Multiple Imputation In The 1992 Survey Of Consumer Finances

Catherine Phillips Montalto,[1] The Ohio State University
Jaimie Sung,[2] The Ohio State University

The 1992 Survey of Consumer Finances consists of five complete data sets because missing data are multiply imputed. The incidence of missing data in the 1992 SCF is addressed and illustrates the difficulty of obtaining financial information from individuals. The value of using all five data sets and the risk of using only a single data set in empirical research are explained. Estimates derived separately from each data set are compared to results using all five data sets to illustrate the extra variability in the data due to imputation. Researchers are encouraged to use information from all five data sets in order to make valid inferences.

KEY WORDS: *inference, missing data, multiple imputation, repeated-imputation inference (RII), Survey of Consumer Finances*

The 1992 Survey of Consumer Finances (SCF) is a rich source of information on assets and liabilities of U.S. households. The wealth information in the SCF is exceptionally good because the survey uses a sample designed specifically to support wealth estimation. An interesting characteristic of the 1992 SCF is that the public use data tape contains five complete data sets, instead of the more common single data set. Persons new to the SCF, who are either interested in practical implications from SCF research or in using the SCF in applied research, may wonder why five complete data sets are provided in the 1992 SCF and what techniques need to be used to analyze data appropriately in the presence of five data sets. Previous research using the SCF has been inconsistent in the treatment of the multiple data sets and in reporting this information.[a]

The 1992 SCF consists of five complete data sets as a result of the procedure used to handle missing data. Missing or incomplete information is common in all survey data. Data can be missing because respondents are unable or unwilling to provide information, or due to errors in data recording and processing which render data unuseable. For the final release of the 1992 SCF public use tapes, missing and incomplete data were imputed, using the multiple imputation technique developed for the SCF (Kennickell, 1991). The multiple imputation technique produces five complete data sets, referred to as "implicates" (Board of Governors of the Federal Reserve System, 1996).

This paper provides information on the incidence of missing data in the 1992 SCF, summarizes alternative methods of dealing with missing data, and describes strengths of multiple imputation techniques for imputing missing data. The value of incorporating information from all implicates of the 1992 SCF in empirical analysis is addressed. The "repeated-imputation inference" (RII) approach, a technique which uses information from all five implicates, is described and used to present point estimates, variance estimates and test statistics for selected variables in the 1992 SCF.

Survey of Consumer Finances

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. Since 1983 the survey has been conducted every three years. The SCF is intended to provide information on the financial characteristics of U.S. households. Detailed information is collected on household assets and liabilities. Information is also

[1] *Catherine Phillips Montalto, Assistant Professor, Consumer and Textile Sciences Department, The Ohio State University, Columbus, OH 43210-1295. Phone: (614) 292-4571. Fax: (614) 292-7536. E-mail: montalto.2@osu.edu.*

[2] *Jaimie Sung, Graduate Student, Consumer and Textile Sciences Department, The Ohio State University, Columbus, OH 43210-1295. Phone: (614) 292-4590. Fax: (614) 292-7536. E-mail: sung.13@osu.edu.*

©1996, Association for Financial Counseling and Planning Education

collected on current and past employment, pension rights, inheritances, income, marital history, household demographic characteristics, attitudes, and numerous other items. A more complete description of the data is given in Kennickell and Starr-McCluer (1994).

The 1992 SCF employs a dual-frame sample incorporating both an area-probability (AP) sample and a special list sample developed from a sample of tax records. The area-probability sample provides good information on financial variables which are broadly-distributed in the population, such as automobile ownership, home mortgages, and credit card debt. The special list sample oversamples households which are more likely to be wealthy, and provides good information on financial variables which are highly correlated with wealth, such as ownership of stocks, real estate investments, and business assets. Of the 3,906 completed cases in the 1992 survey, 2,456 households were part of the area-probability sample, and the remaining 1,450 were part of the list sample.

Reporting Rates in the 1992 SCF

Reporting rates for selected financial items for the weighted full sample of the 1992 SCF are provided in Table 1. (Kennickell, 1996, summarizes response rates for the 1983, 1986, 1989, 1992 and 1995 Surveys of Consumer Finances for the unweighted full sample, the unweighted area-probability sample, and the weighted full sample). Focusing on asset information, households were much more likely to have checking accounts (83%) and to own their homes (59%) than to have stocks (17%) or business assets (13%). Households which reported that they did have a financial item were then asked to report the value of the item. In most cases, respondents provided a dollar value. Respondents who were reluctant to provide dollar values directly were offered a card containing dollar ranges and asked to select the range encompassing the dollar value. In some cases respondents refused to answer or indicated they didn't know. The unwillingness or inability of respondents to provide information, as well as errors in data recording and processing which render data unusable, result in missing data.

The incidence of missing data in the SCF varies widely, but is generally within a range typical of other economic surveys. In the 1992 SCF, the incidence of missing data ranges from very low for data on monthly mortgage payments, rent payments, and credit card balances, to much higher for data on business assets and stocks. Less than 5% of households in the 1992 SCF that had mortgage payments, rent payments, or balances on credit cards were unwilling or unable to provide information on the dollar amount of the item. In contrast, over one-quarter of households with business assets and nearly one-fifth of households with stocks did not know the value of the assets, resulting in much higher rates of missing data.

The response rates in the 1992 SCF illustrate the difficulty which can be encountered in obtaining information, particularly financial information, from individuals. In general, individuals are more willing and able to provide information on financial variables which are more easily quantified and considered less private (e.g. monthly housing payments in contrast to stocks).

Handling Missing Data

The problem created by missing survey data is that data intended by the survey design to be observed are in fact missing. Missing information raises issues of both efficiency and bias for users of the data. Nonresponse to selected survey components means less efficient estimates due to the reduced size of the useable data base. In addition, the useable data base is vulnerable to possible bias since nonrespondents are often systematically different from respondents.

Several standard procedures to address missing data are used in empirical research (Little & Rubin, 1987). A brief description of the most common of these procedures and the limitations follows. Observations with missing data can be eliminated from the sample; this results in less efficient estimates due to the reduced sample size, and assumes no nonresponse bias. Observations with missing data can be matched to observations with complete data on a set of background characteristics, and then the observations with complete information can be weighted to compensate for observations with missing data; this method assumes no nonresponse bias beyond that explained by the measured background characteristics. Missing values can be replaced with the sample mean value of the variable. This too assumes no nonresponse bias, can distort correlations among variables, and understates the

©1996, Association for Financial Counseling and Planning Education

variance because all missing values are replaced with the sample mean. A "hot-deck" approach can be used whereby each observation with missing data is matched on a set of defining characteristics to a similar complete observation within the sample, and the variable value from the complete observation is substituted in place of the missing data. In addition to assuming no nonresponse bias after accounting for the variables used as defining characteristics, this approach can result in underestimation of the true variability in the sample. Multivariate methods can be used to fill-in missing data by single imputation. Single imputation commonly uses regression equations to generate estimates of the values of missing data utilizing information available in the data set.

