List Sample Design for the 1998 Survey of Consumer Finances

Arthur B. Kennickell
Senior Economist and Project Director
Survey of Consumer Finances

Board of Governors of the Federal Reserve System
Mail Stop 153
Washington, DC 20551

Voice: 202-452-2247
Fax: 202-452-5295
Email: akennickell@frb.gov

April 2, 1998

The author wishes to thank Kevin Moore and Amy Stubbendick for outstanding research assistance in the work reported here. Fritz Scheuren has been a guiding source of inspiration at innumerable stages in the SCF project. Louise Woodburn, who was a close colleague on the SCF for many years, has a strong presence behind the work summarized in this paper. Barry Johnson has long been an essential player in the work of selecting and implementing the sample. The author is also grateful to have been able to work with Marty Frankel, Steve Heeringa, and Douglas McManus and absorb so many of their good ideas. Thanks to Gerhard Fries, Martha Starr-McCluer, Annika Sundén and Brian Surette for comments. The opinions contained here are the responsibility of the author alone and do not necessarily reflect those of either the Board of Governors or the Federal Reserve System.

The Survey of Consumer Finances (SCF) is well known as a source of household level data for the U.S., primarily in the areas of portfolio choices and use of financial institutions.[1] Many of the financial behaviors measured by the survey are relatively rare—for example, ownership of closely held corporate businesses. At the same time, the survey is also used extensively to study more broadly distributed behavior—for example, the use of credit cards. To provide reliable coverage of both types of behaviors, the SCF employs a dual frame sample design, including a standard multi-stage area-probability sample (see Tourangeau et al. [1993]) and a list sample designed specifically to over-represent wealthy households. This paper focuses on the selection of the list sample for the 1998 Survey of Consumer Finances (SCF).

The following section provides some general background on the survey and earlier approaches in the design of the list sample. The next section describes in detail the selection of the 1998 sample. A final section describes areas for future research.

I. Background

The SCF is conducted by the Board of Governors of the Federal Reserve System in cooperation with the Statistics of Income Division (SOI) of the Internal Revenue Service. Beginning with the 1983 survey, the first of the current series, the SCF has been conducted every three years. Data for the surveys were collected by the Survey Research Center at the University of Michigan from 1983 to 1989, and by the National Opinion Research Center at the University of Chicago (NORC) beginning with the 1992 survey.

The SCF list sample design has evolved in many ways since the 1983 survey. As our understanding of the underlying processes has deepened, and as more access to detailed information has become available, it has been possible to refine the design of the list sample. One thing that is common to all of the SCF list samples to date is that they were selected using a stratified probability design from the Individual Tax File (ITF) maintained by SOI.[2]

---

[1]For an overview of the 1995 SCF, see Kennickell, Starr-McCluer and Sundén [1997]. Other information about the survey, including the data, is available on the Internet at http://www.bog.frb.fed.us/pubs/oss/oss2/scfindex.html.

[2]See Statistics of Income [1992].

Although the details of the 1983 list sample design still cannot be released to the public, the sample may be described without disservice as a "high income sample," which was stratified by various types and amounts of income.[3] The respondents were selected by SOI, and the names and addresses of those people were transmitted to the Office of the Comptroller of the Currency (OCC), an agency of the Department of the Treasury. The OCC mailed a letter to each respondent along with a postcard which was to be returned if the person *agreed* to participate in the survey. No follow up for unreturned postcards was allowed. Given this strong requirement of active agreement to participate, it is not surprising that the completion rate was only about 10 percent of the initial sample.

Significant changes occurred in the design of the 1989 SCF list sample.[4] A project group combining statisticians from SOI, the main statistician from the vendor, and economists from the Office of Tax Analysis and the Federal Reserve Board was charged with creating a new design that would be broadly more efficient for wealth estimation. Two particularly important changes came out of this collaboration. First, a proxy for a household's wealth—subsequently known as a "wealth index," or WINDEX0 (see figure 1)—was created by grossing up capital income flows observed in the tax data in the 1987 ITF using

Figure 1: Definitions of WINDEX0 and WINDEX1

Assumed rates of return:

$$\text{WINDEX0} = \Sigma\, (1/r_i)\, Y_i$$

where Y_i is a type of capital income in the ITF for the year two years before the survey, and
r_i is the associated rate of return

Estimated model:

$$\text{WINDEX1} = F(\mathbf{Y}, \epsilon; \beta)$$

where **Y** is a vector of characteristics in the ITF (including income flows, filing status, age, etc.) for the year two years before the survey,
ϵ is a random disturbance
β is a vector of parameters estimated by regressing net worth from the survey on F
F is a function relating SCF net worth and **Y**

[3]Indeed, it has only been relatively recently that this author has seen the technical documentation of the design. Heeringa and Curtin [1986] and Avery, Elliehausen and Kennickell [1988] provide some background on the sample.

[4]The 1986 SCF was a reinterview with a subsample of participants in the 1983 survey. The 1989 survey consisted of two major parts: a reinterview with a sample of 1983 participants and a new cross-section sample (including a list sample). The discussion here focuses on the new cross-section sample. For details on the entire 1989 sample, see Heeringa, Conner and Woodburn [1994].

average market rates of return.[5] For example, if the rate of return on assets yielding taxable interest were 5 percent, then the asset value corresponding to the income would be 20 times (1/.05) the amount of interest income. The wealth index was used to stratify the sample, and observations with higher values of the index were sampled more intensively. The second important change was in the initial contact with the respondent. Each sample member was mailed a description of the survey along with a postcard to be returned if the respondent did *not* want to participate in the survey. All cases that did not return a postcard were as actively pursued by interviewers as the cases in the area-probability sample. Largely as a consequence of this change, the completion rate was dramatically higher in 1989.

