Schedule C returns from year-to-year, such that aggregate change reflects actual economic change, rather than change of coverage in the sample -- MAY HAVE IMPROVED STABILITY DUE TO SEPARATE POSITIVE AND NEGATIVE

STRATA

SOI Division Goals

- Meet the needs of our major customers -- OTA, JCT, and the Bureau of Economic Analysis -- MET
- Provide support for SOI published estimates of income, deductions, exclusions, taxes, and other data that it has produced since 1916 -- MET
- Provide a sample of returns that would be appropriate for developing the SOI Tax Model -- MET
- Have coefficients of variation equal to or better than those calculated under the current design -- ALL COEFFICIENTS OF VARIATION IMPROVED EXCEPT THOSE FOR ADJUSTED GROSS INCOME, WHICH ARE ALREADY VERY LOW
- Be sure to sample returns with the largest values for important income or tax fields -- MET
- Continue to include the High Income Nontaxable returns -- MET
- Build a design/framework that is flexible enough to allow small or moderate changes in design from year-to-year -- those which we cannot anticipate specifically but that we anticipate tax policy issues will call for -- MET; THIS WORK WILL CONTINUE THROUGHOUT THE NEXT YEAR
- Provide adequate structure and mechanisms for conducting related periodic studies such as the Foreign Earned Income/Foreign Tax Credit, as needed -- MET

many returns in the interesting category. So, the definition of uninteresting was expanded to include three types of returns: returns with alternative minimum tax preference items but zero alternative minimum tax; returns which were interesting only because of substantial tax exempt interest income; and returns which had predominately Schedule C income, interest, and dividend income.

This second definition would have reduced the total number of sole proprietors in the sample by an estimated 2,284. This was particularly a problem in the \$0-\$30,000 and \$30-\$60,000 strata because so many sole proprietors were being moved to uninteresting. The estimates IRS provides for the Bureau of Economic Analysis needed more sole proprietors for stability.

A third condition was added to the first two definitions: an uninteresting return was reclassified as interesting if its total negative income exceeds 40% of its total positive income. This will move an estimated 524 sole proprietors back into the sample in the \$0-\$60,000 income range.

Goals and Evaluation

The continual sharing of review and analysis with our major customers provided much clearer insight into their needs and how such needs relate to their analyses. A set of goals was eventually developed, to guide the redesign effort. These, too, were fluid and subject to change during the development process. In fact, many goals were shaped by the Redesign Team, itself.

In developing these goals, it should be noted that different interests were at work. Clearly, when designing a sample for our users at Treasury, one would like to use a stratifier that is highly correlated with tax analysis. So, the major task was to develop stratifiers that would give us the best composite of returns for tax policy modeling, while maintaining or improving the advantages in the current design.

The Bureau of Economic Analysis, on the other hand, uses sole proprietorship data by industry for developing national income and product accounts. Early in the design process we met with their staff and elicited their goals, which SOI staff would represent for them in developing the strata redesign.

Finally, throughout the development process, SOI staff maintained a careful vigil to ensure coverage for its basic statistical needs and those of its other users. An evaluating comment describing the success of the design in meeting each goal is provided in bold face type in Figure 7.

FUTURE PLANS FOR IRS

Next year we will be running the new design in production. We will be designing a reduction in the size of our longitudinal panel (currently about 85,000 "nondependent" returns and about 30,000 returns with dependents). The reduction design will take advantage of research developed

in designing the new strata, and it will include a subsample of returns for the Sales of Capital Assets study. These returns will be sent for further processing to code and list all asset transactions for the Tax Year. It will be implemented for Tax Year 1991.

Other plans include future publication of data including combined cross-sectional and longitudinal samples, as well as some publication of tax family data.

ACKNOWLEDGMENTS

The authors would like to extend special thanks to Wendy Alvey and Beth Kilss, for their editorial assistance on this paper, and to Bettye Jamerson and Clementine Brittain, for their help in copy preparation for both the presentation and the paper.

NOTES AND REFERENCES

- [1] See Czajka, J. (1988), Development of a New Income Classifier for a Sample of Individual Tax Returns, Amer. Stat. Assn. Proc., Sec. on Surv. Res. Meth., and Czajka, J. and Walker, B. (1989), Combining Panel and Cross-Sectional Selection in an Annual Sample of Tax Returns, Amer. Stat. Assn. Proc., Bus. and Econ. Stat. Sec.
- [2] Hostetter, S. (1987), Measuring Income for Developing and Reviewing Individual Tax Law Changes: Exploration of Alternative Concepts, Statistics of Income and Related Administrative Record Research: 1988-1989, Internal Revenue Service.
- [3] Internal Revenue Service (1985), Statistics of Income...1985 Individual Income Tax Returns, Pub. No. 1304.
- [4] Nelson, S. (1987), Family Economic Income and Other Concepts Used in Analyzing Tax Reform, Compendium of Tax Research, 1986, Office of Tax Analysis, Treasury.
- [5] Internal Revenue Service (1989), Package X.
- [6] Holik, D.; Hostetter, S.; and Labate, J. (1990), The 1985 Sales of Capital Assets Study, Statistics of Income and Related Administrative Record Research: 1988-1989, Internal Revenue Service.
- [7] See Czajka, J. and Schirm, A. (1990), Overlapping Membership in Annual Samples of Individual Tax Returns, 1990 Amer. Stat. Assn. Proc., Sec. on Surv. Res. Meth., and Schirm, A. and Czajka, J. (1990) Intertemporal Stability in Total Income and the Overlap in Annual Samples of Tax Returns, 1990 Amer. Stat. Assn. Proc., Sec. on Surv. Res. Meth.
- [8] This sample is the longest running longitudinal sample currently available and is identified by a range of social security numbers. For more information on it, see Buckler, W. and Smith, C. (1980) The Continuous Work History Sample (CWS): Description and Contents, Economic and Demographic Statistics, Social Security Administration.