With respect to efficiency, methods of handling missing data which produce data sets with no missing data (i.e. complete data sets) increase the efficiency of estimation by allowing the researcher to use all available data. With respect to bias, methods which incorporate multivariate techniques can incorporate information to adjust for observed differences between respondents and nonrespondents in an effort to reduce nonresponse bias (Little, 1983). However, all methods which replace each missing data point with one imputed value share a common problem -- they ignore the extra variability due to the unknown missing values. Empirical analyses based on data sets with single imputed values systematically underestimate variability because the imputed values are treated as if they were known with certainty.

Multiple Imputation

Multiple imputation is a procedure for handling missing data which provides information that can be used to estimate the extra variability due to the unknown missing values. Multiple imputation uses stochastic multivariate methods to replace each missing value with two or more values generated to simulate the sampling distribution of the missing values.[b] The goal of the imputation process is to obtain the best possible estimates of the true but unobserved values of data which are missing. As more imputed values are generated, the approximation to the true sampling distribution improves. In analysis, the multiply imputed values are averaged to produce the best estimate of what the results would have been if the missing data had been observed, and the variance estimates are corrected for the uncertainty due to missing values. Rubin (1987) provides an extensive discussion of both the theory and practice of multiple imputation. Since the 1989 SCF, multiple imputation has been used to replace each missing value with five values. The end result is five complete data sets, referred to as "implicates". For a description of the imputation procedure used since the 1989 SCF see Kennickell (1991). Kennickell and McManus (1994) discuss relevant issues associated with the multiple imputation of the 1983-89 SCF Panel.

Multiple imputation, like several of the previously mentioned techniques, offers the advantage of increased efficiency in estimation and the ability to incorporate information in an effort to reduce nonresponse bias. The ability to address nonresponse bias is actually enhanced because multiple imputation can incorporate information about nonresponse by modeling either known reasons for nonresponse or uncertainty about the reasons for nonresponse.

The distinct advantage of multiple imputation is that it provides information which can be used to estimate the uncertainty in estimates due to missing values. As a result, multiply imputed data sets provide a basis for more valid inference and tests of significance. As an illustration, consider a representative sample of households and no missing data on family income. The best estimate of mean family income for the population of households is mean family income for the sample. Similarly, the best estimate of the variance of mean family income is the variance of the sample mean. However, in the presence of missing data and the use of multiple imputation techniques to fill in the missing data, the best estimates of mean family income and the variance of mean family income need to average over all the imputed values. In addition, the best estimate of variance needs to incorporate information on the amount of uncertainty in the estimate due to missing data. When the uncertainty due to missing values is ignored, as when missing data are filled in by single imputation or when the additional information available in multiple implicates of a data set is not used, the variance estimate will underestimate the true variance. All inference based on this biased variance estimate risk misrepresenting the precision of the estimates and the significance of relationships. In general, as the proportion of values which are missing and therefore imputed increases, the downward bias in the uncorrected variance estimate increases.

Table 1.
Reporting Rates for Various Items; 1992 SCF, Full Sample, Weighted

Item	Percent of Households That Have the Item		Of Households That Have the Item, the Percent Providing Each Type of Response When Asked to Report the Value of the Item			
					Missing Because:	
	Yes	Unknown	Dollar Value	Range Card Response	Respondent Didn't Know	Other Reasons
Principal residence	58.9	0.0	93.7	0.7	4.8	0.8
Borrowed on mortgage	38.2	0.2	91.1	1.0	4.5	3.4
Owe on mortgage	38.2	0.2	85.8	0.0	11.1	3.1
Mortgage payment	37.9	0.2	95.7	0.5	1.5	2.3
Rent payment	31.3	0.0	96.9	0.3	0.5	2.2
Other real estate	17.9	0.2	89.7	0.2	8.4	1.7
Business assets	13.2	0.0	70.0	0.9	27.2	1.9
Car loan payment	24.6	0.2	90.9	0.4	3.9	4.8
Credit card balance	62.2	0.1	95.7	0.8	2.0	1.6
Checking account	83.2	0.3	87.2	1.7	3.7	7.4
Money market account	11.1	0.4	84.8	0.9	4.5	9.8
Savings account	43.7	0.5	84.3	1.6	4.1	10.1
Certificates of deposit	16.6	0.5	73.1	1.7	7.9	17.3
IRA/Keogh account	23.0	0.4	79.5	2.4	8.1	10.0
Savings bonds	22.1	0.5	84.8	1.7	9.8	3.6
Municipal bonds	2.1	0.6	79.7	1.7	12.4	6.3
Tax-free mutual funds	2.7	0.9	67.4	1.2	15.3	16.1
Stock	16.8	0.5	73.8	1.9	17.9	6.5
Wage income	71.0	2.3	85.6	3.6	4.1	6.7
Business income	11.1	2.2	78.3	1.8	7.9	10.0
Non-tax. interest income	5.1	2.5	72.5	2.0	14.1	11.4
Taxable interest income	38.0	2.4	73.0	2.9	12.3	11.8
Dividend income	16.4	2.7	70.9	1.6	12.7	14.8
Capital gains and losses	7.7	2.7	73.1	1.7	12.3	11.9
Rent and royalties	8.9	2.7	83.0	1.4	5.2	10.4
Unemployment comp.	6.0	2.7	87.3	0.7	3.8	8.2
Transfers	3.6	2.7	87.0	1.1	3.1	8.8

SOURCE: Kennickell (1996). Table 4.

Repeated-Imputation Inference

The relevant question for the empirical researcher using any SCF since the 1989 survey is how to use the information from all five implicates to generate the best point estimates and estimates of variance for parameters of interest. In general, this is achieved by simply combining results across the five complete data sets (i.e. implicates). Combining the results often requires only the calculation of the means and variances of the results from the five separate implicates. This method of inference, based on multiple complete data sets, is referred to as "repeated-imputation inference" (RII) (Rubin, 1987). RII is based on Bayesian theory, and is applicable to linear and nonlinear models, and to models estimated by both least squares and maximum likelihood. A brief description of RII follows; more technical information is provided in Appendix A.

Point Estimate

The best point estimate of a parameter of interest is the average of the point estimates derived independently from each of the five implicates.

Variance Estimate

The best estimate of variance is the average of the variance estimates derived independently from each of the five implicates ("within" imputation variance), plus an estimate of the "between" imputation variance, with an adjustment factor for using a finite number of imputations. The "between" imputation variance is the sum of the squared deviations of the point estimates in each implicate from the overall average point estimate (as described above), divided by the number of implicates minus one.

