

USING SUPPLEMENTARY INFORMATION TO INVESTIGATE NONRESPONSE BIAS

Louise Woodburn, Internal Revenue Service and Steve Heeringa, University of Michigan

This paper reports on a study by the Statistics of Income (SOI) Division of the Internal Revenue Service to investigate a possible bias due to the high rate of nonresponse on the high income supplementary sample for the 1983 Survey of Consumer Finances.

Organizationally, this paper is divided into five sections. First, there is a brief description of the background of the study which has been conducted on a special data set derived from administrative records for those who volunteered for the Survey and a sample of those who did not volunteer. The second section discusses that file. A variety of parametric and non-parametric tests were applied within collapsed stratum to compare distributional differences of selected variables between the two groups. These are described in the third section. Next, there is a brief presentation of Quantile-Quantile charts which were used to compare the volunteers and nonvolunteers. The results of the investigation to date is summarized. Conclusions and a look at future plans are discussed in the final section.

BACKGROUND

In 1983 the Federal Reserve Board sponsored a Survey of Consumer Finances to investigate the financial characteristics of households (hereafter referred to as the Survey). The Survey was conducted by the Survey Research Center (SRC) at the University of Michigan [1]. The sample for the Survey was selected from a dual frame composed of a nationally representative household sample and a "high income" sample. The "high income" component is based on a probability sample of tax filing units whose returns had been previously selected in the 1980 SOI Individual Statistics' program.

The 1980 Statistics of Income program for individuals was based on a stratified random sample of approximately 150,000 tax records selected and compiled from the population of almost 100,000,000 individual returns (Forms 1040 and 1040A) which were filed for the 1980 tax year [2]. Almost from the inception of the income tax, SOI information has been obtained for each tax year and is used by the Internal Revenue Service for research and statistical reporting purposes [3]. Statistical and research uses of the SOI data are closely regulated to guarantee that individuals (and other entities) will remain protected against any and all disclosure of their financial and tax data. For discussions of how SOI prevents the disclosure of taxpayer information see [4-6].

The 1983 Survey research protocol for the high income sample specified that a signed voluntary consent form be obtained before SRC could contact the tax filer to conduct the Study interview. To that end, a single invitation to participate in the Study was mailed to each sampled tax filer. Mailing of the consent package--including a letter of introduction and

(postcard) consent form--was coordinated by the U.S. Comptroller of the Currency. Sample tax filers who received the package were asked to indicate their willingness to participate in the survey by signing and returning the consent form. At the conclusion of this process, names and addresses of consenting high income tax filers were transmitted to SRC.

The direct outcome of the voluntary consent procedure was a very low sample retention rate; only 9 percent of the original sample of tax filer names and addresses resulted in signed permission for contact by the SRC interviewer. It is important to note that the solicitation took place in the Spring of 1983, but the consent packages were mailed to tax return addresses as of an earlier date. This raises the question of exactly what proportion of the consent packet mailings actually reached the intended sample tax filer; nevertheless, even if a generous allowance is made for undelivered consent packets, high income sample respondents' true cooperation rate was disappointingly low. Such a low cooperation rate leads immediately to the question of nonresponse bias in the observed sample data. Indeed, one should assume a bias exists unless proven otherwise. Through analysis of differences in auxiliary variables (which in the ideal case are highly correlated with the survey data) for the respondent and nonrespondent groups, the potential significance of nonresponse bias for survey variables can be evaluated. Auxiliary information for both volunteers and nonvolunteers is available from the administrative SOI records which were used in the design and selection of the high income sample. The groups in this Study will be referred to as volunteers and nonvolunteers to distinguish those who sent in the consent form, i.e., the volunteers, from those who actually responded to the survey [7].

