

THE 1989 SURVEY OF CONSUMER FINANCES A SURVEY DESIGN FOR WEALTH ESTIMATION

**Steven G. Heeringa and F. Thomas Juster, University of Michigan,
R. Louise Woodburn, Internal Revenue Service**

I. Introduction

Researchers have addressed the problem of estimating U.S. household wealth by a number of different methods. National, "macro-level" estimates of the components of U.S. household wealth are available in the National Income and Product Accounts prepared by the Bureau of Economic Analysis (Ruggles and Ruggles, 1982) and from the Federal Reserve's "Flow of Funds" program (e.g., Wilson et al., 1989). "Micro-level" estimates of household wealth have been developed through the use of estate tax multiplier methods (McCubbin, 1987), as direct estimates from household surveys (Projector and Weiss, 1966; Curtin et al., 1989); and by capitalization of income reports from administrative files or tax record systems (Steuerle, 1983). Recently, recognition of the separate strengths and weaknesses of these different methods has led to a call for research into composite approaches which combine the strengths of the individual methodologies (Scheuren and McCubbin, 1987).

This paper presents an overview of the statistical design of the 1989 Survey of Consumer Finances (SCF), placing a particular emphasis on the approaches used to integrate administrative data sources and income capitalization methods into the design stage of this new survey of U.S. household income and wealth. Including the introduction, the discussion is organized into five sections. Section II describes the study objectives and dual frame sample plan for the 1989 SCF. A brief description of the conventional area probability sample component of the 1989 SCF dual frame design is given in Section III. Section IV contains a detailed description of data and procedures used to develop the second and very special component of the dual frame design, a stratified random sample from the 1987 Statistics of Income (SOI) individual data base. The paper concludes in Section V with a summary.

II. The 1989 Survey of Consumer Finances

The 1989 Survey of Consumer Finances is a national study of the financial characteristics of U.S. households covering a wide range of topics including household income, assets, debts, pensions and the use of financial institutions. This new survey is the continuation of a long series of consumer finance studies conducted by The University of Michigan Survey Research Center (SRC) for the Federal Reserve Board, although its special features make it comparable mainly to the 1983 Survey of Consumer Finances. Over time there have been many important developments in the survey methodology for these special studies of income and wealth. Significant among these general advances in survey methods has been the opportunity to implement dual-frame sample designs which incorporate special supplemental samples of households in the upper tail of the income and wealth distribution.

The first survey to include a special supplement of high income households was the 1962 Survey of Financial Characteristics of Consumers (SFCC, Projector and Weiss, 1966). Twenty years elapsed before a sequel to the Projector and Weiss study was fielded. Like its precursor, the 1983 Survey of Consumer Finances (Heeringa and Curtin, 1987) was based on a design which combined a national area probability sample of households and a supplement of high income taxpayers selected from the 1980 Statistics of Income Tax Model data base. The design of the 1989 SCF sample described in this paper has benefitted significantly from lessons learned in the 1983 SCF experience.

The 1989 SCF study specifications call for the completion of approximately $n=2000$ interviews with sample households. Study interviews are conducted in-person and on average last from 60 to 90 minutes depending on the complexity of the sample household's

financial portfolio and on their demographic and employment characteristics. The varied and multi-purpose content of the 1989 SCF interview makes it an extremely useful data set with applications to a broad range of important policy and research questions. The multi-purpose objectives of the 1989 SCF pose a number of complicated and interesting sample design problems.

The larger 1989 SCF data collection program also includes an independent set of n=1800 panel interviews and new cross-section interviews with households originally sampled for the 1983 Survey of Consumer Finances.

III. A 1989 SCF Study Objectives

One class of research objectives to be pursued in the study focuses on household financial characteristics which are distributed evenly among U.S. households. Example analyses from this class of objectives include the investigation of such financial characteristics as annual income, liquid assets, mortgage debt, installment debt, the value of pensions and annuities. This class of analysis objectives is favored by a sampling design in which the sample is allocated proportionately to strata of households with varying income and net worth.

The second of the two general classes of analysis objectives for the 1989 SCF centers on the analysis of financial and non-financial assets which contribute to the household's "wealth" or total net worth. Such wealth-defining assets as stocks, bonds, trusts, real estate, and business holdings tend to concentrate in the upper tail of the household income and net worth distributions. This class of analysis objectives -- the

study of assets and wealth -- argues for a sampling design in which the sample is allocated disproportionately to strata of households with high amounts of income and/or total net worth.

Given these two conflicting objectives, it is clear that the multi-purpose nature of the 1989 SCF presents a problem in the choice of an optimal sample stratification and sample allocation plan. A design plan which is optimal for a household characteristic such as total installment debt may be highly inefficient for studying asset characteristics such as the nature of households' common stock holdings or equity in a privately owned business. The converse is also true: a design that is strictly optimized for the study of wealth may perform poorly for studies of more generally distributed financial characteristics.

III. B 1989 SCF Dual-frame Sample Design

To accommodate these two competing sets of analysis objectives, a dual-frame sample design has been developed for the 1989 SCF. The theory of dual-frame survey design and estimation is presented in Hartley (1974). Heeringa and Curtin (1987) discuss the statistical properties of the general SCF dual-frame design and the comparative strengths and weaknesses of its component frames.

The dual-frame sample design for the 1989 SCF incorporates both a conventional multi-stage area probability sampling of households and a stratified random sampling of medium and high wealth households from a special list frame of U.S. taxpayers. Table 1 (below) outlines the general plan for the

| General net worth stratum | 1983 SCF estimate percent of U.S. households | Approximately net worth range | Total sample | | Area probability sample cases | Taxpayer list (SOI) sample cases |
|---------------------------|--|-------------------------------|--------------|---------|-------------------------------|----------------------------------|
| | | | Cases | Percent | | |
| 1 | 76% | \$0-99K | 950 | 47.5% | 875 | 75 |
| 2 | 22% | \$100-999K | 600 | 30.0% | 252 | 348 |
| 3 | 2% | \$1M+ | 450 | 22.5% | 23 | 427 |
| Total | 100% | | 2000 | 100.0% | 1150 | 850 |

overall stratification of the sample design and the apportionment of the sample size to three general strata and the two sample frames.

The final two columns of Table 1 illustrate the separate roles of the two sample frames in the dual-frame sample design. Representation of households in the lowest net worth stratum will be achieved primarily through the lower cost, high coverage area probability sample component. Conversely, the probability sample from the list frame of taxpayers will bear the burden of representation for the stratum of high net worth (\$1 million and over) households. Both samples will share in the representation of households in the middle range of net worth.

This general sample plan is the result of an extensive program of research into design issues of optimal allocation, weighting and the effects of stratum misclassification. The basic strategy in planning the sample design was to investigate the variance properties of estimates of net worth and major components of net worth that would result from various allocations of the sample to the different

wealth strata, and then to select an allocation that represented the best compromise between the competing objectives. Table 2 summarizes the results of the investigation of alternative 1989 SCF sample designs for measuring net worth and other financial characteristics of U.S. households.

Table 2 describes the effect which optimal allocation for one survey variable has on the precision for other variables of interest in the survey. The statistics in this table are ratios of standard errors and can be interpreted as measures of relative precision for competing sample allocation alternatives. Reading down the columns of the table, the denominator of each ratio statistic is the standard error expected under a design that was optimally allocated for the column variable. The numerator of the ratio is the standard error that is expected for estimates of the column variable for a design that is optimal for the row variable. For example, a design that is optimal for estimating total net worth will result in standard errors for adjusted gross income (AGI) estimates that are 1.17 times greater than expected for a sample allocation which is optimal for estimates of AGI.

| Table 2.--1989 SCF Design: Relative Precision for SCF Analysis Variables Under Optimum Allocations for Design Variable Alternatives | | | | | | | | |
|---|--|---------------|-----------|-------|--------|-----------------|----------------|------------------|
| Design Variable for Optimal Allocation | Relative Precision for Analysis Variables* | | | | | | | |
| | AGI | Liquid Assets | Net Worth | Stock | Trusts | Business Equity | Housing Equity | Installment Debt |
| AGI | 1.00 | 1.07 | 1.33 | 1.65 | 2.98 | 2.75 | 1.01 | 1.06 |
| Liquid assets | 1.08 | 1.00 | 1.22 | 1.49 | 2.65 | 1.45 | 1.08 | 1.26 |
| Net worth | 1.17 | 1.08 | 1.00 | 1.08 | 1.32 | 1.07 | 1.19 | 1.40 |
| Stock | 1.74 | 1.34 | 1.15 | 1.00 | 1.18 | 1.00 | 1.76 | 2.25 |
| Trusts | 1.84 | 1.51 | 1.26 | 1.11 | 1.00 | 1.13 | 1.87 | 2.33 |
| Business equity | 1.74 | 1.32 | 1.14 | 1.00 | 1.19 | 1.00 | 1.75 | 2.24 |
| Housing equity | 1.02 | 1.07 | 1.39 | 1.79 | 3.06 | 2.68 | 1.00 | 1.06 |
| Installment debt | 1.08 | 1.25 | 1.79 | 2.33 | 4.31 | 2.25 | 1.06 | 1.00 |
| 1989 SCF allocation | 1.06 | 1.20 | 1.14 | 1.33 | 1.51 | 1.32 | 1.09 | 1.08 |

*Statistics are ratios of standard errors: Numerator is standard error of column variable under design that is optimal for the row variable.

Denominator is standard error of the column variable under a design that is optimal for the column variable.

The "compromise" sample allocation actually selected for the SCF forms the basis for the final row of Table 2. The ratios in this row represent the relative precision of analysis variables under the 1989 SCF design when compared to a design that was optimal specifically for that variable. These ratios are all greater than 1.0 suggesting that the 1989 SCF allocation is not truly optimal for any members of the set of analysis variables under consideration. The loss in precision (relative to the design optimum) for the individual variables ranges from a minimum of 6% for estimates of AGI to a high of 51% for estimates of holdings in trusts.