Significance Tests
The point estimates and variance estimates derived by RII techniques can be used to construct confidence intervals and conduct significance tests. The adjustments to the test statistics due to the multiple implicates are described in Appendix A.

Example: Using Repeated-Imputation Inference (RII) Techniques

To illustrate use of the five implicates in estimating the amount of uncertainty in estimates due to missing values, examples using the household liquid assets, total household income, age of the respondent, and household size variables in the 1992 SCF are presented. Estimates of the mean and the variance of the mean for each variable are computed using the final non-response adjusted weight provided in the 1992 SCF.[c] The multivariate analysis is conducted using ordinary least squares on the unweighted data.[d] Results derived separately from each implicate are compared with results derived using RII techniques.

Household liquid assets represent the total amount in checking, savings and money market accounts. The household liquid asset variable was chosen as the dependent variable for the multivariate analysis because it is an interesting financial variable which has not been extensively studied, and because the percentage of missing values, and therefore values which are imputed in the final data set, is moderate. Of households that had the item, the percentage of cases for which the amount in the account was missing, either because the respondent didn't know or for some other reason, was 11% for checking accounts, 14% for money market accounts, and 14% for savings accounts (Table 1). Since we want to illustrate use of RII techniques and to compare results of RII techniques with results which ignore variability due to missing values, choice of a variable with a moderate level of missing values provides a nice illustration. As the proportion of missing values increases (decreases), correcting the variance estimate for variability due to missing values will produce relatively larger (smaller) increases in the magnitude of the estimated variance.

Example 1. Estimates of Means and Variances
In the 1992 SCF, 91% of households had liquid assets and over 99% of households reported a positive value for total household income (Table 2). A comparison of the estimates of mean household liquid assets and the variance of mean household liquid assets derived separately for each implicate illustrates the variability in the data due to imputation of missing data. The first implicate produces relatively large estimates of the mean ($12,089) and standard error of the mean ($1,456) compared to the other four implicates. When RII techniques are applied, the best estimate of mean household liquid assets is $11,898, with a standard error of $1,344.

Similarly there is variability in the data due to imputation of missing values for components of total household income. The fifth implicate produces relatively large estimates of the mean ($39,221) and standard error of the mean ($1,316) compared to the other four implicates. When RII techniques are applied, the best estimate of mean total household income is $38,914, with a standard error of $1,294. In contrast, the estimates of the mean and variance of age of the respondent and household size are similar across the five separate implicates since little information on these variables is missing.

Example 2. Regression Model
A linear regression model is used to illustrate the use of RII techniques in a multivariate framework. The specification of the model and the estimation procedure are purposely kept simple in order to focus on the RII methodology. The dependent variable is the dollar value of the household's liquid assets. The independent variables are selected to analyze the effects of income, age of the respondent, and household size on household liquid assets.

In order to reduce heteroskedasticity (unequal variance of the disturbances), the natural logarithm of annual total household income (before taxes) is used instead of the dollar amount.[e] Linear and quadratic terms for age of the respondent are included to allow for nonlinear age effects. Household size is captured with four dummy variables, one each for household size one, two and three, and one variable for households with four or more persons. Interactions terms between income and each of the age and household size variables are used to allow the effect of income to vary with age of the respondent and across households of different sizes. The equation is estimated by ordinary least squares on the sample of households with total household income less than or equal to $100,000 [f]. Ordinary least squares is chosen due to the ease of interpretation of the estimated coefficients.[g]

There are several consistent results across the five separate implicates but also some differences (Table 3). The estimated coefficient on the income term is positive and statistically significant (p<.05) in all five implicates. The linear and quadratic age terms are positive and negative respectively, and statistically significant in all five implicates. All estimated coefficients on the household size dummy variables are positive. The coefficients on the variables for two person households and four person households are statistically significant in all but the second implicate. The coefficient on the variable for three person households is statistically significant only in the fifth implicate. The interaction term between income and age has a negative and statistically significant coefficient in all five implicates; and the interaction term between income and age squared has a positive and statistically significant coefficient in all five implicates. The interaction terms between income and the three household size variables all have negative coefficients. The coefficients on the interaction terms for two person households and four person households are statistically significant in all but the second implicate. The coefficient on the interaction term for three persons households is statistically significant only in the third, fourth and fifth implicates.

There is also variability across the five separate implicates in the magnitude of the estimated coefficients and standard errors. For example, the estimated coefficient for the income term ranges from 41,174 in the fourth implicate to 56,771 in the third implicate. The estimated standard error of the income coefficient ranges from 8,296 in the fourth implicate to 18,900 in the second implicate.

Estimates derived using RII techniques use the information from all five implicates and incorporate imputation variability. Results which were consistent across the five separate implicates are confirmed in the RII results, although the levels of significance are generally more stringent. For example, the coefficient on the income term which is positive and significant in each of the five implicates, is also positive and significant (p<.01) in the RII results. Similar RII results are obtained for age, age squared, and the interaction of the income and age terms. It is informative to examine variables for which results from the five separate implicates were not consistent. For example, the coefficient on the dummy variable for two person households is positive and significant in four of the five implicates, and is positive and significant (p=.04) in the RII results. Similar RII results are obtained for four person households and the interaction terms between income and household size two and income and household size four. The coefficient on the dummy variable for three person households is positive, but significant only in one implicate, and is not significant in the RII results (p=.24). The interaction term between income and household size three is negative and is significant in three implicates, but is not significant in the RII results (p=.18). These results illustrate the risk of basing inference on results of a single implicate. Based on the fifth implicate alone, all eleven independent variables are statistically significant. Based on the second implicate alone, only six of the eleven independent variables are statistically significant. The best estimates are generated through RII techniques which result in nine of the eleven independent variables being significant predictors of household liquid assets.

These examples illustrate the quantitative differences in estimates derived independently from each of the five implicates in the 1992 SCF. Due to these differences, inference based on analysis of only a single implicate of the 1992 SCF risk misrepresenting the magnitude, variability and statistical significance of parameters of interest. Further, the best point estimates and estimates of variance for parameters of interest are generated through RII techniques which use information from all five implicates and incorporate information on the variability due to missing values.

As an example of the combined impact of the differences in estimates, Figure 1 shows the predicted level of liquid assets for one person households with a mean income level ($38,914) for ages ranging from 20 to 80. (The

©1996, Association for Financial Counseling and Planning Education

predicted levels are high because this analysis is unweighted, so they are not representative of the U.S. population.) The difference between Implicate 1 and Implicate 2 in predicted levels of liquid assets is $7,797 at age 20 and $13,135 at age 80. The RII estimates are generally between the extreme values predicted from the Implicate estimates. Clearly, when predictions are desired, the RII method should be used.

Table 2. Estimates of Mean and Variance of the Mean Derived Separately for Each Implicate and Using RII Techniques. 1992 SCF, Full Sample, Weighted.