Building on the collaboration begun with the 1989 survey, a stream of research addressed the reasonableness of assumptions underlying the design of the list sample, the characteristics of nonrespondents, and other related statistical areas.[6] Because this research was still in progress at the time the 1992 SCF sample was selected, only relatively small changes were made in the design—mainly, the number and definition of the list sample strata were changed to reflect more natural breaks in the data. In addition, the size of the list sample was approximately doubled.

In the 1995 survey, a more fundamental change was implemented. Following several years of negotiation and a review by disinterested outside statisticians, contractural agreements allowed for a limited match of 1992 survey data with the corresponding frame data.[7] A file (stripped of all identifiers) was constructed, including limited survey information on wealth, and restricted to use by the present author alone. These data were used to validate the usefulness of the 1992 WINDEX0 as a proxy for actual wealth values. The data were also used to estimate a model-based wealth index, WINDEX1 (see figure 1), for use in sampling for the 1995 survey. For this model, the log of net worth calculated from the 1992 SCF was regressed against a variety of income values and other characteristics in the 1990 ITF. To hedge against possible instability of WINDEX1 over time, the index ultimately used for stratification in 1995, WINDEXM (see figure 2), was the average of

---

[5]See Greenwood [1983] for a similar capitalization of income flows.

[6]See Kennickell and McManus [1993], Frankel and Kennickell [1995], Kennickell, McManus and Woodburn [1996], and Kennickell and Woodburn [1997].

[7]Further background on this is provided in Frankel and Kennickell [1995].

Figure 2: Definition of WINDEXM

WINDEXM = {[WINDEX0-median(WINDEX0)]/IQR(WINDEX0) +
[WINDEX1-median(WINDEX1)]/IQR(WINDEX1)} / 2

versions of WINDEX0 and WINDEX1 adjusted to have the same median and inter-quartile range (75^{th} percentile minus the 25^{th} percentile).

Assembling the SCF and ITF data to reestimate WINDEX1 for the 1998 survey offers an opportunity to evaluate the innovation in stratification for the 1995 survey. Figure 3 presents estimates from a logarithmic regression of net worth (computed from the 1995 SCF) on various combinations of wealth indices (computed from the 1993 ITF). The sample for this estimation excludes families with negative net worth and those that experienced a change in marital status between the time the tax return was filed and the interview was performed. Clearly, both WINDEX0 and WINDEX1 have strong explanatory power for net worth (columns 1 and 2 of the table). When both variables are included in the same regression (column 3), both are still significant, but the incremental addition to the R^2 of WINDEX0 over WINDEX1 is only 0.02 out of 0.73.

Figure 3: Logarithmic Regression of Net Worth Against Various Wealth Indices, 1995 SCF and 1993 ITF

	(1)	(2)	(3)	(4)
INTERCEPT	2.86	0.79	0.73	3.11
	0.23	**0.24**	**0.23**	**0.22**
	12.28	*3.30*	*3.17*	*14.45*
WINDEX0	0.79	.	0.29	.
	0.02	**.**	**0.03**	**.**
	50.16	*.*	*10.22*	*.*
WINDEX1	.	0.95	0.66	.
	.	**0.02**	**0.03**	.
	.	*57.36*	*20.01*	.
WINDEXM	.	.	.	0.78
	.	.	.	**0.01**
	.	.	.	*53.01*
R^2	0.65	0.71	0.73	0.68
N	1431	1431	1431	1431

Standard errors corrected for multiple imputation of the components of net worth are given in bold below each coefficient estimate; the corresponding t-statistics are given in italics.

The regressions summarize information from the entire distribution of outcomes. However, the greatest need for the index is in discriminating cases that are likely to have large wealth values. To address possible differences in predictive ability across the wealth distribution, figure 4 shows average

Figure 4: Distributions of WINDEX0, WINDEX1, and WINDEXM, by Deciles of Net Worth

