Imputation of the 1989 Survey of Consumer Finances: Stochastic Relaxation and Multiple Imputation

Arthur B. Kennickell, Board of Governors,
Federal Reserve System, Mail Stop 180, Washington, DC 20551

KEY WORDS: Multiple imputation, Gibbs sampling, stochastic relaxation

The Survey of Consumer Finances (SCF) is designed to gather detailed information on the financial and demographic characteristics of U.S households. Inevitably in such a survey, some respondents are unwilling or unable to provide all of the information requested of them. In waves of the SCF before 1989, imputations of missing values were made on an ad hoc basis. A decision was made for the 1989 survey to build systematic imputation and editing software that reflects the current state of knowledge in the area and that would be substantially reusable in future waves of the survey.

This paper describes the Federal Reserve Imputation Technique Zeta (FRITZ) developed for the 1989 SCF. In the next section of this paper, I give a description of the structure of the 1989 SCF and evaluate the incidence of missing data. In the third section, I review some of the theory underlying the procedures applied. In the fourth section, I give an outline of the structure of the FRITZ model. The final section provides some statistics on the performance of the model.

1989 Survey of Consumer Finances

The purpose of the SCF is to provide a comprehensive and detailed view of the financial behavior of households. Altogether well over 1,500 distinct variables were collected. Detailed information was gathered on all assets and liabilities attributable to the primary economic unit in the household. Liabilities include credit card debts, installment loans, mortgages, lines of credit, and other loans. Assets include the principal residence, all types of loans made to others, real estate assets, businesses, various types of accounts, including checking, saving, money market, IRA, Keogh, and brokerage accounts, stocks, mutual funds, bonds, and other assets. Detailed information was also collected on the current job of respondents and their spouses, their current and future pension rights, and other demographic characteristics.

To accommodate the many types of information requested from a range of socioeconomic groups, the questionnaire for the SCF is quite long and highly structured. Typically, the design of the instrument is such that questions about dollar amounts are preceded by one or more filter questions. For example, before asking the amount in a respondent's fifth checking account, the person is asked first whether the unit has any checking accounts, and then how many accounts the unit has. Sometimes respondents are asked different sequences of questions depending on the answer to such filter questions.

The sample design of the 1989 SCF is also complex (See Heeringa and Woodburn [1991]). The two major parts of the sample are the overlapping panel cross-section based on the 1983 SCF sample (1803 cases), and a new independent cross-section sample (2000 cases). These samples are based on a dual frame design. One part of this sample was drawn from a standard national area-probability frame and the remainder was selected from a list frame developed from administrative files maintained by SOI for the purposes of oversampling respondents likely to have higher levels of income and wealth. The motivation for the dual frame design was two-fold. First, since an important mission of the survey is to characterize the distribution of financial assets, which are highly concentrated in a small part of the population, an efficient design should over-represent that group. Second, common survey folklore and ongoing analyses of the 1989 SCF support the claim that nonresponse tends to be higher for households with higher levels of wealth. In most area-probability samples, there is no means of making systematic adjustment for this differential nonresponse. The advantage of the list sample used for the 1989 SCF is that nonresponse adjustments can be made based on extensive income information contained in the administrative records that served as the sample frame.

Data for the survey were collected between the months of July 1989 and March 1990 by the Survey Research Center at the University of Michigan. Interviews were largely conducted in person and averaged about 75 minutes.

Before the data were punched, the questionnaires went through several stages of editing. Interviewers performed a consistency check as soon as possible after the interview. In the field office, the questionnaires were examined more closely for evidence of interviewer or respondent error – with particular attention to the possibility of double-counting of assets and liabilities. Further machine editing was performed on the punched data for more complicated logical problems.

Data changes at all stages of editing represent something very close to, if not identical to, imputation. Generally, a conservative approach was taken to changing data in editing. However, when missing pieces of information were obvious in the context of other information in the questionnaire, the data were filled in at this stage. Records were kept of major changes to the data. As one might expect of an interview that was administered to households of all ranges of financial sophistication, editing was substantial and important for the quality of the final product.