Table 3 provides an historical comparison, contrasting the precision of estimates expected under the 1989 SCF design to those obtained under the design for the 1983 Survey of Consumer Finances. The allocation selected for the 1989 SCF actually produces a variance of estimates of net worth that is only slightly higher than observed in the 1983 SCF, despite the fact that the total sample size for the 1989 SCF is only half as large. The chosen allocation also produces estimated variances for common stock and tradeable bond holdings, business equity, real estate investment equity, and trust equity that are lower than the observed 1983 SCF variances for estimates of these assets. Variances of estimates of liquid assets, income, mortgage debt, and installment debt are expected to be higher than in the 1983 SCF.

| Table 3.--Comparison of Standard Errors for the 1983 SCF With the 1989 SCF Sample Allocation (n=2000) | |
|---|---|
| Ratios of Standard Errors (SE) | |
| Variable | RE= $\frac{1989 \text{ SCF SE}}{1983 \text{ SCF SE}}$ |
| AGI | 1.36 |
| Liquid assets | 1.26 |
| Net worth | 0.98 |
| Stock | 0.89 |
| Trusts | 0.60 |
| Business equity | 0.91 |
| Housing equity | 1.40 |
| Installment debt | 1.53 |

The allocation we have selected is clearly a compromise, since it would have been possible to choose an allocation with even lower variances for net worth, common stock and bond holdings, business equity, real estate investment equity and trusts, but only at the expense of substantially larger variances for all the other components of net worth.

IV. Dual-Frame Design: The National Area Probability Sample Component

The national area probability sample of U.S. households for the 1989 SCF will be selected from the Survey Research Center's (SRC) 1980 National Sample design (Heeringa et al., 1986). Under this multi-stage area probability sample design, each household in the coterminous United States receives an equal probability of being selected for interview. By its equal probability nature, the sample that is selected from this sample frame is distributed proportionately to household strata of varying income and wealth levels.

For the 1989 SCF, the conventional area probability approach to the sampling of households has several advantages. The multi-stage area probability frame provides both a high level of coverage of households and permits cost-effective "clustering" of survey households within primary stage sample locations. The major disadvantage to the area probability frame is that a cost effective stratification of the population based on either income or net worth is difficult to achieve. At best, Census data on average household income enables the sampling statistician to assign small area sampling units -- tracts, blocks, enumeration districts -- to broadly defined income strata. However, even within these small areas, 1980 Census measures of household income are highly variable and at this stage in the decade could be completely obsolete.

V. Dual-Frame Design: Taxpayer List Frame

The second sample component of the 1989 SCF dual-frame design is a stratified random sample of tax filing units selected from the 1987 Statistics of Income (SOI) Tax Model data base. The Statistics of Income Tax Model data bases are stratified random

samples of U.S. Individual (Form 1040) tax returns which are selected and compiled annually for research uses within the U.S. Department of the Treasury (Internal Revenue Service, 1987). For the 1989 SCF, a special contractual agreement has enabled the Department of Treasury to provide the Survey Research Center with names and mailing addresses of a stratified subsample of taxpayers whose individual returns were selected for inclusion in the 1987 SOI Tax Model File. The terms of this special agreement are written so as to guarantee privacy rights of the individual taxpayers. Only names, addresses and generalized stratum identifiers for the sample tax payers have been provided to The Survey Research Center.[1]

V. A The 1987 Statistics of Income Sample Frame

The 1987 SOI Tax Model data base contains abstracted tax form data for a stratified random sample of approximately 108,000 1987 Form 1040 tax returns selected from the over 100,000,000+ individual income tax filings for the 1987 tax year. The stratification plan for the original selection of the 1987 SOI Individual Tax Model data base is based on several criteria including the type of tax filer unit -- business, non-business -- and the general amounts of income that are reported. The sampling of tax forms for inclusion in the 1987 SOI Tax Model file is highly disproportionate by income stratum, with considerable oversampling of the higher income strata.

A small definitional problem arises in the use of the 1987 SOI Individual Tax Model File as a sample frame for the 1989 SCF. The 1989 SCF questionnaire is built around the household as the reporting unit, but the elements of the SOI data bases are taxpayer units which may or may not constitute complete households. Most of the high income households selected from the SOI frame for the 1989 SCF interview are expected to constitute a single filing unit. Nevertheless, in the course of the survey interview, respondents are asked if their household contains multiple tax filing units. If multiple taxfilers are present in the interviewed household an appropriate correction is being made to the household's sample selection probability and case-specific analysis weight.

V. B Stratification of the 1989 SCF Sample of 1987 SOI Individual Tax-Filers

Under the general plan for the 1989 SCF dual frame sample design (see Table 1), an expected total of n=850 completed interviews will be taken with respondent households selected from the 1987 SOI Tax Model file.

Working within the general stratification and sample allocation guidelines developed for the dual-frame design as a whole, the sampling plan called for the SOI frame sample to be stratified along one primary and two secondary dimensions. The primary stratifier for the SOI-based sample is the index of net worth for the sample element. A simple capitalization model has been developed and used to produce a relative index of total net worth for each tax filing unit included in the 1987 Statistics of Income Tax Model base. In turn, the index of total net worth was used to assign each element in the sample frame to one of seven explicit net worth strata. Within each explicit net worth stratum, eight secondary strata were formed based on the business/non-business status and AGI level of the frame elements. A stratified random sample of taxpayer units was then selected from each stratum.

V. B.1 Stratification Based on Net Worth Index: The Wealth Model

The data contained in the Statistics of Income sample frame describe annual amounts of taxable income flows to taxpaying units. The primary examples of such income flows include:[2]

- 1) wages and salary income;
- 2) interest earnings (both taxable and not taxable);
- 3) dividends;
- 4) business and farm income (gross and net);
- 5) income from rental property (gross and net);
- 6) income from trusts and partnerships; and
- 7) capital gains.

These income flows are reported for tax purposes and represent returns on personal labor, personal assets, real estate investment property, and in the case of business income, a return on combined inputs of labor and business tangible and intangible assets.

These flows do not constitute direct measures of household assets or wealth holdings. However, through a predictive wealth model, the tax reports of income flows can be capitalized and aggregated to form an index of the total underlying net worth of the taxpaying unit. The predicted values from the wealth model are labeled an index of net worth with the explicit recognition that they constitute a relative as opposed to absolute estimate of the total net worth of taxpayer households. The index is a tool to divide the SOI households into strata representing seven broad ranges of household net worth. The index must be correlated with households' actual net worth; however, the correlation need not be perfect since the index is being used to group households into net worth ranges, not to predict exact net worth of individual households. In this discussion, the equation used to compute the index of net worth will be termed the "wealth model." The general form of the wealth model used in the stratification is:

$$\hat{W}_{hi} = \hat{\beta}_{ho} + \sum_{j=1}^J \hat{\beta}_j X_{ji}$$

where:

\hat{W}_{hi} = Predicted wealth of tax filing unit i in AGI stratum h.

$\hat{\beta}_{ho}$ = Model intercept reflecting assets in the form of housing and personal property equity for AGI stratum h = 1,...H,

$\hat{\beta}_j$ = Model coefficient for income flow j = 1,...J;

X_{ji} = Value of income flow j for tax filing unit i.

The intercept term in the wealth model represents the household's net worth in the form of equity in their home, personal property, and other forms of wealth which do not generate a measurable income flow. The SOI Tax Model data base provides housing-related information in the form of Schedule A itemized deductions for mortgage interest and real estate taxes. On first review, it seems natural to try to use these two items to predict housing values and/or housing equity directly. However, effective use of this data involves a great deal of difficulty and com-

plexity (auxiliary data, imputations). When information on applicable assessment and tax rates can be determined, the real estate tax data provide only an estimate of housing value for itemizers -- not home equity. Many home owners no longer have a mortgage on their property, and many mortgagees do not itemize tax deductions. Likewise, the annual amount of the mortgage payments is a poor indicator of home equity unless the starting date of the mortgage is known. Other authors have addressed this problem through sophisticated procedures for home value and equity computation (Greenwood, 1983).

In developing the wealth model for the stratification of the SOI-based sample, the complexities of the ancillary home and property equity models were avoided by simply assigning each tax filing household an estimated median value of home and personal property equity for households in its particular AGI category. The median home and property equity values for each AGI category were initially estimated from the 1983 SCF data set and adjusted with appropriate inflators to 1989 levels. Table 4 provides the wealth model intercept terms for median home and property equity by AGI category.

Table 4.--Wealth Model Intercept Term: Estimate of the Median Value of Home and Property Equity by AGI Category

| 1987 AGI Category | Estimate of Median Home and Property Equity |
|-------------------|---|
| 0-99K | \$ 16,129 |
| \$100K-199K | \$315,558 |
| \$200K-999K | \$617,496 |
| \$1 Million + | \$979,736 |

The remaining terms in the wealth model represent a capitalization of the income flows reported on the tax return. Before summarizing the components of this model, we turn to an examination of the characteristics of the income flows reported on the 1987 SOI file, the probable relation between these income flows and net worth components, and the topic of capitalization rates.

V. B.2 Wealth Characteristics of U.S. Tax filing Units

For developing the index of taxpayer wealth from the 1987 SOI Tax Model data base, the key elements are: data on income sources, particularly interest and dividend income; data from Schedules C, E, and F, which report income from noncorporate business, partnerships, small business corporations, trusts, rent, royalties, and farms; and data from Schedule D, which reports capital gains. The interest and dividend data are a direct reflection of the financial wealth of the taxpayer, since there must be financial asset holdings corresponding to the interest and dividend income flows. For the Schedule C, E, or F filers, there must be business, farm, real estate, or other assets corresponding to the income flows reported on Schedules C, E, or F. Schedule D data contain direct measures of capital assets that have been sold.

The development of the wealth index model uses data from the 1983 Survey of Consumer Finances, which contains extensive and detailed information on assets and liabilities -- interest-earning assets, common stock and mutual funds shares, equity in business or farms, equity in real estate investments, equity in owned home, etc. In addition, the 1983 SCF also has extensive data on household characteristics and income, from which relationships between various types of asset holdings and household characteristics can be estimated.

In this section of the paper, we first examine the distribution of the U.S. taxpayer population by the amount of their interest and dividend income, which is presumed to reflect the distribution of taxpaying units by their financial wealth. We then contrast taxpayer households reporting some income on Schedules C, E, or F with those taxpayers who report no Schedule C, E, or F income. This contrast will show that most taxpayers with large amounts of wealth are likely to file a Schedule C, E, or F. A closer look at the Schedule C, E, or F tax filer subclass will identify the Schedule E filers as the subclass where there appears to be the greatest concentration of wealth in the form of nonfinancial assets.