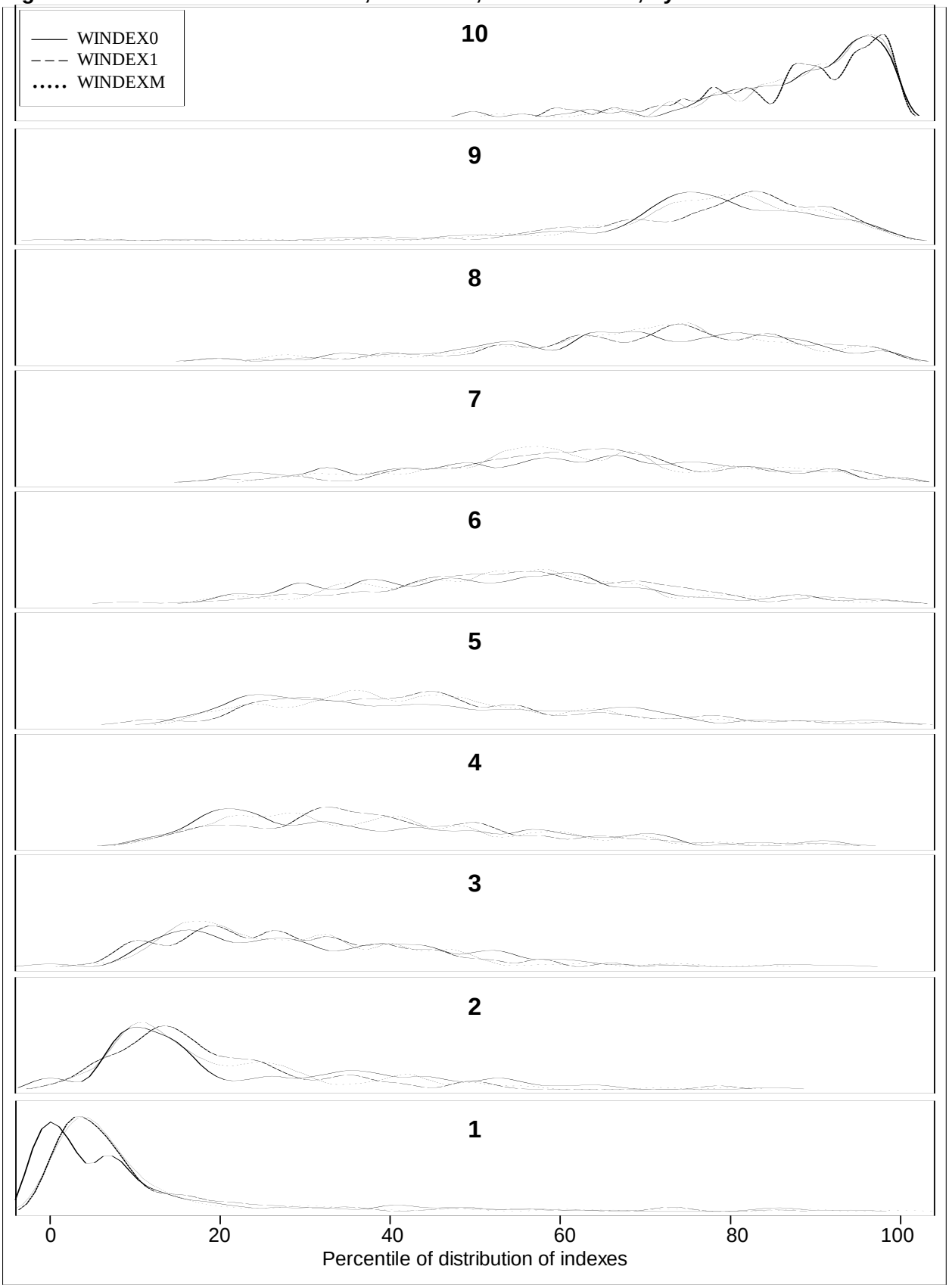
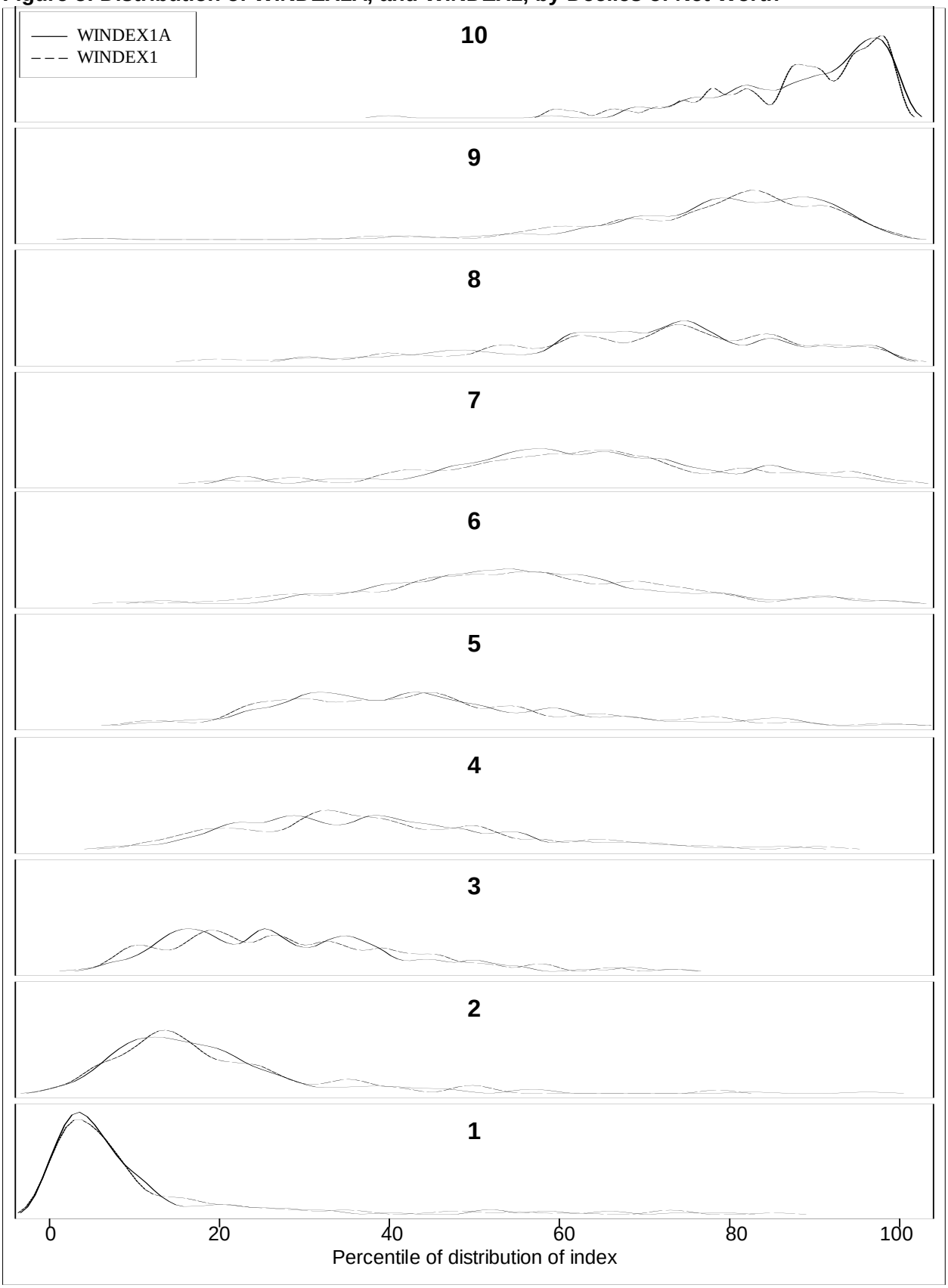


Figure 5: Distribution of WINDEX1A, and WINDEX1, by Deciles of Net Worth



shifted histogram (ASH) estimates of the indices by net worth groups.[8] Each of the horizontal panels in the figure contains an unweighted decile of the net worth distribution for the 1,519 list sample cases.[9] To remove irrelevant location and scale differences among the indices, the horizontal axis for the distributions is given on a percentile basis.

