Interviewers and Data Quality: Evidence from the 2001 Survey of Consumer Finances[1]

Arthur B. Kennickell
Federal Reserve Board, Mail Stop 153, Washington, DC 20551

Key words: Data quality, editing, interviewer effects

Survey data quality might be said to be degraded whenever the information collected differs substantively from the measure the designer of the questionnaire intended. Such differences may occur because the conceptual framework of a respondent differs from the one assumed in the questionnaire design, because the respondent provides–unintentionally or intentionally– an incorrect or otherwise incomplete answer, or because the interviewer fails to follow instructions. In practice, there may be interactions between these sources of error that make them difficult to identify separately. For example, a respondent may not listen carefully to a question that might appear vague for someone in his circumstances, and the interviewer may inappropriately neglect to provide additional information or to probe the initial response.

This paper addresses several aspects of data quality in the 2001 Survey of Consumer Finances (SCF), focusing particularly on the role of field interviewers and following on earlier work reported in Kennickell (1999). Out of a very large set of possible indicators of quality, the paper explores three defined in terms of numbers of: (1) changes to the raw data made as a result of intensive editing; (2) variables determined to be irremediably incorrect and reset to a missing value; and (3) instances in which respondents refused to answer questions or said that they did not know the answer. All three of these measures potentially mix the possibilities of instrument errors, respondent problems, and interviewer errors. Defining the role of interviewers in arriving at a given level of quality is further complicated by the fact that interviewers who are perceived by managers as being particularly good are often given cases that are more challenging in some ways; such interviewers may appear to perform worse than would be the case if they interviewed "easier" respondents. The paper uses a modeling approach to filter out systematic variations not associated with individual interviewers.

Traditionally, field interviewers' performance is evaluated in terms of the number of cases they are able to complete during a field period, and the efficiency with which they do their work. Over time, the application of such a standard through retention and compensation policies shapes the pool of interviewers available to work on surveys as well as interviewers' expectations. Because the quality of data collected in field surveys is most often quite difficult to observe quickly during the period of collection, it would be difficult to develop objective personnel rules to enforce quality standards through regular rewards or punishment. Thus, unless the standard measure of interviewer quality depends on the same interviewer characteristics as those appropriate for producing data quality, one would expect no correlation between such standards. In the analysis presented there, there is at least some weak positive connection.

The structure of the paper is as follows. The first section provides background on the SCF and the relevant technical procedures involved in collecting and processing the data. The next section discusses the role of interviewers. The third section motivates a set of quality measures and looks at the relationship of those measures to the number of cases interviewers completed. A final section concludes and outlines further work.

I. Background on the SCF

The SCF is conducted every three years to gain a basis for research into the structure of the finances of U.S households and their relationships with financial institutions that support that structure.[2] The survey is sponsored by the Board of Governors of the Federal Reserve System (FRB) in cooperation with the Statistics of Income Division (SOI) of the Internal Revenue Service, and since 1992 the data have been collected by NORC at the University of Chicago. Data for the 2001 survey are used in this paper.

The sample employs a dual-frame design, including an area-probability sample and a list sample which oversamples wealthy households.[3] For the 2001 survey, the area-probability sample included about 5,000 cases, of which approximately 2,900 were ultimately interviewed; the list sample included about 5,200 cases, of which approximately 1,500 were interviewed. For the final data set, missing data are multiply imputed using a largely covariance-based approach. The analysis reported here used a preliminary version of the data for which only one imputation is available.

The survey questions cover a wide variety of assets and liabilities, as well as employment history, pensions, income, demographic characteristics, and various attitudes. The 2001 survey instrument was implemented as a CAPI program, which embodies the accumulated experience of earlier survey in detecting and avoiding errors by respondents and interviewers. The median interview length was about 78 minutes, but complex cases required more than two hours. Because of the critical role of dollar values in the SCF, such variables were given particular attention: When respondents provided a complete response, the program returned a confirmation question restating the dollar response in words; when respondents were either reluctant or unable to provide a complete response, the program generated a sequence of

probes for the interviewer to read in an attempt to bound the true value within a range.