Beginning in 1987, U.S. taxpayers were required to report the total amount of both taxable and non-

taxable interest income, which are the flows corresponding to fixed-income assets. Similarly, the value of common stock holdings can be predicted from SOI data on dividend income, and the value of trusts can be estimated from tax file reports of income from trusts. With the possible exception of trusts, there is a reasonably strong relationship between the value of the asset and the amount of interest or dividend income generated by the asset.

Therefore we know from the SOI file that households report a certain amount of interest income, and we know from the 1983 Survey of Consumer Finances data that households in a given income class earn an average rate of return, R percent, on those types of assets, with variance $V(R)$. Updating the 1983 SCF rates of return to 1989 provides a starting point in developing an index of total wealth. Average rates of return in the form of dividends from both publicly traded stock as well as dividends paid by closely held corporations can be estimated using the 1983 SCF data. The variance of the rate return for stocks is larger than for fixed income-yielding assets, since the rate of return on stock in the form of the dividend yield is likely to show more variance in total, and those rates of return are also likely to vary more systematically as a function of income class. Therefore, households with no income reported on Schedules C, E, or F, are relatively easy to manage in terms of indexing wealth. For those households, any financial assets can be reasonably well estimated by capitalizing any dividend or interest income reported and estimating home and property equity.

In contrast to interest, dividend or trust income, procedures for estimating business or farm net equity from SOI reports of business or farm income, or estimates of real estate equity from SOI data elements relating to rental income are complex and subject to large errors. Not only is the estimation problem particularly difficult, but the amount of assets involved is very large; the 1983 SCF estimates that households' net equity in businesses or farms amounts to about \$2 trillion while net equity in real estate holdings amounts to another \$1.5 trillion--a combined \$3.5 trillion out of the total estimated 1983 U.S. household net worth of about \$10.5 trillion.

Taxpayer reports of income on Schedule C (business income), Schedule E (rental income, partner-

ship income, etc.), or Schedule F (farm income) are often poor predictors of the value of the corresponding asset. There are substantial opportunities here for taxpayers to report negative taxable income from business, real estate, or farm investments, and those negative income reports clearly do not correspond to negative net worth.

Analysis of the 1987 SOITax Model file indicates clearly that the income reports on Schedules C, E, and F are very poor indicators of the return on the underlying asset. For example, about 24 percent of sample units report negative Schedule C income. 52 percent of the Schedule E filers report negative Schedule E income.

Tables 5 and 6 display net Schedule C or E income by the dividend and interest income of the taxpayer unit, with net Schedule C or E income ranging from less than -\$100,000 to \$100,000 or more.

One observation from Tables 5-6 is clear. If net Schedule C or E income is used as a classifying variable, households with very high predicted wealth are most likely to be found among taxpayer units reporting large negative net income, least prevalent among households with either small negative or positive net income, with households reporting large positive net income between these two groups. The reason is obvious: It is not possible to have large negative C or E income without having very substantial amounts of assets, and in most cases the large assets will include not only large business or property assets, but large financial assets as well.

Arraying taxpayer units by gross income rather than net income, predicted wealth then becomes monotonic--large amounts of Schedule C or E gross income are associated with large amounts of assets, presumably including both the financial assets displayed in the table as well as the business or property assets underlying the Schedule C or E income. For using these data as inputs into the capitalization model, we have decided to use gross income as the income flow to be capitalized.

A second point of interest in Tables 5-6 is the apparent difference in wealth between units that file

Schedule C and those that file Schedule E. There are about the same number of Schedule C taxpayers as there are Schedule E taxpayers -- almost 13 million for Schedule C, a little less than 14 million for Schedule E. Many Schedule E filers also file Schedule C. While both types of taxfiling units have substantially more financial wealth than households that file neither Schedule C nor E, there is a significant difference in the distribution of financial wealth between Schedule C and Schedule E filers. For example, just over 20 percent of Schedule C filers report zero dividend or interest income, while only 6 percent of Schedule E filers report no income from these sources. In contrast, only about 1 percent of Schedule C filers report dividend and interest income of \$50,000 or more, while over 3 percent of Schedule E filers report interest and dividend income in those categories. In absolute numbers, there are substantially more Schedule E households who have very large holdings of financial assets, in addition to the assets underlying their Schedule E income flows. About 135,000 units filing Schedule C report 1987 annual dividend and interest income above \$50,000, while over 400,000 units filing Schedule E report dividend and interest income above \$50,000. At the other end of the distribution, roughly 9 million Schedule C filers report dividend and interest income of \$1,000 or less, while only 5.5 million Schedule E filers report less than \$1,000 of dividend or interest income.

To summarize the characteristics of households with large predicted wealth based on the SOI data base, a majority report some Schedule C, E or F income, and the great majority of the wealthiest households will report some income on Schedule E. The SOI estimates show about 570,000 units reporting more than \$50,000 of dividend or interest income. Of these, only about 140,000 are households who have no Schedule C, E or F income, while about 135,000 have some Schedule C income and over 400,000 have some Schedule E income (the last two categories are not exclusive, since some households will file both Schedule C and Schedule E). Thus the bulk of the problem of assigning households into a particular net worth or wealth stratum resides in correctly classifying taxpayer units that report some income on Schedule E. The majority of the largest single concentration of wealthy households appears

**Table 5.--Distribution of 1987 Schedule C Taxpayers by: Net Schedule C Income and Dividend and Interest Category
1987 Statistics of Income. Schedule C Filers Only**

| Net Schedule C Income | Dividend and Interest Category | | | | | | | | | | Total |
|-------------------------------------|--------------------------------|------------------------------|------------------------------|----------------------------|----------------------------|-----------------------------|-----------------------------|--------------------------------|--|--|-------|
| | \$0 | \$1-999 | \$1K-9.9K | \$10-24.9K | \$25-49.9K | \$50-99.9K | \$100K+ | | | | |
| <-\$100K Row X Column X | 714.07 6.70X 0.03X | 2122.88 19.92 0.03 | 2375.34 22.28 0.08 | 891.45 8.36 0.17 | 609.29 5.72 0.31 | 1581.18 14.83 1.89 | 2364.78 22.19 4.75 | 10659 100.00X .08X | | | |
| \$-99K to -50K Row X Column X | 758.64 3.98 0.03 | 5093.87 26.72 0.08 | 5793.33 30.39 0.19 | 3457.76 18.14 0.65 | 962.3 5.05 0.49 | 1321.92 6.93 1.58 | 1676.65 8.79 3.37 | 19064.5 100.00X .15X | | | |
| \$-49K to -5K Row X Column X | 109851 16.67 4.10 | 282225 42.84 4.42 | 179146 27.19 5.95 | 45717.4 6.94 8.57 | 21936.5 3.33 11.25 | 12076.7 1.83 14.41 | 7825.1 1.19 15.71 | 658778 100.00X 5.09X | | | |
| \$-4.9K to -1 Row X Column X | 450557 18.87 16.82 | 1263911 52.93 19.80 | 524091 21.95 17.40 | 88322.4 3.70 16.56 | 39401 1.65 20.20 | 16412 0.69 19.58 | 5125.79 0.21 10.29 | 2387822 100.00X 18.46X | | | |
| \$0-5K Row X Column X | 1234750 23.77 46.09 | 2611224 50.28 40.90 | 1083049 20.85 35.96 | 180431 3.47 33.84 | 58887 1.13 30.20 | 14874.9 0.29 17.74 | 10311.5 0.20 20.70 | 5193528 100.00X 40.14X | | | |
| \$5.1K-50K Row X Column X | 867961 20.51 32.41 | 2093342 49.51 32.79 | 1006946 23.82 33.43 | 166224 3.93 31.18 | 52923.2 1.25 27.14 | 25597.2 0.61 30.53 | 14736.4 0.35 29.59 | 4227730 100.00X 32.68X | | | |
| \$50.1K-100K Row X Column X | 12770.6 4.13 0.48 | 100371 32.81 1.57 | 144857 47.36 4.81 | 27466.4 8.98 5.15 | 10950.1 3.58 5.62 | 6725.79 2.20 8.02 | 2740.71 0.90 5.50 | 305881 100.00X 2.36X | | | |
| >\$100K Row X Column X | 1675.69 1.25 0.06 | 26088.2 19.47 0.41 | 65926.9 49.20 2.19 | 20682.9 15.44 3.88 | 9345.11 6.97 4.79 | 5247.12 3.92 6.26 | 5023.71 3.75 10.90 | 133990 100.00X 1.04X | | | |
| TOTAL Row X Column X | 2679039 20.71X 100.00X | 6384378 49.35X 100.00X | 3012185 23.28X 100.00X | 533194 4.12X 100.00X | 195015 1.51X 100.00X | 83836.8 0.65X 100.00X | 49804.7 0.38X 100.00X | 12937451 100.00X 100.00X | | | |