Descriptive Statistic	Implicate					RII Techniques
	First	Second	Third	Fourth	Fifth	
Household liquid assets						
Mean	12,088.57	11,921.83	11,929.12	11,519.49	12,030.49	11,897.90
Std. Error of the Mean	1,456.23	1,419.11	1,330.72	1,226.16	1,153.47	1,344.42
Variance of the Mean	2.121E+6	2.014E+6	1.771E+6	1.503E+6	1.330E+6	1.807E+6
Number of nonzero observations	3,539	3,536	3,538	3,538	3,539	
Percent of nonzero observations	90.6	90.5	90.6	90.6	90.6	
Total household income[‡]						
Mean	38,827.96	38,836.12	38,803.55	38,883.52	39,220.57	38,914.35
Std. Error of the Mean	1,268.74	1,284.32	1,281.30	1.247.48	1,316.05	1,293.83
Variance of the Mean	1.610E+6	1.649E+6	1.642E+6	1.556E+6	1.732E+6	1.674E+6
Number of nonzero observations	3,889	3,889	3,889	3,889	3,889	
Percent of nonzero observations	99.6	99.6	99.6	99.6	99.6	
Age of the respondent						
Mean	48.4642	48.4650	48.4629	48.4577	48.4589	48.4618
Std. Error of the Mean	0.2785	0.2785	0.2783	0.2784	0.2784	0.2784
Variance of the Mean	0.0776	0.0776	0.0775	0.0775	0.0775	0.0775
Number of nonzero observations	3900	3900	3900	3900	3900	
Percent of nonzero observations	99.8	99.8	99.8	99.8	99.8	
Household size						
Mean	2.6112	2.6163	2.6134	2.6142	2.6154	2.6141
Std. Error of the Mean	0.0239	0.0240	0.0240	0.0240	0.0240	0.0241
Variance of the Mean	0.0006	0.0006	0.0006	0.0006	0.0006	0.0006
Number of nonzero observations	3,906	3,906	3,906	3,906	3,906	
Percent of nonzero observations	100.0	100.0	100.0	100.0	100.0	

[‡] Seventeen households in each implicate had negative values for total household income. These negative values were set equal to zero in the calculation of the mean and variance.

Table 3. Regression Model Separately for Each Implicate and Derived Using RII Techniques. Dependent Variable: Household Liquid Assets. 1992 SCF, Sample of Households with Total Household Income Less Than or Equal to $100,000, Unweighted.

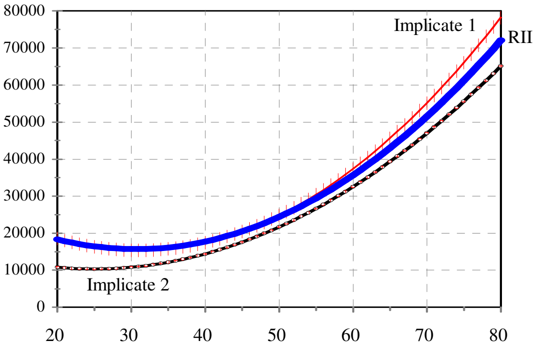
	Coefficient	Std. Error	Variance	t-statistic	p-value	R^2 Model F p-value
First implicate						
Intercept	-533280	99650	9.930E+9	5.351	0.0000	0.0510
Ln(Income)	54305	10220	1.045E+8	5.313	0.0000	13.9985
Age	25285	4206	1.769E+7	6.011	0.0000	0.0000
Age squared	-297.72	42.75	1827.37	6.965	0.0000	
Household size 2	99998	37630	1.416E+9	2.658	0.0079	
Household size 3	51659	39450	1.557E+9	1.309	0.1904	
Household size 4	150020	41120	1.691E+9	3.648	0.0003	
Ln(inc)*age	-2546.4	428.4	183517	5.944	0.0000	
Ln(inc)*age sq	30.65	4.35	18.90	7.050	0.0000	
Ln(inc)*size 2	-10807	3786	1.434E+7	2.854	0.0043	
Ln(inc)*size 3	-6065.7	3952	1.561E+7	1.535	0.1248	
Ln(inc)*size 4	-14918	4090	1.672E+7	3.648	0.0003	

	Coefficient	Std. Error	Variance	t-statistic	p-value	R^2 Model F p-value
Second implicate						
Intercept	-437180	183900	3.38E+10	2.377	0.0175	0.0113
Ln(Income)	43413	18900	3.572E+8	2.297	0.0216	2.9671
Age	21094	7773	6.041E+7	2.714	0.0067	0.0007
Age squared	-249.09	78.97	6235.67	3.154	0.0016	
Household size 2	78733	69400	4.816E+9	1.135	0.2566	
Household size 3	37950	72670	5.281E+9	0.522	0.6015	
Household size 4	92180	75520	5.703E+9	1.221	0.2222	
Ln(inc)*age	-2081.4	793.7	629904	2.622	0.0087	
Ln(inc)*age sq	25.28	8.05	64.84	3.140	0.0017	
Ln(inc)*size 2	-8263.7	6981	4.874E+7	1.184	0.2365	
Ln(inc)*size 3	-4413.5	7279	5.298E+7	0.606	0.5443	
Ln(inc)*size 4	-8311.1	7516	5.649E+7	1.106	0.2688	
Third implicate						
Intercept	-545390	92380	8.535E+9	5.904	0.0000	0.0593
Ln(Income)	56771	9489	9.004E+7	5.983	0.0000	16.4134
Age	25398	3913	1.531E+7	6.490	0.0000	0.0000
Age squared	-296.10	39.94	1594.88	7.414	0.0000	
Household size 2	94017	34950	1.222E+9	2.690	0.0072	
Household size 3	62545	36710	1.347E+9	1.704	0.0884	
Household size 4	167650	38100	1.452E+9	4.400	0.0000	
Ln(inc)*age	-2613.2	399.1	159286	6.547	0.0000	
Ln(inc)*age sq	31.02	4.07	16.55	7.626	0.0000	
Ln(inc)*size 2	-10120	3519	1.238E+7	2.876	0.0040	
Ln(inc)*size 3	-7205.7	3676	1.351E+7	1.960	0.0500	
Ln(inc)*size 4	-16911	3792	1.438E+7	4.459	0.0000	
Fourth implicate						
Intercept	-411510	80810	6.531E+9	5.092	0.0000	0.0442
Ln(Income)	41176	8296	6.882E+7	4.963	0.0000	12.0380
Age	18318	3405	1.159E+7	5.379	0.0000	0.0000
Age squared	-210.70	34.58	1195.85	6.093	0.0000	
Household size 2	94689	30410	9.247E+8	3.114	0.0018	
Household size 3	60333	31900	1.018E+9	1.891	0.0586	
Household size 4	151770	33190	1.102E+9	4.572	0.0000	
Ln(inc)*age	-1801.3	347.2	120567	5.188	0.0000	
Ln(inc)*age sq	21.27	3.52	12.40	6.041	0.0000	
Ln(inc)*size 2	-9878.0	3060	9.361E+6	3.229	0.0012	
Ln(inc)*size 3	-6853.0	3195	1.021E+7	2.145	0.0319	
Ln(inc)*size 4	-15248	3304	1.091E+7	4.615	0.0000	
Fifth implicate						
Intercept	-495040	102400	1.05E+10	4.836	0.0000	0.0461
Ln(Income)	51261	10500	1.102E+8	4.882	0.0000	12.5912
Age	21488	4355	1.896E+7	4.934	0.0000	0.0000
Age squared	-252.81	44.55	1985.06	5.674	0.0000	
Household size 2	133240	38730	1.500E+9	3.440	0.0006	
Household size 3	92993	40560	1.645E+9	2.293	0.0218	
Household size 4	189150	42110	1.773E+9	4.492	0.0000	
Ln(inc)*age	-2186.2	443.8	196990	4.926	0.0000	
Ln(inc)*age sq	26.33	4.54	20.57	5.805	0.0000	
Ln(inc)*size 2	-14493	3895	1.517E+7	3.721	0.0002	
Ln(inc)*size 3	-10550	4064	1.651E+7	2.596	0.0094	
Ln(inc)*size 4	-19416	4190	1.756E+7	4.634	0.0000	