DATA ANALYSIS FILE

A special analysis file to support the investigation of nonresponse properties of the 1983 SCF high income sample was prepared. The special file includes a single record for most members of the volunteer sample of consenting tax filers and a random subsample of the high income sample tax filers who did not respond to the consent solicitation. The random subsample of nonvolunteers was chosen to be of the same size as the volunteer groups within sample strata. In addition to the sample control information (such as the sample stratum code) on the special analysis file, selected items from the 1982 individual tax returns were merged onto each individual record. The 1982 tax data were chosen to correspond to the reference year for the 1983 survey. Included in the set of 1982 tax return data were: total reported wages, total interest earned, and adjusted gross income (AGI).

A summary of the financial variables analyzed

from the data file can be found in Figure 1. The information for each variable is repeated 6 times, each repetition corresponding to a unique collapsed stratum of the original sample of tax filers. The response rate in the different strata varies from 6.5 to 10.5 percent. Strata I-III are nonbusiness strata, IV-VI are business strata. Additionally, the pairs of strata, I & IV, II & V, and III & VI, represent the same size classes of adjusted gross income which increase from I-III. For each stratum, sample sizes, mean values, and sample standard deviations are listed, both for the volunteers and the nonvolunteers.

FIGURE 1. -- 1983 High Income Sample
Descriptive Statistical Comparisons
Volunteers and Nonvolunteers

Stratum and Variable	Sample Size	NONVOLUNTEERS		VOLUNTEERS	
		Sample Mean	Sample Standard Deviation	Sample Mean	Sample Standard Deviation
AGI	143	134,383	134,403	161,670	140,365
I: INT	143	11,836	24,394	14,123	37,258
WAGES	143	95,398	76,363	114,279	112,401
AGI	72	301,607	246,952	256,275	167,231
II: INT	72	38,259	104,337	24,849	45,648
WAGES	72	142,920	157,721	123,259	120,680
AGI	101	593,036	1,276,928	534,849	777,197
III: INT	101	171,657	546,183	91,093	144,840
WAGES	101	228,157	302,060	255,936	456,000
AGI	52	147,596	174,758	119,211	57,711
IV: INT	52	21,513	31,835	8,276	14,581
WAGES	52	69,748	79,653	83,400	64,489
AGI	23	8,789	588,815	244,307	143,504
V: INT	23	36,735	68,129	20,842	32,954
WAGES	23	85,636	107,222	119,086	153,996
AGI	44	591,735	2,126,899	600,318	892,628
VI: INT	44	127,295	209,630	96,587	149,906
WAGES	44	345,912	1,039,815	144,235	224,001

PRELIMINARY ANALYSIS OF DIFFERENCES
IN CENTRAL TENDENCY

Initially, it was desired to use the special data set to test the significance of the differences observed in the central tendencies in the high income sample data set. The first step taken to analyze these differences was to calculate, for three 1982 Form 1040 variables, both the two-sample t-test and the nonparametric Wilcoxon two-sample rank sum test statistics. Due to the small sample sizes, these tests were applied with the understanding that methods more robust to the conditions present in the data would follow. Values for the applicable t-test statistics are included in Figure 2. Also listed in Figure 2 are the significance probabilities for these test statistics under the null hypothesis that there is no difference in volunteer and nonvolunteer distributions. Results significant at the .05 nominal level are noted with a triple asterisk '***'. The use of the nonparametric Wilcoxon test is preferred here since it has more power than the t-test when the underlying distribution is nonnormal or when there are unusually large or

FIGURE 2. -- 1983 High Income Sample
Analytical Statistical Comparisons
Volunteers and Nonvolunteers

Stratum and Variable	t - Statistic	Calculated Degrees of Freedom**	t-test Prob.	Wilcoxon Prob.
AGI	-1.68*	284	.09	.02***
I: INT	0.61	245	.54	.56
WAGES	-1.66	250	.10	.19
AGI	1.29	125	.20	.90
II: INT	.99	97	.32	.20
WAGES	.84*	142	.40	.96
AGI	.39	165	.70	.41
III: INT	1.43	114	.15	.86
WAGES	-0.51	174	.61	.92
AGI	1.11	62	.27	.86
IV: INT	2.73	72	.01***	.00***
WAGES	-.10*	102	.34	.22
AGI	-1.86	25	.07	.00***
V: INT	1.01	32	.32	.55
WAGES	-.09*	44	.40	.88
AGI	-0.02	58	.98	.51
VI: INT	0.79*	86	.43	.90
WAGES	1.26	47	.22	.82

NOTE: * denotes the t-statistics which were calculated using a pooled variance. ** is to clarify that the degrees of freedom were calculated using Satterthwaite's approximation for those t-statistics not computed using a pooled variance. *** has been shown for comparisons that are significantly different at the $\alpha=0.05$ nominal level.

small observations, as in the data being considered.