The achieved sample for the entire 1989 survey includes 3,803 households. Of this number 3,134 have cross-section representation and 1,479 have panel representation. Area-probability and list cases were treated slightly differently in the field. Area-probability cases were approached directly by interviewers, and about 69 percent of these cases were eventually interviewed. The list cases were given a prior opportunity to refuse participation by returning a postpaid card. About a third of the list cases refused participation at this stage. The remainder were approached by interviewers, yielding an overall interview rate for the list sample of about 34 percent. While the interview rate for the list cases is not high according to usual criteria, this figure merely makes explicit the differential nonresponse with respect to income that is hidden in other surveys that have insufficient frame information to reveal the problem. Moreover, in the SCF, we have at least the hope of making systematic adjustments to the sample by estimation of response models using the universe data under the assumption that units are conditionally missing at random.

Every observation in the survey contains at least one piece of missing information – often a very trivial item such as the interviewer ID number. Partial information was available for many items. Respondents who were reluctant to provide dollar values directly were offered a card containing dollar ranges labeled with letters. For total income a more directed "tree" approach was taken to bound income more tightly. Excluding range-card responses, the mean number of missing values per case is 21.6, the median is 11, the 90th percentile is 37, and the total number of missing values to be imputed is 82,125.[1] The mean number of range responses was 3.4 per interview and the total number of such responses was 3,477. For comparison, the maximum possible number of missing values is about 6 million. However, all pieces of missing information are not of equal value in terms of the overall objectives of the survey – e.g., the amount a respondent has in a sixth checking account is usually less important that the total amount of corporate stocks.

Another gauge of severity of the problem, the proportion of missing dollar amounts based on the imputed values, is given below in the discussion of the results of the model.

The structure of missing values is quite complicated. As noted above, the questionnaire is designed so that respondents are led down many question paths with several conditional branches. In addition, a very great number of patterns of missing data appear in the data. For all practical purposes, it is a safe assumption that the overall pattern of missingness for each case is unique. Thus, the imputation of the missing values cannot be addressed routinely using techniques developed for "monotone" patterns of missingness without sacrificing substantial information to achieve monotonicity for subgroups (See Little and Rubin [1987]).

Table 1 provides response rates for a nonrandom selection of survey variables for the panel and cross-section observations taken together. As shown in the table, item nonresponse rates vary widely, but generally within a range that is typical of other economic surveys. One exception is 1988 adjusted gross income, which was missing in over 28 percent of cases. I suspect that this very high level of nonresponse had two important sources. The field period began later than expected after April 15th and ran longer than expected, and respondents were not encouraged to look up data where appropriate.

Review of Imputation Theory

There are numerous ancestors of the missing value techniques reviewed in this section. For a more complete history, I refer the reader to the detailed references in the landmark National Academy volumes (Madow, Olin, and Rubin [1983]), Little and Rubin [1987], and Rubin [1987].

Three strands of literature are particularly relevant to the work reported in this paper: the EM algorithm, multiple imputation, and Gibbs sampling, or stochastic relaxation. A more complete overview of this literature is given in Rubin [1990] and Gelfand and Smith [1990].

The EM algorithm presented in Demster, Laird and Rubin [1977] is intended to estimate parameters in a dataset where some information is only partially observed and direct estimation in the presence of missing information is difficult, but estimation with complete data would be easier. The procedure operates by iteratively simulating the missing data and estimating the parameters. The iteration continues until the parameter estimates are sufficiently close to a fixed point. The intuition of this landmark paper underlies all that is reported here.

Rubin's work on multiple imputation (see particularly Rubin [1987] and references therein) serves as a bridge between EM and the later simulation techniques that involve a structure similar to EM. Briefly, multiple imputation simulates the *distribution* of missing data and, thus, allows a more realistic assessment of variances and a more efficient representation of first moments.

A paper by Tanner and Wong [1987] follows from the methods of EM and ideas of multiple imputation and offers a clear framework for understanding the usefulness of iterative simulation methods in imputation. Tanner and Wong focus on the estimation of a set of parameters where some potential conditioning information is unobserved, but it is easy to extend the argument to estimation of missing data.