	Coefficient	Std. Error	Variance	t-statistic	p-value	R^2 Model F p-value
RII techniques						
Intercept	-484479	134129	1.80E+10	3.612	0.0005	----
Ln(Income)	49385.2	14204.3	2.018E+8	3.477	0.0010	5.5717
Age	22316.4	5978.19	3.574E+7	3.733	0.0006	0.0000
Age squared	-261.28	64.52	4163.40	4.049	0.0004	
Household size 2	100136	49618.6	2.462E+9	2.018	0.0462	
Household size 3	61096.1	51605.2	2.663E+9	1.184	0.2389	
Household size 4	150155	62460.6	3.901E+9	2.404	0.0239	
Ln(inc)*age	-2245.68	627.78	394107	3.577	0.0011	
Ln(inc)*age sq	26.91	6.81	46.34	3.953	0.0007	
Ln(inc)*size 2	-10712.4	5138.29	2.640E+7	2.085	0.0408	
Ln(inc)*size 3	-7017.55	5275.75	2.783E+7	1.330	0.1871	
Ln(inc)*size 4	-14960.9	6603.03	4.360E+7	2.266	0.0358	

N=2875

Figure 1
Illustration of Difference Between Implicates 1 and 2, and RII Method



Assuming household size=1. Based on regression estimates shown in Table 3.

Summary

The 1992 Survey of Consumer Finances contains five complete data sets, instead of the more common single data set, due to the use of multiple imputation techniques to handle missing data. When imputation techniques are used to fill in missing data, there will inherently be extra variability in the data due to the missing values. This variability needs to be incorporated into empirical estimates. "Repeated-imputation inference" (RII) techniques can be used to estimate this variability. Point estimates and estimates of variance derived by RII techniques provide a basis for more valid inference and tests of significance. Inference based on results from a single implicate ignores the extra variability due to missing values and risks misrepresenting the precision of estimates and significance of relationships.

In general, adjusting variance estimates for imputation variability will increase the estimate of variance compared to estimates which ignore this variability. The magnitude of this change is an empirical question -- it could make a large difference in some situations; an unnoticeable change in others. When the proportion of cases with imputed values is high, as for business assets and stocks in the 1992 SCF (Table 1), incorporating information on variability due to imputation is likely to have large effects relative to when only a moderate proportion (i.e. liquid assets) or small proportion (i.e. credit card balance) of cases have imputed values.

The use of RII techniques in analyzing the 1992 SCF is straightforward, and with the use of computers does little to complicate the estimation. Researchers should use RII techniques in order to produce estimates which incorporate the variability in the data due to missing values. Only then can the practical significance (i.e. how much does it matter) of incorporating this imputation variability be evaluated. Future research should carefully study and document the practical importance of imputation variability. The implications of high proportions of imputed values for variables of interest, both as dependent and independent variables in multivariate frameworks, need to be better understood.

Endnotes

a. *The public use tapes of the 1992 SCF were released during Spring 1996, so empirical research using this data set is just beginning to be published. However, numerous studies using the 1989 SCF, which contains five complete data sets produced by multiple imputation, have been published. A review of some of this empirical research revealed two studies which used information from all five data sets for all analysis (Choi & DeVaney, 1995; Kao, 1995), and two studies which used information from all five data sets for descriptive statistics or bivariate analyses, but not in multivariate analyses (DeVaney, 1995c; Xiao, 1995b). Nine studies clearly stated that analyses were performed on the first data set only (DeVaney, 1995a, 1995b; Hong & Swanson, 1995; Kao, 1994; Malroutu & Xiao, 1995a, 1995b; McGurr, 1995; Xiao, 1995a; Zhong & Xiao, 1995). In eleven studies, the treatment of the multiple data sets was not clearly specified (Drollinger & Johnson, 1995; Hatcher, 1995; Hong & Yu, 1995; Kokrda & Cramer, 1995; Liao, 1994; Steidle, 1994; Xiao, Malroutu & Olson, 1995; Yieh & Widdows, 1995; Yu & Kao, 1994; Zhong, 1994; Zhou,1995).*

b. *In a stochastic imputation method the estimating equation includes a non-zero residual term. This residual term captures the residual (within-class) variance, thus producing better estimates of the standard deviation and the distribution of variables of interest.*

c. *Descriptive statistics are computed using the SCF weight variable in order to produce unbiased estimates of means and variances generalizable to the population of U.S. households. Variance estimates are corrected for imputation variability using the described RII technique. However, the variance estimates are not corrected for variability due to the complex sample design (i.e. sampling variability). Kennickell, McManus & Woodburn (1996) show that imputation variability is small (though not unimportant) relative to sampling variability for their estimates of concentration ratios. The 1992 SCF contains information (i.e. bootstrap replicates) which can be used to correct variance estimates for sampling variability. Variance estimates which are corrected for both imputation variability and sampling variability provide the best basis for valid inference. Montalto and Sung (1996) discuss use of the bootstrap replicates to estimate variability due to sampling and present empirical results for selected variables in the 1992 SCF.*

d. *The regressions reported in this paper are unweighted and the standard errors of the estimated coefficients are corrected for imputation variability using the described RII techniques. This method produces unbaised parameter estimates and valid estimates of the standard errors for testing the statistical significance of individual independent variables. In contrast, when weighted OLS is used, the parameter estimates will be unbiased, but weighted OLS will not correct the standard errors for the complex sample design. Therefore, inference based on weighted OLS results will not be valid. Statistical packages specifically designed for analysis of complex survey data are available, for example SUDAAN®. These programs compute robust variance estimates which fully account for unequal weighting and stratification.*