As expected given the small sample sizes and the large sample variances, very few of the differences proved significant using these tests; however, preliminary normal theory F-tests for equality of variances suggest that strongly significant differences exist between the volunteers and nonvolunteers in most sample strata. There was a need to gain an understanding of the differences, not only in the means and variances, but also in the distributional shape corresponding to the volunteers and nonvolunteers. In order to confirm (refute) these preliminary results, an investigation using tests more robust to the non-normality which exists in the variables being compared was begun. To start this in-depth investigation, Q-Q (Quantile-Quantile) plots [8] were used to compare the two groups. In the remainder of this paper, the Q-Q plot analysis to date will be summarized and the additional analysis planned will be documented.

Q-Q PLOT ANALYSIS

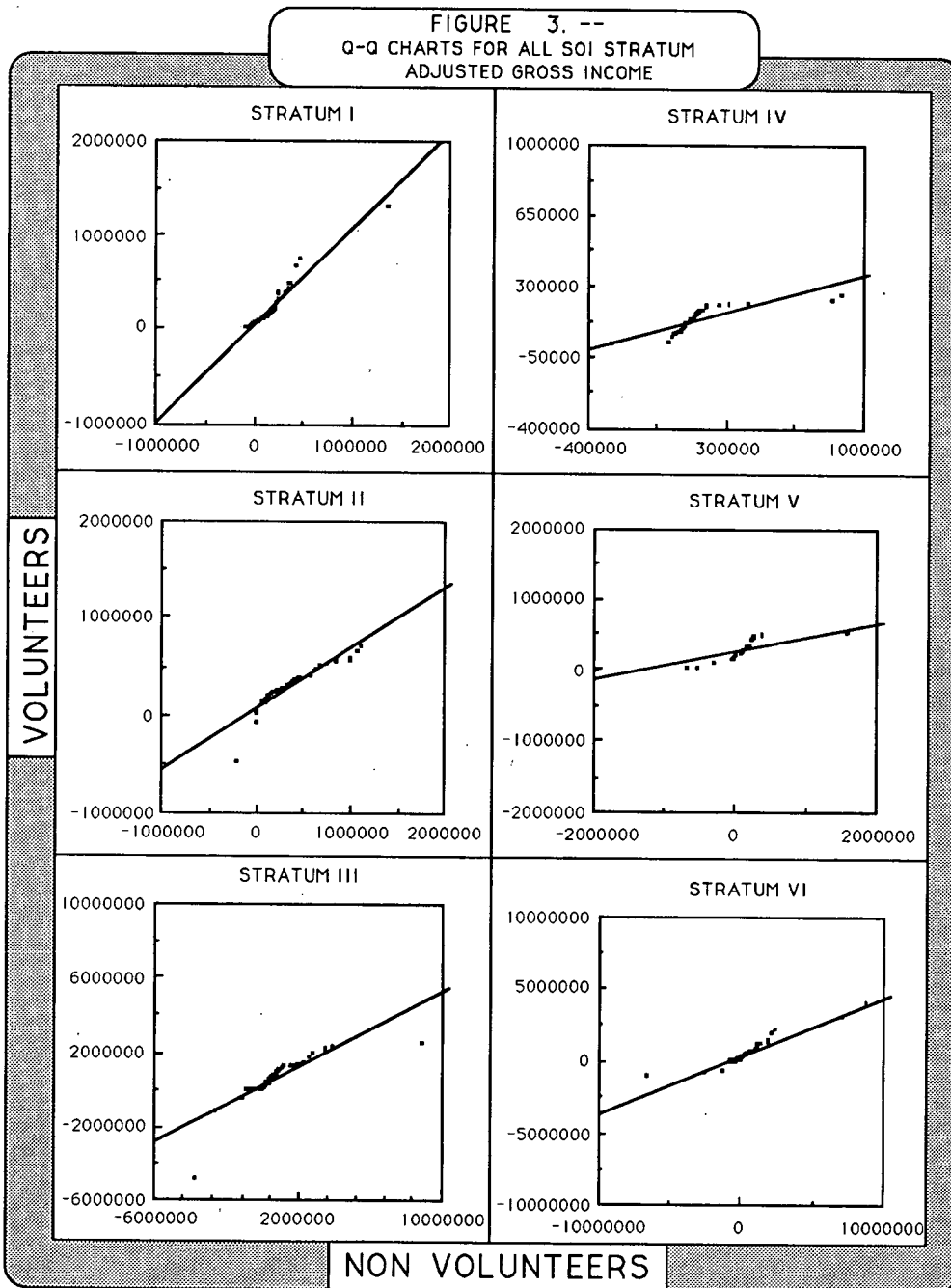
In a quantile-quantile plot, the distributions of two groups are compared by plotting quantiles, or more generally, selected percentiles, from one group against the other. If the underlying distributions are the same, then the plot will be a straight line with the following equation:

$$F^{-1}(y) = (\mu_y - \frac{\sigma_y}{\sigma_x} \mu_x) + \frac{\sigma_y}{\sigma_x} F^{-1}(x)$$

The slope is the ratio of the scales of the underlying distributions and the y-intercept reflects a scaled difference in means. If the two groups are from the same distribution, then this line will have unit slope and pass through the origin. If the variances of the 2 groups are the same, the y-intercept of the fitted line will represent the difference between the means. Since by design the samples in the strata we are considering are of the same size and that size is not too large, we can simply plot the ordered data for each group. We can now visualize the differences between the variances of the volunteers and nonvolunteers to

determine whether the distributions of the two are similar.

Q-Q charts for all strata comparing AGI are shown in Figure 3. The volunteers are graphed versus the nonvolunteers. Note that the scales are necessarily different between strata to accommodate the various range of values. For each stratum, the x and y axes are of the same scale. Thus, a 45-degree line represents the line of unit slope and y-intercept zero; that is, the line that would be formed if the distributions of the volunteers and nonvolunteers were identical. The line shown on the graphs is the least squares regression



line. In most instances, these lines have y-intercept near zero, supporting the preliminary analysis of central tendency which showed no significant difference in most cases. To analyze the differences in the variances of the two groups, the slopes of the least squares line are considered. A line with slope less than 1 reflects a larger variance for the nonvolunteers. As shown in Figure 3, for AGI, more variance is shown by the nonvolunteers than the volunteers in the business strata, IV-VI. Only the lowest AGI/nonbusiness category, strata I, reveals nearly equal variance for the two groups. Although there are extreme cases in several strata, these are not isolated cases. A

definite trend of longer tails in the distribution of the nonvolunteers is reflected.

The corresponding Q-Q charts for interest income are shown in Figure 4. A similar pattern is seen in the slope of the lines for the different strata; however, in this case the inherent lower bound of zero prevents any distortion in the lower tails of the distribution. Again, for the business strata, a larger variance for the nonvolunteers is observed. The least squares lines for strata II and III also dictate a larger variance for the nonvolunteers; however, in these instances, isolated extreme cases greatly influence the fit of the line. Indeed, in stratum II, an analysis