Papers by Geman and Geman [1984] and Li [1988] provide useful approaches for dealing with more complex data structures. These papers describe an iterative Markovian procedure of successive simulation of the distribution of variables conditioned on both observed data and distributions of variables previously simulated in the same iteration. The method is typically referred to as stochastic relaxation or Gibbs sampling. Although convergence is reported to be slow for large numbers of variables, Geman and Geman show that under regularity conditions, the process converges and that the simulated distribution of X_u moves closer to the true latent distribution with each iteration. This approach is discussed further below in the development of the FRITZ model.

Description of FRITZ

In past SCFs, imputation had been performed on an ad hoc basis, with significant and very frequent intervention by analysts at the level of individual imputations well beyond the editing stage. While the effort involved in the development of FRITZ has been great, I believe that much of the core set of procedures can be reused for future SCFs as well as for other purposes.

There is a continuum of changes to the respondents' answers from the point of interviewer recording, through primary data editing, to statistical imputation. Virtually all imputations made after the primary editing stage are model-based, though a small number of documented cases have been imputed judgmentally – typically variables that would be quite cumbersome to impute, but which are resolved with very high probability upon inspection. Judgmentally imputed variables are flagged as such in a set of flag variables maintained for each main variable.

FRITZ was designed to handle the great majority of statistical imputations. Although the procedure is iterative and involves multiple imputations, for relative transparency of exposition, it will be convenient to act at first as though the model were the more usual case of single imputations computed without iterations. The general procedures applied in the first iteration are used in all later iterations. Special problems induced by the mixture of panel and cross-section data will only be presented later in the discussion.

Basic Procedures in the First Iteration

Let the potential set of variables collected for a given case r (r=1 to R) be denoted by X^r where X^r is a vector of N variables. Additionally, let X^r_g (of rank N_g) and X^r_m (of rank $N_m = N - N_g$) denote, respectively, the partitioning of X^r into variables that are available and those missing for some reason. The goal of the imputation process is to obtain a good estimate of $F(X_m, X_g)$. Multiple imputation allows the dataset itself to stand as a proxy for that distribution.

Using a variation on the technique of Gibbs sampling (or stochastic relaxation), FRITZ proceeds through the variables to be imputed in a predetermined sequence making imputations variable-by-variable. In the process, the information set available for imputation of each case expands as imputation proceeds through the sequence of variables. Imputed variables are treated exactly like reported variables within each iteration. That is, in the first iteration we estimate

$$F(\beta_1 | X_g)$$
$$F(X_1 | X_g, \beta_1)$$
.....
$$F(\beta_n | X_g \cup X_{m<n})$$
$$F(X_n | X_g \cup X_{m<n}, \beta_n)$$
.....
$$F(\beta_N | X_g \cup X_{m<N}),$$
$$F(X_N | X_g \cup X_{m<N}, \beta_N),$$

where $X_{m<n}$ denotes the missing values imputed in the sequence before variable n and where the parameters of the distribution are estimated from reported and simulated data in the previous iteration, and where β_j is an intermediate parameter vector corresponding to the "M" stage of EM.

In the FRITZ system, there are four types of model-based imputations: imputation of continuous variables, binary variables, and polychotomous variables, and nonparametric imputation. Unfortunately, theory does not offer much help in finding the "true" functional form of F. In the case of most continuous-variable imputations, it is assumed implicitly that the variables with missing values can be taken to have a conditional distribution of the form

$$F(G(a)|H(b)) \sim \text{Normal}(\mu_a, \sigma_a^2),$$

where a is a variable with missing values, b is a set of conditioning variables, and G and H are transformations of a and b, respectively. This assumptions amounts to assuming that

$$G(a)=H(b) + \varepsilon_a, \text{ where } \varepsilon_a \sim N(0,\sigma_a^2).$$

Typically H is assumed to be multiplicative in b and the transformations G and H are taken as log transforms, implicitly yielding the linear model,

$$A=\text{constant}+\beta_1 B_1 +...+\varepsilon_A ,$$

where the capital letters indicate the log transform. The great benefit of this assumption is that a relatively simple covariance matrix of the variables forms a sufficient statistic for imputation and, thus, the simulation of A is straightforward.