Displaying the data in this way shows that all of the indices do a good job of distinguishing the very wealthiest groups from the bottom groups. However, there is considerable spread in all of the indices even at the very top and bottom, and the ability of the indices to discriminate observations in the fifth through eighth wealth deciles is not as strong as one might like. Some of this dispersion is accounted for by households that changed composition between their filing of a 1993 tax return and their participation in the 1995 SCF; fluctuations in income are doubtlessly important (see Kennickell and McManus [1993]); and the variability of net worth due to imputation is also a contributing factor. Looking at the relative performance of the indices, WINDEX1 does appear somewhat more peaked on average than WINDEX0, but the differences are small. This fact is surprising, particularly given that when the merged data are used to actually estimate the rates of return for WINDEX0 (i.e., the r_i in figure 1) using least squares, the estimates are significantly different from the assumed values.[10]

One might expect there to be a high level of variability in the predictive power of WINDEX1: it is based on coefficients estimated using 1990 tax data and 1992 survey data, which are applied to 1993 tax data to predict wealth in the 1995 survey. Thus, the earlier structure of rates of return is imbedded in the model coefficients for WINDEX1, and the underlying relationships may have

---

[8]Missing data in the SCF are multiply imputed (Kennickell [1991]). For this figure and for figure 5, the net worth estimate for each observation has been averaged over the imputations.

[9]The unweighted decile points of the net worth distribution in the list sample are about: 10th percentile, $82 thousand; 20th percentile, $280 thousand; 30th percentile, $560 thousand; 40th percentile, $1.0 million; median, $1.7 million; 60th percentile, $2.9 million; 70th percentile, $5.4 million; 80th percentile, $11 million; and 90th percentile, $28 million. Obviously, these points are considerably higher than those for the population. For example, the weighted estimate of the median net worth for the full 1995 SCF is about $56 thousand.

[10]Only the estimated coefficient on taxable interest—14.6, implying a rate of return of about 6.9 percent—is similar to the value used in computing WINDEX0. Some others are negative or too small to be meaningful.

changed in important ways by 1995.[11] To address this question, figure 5 shows the distribution of WINDEX1 along with the values predicted from a model estimated from the match of the 1995 SCF data and the 1993 ITF data (WINDEX1A).[12] As in figure 4, the plots are shown separately for each of the deciles of the net worth distribution in the list sample. Contrary to expectation, it is difficult to argue that the differences in predictive ability are strong. Nonetheless, it may still be important to reestimate the model to protect against the possibility of changes in both rates of return and the income categories included on tax returns.

II. 1998 List Sample Selection

The 1996 ITF serves as the frame for the 1998 SCF list sample. The ITF is itself a sample of annual tax returns (1040) filed at any time during the year following the tax year: in the current case, the returns are ones filed at any time during 1997. The ITF sample is stratified by types of income received and selection rates vary by level of income.[13] In 1996, this file includes only about 126 thousand observations to represent about 121 million returns. Sampling variability in the ITF introduces complications into the use of the file as a frame for the selection of the list sample: generally frames are treated as nonstochastic. Fortunately, because the list sample is strongly tilted toward selecting relatively wealthy households and because selection rates tend to be very high for such cases in the ITF, the problem does not appear to be a serious one and it is ignored in the analysis of the data.[14] The selection weight of the ITF cases is treated as the initial size measure in the SCF sample selection.

The ITF also contains returns that are filed from places outside the 50 states and the District of Columbia. Because the reference population in the SCF is domestic households, all observations filed from foreign addresses and APO addresses are deleted. The 1996 file also includes returns for

---

[11]Because the model is a reduced form, the structure of rates of return is potentially implicit in all the coefficients of WINDEX1, and thus not open to simple adjustments.

[12]The construction of coefficients used in computing WINDEX1A is discussed more fully below in the exposition of the 1998 sample design. This model is somewhat more elaborate than the one originally estimated for the 1995 sample selection. The principal difference is the inclusion in the new model of additional controls for types of negative income.

[13]See Statistics of Income [1996] for a discussion of the ITF design.

[14]See Kennickell and McManus for additional discussion.

years before 1996; most of these cases are ones where the taxpayer has filed an amended return for an earlier year, but some are returns filed late. Before any selection is preformed, the observations are first sorted by tax year of returns and taxpayer ID, and only the return filed for the most recent year is retained.

There are some differences in the target populations in the SCF and editions of the ITF used for sampling. The target population for the SCF is the set of households in existence at the time of the survey with at least one household member aged 18 or older.[15] In contrast, the target population for the ITF is all Federal tax returns filed in the year preceding the SCF. The timing differences do not generate many practical problems in administering the survey, and adjustments at the weighting stage appear to be a sufficient way of dealing this problem (see Kennickell and Woodburn [1997[). However, the differences in the units of observation—tax returns versus households—and the differences in age coverage require adjustments at the sampling stage.

Multiple individuals within a household might file a tax return, and thus ultimately a given household could have multiple chances of being selected into the SCF sample. Some adjustment must be made to the weights of such cases to correct for the multiplicity of selection possibilities. Although there is no information in the version of the ITF used for sampling that can be used to create accurate "tax families" of taxpayers living at the same address, we do know whether a given taxpayer is married and filing a separate return. We assume that all persons whose filing status is "married filing separately" have a spouse with identical characteristics, and we divide the ITF weight by two.[16]

---

[15]The core information in the SCF is collected for the "primary economic unit" (PEU). The PEU is intended to represent the economically dominant person or couple (who may be married or living as partners) and all other people in the household who are financially dependent or interdependent with this individual or couple. More precisely the PEU includes in order of precedence: a single individual who lives alone; two individuals who are married or living as partners and any dependents; the individual who owns the housing unit or whose name is on the lease (choosing the one nearest to age 45 when there are multiple such individuals), that individual's spouse or partner, and any other people dependent on that couple or individual; when no one owns or rents, the person whose age is nearest to 45, that person's spouse or partner, and any other people dependent the individual nearest to age 45 or that person's spouse or partner.