Table 6.--Distribution of 1987 Schedule E Taxpayers by: Net Schedule E Income and Dividend and Interest Category
1987 Statistics of Income. Schedule E Filers Only

| Net Schedule E Income | Dividend and Interest Category | | | | | | | | | | Total |
|--|--------------------------------|------------------------------|------------------------------|------------------------------|----------------------------|----------------------------|----------------------------|--|--|--|--------------------------------|
| | \$0 | \$1-999 | \$1-10K | \$10-25K | \$25-50K | \$50-100K | >\$100K | | | | |
| <\$100K Row X Column X | 370.88 0.66X 0.04X | 3706.16 6.61 0.08 | 10781.6 19.24 0.20 | 9861.79 17.59 0.62 | 9357.12 16.69 1.38 | 7748.99 13.83 2.82 | 14223.3 25.38 9.49 | | | | 56049.9 100.00X .41X |
| \$-100K to \$-50K Row X Column X | 1354.13 1.72 0.16 | 6085.42 7.72 0.13 | 27493.4 34.88 0.51 | 15565.5 19.75 0.97 | 11521.7 14.62 1.70 | 8377.9 10.63 3.05 | 8427.43 10.69 5.62 | | | | 78825.6 100.00X .58X |
| \$-50K to \$-5K Row X Column X | 99200.5 4.34 11.99 | 779033 34.10 16.84 | 971181 42.52 17.86 | 230610 10.10 14.39 | 114164 5.00 16.82 | 59360.4 2.60 21.64 | 30726.9 1.35 20.51 | | | | 2284276 100.00X 16.80X |
| \$-5 to 0 Row X Column X | 278425 6.02 33.65 | 1939173 41.92 41.91 | 1757790 38.00 32.33 | 423933 9.16 26.45 | 153835 3.33 22.67 | 55033.8 1.19 20.06 | 17520.8 0.38 11.69 | | | | 4625711 100.00X 34.02X |
| \$0-5K Row X Column X | 303775 6.91 36.95 | 1420260 32.10 30.70 | 1834007 41.45 33.73 | 578893 13.08 36.12 | 197935 4.47 29.16 | 63267.4 1.43 23.06 | 24905.5 0.56 16.62 | | | | 4425044 100.00X 32.54X |
| \$5-50K Row X Column X | 138975 7.71 16.79 | 438868 24.35 9.49 | 717313 39.80 13.19 | 282121 15.65 17.60 | 147242 8.17 21.69 | 52805.6 2.93 19.25 | 24873.2 1.38 16.60 | | | | 1802198 100.00X 13.25X |
| \$50-100K Row X Column X | 2627.07 1.48 0.32 | 29792.7 16.76 0.64 | 68275.4 38.41 1.26 | 31048 17.47 1.94 | 25298.2 14.23 3.73 | 11322 6.37 4.13 | 9376.66 5.28 6.26 | | | | 177740 100.00X 1.31X |
| \$100K+ Row X Column X | 750.52 0.51 0.09 | 9648.6 6.55 0.21 | 50766.1 34.45 0.93 | 30607.2 20.77 1.91 | 19357.4 13.14 2.85 | 16455.3 11.17 6.00 | 19780.6 13.42 13.20 | | | | 147366 100.00X 1.08X |
| TOTAL Row X Column X | 827478 6.09X 100.00X | 4626568 34.03X 100.00X | 5437608 39.99X 100.00X | 1602640 11.79X 100.00X | 678710 4.99X 100.00X | 274371 2.02X 100.00X | 149836 1.10X 100.00X | | | | 13597212 100.00X 100.00X |

to be Schedule E filers who report partnership income, as opposed to rental income, trust income, or royalty income.

Although the data from the 1987 SOI tax model clearly suggest that the nonfinancial assets associated with Schedule E filers is a much more important source of wealth than the nonfinancial assets associated with Schedule C or F filers, that conclusion depends entirely on the assumption that interest and dividend income (and the financial wealth associated with those income flows) is a good predictor of the existence of wealth in the form of real estate assets, business assets, and farm assets. One obvious difficulty is that financial assets may be a good predictor for some types of nonfinancial wealth, but not for others.

In particular, it is plausible that financial assets might well be a good index of wealth in the form of real estate partnerships, small business corporation interests, etc., simply because these represent the

kind of portfolio diversification that many wealthy individuals might try to achieve. But it may not follow that the absence of large financial assets for most Schedule C filers is also associated with relatively modest wealth in the form of business or farm equity. It is certainly possible that many households with a management interest in a business have most of their total wealth in that form, and that their portfolios are not diversified at all.

The 1983 SCF data can be used to explore the relationship between financial assets and holdings of nonfinancial wealth in the form of business equity, farms or other real estate.

Based on the 1983 SCF data, Table 7 categorizes households by the amount of their equity in businesses or farms, or the amount of their equity in real estate investment holdings, and tabulates financial asset holdings for households in these net equity categories. The suspicion that large interest and dividend incomes, hence financial assets, mean

| Table 7.--Relationship Between Financial Assets Holdings, Business Ownership, and Real Estate Investments | | | |
|--|------------|--|--|
| A. Households with Equity in Business, No Real Estate Investments (File Schedule C, not E) | | | |
| Business Equity Category | n | Mean Amount of Business Equity | Mean Amount of Financial Assets (000) |
| < \$50K | 213 | 17.4 | 36.2 |
| \$50-99K | 60 | 67.0 | 52.7 |
| \$100-249K | 78 | 144.2 | 85.6 |
| \$250-999K | 85 | 393.3 | 116.7 |
| \$1M or more | 70 | 2035.5 | 370.8 |
| Total | 506 | 173.9 | 69.6 |
| B. Households with Real Estate Investments, No Business Equity (File Schedule E, not C) | | | |
| Real Estate Investment Category | n | Mean Amount of Real Estate Investment (000) | Mean Amount of Financial Assets (000) |
| < \$50K | 33 | 316.7 | 21.9 |
| \$50-99K | 60 | 66.3 | 33.8 |
| \$100-249K | 32 | 131.5 | 174.9 |
| \$250-999K | 35 | 422.5 | 207.1 |
| \$1M or more | 71 | 655.3 | 1409.4 |
| Total | 467 | 77.1 | 43.0 |
| Source: 1983 Survey of Consumer Finances | | | |

somewhat different things for Schedule E and Schedule C filers is borne out by these data. While households with large real estate investments also have very large financial investments, suggesting portfolio diversification, that is less true for households with large amounts of business equity. Consequently, Schedule C and F filers who own businesses and farms present a particular problem in the indexing of wealth based on taxable income flows.

V. B.3 Capitalization Rates

Data from the 1983 Survey of Consumer Finances can be used to examine the capitalization rate for taxable and interest-earning assets as well as taxable dividends, and to estimate capitalization rates for business or rental income.

The basic data from the 1983 SCF that bear on these issues are displayed below in Table 8. Here, we show the income reported on the SCF from particular types of assets, the amount of such assets for households who report both owning the asset and receiving some income, and the amount of assets for households who report owning the asset but fail to report any income. Data are shown for five types of assets -- taxable interest income, taxable dividends, nontaxable financial assets, business income, and income from rentals and trusts. The data are displayed by 1983 SCF income class, and average rates of return by income class are computed for households reporting both ownership of the asset and receipt of an income flow.

A number of characteristics of the basic data should be noted. While many households report having an asset but not receiving any associated income flow, most assets have income flows associated with them. That shows up in terms of two types of relationships in the data. First, households reporting both income and assets are a much larger fraction of households reporting ownership of the asset as we go up the income scale -- that is, a larger weighted proportion of households report both income and assets than report just the asset at higher income levels. Second, the mean amount of assets owned by households who report both assets and income is much larger than the mean amount of assets held by households who report the asset but do not report an

associated income flow. In brief, households who fail to report income from assets have smaller average asset holdings. The net result is that the great bulk of total assets are held by households who also report an associated income flow.

Interestingly, this general relationship holds for the first four panels in Table 8 (nontaxable financial assets, taxable interest-yielding assets, taxable common stock, and equity in rental property and trust), but does not appear to hold for households reporting equity in business assets. Here, while it is still true that the great bulk of assets are held by households who report both asset amounts and income flows, there is no income pattern at all to this relationship -- just about the same weighted proportion of households report holding both assets and having income as report just owning the assets. The category that comes closest to having a similar pattern (i.e., a weak income relationship) seems to be holdings of taxable common stock, where well over half of all income groups report both the asset and the income.

The data in Table 8 reflect in part the failure of households who actually own assets to report income. That appears to be more common when the income flows are very small (taxable interest-yielding asset flows), or when the income flows are not easily observed -- nontaxable financial assets, where households did not (in 1983) receive tax forms from financial institutions. But that pattern may also reflect timing differences between asset holdings and income flows. The assets were reported as of the date of the 1983 SCR, which was conducted in the Spring of 1983. But the income flows are for the calendar year 1982. Hence, there must be some households who owned assets in 1983 but did not own them during calendar year 1982, hence had no income to report. Similarly, there are households in the survey who reported income but no assets, presumably reflecting assets owned in 1982 when the income was received, but no longer owned in 1983 when the survey was conducted.

Table 8 also contains data on mean rates of return, which can in principle be used to estimate capitalization rates for the various types of assets. These data are generally consistent with the rates of return that

Table 8.--1983 SCF Asset Data: Percentage of Cases Reporting Both Asset and Income from Assets, and Estimated Mean Rates of Return, by Income Class

| 1983 SCF Income Class | 1: Households Reporting Both Asset Amounts and Income From Assets | | | | 2: Households Reporting Assets but No Income | | Proportion Reporting No Income From Asset |
|--|--|-------------------------------|-------------------------------|------------------------|---|------------------------|--|
| | Mean Income From Asset (000) | Mean Asset Amount (000) | Mean Rate of Return (%) | Weighted N (000) | Mean Asset Amount (000) | Weighted N (000) | |
| Panel A: Nontaxable Financial Assets | | | | | | | |
| <\$10K | .22 | 2.87 | 7.7 | 87 | 4.95 | 306 | .22 |
| \$10K-24.9K | 1.17 | 16.35 | 7.2 | 929 | 6.64 | 2,053 | .31 |
| \$25K-49.9K | 1.24 | 10.17 | 12.1 | 2,036 | 4.95 | 4,406 | .32 |
| \$50K-74.9K | 1.74 | 20.20 | 8.6 | 1,126 | 11.07 | 1,609 | .41 |
| \$75K-99.9K | 2.96 | 33.38 | 8.9 | 391 | 14.39 | 502 | .44 |
| \$100K-199.9K | 4.40 | 98.89 | 4.4 | 421 | 37.70 | 475 | .47 |
| \$200K-499.9K | 19.91 | 279.40 | 7.1 | 325 | 50.06 | 173 | .65 |
| ≥\$500K | 47.85 | 701.89 | 6.8 | 72 | 237.92 | 23 | .76 |
| Total | 3.43 | 47.25 | 7.3 | 5,386 | 9.85 | 9,548 | .36 |
| Panel B: Taxable Interest-yielding Assets | | | | | | | |
| <\$10K | .60 | 6.98 | 8.5 | 4,396 | 1.67 | 4,818 | .48 |
| \$10K-24.9K | 1.56 | 16.39 | 9.5 | 12,344 | 3.43 | 10,081 | .55 |
| \$25K-49.9K | 2.10 | 22.13 | 9.5 | 12,763 | 5.07 | 8,027 | .61 |
| \$50K-74.9K | 4.61 | 37.46 | 12.3 | 3,446 | 16.14 | 1,117 | .76 |
| \$75K-99.9K | 4.21 | 46.36 | 9.1 | 1,166 | 19.62 | 314 | .79 |
| \$100K-199.9K | 11.26 | 96.98 | 11.6 | 949 | 38.51 | 130 | .88 |
| \$200K-499.9K | 16.20 | 181.22 | 8.9 | 462 | 100.90 | 66 | .88 |
| ≥\$500K | 88.76 | 668.59 | 13.3 | 101 | 104.20 | 4 | .96 |
| Total | 2.71 | 26.44 | 10.2 | 35,627 | 4.87 | 24,559 | .59 |