In practice, we can be almost certain that the variables we observe are a subset of the appropriate vector B. At the least there are likely idiosyncratic factors for every observation that would be extremely difficult to represent as a reasonably small set of variables even in principle. Once we face the fact that all of B is not known, a potential problem of nonignorable nonresponse arises – that is, conditional on the observed variables the set of nonrespondents for a given item may be a nonrandom subset of the whole sample (See Little [1983]).

In FRITZ an agnostic approach is taken to the set of observed variables chosen to proxy for B. In principle, it might be desirable to take the conditioning set as a series expansion of the function G involving all variables available for each observation. In practice, degrees of freedom limit the number of variables, interaction terms, and higher order terms that can feasibly be included. If one can assume that there is some "local structure" to the distribution of the data (what Geman and Geman call a "cliques") then one can use subsets and summaries of the complete set of information. In any event, no attempt is made in FRITZ to exclude variables that have no obvious "structural" interpretation – the underlying model is a pure reduced form. Most often, the maximal set of conditioning variables for a given case is on the order of 200 or more variables, frequently including a number of recoded variables particularly relevant for a given imputation. Typically included in the set of variables used is a group of interviewer observations on respondents' level of suspicion before and after the interview, their level of interest, etc. The data indicate a reasonable variation in the amount of information reported for all levels of these variables. The hope is that these variables will be correlated with unobserved characteristics of item nonrespondents and, thus, mitigate the potential for nonignorable nonresponse bias. While there is no guarantee that such an approach eliminates – or even reduces – possible response bias, such a strategy may be the best practical insurance against bias. Our means for testing this assumption are very limited.

Operationally, FRITZ looks at a given case, determines whether the current variable in the sequence should be imputed, determines which variables in the conditioning set are available either as reported values or previously imputed values, and computes a randomized imputation. As noted earlier, the combinations of missing values varies widely over all cases so that virtually every case involves a different "regression." Thus, $\hat{A}_j$, the imputed value of variable A for observation j is drawn according to

$$\hat{\beta}_A = [B_{g(i)}'B_{g(i)}]^{-1} [B_{g(i)}'A] \quad \text{and}$$

$$\hat{A}_j \sim F(A|B_{g,j},\hat{\beta}_A) ,$$

where $B_{g(i)}$ denotes the set of values of all observations for variables included in $B_{g,j}$, the set of all available (reported and already imputed within the iteration) values for case j.

An "improper" imputation is made by drawing a value from the distribution implied taking the model coefficients β to be fixed and assuming that ε_A is distributed normally with mean zero and variance given by $A'A-A'B_{g(i)}[B_{g(i)}'B_{g(i)}]^{-1}B_{g(i)}'A$, where the relevant moments are computed as described below. The allowed distribution of ε_A may be truncated or otherwise altered using prior information or editing rules. Because the inversion of a large matrix is usually involved for each such imputation, this method is quite time-consuming.

The moment matrix for the continuous and binary imputations is computed for the appropriate sub-population – e.g., the covariance matrix needed for the imputation of the amount of holdings of certificates of deposit is computed using only households that actually have such instruments. Conveniently, a moment matrix computed using the maximal set of conditioning variables allowed will suffice for every case. The software automatically selects a submatrix for each case corresponding to the conditioning variables available. In the first iteration, the covariance matrix for the imputations is computed without weights and using all non-missing pairs of variables for each observation. As is well-known, this method of calculation allows the possibility that the covariance matrix may no longer be positive definite, implying a negative value for ε_A^2. In practice ε_A^2 is rarely estimated to be negative. For convenience at the first stage, ε_A^2 is given a floor of zero. The alternative of using only cases with full information would usually too drastically reduce the number of observations available for the calculation.