e. *Disturbance terms are heteroskedastic when they have different variances. Heteroskedasticity usually arises in cross-section data where the scale of the dependent variable and the explanatory power of the model vary across observations. In our example, there is likely to be greater variation in the level of household liquid assets among high-income households than low-income households due to the greater discretion allowed by higher income. In response, the income variable is transformed to the natural log value in an attempt to make the variance more homogeneous on the transformed scale.*

f. *Households with total household income greater than $100,000 are omitted from our multivariate analysis to eliminate the influence of these households which were over sampled in the SCF.*

g. *In the 1992 SCF, 9.4% of the households reported a value of zero for household liquid assets. When the dependent variable is truncated at zero for some observations, parameters estimated by ordinary least squares are biased and inconsistent. The bias in the OLS coefficient estimates can be corrected by dividing each estimated coefficient by the proportion of the sample observations which are not truncated at zero (Greene, 1981). Therefore, the coefficients reported in Table 3 for each of the five implicates would be divided by 90.6 to correct for the bias. Greene (1981) also provides the appropriate correction for the other OLS structural parameters, including the intercept and the standard errors.*

Appendix A
Repeated-Imputation Inference

To analyze any SCF since the 1989 survey taking advantage of the information provided by the multiple complete data sets, separate analyses are performed on each of the five complete data sets (implicates) and the results are combined. Combining the results requires the calculation of the means of the results from the five separate implicates, plus the calculation of the "between" imputation variance, as described below. (Notation closely follows Rubin, 1987.)

The quantity of interest, Q, may be a scalar or a vector, and may represent simple descriptive statistics, such as means, proportions or totals, or more complicated estimators, such as regression coefficients. Let Q_1, Q_2, Q_3, Q_4, and Q_5 represent the point estimates and U_1, U_2, U_3, U_4, and U_5 represent the variance estimates from the first, second, third, fourth and fifth implicates, respectively. The best point estimate of the variable of interest is the average of the five separate point estimates

$$\overline{Q}_m = \frac{\sum_{i=1}^{m} Q_i}{m} \qquad (1)$$

Since there are five implicates in the 1992 SCF, m equals five. For simple statistics, like means, the average of the mean of each of the five implicates is the same as the mean calculated over all five implicates (N=19,530).

To estimate the total variance of the point estimate it is necessary to calculate both the average of the five separate variance estimates (the "within" imputation variance), and the variance due to imputation of missing values (the "between" imputation variance). The average "within" imputation variance is estimated by

$$\overline{U}_m = \frac{\sum_{i=1}^{m} U_i}{m} \qquad (2)$$

The "between" imputation variance is estimated by

$$B_m = \frac{\sum_{i=1}^{m} \left(Q_i - \overline{Q}_m\right)^t \left(Q_i - \overline{Q}_m\right)}{m - 1} \qquad (3)$$

where *t* indicates the transpose when *Q* is a vector.

Thus the total variance of the point estimate is given by

$$T_m = \bar{U}_m + \left(1 + m^{-1}\right) B_m \tag{4}$$

and the standard deviation of the point estimate, defined as the square root of the variance, is given by

$$Std\ \ dev = \sqrt{T_m} \tag{5}$$

The total variance is the sum of the average "within" imputation variance, and the "between" imputation variance weighted by an adjustment factor for using a finite number of imputations. The adjustment factor is inversely related to the number of implicates. As the number of implicates increases, the adjustment factor decreases in size, thus reducing the relative importance of the "between" imputation variance in the estimate of total variance. For example, the adjustment factor is 1.5 when there are two implicates, 1.2 when there are five implicates, and 1.01 when there are 100 implicates.

Example A1. Estimates of Descriptive Statistics for Scalar Q

Table A1 summarizes the use of RII techniques to derive estimates of the mean, variance and standard deviation of the household liquid assets, total household income, age and household size variables from the 1992 SCF. The RII techniques average the results from separate analyses on each of the five implicates, and correct the variance estimate for the uncertainty due to missing values. (The results of the separate analysis on each of the five implicates are summarized in Table 2).

Statistical Inference

When the quantity of interest, Q, represents estimated parameters, such as regression coefficients, we are often interested in conducting hypothesis tests and interval estimates of single parameters, as well as joint hypothesis tests on sets of parameters (i.e. vectors).

Confidence intervals for scalar Q For a single parameter, the point estimate defined in equation 1 and the estimate of the standard deviation (i.e. the standard error) defined in equation 5 can be used to construct the standard confidence interval

$$\bar{Q}_m \pm t_v(\alpha/2)\sqrt{T_m} \tag{6}$$

where t_v (α/2) is the upper 100 α/2 percentage point of the student *t* distribution with *v* degrees of freedom. In the case of repeated-imputation inferences, the degrees of freedom is given by

$$v = (m-1)\left(1 + r_m^{-1}\right)^2 \tag{7}$$

where r_m is the relative increase in variance due to nonresponse

$$r_m = \frac{\left(1 + m^{-1}\right) B_m}{\bar{U}_m} \tag{8}$$

The statistic, r_m, is the ratio of the "between" imputation variance to the average "within" imputation variance, with an adjustment factor for using a finite number of imputations. Notice that as the relative increase in variance due to nonresponse becomes larger, the degrees of freedom become smaller (i.e. r_m and *v* are inversely related), resulting in more conservative significance levels. The statistic, r_m, is positively related to the fraction of information about the parameter, Q, which is missing

$$\gamma_m = \frac{r_m + 2/(v+3)}{r_m + 1} \tag{9}$$

Thus, as the fraction of information about a parameter of interest that is missing increases, the relative increase in variance due to nonresponse will increase, reducing the degrees of freedom and the level of significance of tested relationships.

Table A1. Scalar Q: Estimates of the Mean and the Variance of the Mean Derived From "Repeated-Imputation Inference" Techniques; 1992 SCF, Full Sample, Weighted.