Earlier experience in controlling data quality made clear the importance of allowing interviewers to record less structured data easily as events require during an interview. Interviewer training emphasized the need to record such information to explain problems or unusual situations in the interview. There were three ways an interviewer could provide such information. First, throughout the questionnaire, an interviewer could open a "comment box" and enter an unlimited amount of text. Second, at every point in the questionnaire where there was a nontrivial fixed code frame, the program allowed for open-ended responses, and interviewers often used such fields to record comments of their own or extended comments from respondents. Finally, interviewers were required to complete a "debriefing" interview for each case they completed. As noted in more detail below, these three types of information were key drivers of the data editing.

During the field period, interviewers were required to dial in to an NORC computer daily to transmit and receive data. After minimal initial processing, NORC bundled the remaining data for transmission every two to three weeks to the SCF staff at the FRB.

Upon receipt of each wave of data, cases were initially subjected to a variety of additional software tests designed to test constraints too complex to implement during an actual interview and to search for patterns of data errors that had been detected by other means in earlier SCFs or earlier waves of the 2001 survey.[4] Every violation or suspected violation of the interview protocol detected at this point was investigated and resolved, where possible. Subsequently, the data were reviewed, guided by interviewer comments, responses to open-ended questions, and responses to the interviewer debriefing questionnaire. Highly skilled editors reviewed each specially formatted worksheet and its accompanying data for each case in detail. When a clear pattern of errors emerged from this review, the software-based error checks were expanded to look for additional instances of the pattern. Often such checks revealed errors not obvious in the worksheets for other observations.[5]

Edit-driven changes to the data were guided by a "case law" approach. Where similar circumstances had been encountered in past surveys or earlier in the 2001 survey, the data were treated as identically as possible. Where a problem could not be directly encompassed by existing procedures, the project director served as "judge" in developing appropriate general principles and applying them to the specific case. Two factors were key in any decision to change the data: a preponderance of evidence that the raw data were erroneous, and a clear sense that change was in the best interests of future structural analysis. In the best outcome where change was in order, there was sufficient information simply to correct the data. In other cases, clearly erroneous information was set to missing. When the evidence indicated that the data were inconsistent, but the nature of the resolution was doubtful, the inconsistencies were often allowed to remain. All data changes resulting from editing were flagged by a shadow variable.

Because the computer-generated edit checks were more systematic than the ones based on more detailed individual review, they could also move to resolution more quickly. The computer-generated edit checks were normally completed within a week of the receipt of data. Thus, the results of this process might reasonably be used to generate feedback to individual interviewers and the field management. In contrast, the detailed editing was running about three deliveries behind by the end of the field period; thus, without a significant increase in resources devoted to editing, such information would not have a practical use in monitoring data quality.

II. The role of field interviewers

In most field surveys, interviewers are a very critical link between the analytical objectives of survey designers and respondents. An interviewer's task is complex. They need to find and contact the respondents, often performing sample maintenance of various sorts in the process. They have to persuade respondents that the survey is worthwhile and deal with qualms about confidentiality and other such inhibitions. Even the basic task of administering a questionnaire goes far beyond simple reading, listening, and recording. Interviewers must read clearly and according to protocol, explain questions where respondents are confused, probe to guide respondents to answer questions that might have uncertain answers or that raise particular sensitivities, maintain a rapport with the respondent while remaining emotionally detached, listen to and understand responses, and record answers within a specialized framework.[6] A particularly important function of the interviewer is to limit the "distribution of understanding" among respondents that inevitably remains even after the most thorough cognitive testing of an instrument. Because respondents may differ in very many ways, interviewers must show great flexibility in how they approach all tasks involving interactions with respondents. They must also keep records of their work and perform the accounting necessary for them to be paid.

For the 2001 SCF, 255 interviewers were trained. In addition, 5 of the NORC managerial staff for the project completed at least one interview. An attempt was made to include interviews who had worked on earlier rounds of the SCF as well as those seen as having had strong performance records on other studies, and some available evidence suggests that the effort was successful.