[16]Kennickell and McManus [1993] show that the problems induced by more complex households are least important among wealthier households.

Another difference in the target populations is that the ITF includes taxpayers of all ages, but only households where the head of the household (or that person's spouse or partner) is aged 18 or older at the time of the interview is eligible for the SCF. For the person listed as the primary taxpayer in the ITF, the file contains a two-digit variable indicating the year of birth matched from Social Security records. Age is missing for about 1,000 cases, and these cases are all assumed to be aged 45 for purposes of sample selection. There is some ambiguity about some other cases with birth years between 86 and 96: these people could be either very old or very young. Age assignments for these cases were made as follows: if the return claimed a deduction for being over age 65, had a filing status other than single, or contained positive wage or pension income, then the person was assumed to have been born in the 19th Century; if the return contained a form 8615 (minor child), it was assumed that the person was a child. Based on the number of such cases remaining after these assignments, it appeared extremely unlikely that more than a very small number of such people were adults, and all of them were assumed to be children. Filers with a coded birth year between 96 and 99 were assumed to have been born in the 19th Century, and those with are year smaller than 86 were assumed to have been born in the 20th Century (a zero was assumed to indicate 1900). All returns with an age of the primary filer less than 16 (about 18 at the time of the survey) were discarded, leaving about 116 thousand observations at this stage of selection.

Following the design of the 1995 SCF, the sample is stratified by WINDEXM, which is defined in terms of WINDEX0 and WINDEX1, each updated for the new sample. The 1995 specification of WINDEX0 is given in figure 6. The coefficients of the model for WINDEX1 are estimated using a special matched file of 1995 SCF data and 1993 ITF data, and predicted values are computed using the 1996 ITF.[17] The model is specified in log terms and estimated using a robust regression technique. The variables included in the model are listed in figure 7.[18] Further operations

[17]These same coefficients were used to calculate WINDEX1A.

[18]Table A1 in the confidential appendix gives the estimated coefficients along with standard errors corrected for multiple imputation.

Figure 6: Definition of WINDEX0, 1998 SCF List Sample

WINDEX0 is defined as the sum of the following:

Housing equity:
- Median housing value in the 1995 SCF by income groups:

Income ($ thou.)	*Median house value ($ thou.)*
under 60	30
60-120	125
120-250	188
250-1,000	350
1,000-5,000	750
5,000 or more	900

- Multiply by (156.9/152.4) to adjust for inflation (CPI)

Taxable interest income
- Divided by 0.0750
- Rate on corporate bonds, seasoned issues, all industries
- December 1996 *Federal Reserve Bulletin*, table I.32, line 33

Non-taxable interest income
- Divided by 0.0538
- Rate on Aaa state and local notes and bonds
- December 1996 *Federal Reserve Bulletin*, table I.32, line 30

Dividend income
- Divided by 0.0201
- Dividend-price ratio, common stocks
- December 1996 *Federal Reserve Bulletin*, table I.32, line 39

Absolute value of rents and royalties
- Divided by 0.0692
- Assume follows effective mortgage yield
- December 1996 *Federal Reserve Bulletin*, table I.53, line 7

Absolute value of other types of business, farm, and estate income
- Divided by 0.0487
- Assume average of interest and dividend rates

Sum of absolute values of long term, short term, and other capital gains

Multiply index by two if the filing status is "married filing separately"

Income data are taken from the 1996 ITF

on WINDEX1 take place on the exponentiated predicted value of the model. For returns with the filing status "married filing separately," values of WINDEX0 and WINDEX1 are doubled to account for the wealth of both partners.

To remove irrelevant location and scale differences between WINDEX0 and WINDEX1, both are standardized using weighted estimates of their medians and interquartile ranges. The weights used in the estimates are the ITF size measures adjusted for filing status, and the population included is the set of cases remaining at this stage of selection.

The ultimate stratifying variable, WINDEXM, is defined as the average of the standardized values of WINDEX0 and WINDEX1. It is desirable to keep the definitions of the strata defined in terms of WINDEXM as comparable as possible to those used in earlier surveys. Rather than devise a way to translate the units of WINDEXM into meaningful dollar terms, beginning with the 1995 design, the strata are defined in terms of percentiles of the distribution of the index. The exact boundaries of the eight strata cannot be revealed to the public, but it is possible to say that even the

Figure 7: Coefficients of WINDEX1, 1998 SCF List Sample

Have taxable interest		Log(long-term losses)	
Log(taxable interest)	*	Have short term losses	
Have nontaxable interest	+	Log(short term losses)	+
Log(nontaxable interest)	*	Have estate income	
Have dividends		Log(estate income)	
Log(dividends)	*	Have pension income	
Have gross Schedule C income		Log(pension income)	
Log(gross Schedule C income)		Have royalties	+
Have partnership/s-corp income		Log(royalties)	*
Log(partnership/s-corp income)	+	Have real estate tax deduction	*
Have Schedule C receipts	+	Log(real estate tax deduction)	*
Log(Schedule C receipts)	+	Have itemized deductions	+
Have negative Schedule C income		Log(itemized deductions)	
Log(negative Schedule C income)		Log(expanded income)	
Have schedule E income		Log(expanded income)**2	*
Log(schedule E income)		Have negative expanded income	+
Have farm income		Log(negative income)	*
Log(abs(farm income))		Filing status head of household	
Have negative farm income		Filing status single	
Log(negative farm income)		Filed from Northcentral region	
Have gross farm income		Filed from Southern region	
Log(gross farm income)	*	Filed from Western region	+
Have capital gains or losses		Log(age primary filer)	*
Log(abs(gains and losses))		Log(age primary filer)**2	*
Have capital losses		Intercept	*
Log(capital losses)			
Have long-term losses			