Descriptive Statistic	Explanation	Formula/Notation	Numeric Result			
			Household Liquid Assets	Total Household Income	Household Size	Age of Respondent
Point estimate	average of the five separate point estimates	$\overline{Q_m} = \frac{\sum_{i=1}^{m} Q_i}{m}$	11,897.90	38,914.35	2.6141	48.4618
Average "within" imputation variance	average of the five separate variance estimates	$\overline{U_m} = \frac{\sum_{i=1}^{m} U_i}{m}$	1.74785E+6	1.63782E+6	0.0006	0.0775

"Between" imputation variance	sample variance in the estimates from the five implicates	$B_m = \frac{\sum_{i=1}^{m} \left(Q_i - \overline{Q_m}\right)^2}{m-1}$	49,668.50	30,144.45	3.8805E-6	1.05E-5
Descriptive Statistic	Explanation	Formula/Notation	Numeric Result			
			Household Liquid Assets	Total Household Income	Household Size	Age of Respondent
Total variance	sum of the average "within" and "between" variance weighted by an adjustment factor	$T_m = \overline{U_m} + \left(1 + m^{-1}\right) B_m$	1.80745E+6	1.67399E+6	0.0006	0.0775
Std. error of the mean	square root of the variance	$Std\ dev = \sqrt{T_m}$	1,344.41	1,293.83	0.0241	0.2784

Testing the null hypothesis for scalar Q The test statistic for the null hypothesis that the point estimate of the scalar Q equals some value, Q_o, is given by

$$t = \frac{\overline{Q_m} - Q_o}{\sqrt{T_m}} \qquad (10)$$

which has a *t* distribution with *v* degrees of freedom, with *v* defined in equation 7. When we are interested in whether a parameter, Q, is significantly different from zero, a common test for estimated parameters in a regression model, the test statistic simplifies to

$$t = \frac{\overline{Q_m}}{\sqrt{T_m}} \qquad (11)$$

Single linear constraints can also be tested with an *F* statistic given by

$$F = \frac{\left(\overline{Q_m} - Q_o\right)^2}{T_m} \qquad (12)$$

which has an *F* distribution with one and *v* degrees of freedom, with *v* defined in equation 7. The resulting *F* statistic will be the square of the *t* statistic defined in equation 10. This reflects the general result that the square of a *t* statistic is an *F* statistic with degrees of freedom one and the degrees of freedom for the *t* test. When we are interested in testing whether a single coefficient is equal to a specific value, it is usually easier to perform a *t* test than an *F* test.

Testing the null hypothesis for a k-dimensional vector Q In many situations we are interested in a set of parameters rather than an individual scalar, and Q becomes a k-dimensional vector. For example, in a regression model we are interested in the overall significance of the regression equation. This is a joint test of the hypothesis that all of the coefficients except the intercept term are zero. The test statistic for a joint hypothesis when the number of implicates is modest relative to the number of variables (according to Rubin, 1987, when m < 5k), is given by

$$\tilde{D}_m = \frac{\left(1 + r_m\right)^{-1}\left(Q_o - \overline{Q}_m\right)\overline{U}_m^{-1}\left(Q_o - \overline{Q}_m\right)^t}{k} \qquad (13)$$

which has an *F* distribution with k and (k + 1)*v*/2 degrees of freedom; *v* is given in equation 7 with r_m generalized to be the average relative increase in variance due to nonresponse

$$r_m = \frac{\left(1 + m^{-1}\right) Tr\left(B_m \overline{U}_m^{-1}\right)}{k} \qquad (14)$$

where Tr(A) is the sum of the diagonal elements in the k x k matrix A.

A test statistic for a joint hypothesis test on a set of parameters can also be computed from the m complete-data χ^2 statistics and the value of r_m as defined in equation 14. This test statistic is asymptotically equivalent to the test statistic in equation 13 (Rubin, 1987, pp. 78-79). This test statistic is given by

$$\hat{D}_m = \frac{\frac{\overline{d_m}}{k} - \frac{m-1}{m+1} r_m}{1 + r_m} \qquad (15)$$

$$where\ \overline{d_m} = \frac{\sum_{l=1}^{m} d_{*l}}{m} = average\ of\ the\ five\ \chi^2\ statistic$$

which has an *F* distribution with k and (k + 1)*v*/2 degrees of freedom, with *v* defined in equation 7. The test statistic in equation 15 depends on the scalar χ^2 statistics and the scalar r_m and is therefore more easily computed than the test statistic in equation 13 which depends on k x k matrices. Nonlinear models often generate χ^2 statistics for testing the joint hypothesis that all of the coefficients except the intercept term are zero.

Example A2. Inference for regression coefficients
A linear regression model is used to illustrate the use of RII techniques in a multivariate framework. Household liquid assets is regressed on an intercept term, and eleven independent variables. The regression model is estimated separately on each of the five implicates. (These regression results are provided in Table 3).

The quantity of interest, Q, is now a k-dimensional vector (k=12), and the variance estimates are now kxk variance-covariance matrices. The relevant statistics for total household income and age of the respondent are the second and third elements in the vector Q and the second and third diagonal elements in the variance-covariance matrices. In the following example, numeric results are shown only for these elements to simplify the illustration.

Equation 1, with Q_i a 1xk vector, is used to calculate the average of the results from the five implicates which produces the best point estimates of the structural parameters

$$\overline{Q_m} = \begin{bmatrix} q_1 & 49385.34 & 22316.44 & q_4 & \ldots & q_k \end{bmatrix}$$

Equation 3 is used to calculate the variance-covariance matrix

$$B_m = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \ldots & b_k \\ b_{21} & 4.633E{+}7 & b_{23} & \ldots & b_{2k} \\ b_{31} & b_{32} & 9.119E{+}6 & \ldots & b_{3k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ b_{k1} & \ldots & \ldots & \ldots & b_{kk} \end{bmatrix}$$

Equation 2 is used to calculate the average of the variance-covariance matrices from the five implicates

$$\overline{U_m} = \begin{bmatrix} u_{11} & u_{12} & u_{13} & \ldots & u_{1k} \\ u_{21} & 1.462E{+}8 & u_{23} & \ldots & u_{2k} \\ u_{31} & u_{32} & 2.480E{+}7 & \ldots & u_{3k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ u_{k1} & \ldots & \ldots & \ldots & u_{kk} \end{bmatrix}$$

Equation 4 is used to calculate the total variance

$$T_m = \begin{bmatrix} t_{11} & t_{12} & t_{13} & \ldots & t_{1k} \\ t_{21} & 2.018E{+}8 & t_{23} & \ldots & t_{2k} \\ t_{31} & t_{32} & 3.574E{+}7 & \ldots & t_{3k} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ t_{k1} & \ldots & \ldots & \ldots & t_{kk} \end{bmatrix}$$

The relative increase in variance due to nonresponse is, from equation 8, 0.38 and 0.44 for the estimated coefficients on total household income and age of the respondent, respectively, implying degrees of freedom, from equation 7, of 53 and 43, respectively.

The test statistic in equation 11 can be used to test whether each estimated coefficient is significantly different from zero. For total household income, the t-statistic is 3.477; with 53 degrees of freedom, the p-value is 0.0010. Similarly, the t-statistic for age of the respondent is 3.733; with 43 degrees of freedom, the p-value is 0.0006.