A more serious problem in the covariance estimation is that induced by the presence of very influential cases. Typically this has been a problem in cases where there are coefficients of conditioning variables that are identified by a very small number of observations. In such cases as have been detected, the set of conditioning variables has been reduced to exclude the offending variables. Unfortunately, I have not had either computer power or staff resources to explore this dimension systematically. FRITZ writes out information about imputations as it proceeds and such problems detected to date have been found through inspection of the model output. One sign of problems is the frequent inability of a given model to impute a value strictly within the bounds imposed by the constraints (either determined through edit rules, or from range card estimates). The most desirable approach would be to use robust estimation techniques for the covariance matrix. This will be an important line of research for this project in the future.

For binary variables, it is assumed that the same model holds as in the continuous case. This amounts to the somewhat more suspect assumption that the linear probability model applies. Problems with the linear probability model are well-known. The model fails to account for the information implied by the fact that probabilities must be in the closed interval [0,1] and, because the model is heteroskedastic, produces inefficient estimates of the model parameters. Much better from a theoretical point of view would be to pose the relationship as a probit, logit or other such explicit probability model. There is no low-dimensional set of summary statistics that would apply to all subsets of conditioning variables for such models. Given the great array of patterns of missing data, virtually every observation would require a different model and additional passes through the data. As Rod Little has pointed out, the discriminate model uses the same set of input statistics as the linear probability model, but has the advantage that outcomes are constrained to lie between zero and one. In the on-going revision of the FRITZ model, the discriminant function approach is being implemented.

Given an estimated probability from the linear probability model, a draw is made from the implied binomial distribution to determine the outcome. Some key polychotomous imputations are structured as the sequential prediction of binary choices. The input covariance matrix is computed exactly as in the continuous variable case above.

Less critical polychotomous variables are imputed using a type of randomized hotdeck. Cases are arrayed as a multidimensional frequency table using a number of classifying variables. The imputation procedure randomly selects a value for the missing variable from the appropriate conditional cell. A minimum number of cases is required in each cell. If that minimum is not achieved, there are rules for collapsing adjacent cells. Very closely related to this simple hotdeck procedure is a non-parametric regression technique. Essentially, the difference is that continuous variables are allowed in the frequency table and the collapsing procedures select a slice of specified size from the joint distribution of the variables.

Higher-Order Iterations

In the first iteration, the goal is to estimate a reasonable set of starting values for further iterations. At the end of the first iteration, we have one copy of the dataset with all missing values filled in. From the second iteration and on, the initial dataset containing missing values is replicated 3 to 5 times, and the missing values are filled in based on statistics computed using the completed dataset from the prior iteration. In the second iteration, the covariance matrices and other such basic statistics needed for imputation are estimated from the reported data and the single imputations from the first iteration. In higher-order iterations, these statistics are pooled across the imputation replicates.

The number of replicates increases from one replicate in the first iteration to five in the third and higher iterations. Given the complex tree structure of the data, it is an open question how many replicates may ultimately be needed to reflect adequately the variation due to imputation.

If the assumptions we have made do not move us too far from the requirements of the underlying theory, at each iteration, FRITZ will move closer to the true latent posterior distribution of the data. For convenience, we define convergence in terms of changes in the implied distribution of wealth, rather than as a function of all possible variables. In many applications, Gibbs sampling is known to converge slowly. Unfortunately, this may be a severe limitation in this application. The first iteration of FRITZ requires at least 11 days – largely due to the number of matrices that must be inverted – on a fairly fast Solbourne minicomputer computer dedicated to the project. Subsequent iterations can take 3 weeks or longer. The amount of time required places particularly strains on our ability to debug such complex software. For this paper, only the output of the first two iterations is available.

Some Results from the Model

As noted earlier, short of adding up all missing values equally, it is difficult to find a universally applicable single measure of the information missing due to nonresponse. After imputation, other metrics are available. One such compelling measure is the proportion of dollars imputed for various items. Table 2 provides an estimate of the unweighted percentage of dollars that were imputed for selected items for the panel and cross-section cases together. Weighted percentages might be more informative here, but sampling weights are at such a stage that I do not believe such estimates would be reliable. Weighted and unweighted estimates will be provided later for the panel and cross-section separately.

An estimated 19 percent of total net worth in the sample was imputed, with 4.9 percentage points of that amount imputed using range information. In the case of total income, 35.2 percent of dollars was imputed with an amazing 30.5 percentage points of this amount constrained by range estimates. Most of the other figures reported lie somewhere between these cases.