Adjusted $R^2 = 0.72$

+ indicates that the estimate is significant at the 5 percent level; * indicates that the estimate is significant at the 1 percent level.
Standard errors used in the significance test are corrected for multiple imputation
All dollar values are taken as absolute values with a floor of one.

fourth stratum is well above the 99th percentile of the population distribution of WINDEXM.[19] The highest stratum (stratum 8) is excluded from the sample. The exclusion argument is based on the assumption that although such people are very wealthy, they are very small in number and unlikely to participate in any type of survey.[20] Moreover, disclosure risks would make it impossible to release information collected from such cases except in severely altered form that would limit the usefulness of the data.

[19]For readers within the SCF project group, figure A2 in the classified appendix specifies the stratum boundaries.

[20]The final analysis weights are adjusted to account for cases in all eight strata. See Kennickell and Woodburn [1997].

A small number of cases in the lowest wealth index stratum had either very large positive income or large absolute negative income. Close inspection revealed that these observations have small weights and types of income that would be difficult to incorporate into the wealth indices. To avoid the possibility of extreme outliers at the analysis stage, the definition of stratum two was expanded to include cases originally in stratum one that had expanded income of more than $500,000 or less than -$25,000.[21] Approximately 50 observations are affected by this decision, and their total measure of size is comparably small.

For reasons associated with cost and administration of the survey, the part of the ITF remaining at this stage of selection is reduced to include only taxpayers who live in the primary sampling units (PSUs) in the 1990 NORC area-probability sample.[22] Implicitly, this approach accepts the probability selection mechanism for the PSUs as appropriate for the list sample. In principle, serious inefficiencies could be introduced into the sample by retaining areas selected on the basis of the size of their overall populations and using these weights to adjust the ITF size measures: the distribution of wealthy households may look quite different from that of the general population. Fortunately, other research indicates that this problem is manageable.[23]

ZIP code information, which in theory represents the home address of the primary filer, is available for every case in the ITF. Using auxiliary data, every ZIP code was associated with a state-county (FIPS) code, and only those in the FIPS codes included in the PSUs are retained. The initial size measure of these observations is increased by the inverse of a smoothed version of the probability

---

[21]Expanded income is defined as follows: form 6251 other adjustments + PRF amount + incentive stock options + tax-exempt interest + tested Social Security - taxable Social Security + AGI - investment interest - non-limited Schedule A miscellaneous deductions - other gambling amounts - Max(0, foreign earned income exclusion) - deductions from excluded income.

[22]See Tourangeau et al. [1993] for a discussion of the area-probability sample design.

[23] Frankel and Kennickell [1995] investigate this problem by comparing a hypothetical sample designed to provide an optimal area-probability sample of relatively wealthy households to the actual sample. They show that there are important differences between the two approaches, but at least for the current set of PSUs, the problem is not so serious that it cannot be addressed at the weighting stage.

of selection of the PSUs in the area-probability sample.[24] After deleting cases outside the sample PSUs, approximately 73 thousand observations remain eligible for selection at this stage.

Analysis of earlier surveys—particularly the examination of estimated variances after nonresponse adjustment—has led to the determination of a minimum number of observations in each stratum required for the reliability of key estimates. For the 1998 SCF, this required minimum number of cases to be completed within each stratum is fixed by prior agreement with NORC.[25] Within each stratum, the difference between the sample size and the number of completed cases comprises cases that refused participation in the survey, those that are ineligible (e.g., deceased or living out of the country for the duration of the survey), and cases deleted in a mandatory review of the sample (see figure 8).[26]

Figure 8: Decomposition of Sample Size

$S_i = C_i + NR_i + IE_i + DE_i$

where the subscript indicates the stratum,
S_i is the sample size,
C_i is the number of completed interviews,
NR_i is the number of nonrespondents,
IE_i is the number of ineligible cases, and
DE_i is the number of deleted cases.

In addition to achieving the targets, another important goal is to face the interviewers with a challenge that is comparable to that in the previous survey.[27] The most direct way of maintaining the level of difficulty would be to reproduce in 1998

---

[24]Smoothing at this stage avoids introducing large variations in the size measures. For PSUs that were originally selected with probability greater than .84, the selection probabilities were rounded to 1.00, those remaining greater than .60 were rounded to .70, those greater than .40 were rounded to .50, those greater than .20 were rounded to .30, and those smaller than .2 were rounded to .10. The smallest PSU selection probability is 0.006.

[25]For various reasons, the number of cases completed across the strata has varied in the past both above and below the target numbers.

[26]An unusual aspect of the SCF is the requirement that cases be reviewed before their release to the field. The motivation for this review is to minimize the likelihood that a survey participant might be identified by a malicious data user. Although strenuous efforts are made to modify the part of the data ultimately made available to the public in order to protect the identity of the survey participants, it might not be possible to alter the data of particularly well known people without overly compromising the analytic usefulness of the information. Thus, a number of cases are removed from each stratum at this final stage. These cases are treated as nonrespondents at the weighting stage.

[27]This goal reflects two faactors. First, the survey cost is a function of the expected level of effort, and consistency of the challenge was assumed in writing the survey contract. Second, if the pressures to obtain completed cases varies, there may be selection bias.