Table 8.--1983 SCF Asset Data: Percentage of Cases Reporting Both Asset and Income from Assets, and Estimated Mean Rates of Return, by Income Class--Continued

| 1983 SCF Income Class | 1: Households Reporting Both Asset Amounts and Income From Assets | | | | 2: Households Reporting Assets but No Income | | Proportion Reporting No Income From Asset |
|---|--|-------------------------------|-------------------------------|------------------------|---|------------------------|--|
| | Mean Income From Asset (000) | Mean Asset Amount (000) | Mean Rate of Return (%) | Weighted N (000) | Mean Asset Amount (000) | Weighted N (000) | |
| Panel C: Taxable Common Stock | | | | | | | |
| < \$10K | .40 | 5.47 | 7.3 | 552 | 5.99 | 286 | .66 |
| \$10K-\$24.9K | .84 | 11.38 | 7.4 | 2,617 | 6.31 | 1,731 | .60 |
| \$25K-49.9K | 1.60 | 21.72 | 7.4 | 3,473 | 10.06 | 2,766 | .56 |
| \$50K-74.9K | 2.55 | 46.71 | 5.4 | 1,777 | 14.57 | 531 | .77 |
| \$75K-99.9K | 3.47 | 78.80 | 4.4 | 603 | 10.16 | 283 | .68 |
| \$100K-199.9K | 10.24 | 250.31 | 4.1 | 716 | 25.06 | 47 | .94 |
| \$200K-499.9K | 31.07 | 761.52 | 4.1 | 425 | 89.31 | 43 | .91 |
| ≥ \$500K | 96.32 | 1566.60 | 6.1 | 104 | 664.96 | 1 | .99 |
| Total | 4.41 | 88.22 | 5.0 | 10,269 | 9.98 | 5,689 | .64 |
| Panel D: Rental Property and Trusts (Equity) | | | | | | | |
| < \$10K | 1.40 | 25.66 | 5.4 | 247 | 34.54 | 1,368 | .15 |
| \$10K-24.9K | 3.27 | 55.99 | 5.8 | 1,410 | 29.35 | 3,587 | .28 |
| \$25K-49.9K | 4.98 | 97.16 | 5.1 | 2,327 | 31.08 | 4,232 | .35 |
| \$50K-74.9K | 11.19 | 207.73 | 5.4 | 739 | 67.37 | 1,341 | .36 |
| \$75K-99.9K | 8.91 | 253.30 | 3.5 | 258 | 45.38 | 495 | .34 |
| \$100K-199.9K | 20.12 | 347.59 | 5.8 | 292 | 133.73 | 362 | .45 |
| \$200K-499.9K | 26.75 | 1643.54 | 1.6 | 193 | 1,794.84 | 188 | .51 |
| ≥ \$500K | 143.81 | 3454.32 | 4.2 | 58 | 718.51 | 21 | .73 |
| Total | 8.41 | 208.02 | 4.0 | 5,525 | 68.53 | 11,593 | .32 |

Table 8.--1983 SCF Asset Data: Percentage of Cases Reporting Both Asset and Income from Assets, and Estimated Mean Rates of Return, by Income Class--Continued

| 1983 SCF Income Class | 1: Households Reporting Both Asset Amounts and Income From Assets | | | | 2: Households Reporting Assets but No Income | | Proportion Reporting No Income From Asset |
|--|--|-------------------------------|-------------------------------|------------------------|---|------------------------|--|
| | Mean Income From Asset (000) | Mean Asset Amount (000) | Mean Rate of Return (%) | Weighted N (000) | Mean Asset Amount (000) | Weighted N (000) | |
| Panel E: Business Assets (Equity) | | | | | | | |
| <\$10K | 3.07 | 71.40 | 4.3 | 624 | 284.54 | 249 | .71 |
| \$10K-24.9K | 10.64 | 57.41 | 18.5 | 24,420 | 96.21 | 641 | .79 |
| \$25K-49.9K | 19.84 | 103.25 | 19.2 | 3,263 | 59.12 | 935 | .78 |
| \$50K-74.9K | 35.68 | 176.23 | 12.9 | 1,140 | 203.91 | 251 | .82 |
| \$75K-99.9K | 39.74 | 221.23 | 17.9 | 391 | 98.46 | 101 | .79 |
| \$100K-199.9K | 75.71 | 458.06 | 16.5 | 545 | 350.88 | 164 | .77 |
| \$200K-499.9K | 136.17 | 1,1013.94 | 13.4 | 271 | 1,223.37 | 120 | .69 |
| ≥\$500K | 344.73 | 3,387.13 | 10.2 | 76 | 4,365.24 | 18 | .81 |
| Total | 28.98 | 195.14 | 14.8 | 8,732 | 210.15 | 2,480 | .78 |

were observable in 1982: Nontaxable financial assets have substantially lower rates of return than taxable interest-yielding assets, and taxable common stock shows a lower dividend yield than the interest return on fixed-yield assets. There appears to be an income pattern to the results--rates of return generally seem to be lower for the higher income categories, although that is not universally true. There may in addition be an age effect.[3]

For the difficult-to-value assets (equity in rental property and in business), the estimated rates of return are much more difficult to interpret. For rental property, the estimated rates of return are quite low, which may be realistic if one is thinking of nominal rates of return and not total return including capital gains. For business equity, rates of return are extremely high, which may reflect some contamination of the income from business assets with wage-and-salary income for the owner of the business. In both cases -- equity in rental property and equity in business -- rates of return appear to decline with income level.

V. B.4 Estimating the Wealth Model

The capitalization equation actually used to assign an index of net worth to each SOI data base taxpayer is shown in Table 9. The parameter values in the wealth model are assigned on the basis of information derived partly from the structure of the 1983 Survey of Consumer Finances, partly from current market rates of return, partly from the judgments of financial and tax experts.

| Table 9.--The Wealth Model: Coefficients for Income Flows | | |
|---|-----------------------|-------------------|
| Income Flow Factor | Capitalization Factor | Model Coefficient |
| Intercept term | [1] | -- |
| Taxable interest | .10 | 10 |
| Nontaxable interest | .07 | 14.28 |
| Dividends | .05 | 20 |
| Gross rental income | .10 | 10 |
| Gross business and farm income | .15 | 6.66 |
| Partnership, estate trust | .15 | 6.66 |
| Long-term capital gains | 1.00 | 1.00 |

[1] See Table 4.

In terms of specific capitalization factors and model coefficients, the first three items in Table 9 (taxable interest, nontaxable interest, and dividends) are rounded approximations to prevailing market rates of return on these kinds of assets. The gross property income capitalization factor is substantially larger than the rate of return shown by the 1983 SCF data, but it is not at all clear what respondents to the 1983 SCF had in mind when reporting rental or property income. The factor actually used represents the best judgment of experts in that area. For business and farm income, we used a capitalization factor much like the average yielded by the SCF comparison of business income with business assets. Here, the fact that business income is probably contaminated by the inclusion of a certain amount of labor income is irrelevant, since the only issue here is to compute a realistic value of business equity from reported income. The capitalization factor for partnership and estate trusts is again based on the judgment of knowledgeable experts.

Tax reports of long-term capital gains posed an interesting problem. Clearly, the report of a long-term gain indicates the presence of assets whose combined value equals or exceeds the amount of the reported gain. If the reported gains originate primarily with the sale of stock, then it is possible that the value of the underlying asset may already be captured by the capitalization of dividend income, assuming the proceeds were immediately reinvested. A simple test of the degree of association between taxpayer annual dividend income and tax reports of long-term capital gains found the relationship to be insignificant, although that might be because the proceeds were reinvested but not in stock. In any event, we decided to simply add the total amounts of long-term capital gains to the wealth model estimate of total net worth without attempting a compensatory adjustment of the capitalization applied to reported dividends or other income flows.

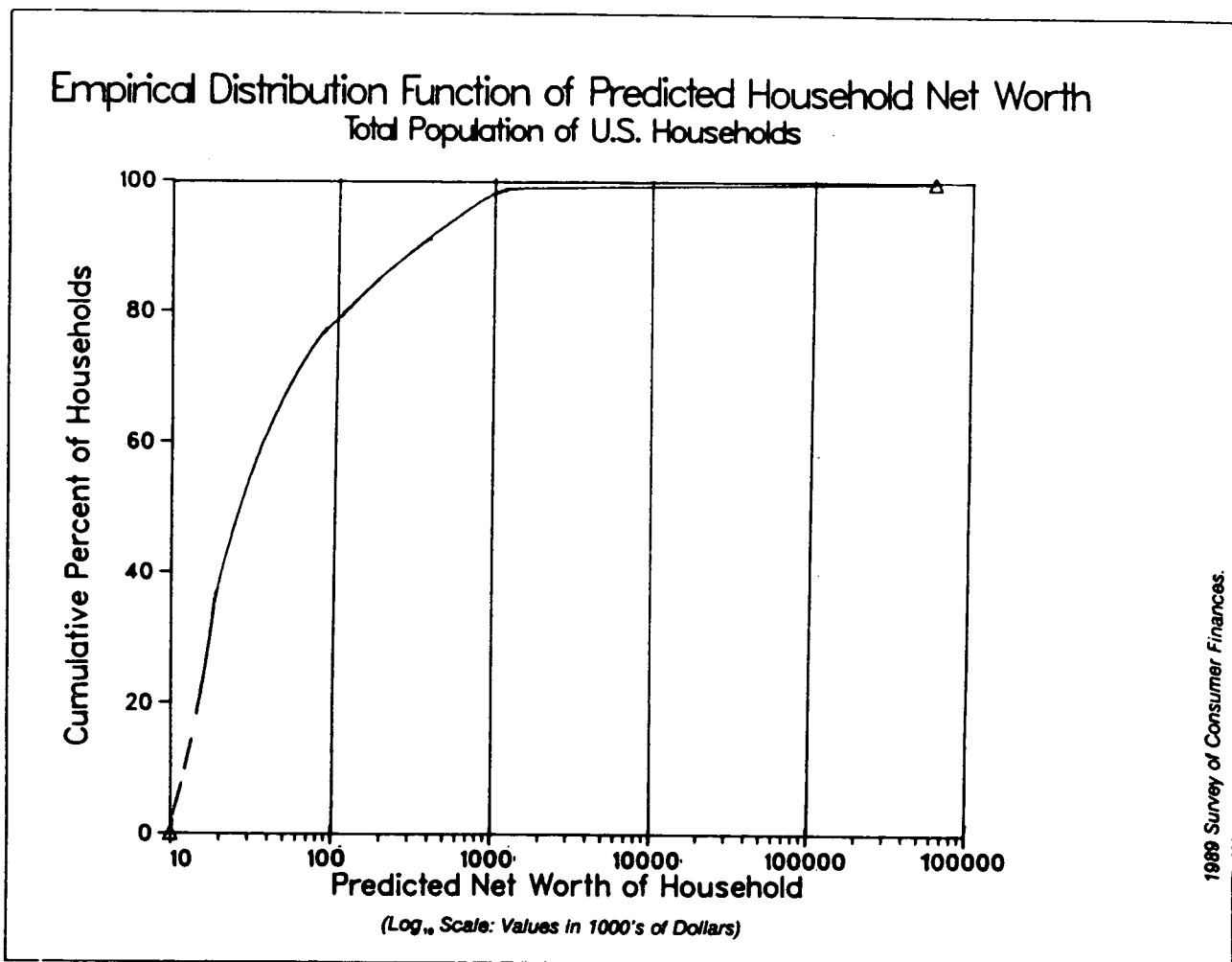
Using the simple capitalization equation defined in Table 9, the net worth index was computed for each taxpayer household in the SOI sample frame. If the index of net worth for each SOI data base element is assigned the correct sample selection weight, the simple expansion estimate of total 1987 U.S. household net worth is 10.2 trillion dollars -- slightly less than the 1983 SCF estimate of 10.5 trillion dollars--suggesting that the capitalization rates are a bit low.