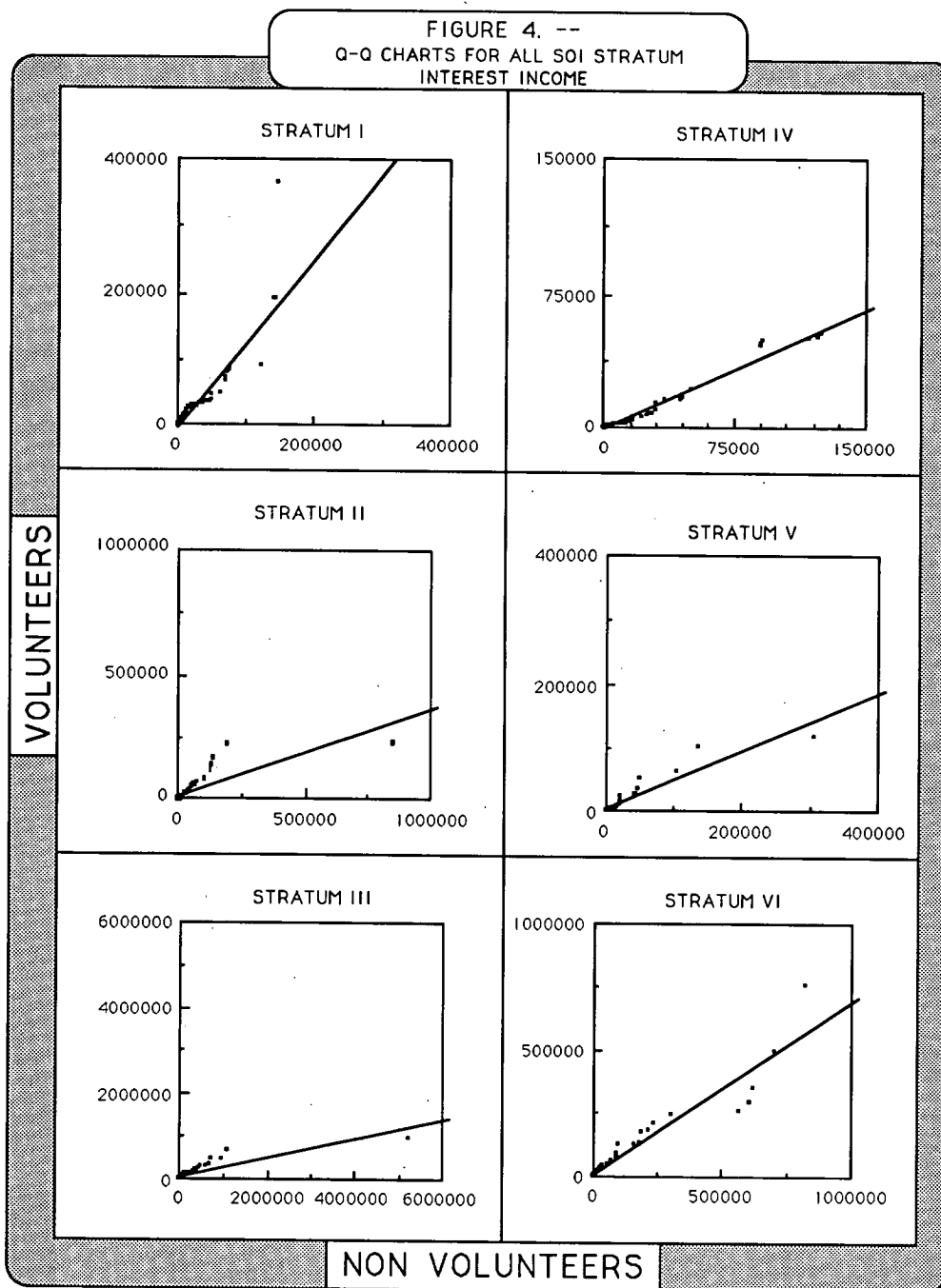