From equation 9 the fraction of missing information for total household income is

$$\frac{0.38 + 2\,/\,(\,53 + 3\,)}{0.38 + 1} = 0.3016$$

and for age of the respondent is

$$\frac{0.44 + 2\,/\,(\,43 + 3\,)}{0.44 + 1} = 0.3366$$

Equation 13 can be used to test the overall significance of the regression equation. Since this is a joint test of the hypothesis that the coefficients on the eleven independent variables are all equal to zero, information on the intercept term is irrelevant and Q is an 11-dimensional vector and U is a 11x11 matrix. For this joint hypothesis test, r_m is generalized to be the average relative increase in variance due to nonresponse which from equation 14 equals 0.4781. The test statistic equals 5.5717 and has an *F* distribution with k = 11 and (k+1)v/2 = 230 degrees of freedom. The test statistic is large, and the associated p-value is 0.0000.

SAS code for an example of using the techniques described in this article is available at :
http://hec.osu.edu/people/shanna/imput.htm

References

Board of Governors of the Federal Reserve System. (1996). *Codebook for 1992 SCF.* Washington, D.C.: Author.

Choi, H. N. & DeVaney, S. A. (1995). Determinants of bank and retail credit card use. *Consumer Interests Annual, 41,* 148-154.

DeVaney, S. A. (1995a). Emergency fund adequacy among U.S. households in 1977 and 1989. *Consumer Interests Annual, 41,* 222-223.

DeVaney, S. A. (1995b). How well off are older men and women: Evidence from the 1989 Survey of Consumer Finances. *Family Economics and Resource Management Biennial, 1,* 121-128.

DeVaney, S. A. (1995c). Retirement preparation of older and younger baby boomers. *Financial Counseling and Planning, 6,* 25-33.

Drollinger, T. L. & Johnson, D. P. (1995). Life cycle, financial and attitudinal characteristics of charitable donors. *Consumer Interests Annual, 41,* 106-111.

Greene, W. H. (1981). On the asymptotic bias of the ordinary least squares estimator of the Tobit model. *Econometrica, 49*, 505-513.

Hatcher, C. B. (1995). Wealth, reservation wealth, and the decision to retire. *Consumer Interests Annual, 41*, 244-245.

Hong, G. S. & Swanson, P. M. (1995). Comparison of financial well-being of older women: 1977 and 1989. *Financial Counseling and Planning, 6*, 129-138.

Hong, G. S. & Yu, J. (1995). Life cycle stages and the use of home equity lines of credit. *Proceedings of the Association for Financial Counseling and Planning Education 13th National Conference*, New Orleans, LA, 147-159.

Kao, Y. E. (1994). Consumer choice between adjustable rate mortgages and fixed rate mortgages. *Consumer Interests Annual, 40,* 202-209.

Kao, Y. E. (1995). Probability of receiving an inheritance and leaving a bequest: Evidence from the 1989 Survey of Consumer Finances. *Consumer Interests Annual, 41,* 248-254.

Kennickell, A. B. (1996). Using range techniques with CAPI in the 1995 SCF. Proceedings of the Section on Survey Research Methods, American Statistical Association. Chicago, Illinois. Table 4.

Kennickell, A. B. (1991). Imputation of the 1989 Survey of Consumer Finances: Stochastic relaxation and multiple imputation. Proceedings of the Section on Survey Research Methods, American Statistical Association. Atlanta, Georgia.

Kennickell, A. B. & McManus, D. A. (1994). Multiple imputation of the 1983 and 1989 waves of the SCF panel. Proceedings of the Section on Survey Research Methods, American Statistical Association. Toronto, Canada.

Kennickell, A. B., McManus, D. A. & Woodburn, R. L. (1996). Weighting design for the 1992 Survey of Consumer Finances. Mimeo, Board of Governors of the Federal Reserve System.

Kennickell, A.B. & Starr-McCluer, M. (1994). Changes in family finances from 1989 to 1992: Evidence from the Survey of Consumer Finances, Federal Reserve Bulletin (October), pp. 861-882.

Kokrda, E. & Cramer, S. (1995). Factors affecting retirement savings of women in two age groups. *Family Economics and Resource Management Biennial, 1,* 115-120.

Liao, S. (1994). Expectations for the future, attitudes toward credit and the use of consumer loans. *Consumer Interests Annual, 40,* 164-169.

Little, R. J. A. (1983). The ignorable case (Chapter 21) and The nonignorable case (Chapter 22). In W. G. Madow, I. Olin and D. B. Rubins (Eds.), *Incomplete Data in Sample Surveys, Volume 2.* New York: Academic Press.

Little, R. J. A. & Rubin, D. B. (1987). *Statistical Analysis with Missing Data.* New York: Wiley.

Malroutu, Y. L. & Xiao, J. J. (1995a). Financial preparation for retirement. *Consumer Interests Annual, 41,* 49-54.

Malroutu, Y. L. & Xiao, J. J. (1995b). Perceived adequacy of retirement income. *Financial Counseling and Planning, 6,* 17-23.

McGurr, P. T. (1995). Determinants of credit card holding and usage by older Americans. *Proceedings of the Association for Financial Counseling and Planning Education 13th National Conference,* New Orleans, LA, 106-117.

Montalto, C. P. & Sung, J. (1996). Estimating sampling variability in the 1992 Survey of Consumer Finances. Unpublished manuscript, The Ohio State University.

Rubin, D. B. (1987). *Multiple Imputation for Nonresponse in Surveys.* New York: John Wiley Sons.

Steidle, R. E. P. (1994). Determinants of bank and retail credit card revolvers: An application using the life-cycle income hypothesis. *Consumer Interests Annual, 40, 170-177.*

Xiao, J. J. (1995a). Family income, life cycle, and financial asset ownership. *Proceedings of the Association for Financial Counseling and Planning Education 13th National Conference*, New Orleans, LA, 188-202.

Xiao, J. J. (1995b). Patterns of household financial asset ownership. *Financial Counseling and Planning, 6,* 99-106.

Xiao, J. J., Malroutu, L. & Olson, G. I. (1995). The impact of banking deregulation on family checking ownership and balances. *Family Economics and Resource Management Biennial, 1,* 137-138.

Yieh, K. & Widdows, R. (1995). Households showing financial characteristics of potential bankrupts. *Consumer Interests Annual, 41,* 155-160.

Yu, J. & Kao, E. (1994). Determinants of the use of home equity lines of credit and second mortgages. *Consumer Interests Annual, 40,* 186-193.

Zhong, L. X. (1994). Factors associated with bond and stock holdings. Consumer Interests Annual, 40, 359-360.

Zhong, L. X. & Xiao, J. J. (1995). Determinants of family bond and stock holdings. *Financial Counseling and Planning, 6,* 107-114.

Zhou, H. (1995). Determinants of credit card holding and usage by older Americans. *Proceedings of the Association for Financial Counseling and Planning Education 13th National Conference*, New Orleans, LA, 106-117.

©1996, Association for Financial Counseling and Planning Education