There are also some wealth elements not represented at all by any of the flows from the tax file. The true 1989 wealth total is probably a lot higher than the estimate of 10.2 trillion, and the difference is probably too large to be accounted for by omitted items.

Figure I plots the total population cumulative distribution of the predicted net worth values which resulted from the application of the model to the 1987 SOI Tax Model data base. Again, there is general consistency with the 1983 SCF, although the simple wealth model assigns 78% of U.S. taxpayers to the <\$100K net worth category -- probably too many. The 1983 SCF estimate was 76% of households with <\$100K net worth.

The next step in the stratification process for the SOI-based sample component was to develop a more detailed stratification of the SOI tax filers based on the value of their total net worth index. The outcome of this step is summarized in Table 10. Seven net worth substrata of SOI frame elements have been explicitly defined. The seventh net worth stratum (labeled 3D) contains SOI frame elements whose net worth index exceeds \$250,000,000 dollars (roughly equal to the Forbes 400 [4] list cutoff). In developing the sample for the 1989 SCF, these extremely wealthy households were excluded from the sample selection and the survey data collection. The choice of boundary values for defining the six remaining non-censored net worth strata is based on an

Figure I



| Table 10.--SOI-based Sample Stratification: New Worth Dimension | | | | | |
|--|----------------------------|---|----------------------------|--|--|
| 1989 SCF Sample General Net Worth Strata | | SOI-based Sample Design Primary Strata | | | |
| General Design Stratum | Net Worth Range | SOI-based Sample Primary Stratum | Net Worth Range | Sample Allocation to Primary Strata | |
| 1 | \$0-\$99K | 1 | \$0-\$99K | 75 | |
| 2 | \$100K-\$999K | 2A | \$100K-\$499K | 125 | |
| | | 2B | \$500K-\$999K | 225 | |
| 3 | \$1 Million + | 3A | \$1M-\$2.49M | 191 | |
| | | 3B | \$2.5M-\$9.99M | 128 | |
| | | 3C | \$10.0M-\$250M | 107 | |
| | | 3D | \$250M + | Censored | |

application of the optimal stratification guidelines proposed by Dalenius and Hodges (1959). The proposed sample allocation to each of the six explicit net worth strata was performed in accordance with standard procedure for Neyman (optimal) allocation of the sample based on the stratum sizes and the variances of the predicted net worth values within each of the six strata (Cochran, 1977). Figure II illustrates cumulative distributions of the net worth index values for cases assigned to the six net worth strata.

V. B.5 Secondary Stratification by AGI Level and Business/Non-business Status

The secondary dimension of the 1989 SCF stratification plan for the SOI-based sample was constructed through a collapsing of the original 1987 SOI strata to form eight combined strata -- four business and four non-business -- which represent varying levels of adjusted gross income. Table 11 below summarizes the general definitions of these eight secondary strata.[5]

For the general study of household assets and wealth, household AGI is not in and of itself a complete stratifier. In the 1983 SCF, the correlation of household AGI and Net Worth was estimated at $r=.50$; however, for restricted ranges of AGI the

correlations can be considerably lower. (See Table 12.) The value of using AGI level as a secondary stratifier is that while household AGI may not be very highly correlated with total net worth, there is good evidence from the 1983 SCF and other data sources that AGI level does influence the particular choices of investments and assets which contribute to household wealth. Thus, secondary stratification by AGI level is important for improving the precision of analysis of the composition of household's wealth.

Likewise, from analysis of the 1983 SCF, net worth characteristics of households with significant amounts of business or farm income are known to

differ from those of "non-business" households. The original strata definitions used in the 1987 SOI Tax Model reflect the business/non-business status of tax filing households and this basic distinction is maintained in the combinations of SOI strata which have been used in the selection of the 1989 SCF sample.

Figure III presents the empirical distributions of the predicted net worth values within each of the eight (8) secondary strata of SOI Tax Model elements. As the figure indicates, a secondary AGI-based stratum can span more than one primary net worth stratum. In the final selection of the stratified

| Table 11.--1989 Survey of Consumer Finances | | |
|---|-----------------------|----------------------------------|
| Secondary Stratification of the SOI-based Sample by Adjusted Gross Income and Business/Non-Business Status | | |
| 1989 SCF Secondary Stratum | Status | Approximate AGI Range |
| 11 | Non-Business/Non-Farm | \$0-\$99,999 |
| 12 | Non-Business/Non-Farm | \$100K-\$199K |
| 13 | Non-Business/Non-Farm | \$200K-\$999K |
| 14 | Non-Business/Non-Farm | \$1M and higher |
| 21 | Business and Farm | \$0-99,999 |
| 22 | Business and Farm | \$100K-\$199K |
| 23 | Business and Farm | \$200K-\$999K |
| 24 | Business and Farm | \$1M and higher |

| Table 12.--Estimated Correlation Between Adjusted Gross Income (AGI) and Major Income Variables and Net Worth* | | | | | | |
|---|--------------------------------------|-------------------------------------|--|--------------------------|-----------------------------|-----------------------------|
| Variable | Total Sample (n=4103) | Adjusted Gross Income | | | | |
| | | 1983 SCF Area Sample | 1983 SCF High Income Categories | | | |
| | | | <100K (n=3632) | Total (n=471) | 100-199K (n=182) | 200-499K (n=190) |
| Wages and Salary | .4552 | .7221 | .2214 | .2730 | .1930 | .0099 |
| Profession, Business | .4546 | .3758 | .2801 | .1575 | .0851 | .0450 |
| Nontaxable Interest | .4934 | .2278 | .3970 | .0676 | .1465 | .2791 |
| Taxable Interest | .6123 | .3217 | .5394 | .0588 | .0785 | .3882 |
| Dividends | .4884 | .2321 | .3657 | .0940 | .2261 | .0059 |
| Interest on Bonds | .6520 | .2107 | .6653 | .0093 | .1029 | .6758 |
| Rent and Trusts | .4796 | .1704 | .4724 | .0650 | .1330 | .3889 |
| Net Worth | .4997 | .1802 | .4007 | .0908 | .1665 | .3036 |

*Correlation estimates (unweighted) from the 1983 Survey of Consumer Finances.