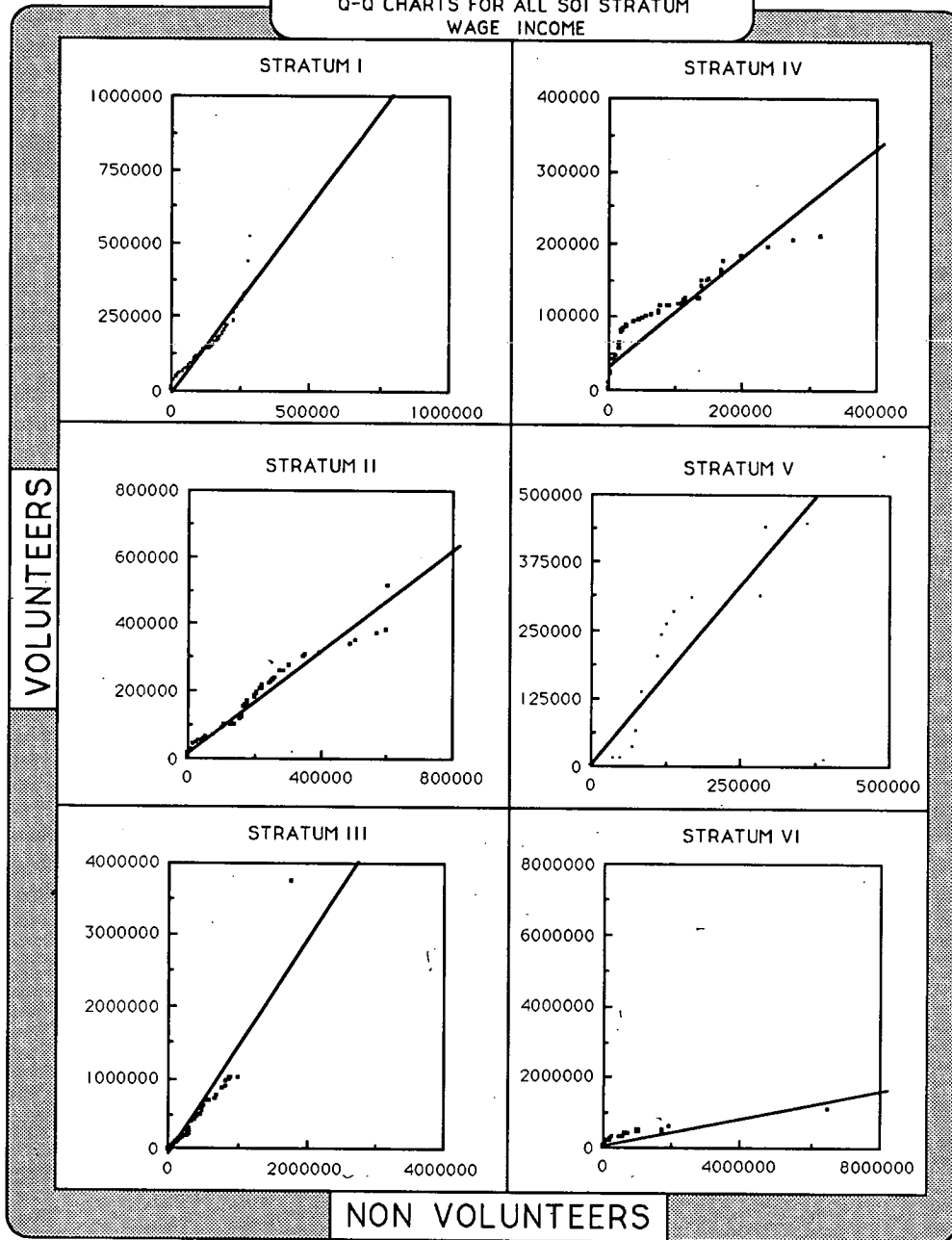