Over the course of the field period, interviewers were monitored and supported by 14 regional managers, who in turn were overseen by 3 higher-level area managers as well as the central office staff. All of the managerial staff

were able to use a case management system that allowed for straightforward queries on the progress of individual interviewers in completing interviews and on the actions taken in the effort to complete each interview.

Except in unusual circumstances, the quality of interviewers' performance was evaluated in terms of numbers of completed interviews and costs per case. A very concerted effort was made to minimize costs and to eliminate interviewers who did not show sufficiently high completion rates. Because it was made clear to everyone that high completion rates and low costs were the key axes of "natural selection," it is reasonable to assume that the interviewers who were retained adapted themselves successfully to these criteria.

Of the 260 people who worked at completing interviews, only 207 completed at least one case. Among those completing any cases, completion is heavily concentrated: the bottom half of the group completed 22.4 percent of the cases, while the 10 percent most productive interviewers completed 27.0 percent of the cases. The most productive interviewer alone completed over 4 percent of the cases, while 24 interviews who did at least one interview completed fewer than 5 cases.

An important question is, do other aspects of quality "travel on the same chromosome" as high completion rates? Several arguments support the idea that the connection might be weak. A priori, it would be easy to believe that the skills necessary to locate and persuade a respondent sufficiently to agree to participate in an interview might be different from those required for a sustained engagement through the administration of a highly structured technical protocol. In practice, given the lack of active monitoring of the quality of the data collected, it is reasonable to assume that interviewers ultimately evolved their own standards for acceptable behavior subject to the need to maintain an acceptable level of productivity. Some interviewers may simply have extend in their own ways the information they received in training, in order to deal with situations in the field that appeared idiosyncratic to them. Absent further guidance, a great variety of even such well-intentioned rules could have evolved. Much more troubling is that the available evidence indicates that some interviewers almost surely did not always follow even the direct instructions given in training. If interviewers who intentionally or unintentionally violate expected interview protocol are productive in the conventional sense, the implications for data quality may be very serious.

III. Measures of data quality

Three measures of data quality are developed here. The first is defined in terms of the number of changes of any sort made to the survey data. In the 2001 SCF, 52,201 such changes were made, affecting 2.67 percent of the set of all eligible variables and 63.9 percent of all cases.[7] The second measure characterizes the extent to which variables were set to missing as a result of editing. This measure includes instances where specific responses were set to missing, as well as instances where a variable that was inapplicable before editing became indeterminate as a result of editing. There were 14,647 such new missing values, affecting 0.75 percent of the set of eligible variables and 31.3 percent of all cases. The final measure is defined in terms of the number of variables reported by the respondent as a missing data response–either because the respondent did not know the answer to the question or because the respondent did not wish to answer–and did not subsequently provide a range answer. There were 41,026 such instances, affecting 2.10 percent of all eligible variables and 89.9 percent of all cases. Because dollar values have a particularly important place in the SCF, it is worthwhile to highlight data quality for such variables separately. Despite the tools built into the CAPI program, more than 19 percent of all data changes due to editing were made to dollar variables, almost 23 percent of the missing variables created by editing, and over 29 percent of the originally missing data items.

Despite considerable efforts to improve data quality on several fronts–interviewer hiring, questionnaire design, and interviewer training–error rates for the first and third measures have increased since the 1998 SCF, while the rate for missing values created as a result of editing remained about the same.

As shown in figure 1, there is a substantial dispersion in the relationship between numbers of completed cases by interviewers and their average percent of eligible variables subjected to editing. The corresponding plots for the other error rates and for the rates plotted for dollar variables alone appear very similar. The first two columns of table 1 show there is a small negative Pearson correlation between completed cases and data quality problems for all instances except the cases of initially missing values for dollar variables. The Spearman rank correlation is weaker, and in the case of initially missing dollar variables, it shows a positive correlation.