Figure 9: Determination of 1998 List Sample Size

$$S_{i,1998} = \{[T_{i,1998} * S_{i,1995} / T_{i,1995}] + [T_{i,1998} * S_{i,1995} / C_{i,1995}]\} /2$$

where i indicates the sample stratum,
$S_{i,j}$ is the sample size in year $j \in (1995, 1998)$,
$T_{i,j}$ is the minimum target number of completed cases in year j, and
$C_{i,1995}$ is the number of completed cases in 1995.

the 1995 ratio of the sample size in each stratum to the minimum number of completed cases. However, there are two complicating factors that affect the feasibility of meeting the minimum number of cases. First, there is a random component in the fraction of both ineligibles and deletions. Unfortunately, insufficient evidence is available to make any substantial adjustments to the sample size based on variations in ineligibles and deletions over time, though in all likelihood, the variance of these components is not very important. Second, at the end of the 1995 survey, it was clear that for some strata it would not have been possible to raise the number of completed cases without an unreasonable level of effort. One approach that would incorporate the information about the feasibility of obtaining cases would be to gross up the target number of cases by a factor equal to the inverse of the completion rate in 1995. However, this approach might also unduly accommodate what could have been random fluctuations in the level of difficulty within each stratum. The final approach selected (see figure 9) is to average the 1998 target grossed up by the ratio of the sample to the target in 1995 (which would be exactly the 1995 sample if the targets were identical), and the 1998 target grossed up by the inverse of the completion rate in 1995. The detailed components of this calculation are given for readers in the SCF project group by figure A3 in the confidential appendix.

The final sample selection is performed using a probability proportional to size (PPS) technique with systematic sampling (Kish [1965]). The size measure of each case at this stage is the original ITF weight adjusted for filing status and the PSU selection probability. Within each stratum except the highest ones, the size measures vary between a small number (reflecting near-certainty selection into the ITF) and figures in the thousands. In the case of the lower list sample strata, the small size measures reflect a decision by SOI to oversample certain types of cases to support the formal modeling activities at the Office of Tax Analysis in the Treasury. As one proceeds to the higher strata, observations with large size measures tend to be a progressively smaller fraction of the whole; these cases are often ones that are particularly inflated by the small probability of selection

associated with some non-urban PSUs. The range of the underlying (unadjusted) ITF weights is substantially smaller, though it is still substantial. If we applied PPS selection at this stage, many of the cases with large size measures would be selected with certainty. Although this approach is formally unbiased, it may introduce some inefficiencies connected with the geographic distribution of wealthy families discussed above. Moreover, because the ITF weights are based on criteria other than predicted wealth, some of the variation in those weights may be irrelevant for the SCF sample. To hedge against inefficiencies at this stage, cases within each stratum are divided into three groups of equal size according to the magnitude of their size measure, and within each group, every case is given the average size measure of the group.

Figure 10: Definitions of Implicit Substrata, 1998 SCF List Sample

Primary level of implicit stratification: age
- less than 35
- 35-49
- 50-64
- 65 and older

Secondary level of implicit stratification: financial income ($1987 dollars)
- 100 or less
- 101-1,000
- 1,001-5,000
- 5,001-10,000
- 10,001-25,000
- 25,001-50,000
- 50,001-100,000
- 100,001-500,000
- more than 500,000

Notes

1. 1987 dollars are chosen to be comparable to classifications used in the analysis of the 1989 SCF sample.
2. Age is that of the principal filer.
3. Financial income is the sum of taxable and nontaxable interest income and dividend income.

Prior to sampling, two levels of implicit substratification—age of the primary filer and total financial income—are imposed within each WINDEXM stratum. The definitions of the implicit strata are given in figure 10. The implicit substratification is imposed by sorting by age categories in descending order within WINDEXM strata, and by sorting the second level in alternating ascending and descending order ("head-to-head and tail-to-tail") within the age substrata. Below the level of the second implicit substratum, the order of observations is determined by the original sequence number of the returns.

To determine the sampling interval for the systematic sample of each stratum, the total size measure of the cases remaining in the sample for each stratum at this stage is divided by the final sample size for that stratum. If there are units with a measure of size exceeding the sampling interval, such cases would be selected with

certainty. In this case, an effect of the averaging of size measures described above is that no observation is selected with certainty at this stage.

Systematic PPS selection is preformed as follows: Beginning at a randomly selected starting point between one and the length of the sampling interval, the size measures of cases within a stratum are cumulated in the sort order implied by the implicit stratification until the cumulated size equals or exceeds the sampling interval, at which point the case whose size measure was added last is selected into the final sample. To select the next case, the sampling interval is subtracted from the cumulated size measures, and the cumulation process continues until the total again equals or exceeds the size measure. The procedure is repeated until the required number of cases has been selected.[28] The cases selected in the bottom six strata include a replicate structure to allow some flexibility during the field period: Every eleventh selection in these strata is marked as part of a "supplemental" sample to be released to the field in the event of serious difficulties in obtaining the minimum number of cases in each stratum. Across all strata and replicates, a total of 5,642 cases were selected into the list sample.

III. Final Comments and Further Research

The list sample is a critical part of the value added by the SCF over other surveys. The use of the ITF provides a means of obtaining a sufficiently large number of wealthy households for meaningful analysis in the survey. At least as importantly, this information also provides a way of addressing the serious differential nonresponse across and within wealth groups. Nevertheless, there is still room for improvement.