Figure II Empirical Distribution of Predicted Net Worth Within Primary Strata

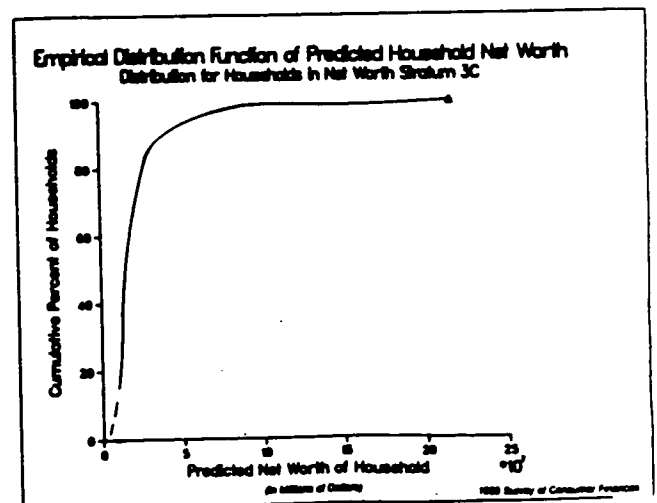
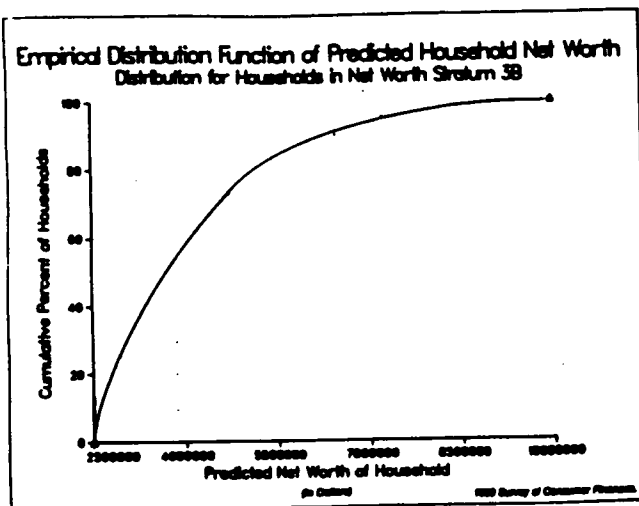
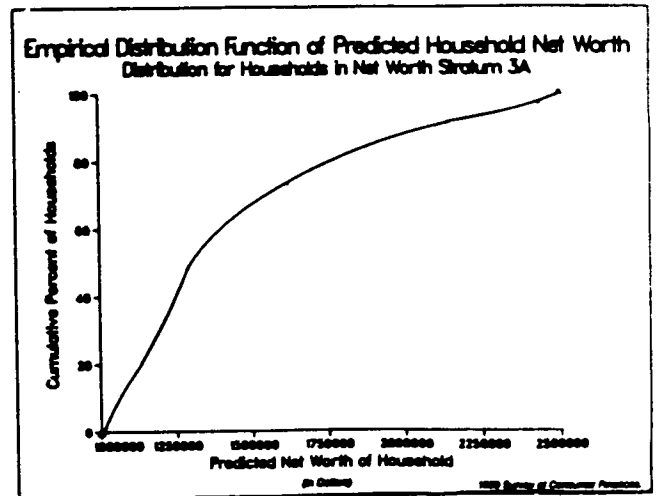
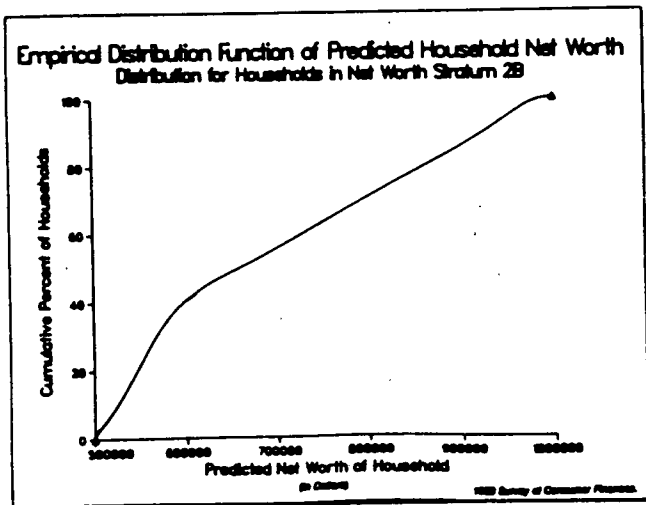
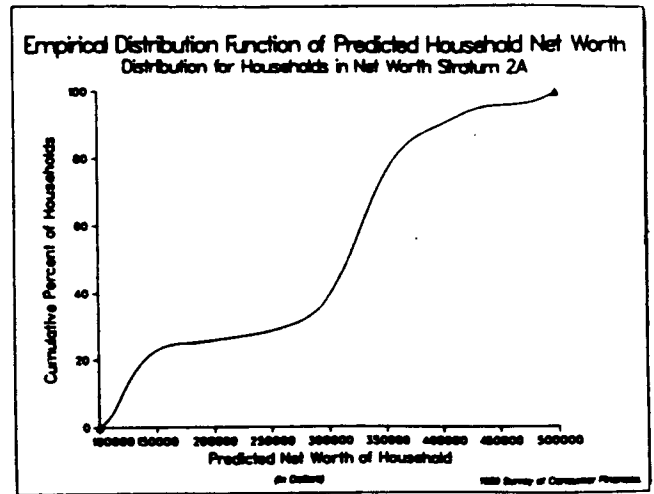
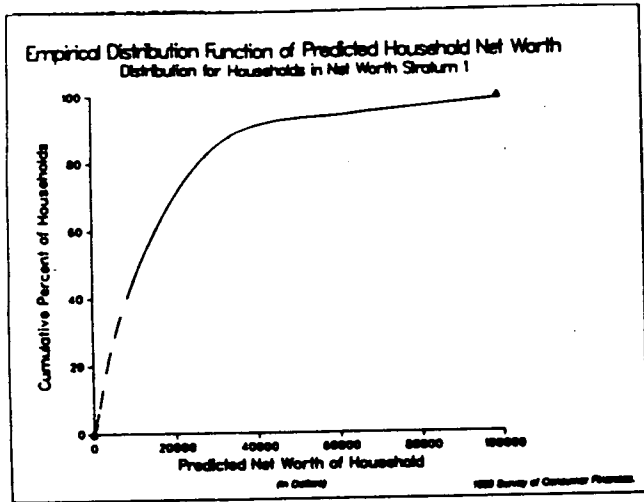
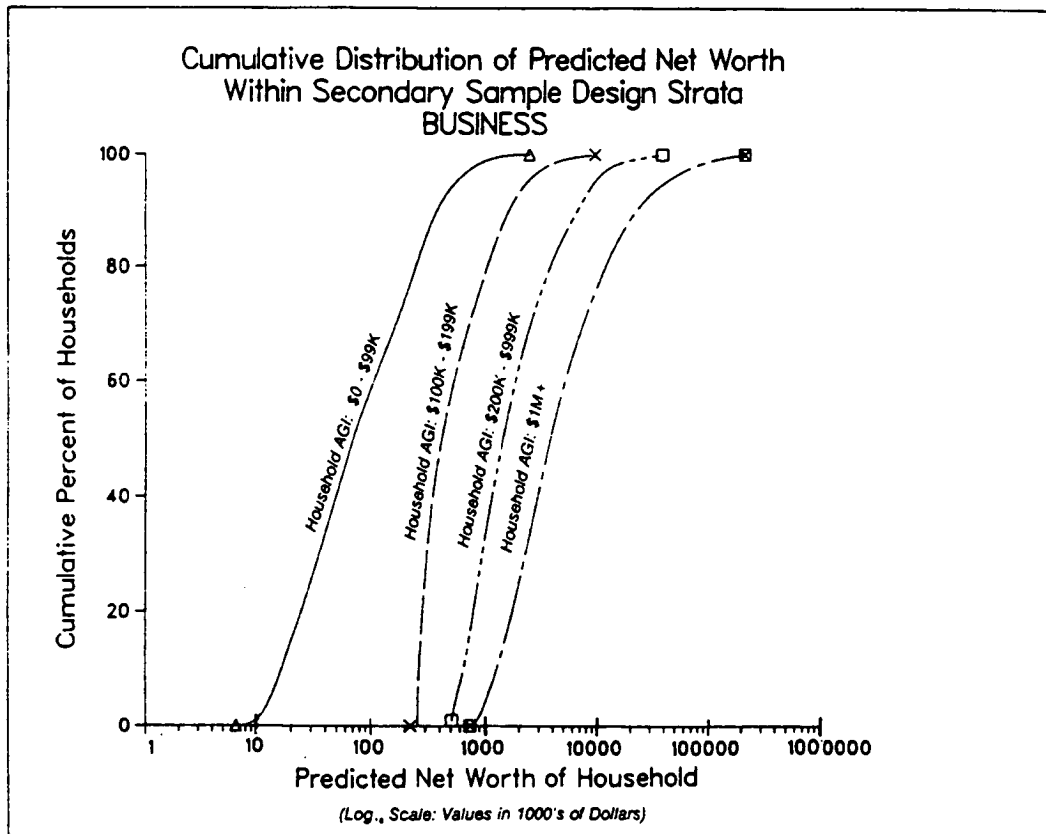
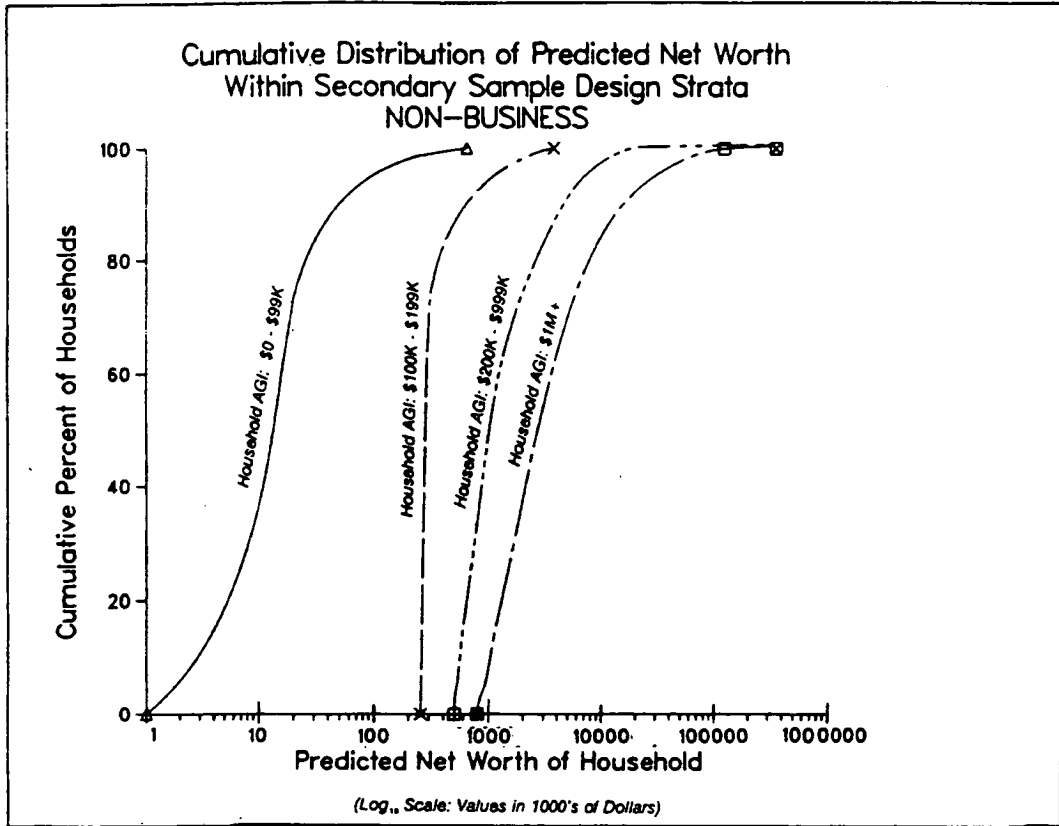


Figure III



random sample of SOI taxpayer elements for the 1989 SCF, specific sample allocations were made to each primary stratum (see Table 10). Within a primary net worth stratum, the sample selections were permitted to distribute proportionately across the AGI/Business:Non-business secondary strata.