Table 2 also displays the coefficient of variation due to imputation for components of household net worth and other variables based on data from the second iteration of FRITZ. As might be expected, the model performs better in terms of predicting higher-order aggregates than in terms of individual assets. For example, while the variation for money market accounts is 7.6 percent, the total variation in net worth is only 0.3 percent.

Because only the first two iterations of the model are currently available, it is impossible to say very much about the empirical convergence properties of FRITZ. However, it does appear from the data that are currently available that the difference in the cumulative distribution of net worth (a key variable) is virtually unchanged between the first two iterations.

Endnote

1. If one looks only at dollar amounts of financial assets (checking, money market, savings, and other such accounts, certificates of deposits, stocks, mutual funds, bonds, and trusts), out of a maximum of 136,908 data items, 3350 are missing, the mean number missing per case is 0.9 and the median number is zero.

Acknowledgments

Opinions expressed in this paper are the responsibility of the author alone and do not necessarily reflect the views of the Board of Governors of the Federal Reserve System. The author gratefully acknowledges the support for this work by staff in the Division of Research and Statistics including Edward Ettin, David Jones, Myron Kwast, and Michael Prell, the advice and support of Fritz Scheuren, comments by Roderick Little and Donald Rubin on an earlier description of the imputation model presented here, the patience and creativity of Louise Woodburn in serving as sounding board for many of the ideas, Gerhard Fries for valuable help in debugging the software, and Robert Denk for his careful work in managing the SCF database. Finally, thanks are due also to Robert Avery, the author's predecessor as SCF project director, who nourished an environment where data collection and imputation are taken seriously.

Bibliography

Demster, A.P, N.M Laird, and D.B. Rubin [1977] "Maximum Likelihood From Incomplete Data Via the EM Algorithm," *Journal of the Royal Statistical Society*, Ser. B, v. 39, pp. 1-38.

Gelfand, Alan E. and Adrian F.M. Smith [1990] "Sampling-Based Approaches to Calculating Marginal Densities," *Journal of the American Statistical Association*, Vol. 85, No. 410, pp. 398-409.

Geman, Stuart and Donald Geman [1984] "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. PAMI-6, No. 6 (November), pp. 721-741.

Heeringa, Steven G. and R. Louise Woodburn [1991] "The 1989 Surveys of Consumer Finances Sample Design Documentation," mimeo, Survey Research Center, University of Michigan, Ann Arbor.

Herzog, Thomas N. and Donald B. Rubin [1983] "Using Multiple Imputations to Handle Nonresponse in Sample Surveys" (Chapter 15), in *Incomplete Data in Sample Surveys*, New York: Academic Press.

Judge, George G., W.E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee [1985] *The Theory and Practice of Econometrics*. New York: Wiley.

Li, Kim-Hung [1988] "Imputation Using Markov Chains," *Journal of Statistical Computing and Simulation*, Vol. 30, pp. 57-79.

Little, Roderick J.A. [1983] "The Ignorable Case" (Chapter 21) and "The Nonignorable Case" (Chapter 22), in *Incomplete Data in*

Sample Surveys, New York: Academic Press.

__________ and Donald B. Rubin [1987] *Statistical Analysis with Missing Data*, New York: Wiley.

Madow, William G, Olin, and Donald B. Rubin (editors) [1983] *Incomplete Data in Sample Surveys*, New York: Academic Press.

Projector, Dorothy S. and Gertrude S. Weiss [1966] *Survey of Financial Characteristics of Consumers*, Washington: Board of Governors of the Federal Reserve System.

Rubin, Donald B. [1987] *Multiple Imputation for Nonresponse in Surveys*, Wiley: New York.

__________ [1990] "EM and Beyond," mimeo Department of Statistics, Harvard University (forthcoming in *Psychometrica*).

Tanner, Martin A. and Wing Hung Wong [1987] "The Calculation of Posterior Distributions by Data Augmentation,"(with comments) *Journal of the American Statistical Association*, Vol 82, No. 398, pp. 528-550.