One might argue that some of the survey questions are error prone and that the interviewers should not be penalized in accounting for error from such questions. In an attempt to purge the comparison of quality measures of the influence of questions that were relatively commonly problematic, an experiment was conducted wherein data quality problems for particular variables were down-weighted in proportion to the frequency with which those variables had data quality problems across all interviews. The descriptive plots and correlations were remarkably little changed by this exercise. Thus, the analysis presented below ignores potential question-specific problems and implicitly attributes any error from this source to interviewers.

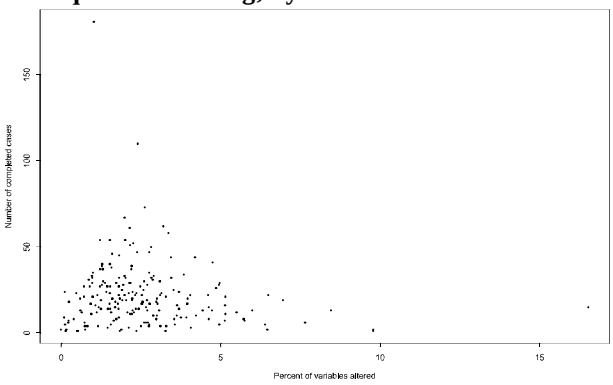
Almost certainly, the most important problem with the simple treatment of the relationship between data

Table 1: Pearson and Spearman correlations across interviewers of number of completed cases with mean percent of eligible variables edited, set to missing in editing, or initially reported by the respondent as a missing value; raw correlations and correlations filtered through OLS and Poisson models to remove variations in respondents; all variables and dollar variables only.

Mean value of:	Raw correlations		OLS-filtered		Poisson-filtered	
	Pearson	Sprmn	Pearson	Sprmn	Pearson	Sprmn
All types of edits						
All variables	-0.11	-0.05	-0.13	-0.16	-0.12	-0.13
	(0.10)	(0.44)	(0.07)	(0.03)	(0.10)	(0.07)
Dollar variables	-0.17	0.13	-0.19	-0.17	-0.10	-0.06
	(0.01)	(0.07)	(0.01)	(0.01)	(0.17)	(0.41)
New missing values						
All variables	-0.05	-0.09	-0.09	-0.11	-0.07	-0.02
	(0.44)	(0.18)	(0.20)	(0.13)	(0.32)	(0.78)
Dollar variables	-0.11	0.00	-0.13	-0.13	-0.17	-0.14
	(0.10)	(0.98)	(0.07)	(0.08)	(0.02)	(0.05)
Initial missing values						
All variables	-0.01	0.00	-0.05	-0.13	-0.03	-0.10
	(0.78)	(0.96)	(0.46)	(0.07)	(0.64)	(0.15)
Dollar variables	0.11	0.17	0.04	0.01	0.08	0.11
	(0.11)	(0.01)	(0.56)	(0.92)	(0.29)	(0.12)

Note: p-values are given in parentheses.

Figure 1: Number of completed cases vs. average percent of all types of variables altered as a consequence of editing; by interviewer.



quality and completed cases is that the assignment of cases to interviewers is not random. Interviewers who were successful in completing cases tended to be given cases that were more difficult to convince to participate, and those cases might also have been more likely to have data quality problems. Modeling may allow a separation of respondent-specific and interviewer-specific effects. Suppose that data quality for a given case, **Q**, and the propensity to complete the case, **C**, are given by the linear models specified as equations 1 and 2 below. Neither the vectors of respondent characteristics **X** and **F** nor the vectors of interviewer characteristics **Y** and **G** need be the same. Although there might be additional interaction effects–for example, some people may be more cooperative with interviewers of one gender or the other–such possibilities will be ignored here. The interviewer-specific error terms, ν and υ are each taken to have a fixed idiosyncratic component (denoted by a bar) and an iid mean zero random deviation (denoted by a tilde). The respondent-specific error terms ϵ and ξ are assumed to be purely iid mean zero random.