VI. Summary

VI. A Implications of the 1989 SCF Design for Data Analysts

The 1989 SCF dual-frame probability sample design that has been described in the preceding sections is fully compatible with sample-based or "design-based" methods of estimation and inference. In addition, the structure of the sample design (through stratification and disproportionate allocation) should provide econometricians and other model-based analysts with a data set that is efficient and robust for their purposes.

As described in the preceding section, the 1987 Statistics of Income Tax Model data base has an essential role in the design stage of the 1989 SCF. Aggregate data from the SOI program will also contribute to the estimation and analysis phase of the survey, but the part that it will be allowed to play will be somewhat restricted in order to guarantee the individual respondents' right to privacy and nondisclosure of their tax data. By the terms of the research agreement between the Survey Research Center and the federal government sponsors of the 1989 SCF, neither party will perform an exact match of household survey responses to tax data from the SOI Tax Model data base. Furthermore, selective top coding and other protective procedures will be applied to the public release versions of the final data set to guarantee that the identity of survey households is protected against disclosure.

Although an exact micro-level linkage of the survey responses and Tax Model data is precluded, there are several ways in which the aggregate level data from the Statistics of Income data bases can be incorporated into the 1989 SCF analysis without a risk of disclosure or breach of the confidentiality promise made to the individual survey participants. Aggregate level statistics can be computed from the SOI Tax Model data base and employed as

poststratification controls in the development of the case weights required for descriptive analysis of the 1989 SCF data set. For example, Higgins and Fay (1988) report significant improvements in the precision of Survey of Income and Program Participation (SIPP) estimates of household income characteristics when estimates of total numbers of households by AGI category from the Internal Revenue Service (IRS) Individual Master File are incorporated into the poststratification weight. For the 1989 SCF, the poststratification uses of the SOI Tax Model estimates could be extended beyond simple controls by AGI category, to include controls on total numbers of households with selected earnings and asset characteristics such as the existence and general amounts of business or property income, the presence or absence of dividend income, etc.

VI. B Relationship of the 1989 SCF to Other Programs of Wealth Estimation

The 1989 SCF is one of several ongoing programs of research on the distribution and characteristics of wealth in the U.S. population. While there are high expectations for the 1989 SCF data product itself, the greatest long term benefit of the survey may be realized when the information that it provides is integrated with and supplemented by data and results from these other research programs. Even though exact match linkages of the 1989 SCF and SOI Tax Model data bases are not a possibility, it is very likely that parallel, aggregate level analyses of these two data sources will yield an improved income flow capitalization model for the SOI Tax Model data.

Age specific analysis of the 1989 SCF data on the net worth of households may provide new insights for strengthening the estate multiplier program of wealth estimation. (McCubbin, 1987). Conversely, the tremendous data resources of the estate multiplier program and the Intergenerational Wealth Study (Medve, 1987) can be used in confirming or refining the survey-based and capitalization models used in the other programs.

Integrated approaches which combine both survey data and other wealth-estimation methods are not necessarily new. A precedent is found in important work by Greenwood (1983). Working with a special

merged file of Current Population Survey (CPS) and federal income tax return data, Greenwood used income-capitalization methods to estimate the total value of each sample household's assets in the form of interest-bearing debts instruments and corporate stock. Separately, estate tax data were used to model the regression relationship between holdings of these two classes of financial assets and total reported financial and non-financial net worth of the deceased. This predictive regression model was then applied to each household in the special CPS sample to estimate total net worth as a function of the household's capitalized estimates of interest bearing investments and corporate stock.

In her paper, Greenwood uses the CPS sample data primarily as a means for providing a representative framework around which to build the income-capitalized estimates of selected assets (using merged tax return information) and subsequently the regression predictions of corresponding total net worth. Having attached predicted values of total net worth to each sample case, the CPS-sample could then be used to develop designed-based estimates of total net worth for the total U.S. household population and its subclasses.

Similar methods, involving the 1989 SCF, will most certainly be investigated.

APPENDIX A

1989 SCF designated respondents who are selected from SOI Tax Model frame have been sent a special "consent package" approximately three weeks in advance of a contact by the Survey Research Center's interviewer. The consent package contains a letter of explanation and introduction from the Director of SRC, a supporting letter from Dr. Alan Greenspan, Chairman of the Federal Reserve, and a franked post-card which the respondent is instructed to mail back to the Survey Research Center if he or she decides not to participate in a study interview. This "passive consent" procedure differs from the "active consent" procedure used in the 1983 SCF where the designated respondent returned the post card only if he or she agreed to be contacted for a study interview. The passive consent procedure developed for the 1989 SCF significantly improved

cooperation on the part of sampled individuals while at the same time guaranteeing the right of privacy and minimizing the inconvenience for individuals who choose not to participate.

For SOI frame respondents who consented to participate in the study, all other survey procedures including interviewer contact and questionnaire materials were identical to those used in the area probability component of the sample design.

BIBLIOGRAPHY

- Cochran, W.G., *Sampling Techniques*, 3rd ed., New York, John Wiley and Sons, 1977.
- Curtin, Richard T.; Juster, F. Thomas; and Morgan, James N., "Survey Estimates of Wealth: An Assessment of Quality," in Robert E. Lipsey and Helen S. Tice (eds.), *The Measurement of Saving, Investment, and Wealth*, Chicago, The University of Chicago Press, 1989.
- Dalenius, T., and Hodges, J.L., "Minimum Variance Stratification," *Journal of the American Statistical Association*, Vol. 54, 1959, pp. 88-101.
- Greenwood, Daphne, "An Estimation of U.S. Family Wealth and its Distribution From Micro-Data, 1973," *Review of Income and Wealth*, March 1983, pp. 23-44.
- Hartley, H.O., "Multiple Frame Methodology and Selected Applications," *Sankhya*, Vol. 36, 1974, pp. 99-118.
- Heeringa, Steven G., and Curtin, Richard T., "Household Income and Wealth: Sample Design and Estimation for the 1983 Survey of Consumer Finances," in Wendy Alvey and Beth Kilss (eds.), *Statistics of Income and Related Administrative Record Research: 1986-1987*, Selected Papers Given at the 1986-1987 Annual Meetings of the American Statistical Association, Washington, D.C., Department of the Treasury, Internal Revenue Service, November 1987.
- Heeringa, Steven G., Connor, Judith H., and Darrah, Doris C., "1980 SRC National Sample: Design

- and Development," Ann Arbor, The University of Michigan, Institute for Social Research, 1986.
- Heeringa, Steven G., and Woodburn, Louise, *Sample Design Documentation for the 1989 Survey of Consumer Finances*, Ann Arbor, The University of Michigan, Institute for Social Research.
- Higgins, Vicki J., and Fay, Robert E., "Use of Administrative Data in SIPP Longitudinal Estimation," *Proceedings of the Survey Research Methods Section*, American Statistical Association, Washington, D.C., 1988.
- Internal Revenue Service, *Statistics of Income - 1987, Individual Income Tax Returns*, Washington, D.C., U.S. Government Printing Office.
- McCubbin, Janet, "Improving Wealth Estimates Derived From Estate Tax Data," in Wendy Alvey and Beth Kilss (eds.), *Statistics of Income and Related Administrative Record Research: 1986-1987*; Selected Papers Given at the 1986-1987 Annual Meetings of the American Statistical Association, Washington, D.C., Department of the Treasury, Internal Revenue Service, November 1987.
- Medve, Kathy, "Intergenerational Wealth Study," in Wendy Alvey and Beth Kilss (eds.), *Statistics of Income and Related Administrative Record Research: 1986-1987*, Selected Papers Given at the 1986-1987 Annual Meetings of the American Statistical Association, Washington, D.C., Department of the Treasury, Internal Revenue Service, November 1987.
- Projector, Dorothy S., and Weiss, Gertrude S., *Survey of Financial Characteristics of Consumers*, Federal Reserve Technical Paper, Washington, D.C., Board of Governors of the Federal Reserve System, 1966.
- Ruggles, Richard, and Ruggles, Nancy D., "Integrated Economic Accounts for the United States, 1947-1980," *Survey*, May 1982.
- Scheuren, Fritz, and McCubbin, Janet, "Piecing Together Personal Wealth Distributions," Wendy Alvey and Beth Kilss (eds.), *Statistics of Income and Related Administrative Record Research: 1986-1987*, Selected Papers Given at the 1986-1987 Annual Meetings of the American Statistical Association, Washington, D.C., Department of the Treasury, Internal Revenue Service, November 1987.
- Steuerle, C. Eugene, "The Relationship Between Realized Income and Wealth," *Statistics of Income Bulletin*, Internal Revenue Service, Spring 1983, pp. 29-34.
- Strudler, Michael, "General Description Booklet for the 1982 Individual Tax Model file," Washington, D.C., Internal Revenue Service, Statistics of Income Division, 1983.
- Wilson, John F.; Freund, James L.; Yohn, Jr., Frederick O.; and Lederer, Walter, "Measuring Household Saving: Recent Experience from the Flow-of-Funds Perspective," in Robert E. Lipsey and Helen S. Tice (eds.), *The Measurement of Saving, Investment, and Wealth*, Chicago, The University of Chicago Press, 1989.

Footnotes

- [1] No detailed financial data or other information from the individual tax return records will ever be shared with the Survey Research Center. Conversely, before the 1989 SCF data are released to the sponsoring agencies and the general public, all identifying codes or variables which might permit an exact match to a tax return of other administrative record will be suppressed.
- [2] Individual income tax filers report wages and salary income on IRS Form 1040. Taxable interest, nontaxable interest and dividends are also declared on the 1040. Schedule C is used to report business income and farm income is reported on Schedule F. Individual filers who have either short- or long-term capital gains report such gains by filing Schedule D. Rental income, royalties, income from partnerships, estates, trusts are reported on Schedule E.

- [3] Regression analysis on the 1983 SCF data indicates that there is both an asset amount effect and an income effect, but no significant age effect. In general, rates of return are higher for larger asset amounts, lower for higher income categories. Since the two are positively correlated, the net effects are roughly the sum of the coefficients. It appears that fixed-income yields, on balance, rise with income, while dividend yields fall with income (presumably because the rich want capital gains and not dividends).
- [4] The Forbes 400 is a list of the presumed 400 wealthiest individuals in the U.S. The list is published annually by *Forbes Magazine*.
- [5] 28 "Business" is defined as taxpayer units who report 1987 income from a personal business (Schedule C) or farming operation (Schedule F).