FIGURE 5. --
Q-Q CHARTS FOR ALL SOI STRATUM
WAGE INCOME



without the most extreme case would show nearly equal variances. Similar Q-Q charts for the Wage variable were considered. For wages, the difference in variances is not as extreme as for AGI or interest, except for the top business stratum. (Figure 5)

CONCLUSIONS AND FUTURE ANALYSIS

The conclusion based on the analyses to date of the available Form 1040 variables is that the very high level of nonresponse to the original 1983 SCF consent solicitation had only a modest impact on the representativeness at the mean of the high income sample respondents. Among the

three test variables, differences in mean computed AGI (from the tax report) are most likely to be significant. Whether or not the impacts on true survey variables are of similar nature and magnitude cannot be stated with certainty. However, barring the extreme condition where household reports of income and assets from the Survey are poorly correlated with the amounts on their tax return, it is fair to assume that nonresponse bias does not threaten the basic utility of analysis at the mean of the 1983 SCF high income sample data. This is not to say that caution is not needed in the analysis of the high income sample data set. The higher variance observed in the

nonvolunteer group in the majority of the strata should be reflected in the total survey error calculated for the high income sample data.

The investigation, to date, into the nonresponse bias of the high income sample of the Survey is just the beginning of an in-depth analysis to determine what sort of differences exist between the volunteers and nonvolunteers, and how adjustments can be made for such found differences. It is planned to follow-up the Q-Q analysis by determining if the slope of the least squares line is significantly different from 1 and to more completely analyze the differences in distributional shapes. Nonparametric tests are preferred here to accommodate the non-normality of the sample data.

The concern of extreme cases raised by the effects of the isolated cases observed for interest income must be addressed. One possible explanation for such extremities is imposed by the structure of the data and the SOI sample design which implies that the nonvolunteers have more aggressive financial habits and thus more volatile financial realizations.

Another consideration in the nonresponse bias investigation is the longitudinal characteristics of the volunteers and nonvolunteers. The 1983 Survey was followed up in 1986, and again in the fall of 1989.

As the detailed analysis of nonresponse in the 1983 SCF high income sample progresses, the goal will be to increase the accuracy of both cross-sectional and longitudinal analyses of these high income sample data through development of an improved nonresponse adjustment weighting strategy.

ACKNOWLEDGMENTS

The authors would like to thank Beth Kilss and Wendy Alvey for their help in preparing this paper for publication and for their invaluable assistance on the presentation of this paper at the ASA meetings. Additionally, the authors thank Joan Kahn for her input. Finally, many thanks to Lisa Bright for helping with the preparation of the text.

NOTES AND REFERENCES

[1] For a description of the design and estimation procedures of the 1983 Survey, see Heeringa, Steven G. and Curtin, Richard

T., "Household Income and Wealth: Sample Design and Estimation for the 1988 Survey of Consumer Finances," Statistics of Income and Related Administrative Record Research: 1986-1987, Internal Revenue Service, November 1987, pp. 3-24.

An evaluation of the Survey data is provided by Avery et al., "Measuring Wealth with Survey Data: An Evaluation of the 1983 Survey of Consumer Finances," Review of Income and Wealth, 34, pp. 339-369.

[2] For a description of sampling procedures for the 1980 SOI program, see Individual Income Tax Returns, 1980, Internal Revenue Service, September 1982, pp. 11-14.

[3] Details on the annual SOI program for individuals can be found in: Coleman, Mike, "Statistics of Income Studies of Individual Income and Taxes," Statistics of Income Bulletin, Vol. 7, No. 3, pp. 21-37.

[4] Strudler, Michael, Oh, H. Lock and Scheuren, Fritz, "Protection of Taxpayer Confidentiality with Respect to the Tax Model," Statistics of Income and Related Administrative Record Research: 1986-1987, Internal Revenue Service, November 1987, pp. 279-286.

[5] Spruill, Nancy, "The Confidentiality and Analytic Usefulness of Masked Business Microdata," 1983 Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 602-607.

[6] Wilson, Oliver and Smith, William J., Jr., "Access to Tax Records for Statistical Purposes," 1983 Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 595-601.

[7] The IRS does not have access to identifiable survey data. Indeed no linkage of survey and administrative data was possible given the structure of the study, including assurances given respondents. Thus we cannot compare the income measures from the survey to similar measures from the administrative file on an individual basis.

[8] For a description of Q-Q analysis, see Hoaglin, David C. et al., Exploring Data Tables, Trends, and Shape, John Wiley and Sons, Inc., 1985, pp. 432-442.