Ideally, we would like to compare the total interviewer-specific effects in controlling data quality and in completing cases as given by equations 3 and 4 respectively. The extensive amount of data available for completed cases makes it seem reasonable to estimate the model given in equation 1 in order to extract the effect given by equation 3. The argument for estimating the quantity in equation 4 is more difficult for three reasons. First, little beyond location and stratum identifiers is available for modeling completion across all cases. Second, because only completion or non-completion is observed for each case (rather than the latent variable **C**), it is necessary to estimate equation 2 using a limited-dependent variable technique, such as probit or logit; in such nonlinear models, it may be difficult to get robust estimates of the large number of coefficients necessary to identifying the interviewer-specific effects. Third, because a given case may have been assigned to several interviewers before it was assigned a final status as a completed case or a nonrespondent, the simple model is not strictly appropriate. It turns out that this third problem is a sufficiently serious one to make meaningful estimation of equation 2 doubtful: In a number of instances, the data indicate that interviewers completed 100 percent of the cases where they are listed as the final "interviewer of record." It appears that all of these interviewers also had involvement at an earlier stage with other cases that were not ultimately completed, but where another interviewer was recorded as the final one. For other interviewers who had completion rates of less than 100 percent by this definition, the problem varies in its seriousness. Unfortunately, there is no clear way to allocate shares of success at this level to individual interviewers without imposing a much more complicated model that would be difficult to estimate from the

1. $Q^{ri} = X^r \alpha + Y^i \beta + \varepsilon^r + \nu^i, \quad \nu^i = \bar{\nu}^i + \tilde{\nu}^i$
2. $C^{ri} = F^r \varphi + G^i \gamma + \xi^r + \upsilon^i, \quad \upsilon^i = \bar{\upsilon}^i + \tilde{\upsilon}^i$
3. $E\left[Q^i\right] - E\left[X^r\right]\alpha = Y^i \beta + \bar{\nu}^i$
4. $E\left[C^i\right] - E\left[F^r\right]\varphi = G^i \gamma + \bar{\upsilon}^i$

available data.

Although it is not feasible here to purge observed case completion data of differential difficulty, it may be that the simple number of completed cases is still a good indicator of an interviewer's ability to complete cases. Some interviewers may have had only relatively easy assignments, and some areas may be more difficult than others. But interviewers who were successful in completing cases were offered additional work, and those who were willing to travel were effectively unconstrained in terms of the number of cases they could add to their caseloads. Thus, interviewers with larger numbers of completed cases would tend strongly to be ones who were viewed by managers as having higher levels of ability.

A model was estimated using both an OLS model of the error rate and a Poisson model of the error counts.[8] The explanatory variables included sample and interview characteristics (dummy variables for strata of the list sample, dummies for regions of the country, and a dummy for self-representing PSUs, a dummy for whether the interview was completed by telephone, a dummy for whether a fee was paid to the respondents, the logarithm of the maximum of 1 and the amount of any fee paid), respondent characteristics (age, education, work status, health, race, length of time living in the area, marital status, household size, dummy for home ownership, reported "normal" income, total assets, total debts, the financial planning horizon, and a dummy for whether the household made charitable contributions of $500 or more in the past year), some interviewer assessments of the interview (a dummy for whether the respondent was suspicious after the interview was completed, a dummy for whether the respondent used records, and indicators for the level of the respondent's interest in the interview, for whether the respondent had a good understanding of the survey questions, and for whether the respondent was able to express himself clearly), and a dummy variable for each interviewer. The interview-specific dummy variables capture the full effect of $Y^i\beta + \nu^i$. The models were estimated using the 4,384 observation for which the interviewer of record completed at least 4 cases.

The most interesting estimates for present purposes are the interview-specific effects. In all the models, the joint hypothesis that the interviewer effects are all equal to zero is rejected at less than the 1 percent level. Using other indicators of interviewers' productivity in gaining cooperation and the quality of the data collected in a Belgian survey, Loosveldt *et al.*(1999) also find significant interviewer effects, but no significant relationship between these terms in the context of a model with limited controls for respondent characteristics.

Plots of average error rates simulated from these models differ little from those for the actual rates. The estimates shown in the last four columns of table 1 suggests that there is a weak correlation between the simulated indicator and the number of cases an interviewer completed. Viewed across all variables, the Pearson and Spearman correlations are relatively small, but uniformly negative. However, when the correlation is restricted to dollar variables, the results are more mixed: The correlations are negative for the ratio for all edits and for missing values created by editing; the correlation is positive for the proportion of originally missing dollar variables, possibly reflecting failure of the model to control completely for assignment effects.