Woodburn, R. Louise [1991] "Using Auxiliary Information to Investigate Nonresponse Bias," paper presented at the annual meetings of the American Statistical Association, Atlanta.

Table 1

Item Nonresponse Rates, Selected Items, Percent*
1989 Survey of Consumer Finances, Panel and Cross-Section, Unweighted

Item	Don't know	Not available	Unknown whether have item	Range response	*Memo item:* percent of all all cases inap.
Balance on bank credit cards	0.6	1.2	0.0	0.8	30.9
Value of own home, excl. mobile homes	1.6	1.2	0.0	0.6	29.6
Amount outstanding on mortgage on home	3.2	2.0	0.1	1.2	58.5
Have any owned cars	0.0	0.0	0.0	0.0	0.0
Number of owned cars	0.0	0.2	0.0	0.0	11.5
Value of 1st business with mgt. role	15.0	3.1	1.1	4.8	73.7
Have checking acct.	0.0	0.2	0.0	0.0	0.0
Number of chkg. accts.	0.0	0.2	0.3	0.0	11.6
Amt. in 1st chkg. acct.	1.4	4.6	0.3	2.5	11.6
Amount of CDs	3.1	7.8	1.8	5.7	73.9
Amt. of savings bonds	4.9	2.3	2.5	2.5	76.3
Amount of stock, excl. mutual funds	5.4	5.1	1.5	5.4	65.4
Cash value of life insurance	31.7	1.6	4.5	1.9	53.0
Wage for respondent currently working	0.9	4.3	1.1	2.1	42.3
Balance in 1st defined contribution pension plan for respondent	16.7	1.8	6.2	2.7	84.2
Total family income	2.1	1.7	14.6	0.0	0.0
Filed 1988 tax return	0.2	1.0	0.0	0.0	0.0
Amount of 1988 adj. gross income	29.0	6.3	1.4	5.2	13.4
Amount of 1st inheritance	5.9	4.4	3.3	3.6	68.1
Amount of 1988 charitable contrib.	1.6	1.9	2.5	3.0	48.9
Wage income for non-primary unit members	30.3	2.9	5.1	5.9	90.1

* Computed as a percent of cases either where response was appropriate or where it was unknown whether response is appropriate.

Table 2

Proportion of Total Dollar Value Imputed,
Coefficient of Variation Due to Imputation, Various Items
1989 Survey of Consumer Finances, Panel and Cross-Section, Unweighted

Item	Proportion of total dollars imputed using range information	Proportion of total dollars imputed without range information	Coefficient of variation due to imputation
Checking accounts	3.1	11.8	0.039
IRA and Keogh accounts	10.9	4.2	0.013
Money market accounts	4.2	16.3	0.076
Savings accounts	3.6	13.8	0.056
Certificates of deposit	5.4	8.0	0.014
Corporate stock	13.2	15.5	0.056
Mutual funds	7.5	15.6	0.087
Savings bonds	3.6	41.7	0.026
Other bonds	3.9	8.3	0.042
Trust assets and annuities	7.5	6.0	0.024
Cash value of life insurance	1.8	19.0	0.033
Notes held	0.8	15.4	0.037
All financial assets	7.1	12.0	0.005
Principal residence	3.3	2.2	0.003
Other real estate	5.5	2.9	0.016
All businesses	22.2	6.3	0.066
Vehicles	2.6	0.3	0.001
Misc. assets	9.8	5.0	0.011
Total assets	5.3	12.9	0.005
Credit card debt	6.0	4.2	0.012
Consumer debt	0.1	4.2	0.000
Principal residence mortgage	0.7	6.3	0.002
Other mortgages	4.4	5.8	0.036
Lines of credit outstanding	0.9	3.4	0.007
Misc. debt	12.2	7.1	0.028
Total debt	3.8	6.3	0.030
Net worth	4.9	14.1	0.003
Total income	30.5	4.7	0.010
Adjusted gross income	15.6	38.6	0.036
Total inheritances received	6.6	19.5	0.117
Total charitable contributions	4.3	2.6	0.009