As is clear from the discussion in this paper, the classification of households by wealth in the list sample is imperfect. The evaluation of the 1995 list sample suggests that the misclassifications do not appear to be strongly driven by time variation in the model coefficients for the WINDEX models as might have appeared likely a priori. There are at least three other possible explanations. First, income of wealthy families may be quite variable over time. Some investigation of this issue

---

[28]As described in more detail in the confidential appendix, the method actually applied differs slightly in ways that minimize the likelihood of including a respondent that had been selected for the 1995.

was made using earlier ITFs (see Kennickell and McManus [1993]). If the ITF were a panel, the comparison would have been reasonably straightforward. In fact, the ITF sample is selected using a procedure related to Kefitz sampling (see Kefitz [1951]). In the application of this procedure to the ITF, there is an upward longitudinal bias in the observed income changes for cases included in successive files. Thus, a more comprehensive match of return data is needed to assess the effect of income variability on the predicted value of the wealth indices. There is a reasonable hope that research on this front can proceed given sufficient effort. A second possibility is that there are types of capital that do not generate observable income flows. In some cases, this problem may be just a more extreme example of the first possibility—e.g., the case of a tree farm that is cut only once every twenty years. Some sense of the magnitude of this problem might be obtained by modeling the residuals of the wealth index model in terms of observable survey data. Finally, there may be accidental—or intentional—misreporting in the ITF. However, such research is not the province of the SCF, and no information from the SCF can be made available to anyone else to conduct such research. Frequent efforts are made to address possible misreporting on the survey side, but the means of identifying such problems is very limited.

As strongly as the SCF oversamples wealthy families, the number of observations on some narrowly held items is still inadequate for some purposes. An increase in the sample size would do much to reduce such problems, but three factors stand in the way. First, the SCF is a very expensive survey already. Second, the quality of the surveys so far has depended in important ways on careful attention to the data at all stages of processing. Even with a larger staff, it is inevitable that such attention would be diluted with larger sample sizes. It is possible that further advances in automation will weaken this constraint. Finally, if the goal is to interview wealthy families, there is the problem related to the number of such families in the population. Because types of wealth are very concentrated, an increase in sample size would inevitably cause respondents to be selected into multiple samples over time. The SCF is a long and difficult interview, and respondents are performing a patriotic act when they complete it once. It is too much to expect that there would be no selection bias if a significant set of families remained in the sample from survey to survey.

Nonresponse remains a problem in the survey. There has been much research in the SCF on the characteristics of nonrespondents and on adjustments for nonresponse at the weighting stage

(Kennickell and Woodburn [1997], Kennickell and McManus [1993], and Kennickell [1997]). However, as Arnold Zellner has noted forcefully, there is no substitute for actual interviews. Although it is unlikely that we will find a dramatic breakthrough in reducing nonresponse, it is very important to continue to develop incremental techniques that will maintain and possibly improve current efforts on the margin. Interviewers are critical in this effort, and no energy spent in convincing interviewers of the importance of the survey is wasted. If that group does not feel the importance of the work, they will not be able to communicate that information to respondents. Better work on explanatory materials for respondents may help. We routinely offer respondents letters from Chairman Greenspan and the project director at NORC explaining the survey, and a range of articles and brochures. For technologically inclined respondents, the SCF Internet site

http://www.bog.frb.fed.us/pubs/oss/oss2/scfindex.html

offers descriptive and methodological information along with all of the public data from past surveys. More creativity is needed to reach people with compelling reasons why everyone selected should participate in the survey.

Bibliography

Avery, Robert B., Gregory E. Elliehausen, and Arthur B. Kennickell [1988] “Measuring Wealth with Survey Data: An Evaluation of the 1983 Survey of Consumer Finances,” *Review of Income and Wealth* (December), pp. 339-369.

Heeringa, Steven G. and Richard T. Curtin [1986] “Household Income and Wealth: Sample Design and Estimation for the 1983 Survey of Consumer Finances,” mimeo, Survey Research Center, University of Michigan.

Frankel, Martin and Arthur B. Kennickell [1995] "Toward the Development of an Optimal Stratification Paradigm for the Survey of Consumer Finances," paper presented at the 1995 Annual Meetings of the American Statistical Association, Orlando, FL.

Greenwood, Daphne [1983] “An Estimation of U.S. Family Wealth and its Distribution from Micro-Data, 1973,” *Review of Income and Wealth*, March 1983, pp. 23-44.

Heeringa, Steven G., Judith H. Conner and R. Louise Woodburn [1994] “The 1989 Surveys of Consumer Finances Sample Design and Weighting Documentation,” working paper, Survey Research Center, University of Michigan, Ann Arbor, MI.

Keyfitz, N [1951] “Sampling with Probabilities Proportional to Size: Adjustment for Changes in the Probabilities,” *Journal of the American Statistical Association*, v. 46, pp. 105-109.

Kennickell, Arthur B. [1997] “Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances,” *Proceedings of the Section on Survey Research Methods*, 1997 Annual Meetings of the American Statistical Association, Anaheim, CA.

_____, Martha Starr-McCluer, and Annika Sundén [1997] “Family Finances in the U.S.: Recent Evidence from the Survey of Consumer Finances,” *Federal Reserve Bulletin*, vol. 83 (January), pp. 1-24.

_____ and R. Louise Woodburn [1992] "Estimation of Household Net Worth Using Model-Based and Design-Based Weights: Evidence from the 1989 Survey of Consumer Finances," mimeo, Board of Governors of the Federal Reserve System.

_____and _____ [1997] “Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth,” working paper, Board of Governors of the Federal Reserve System, Washington, DC.

_____ and Douglas A. McManus [1993] "Sampling for Household Financial Characteristics Using Frame Information on Past Income," *Proceedings of the Section on Survey Research Methods*, 1993 Annual Meetings of the American Statistical Association, San Francisco, CA.

_____, _____, and R. Louise Woodburn [1996] “Weighting Design for the 1992 Survey of Consumer Finances,” working paper, Board of Governors of the Federal Reserve System.

Kish, Leslie [1965] *Survey Sampling*, John Wiley, New York.

Statistics of Income Division, Internal Revenue Service [1992] “Individual Income Tax Returns, 1990.”

_____ [1996] “SOI Computer Selection Specifications for Tax Year 1996,” mimeo, Statistics of Income Division.

Tourangeau, Roger, Robert A. Johnson, Jiahe Qian, Hee-Choon Shin, and Martin R. Frankel [1993] “Selection of NORC’s 1990 National Sample,” working paper, National Opinion Research Center at the University of Chicago, Chicago, IL.