Conclusions

This paper focuses on the role of interviewers in maintaining data quality. Standards for field interviewers typically are set in terms of completing interviews and maintaining low costs. Unfortunately, estimates reported in the paper suggest that the connection between high numbers of completed interviews and high quality data are weak at best. This result calls for a broad reexamination of interviewing.

Survey data may be degraded in several ways–through bad questionnaire design, poor respondent engagement, and interviewer failure. Interviewers potentially have as critical a role in mitigating the first two as they have in the latter. Much work has been done on questionnaire design, particularly through cognitive psychology. While it is clear that much more such work needs to be done, experience suggests that there is likely a point in the refinement of any questionnaire design where improvements for one part of the population are offset by deteriorations for another part. Interviewers who understand their task, who listen to the respondent, and who understand the questionnaire and auxiliary help materials should often be able to bridge such problems. In addition, effective interviewers are able to engage the attention of respondents who are initially resistant or uninterested. Both such activities call for a substantially higher skill level than the basic expectation that interviewers merely follow instructions. It appears that the level of skill needed is not broadly recognized.

Data quality problems that are avoidable in principle may persist either because of shirking or because interviewers genuinely, but incorrectly, believe they are performing as required. Although one hopes the former is not dominant, it is a possibility that deserves as serious attention as is normally given to the possibility of falsification of entire interviews. Whatever the source of quality problems, there is not likely to be an improvement without systematic efforts to observe data quality and to use that information to feed back evaluations to interviewers. Thus, a high priority should be to develop systems capable of generating sufficient characterizations of data quality quickly enough to use in routine monitoring interviewers. If nothing else more comprehensive is feasible in the short term, at least a systematic review of a sample of each interviewer's cases could be undertaken, as is done to test for falsification.

At the beginning of modern survey research, two populations played a critical role in the creation of the field: graduate students, and to a larger extent, married women who did not "need to work." Graduate students remain to a limited degree, but the labor market for women of all types has changed radically. At the same time, wages for interviewers have remained relatively low, and though many very admirable people continue to work as interviewers, there are unmistakable signs of overall deterioration. Urgent attention needs to be given to a study of the labor markets from which interviewers are selected and the price sensitivity of people with the skills necessary to function well as interviewers.

Another possibility that should at least be considered is breaking up the current work of field interviewers into more specialized tasks. Locating, persuading, and interviewing are obvious sub-tasks to consider. Locating respondents sometimes involves negotiations with gatekeepers or similar people, but there is a core aspect that is technical and well-ordered. Convincing strangers to do an interview is often stressful and it requires flexible thinking to accommodate the perspective of the potential respondents in a way that is not easily scripted. Effective administration of a questionnaire requires the ability to engage respondents in a neutral way while managing a technical interview protocol and remaining conscious of the intent of the survey questions and the way that the respondents' answers relate to them and the available framework for recording answers. Where respondents are not fully convinced to do an interview, questionnaire administration may also retain some of the character of the initial negotiation. In the case of part of the sample for the National Survey of America's Families (Safir, Scheuren, and Wang (2001)) that could not be reached by telephone, field interviewers were only used to negotiate agreement to participate in the survey. Once agreement was reached, the interviewer gave the respondent a cellular telephone set to call a telephone center where interviewers were given training in questionnaire administration and where it was straightforward to perform direct monitoring of interviews. Although the intention in this survey was to avoid differential mode effects, this case should be studied closely and other such possibilities examined.

Interviewers are the foundation of a vast amount of important social research. Unless survey managers work actively to improve standards among the people who do this work, analysts may soon find themselves with no data–or worse, with unreliable data that are treated as reliable. The situation is not hopeless, but the required adjustments are likely to be both painful and costly. Delay is not a reasonable alternative.

Bibliography

Bledsoe, R.M. and G. Fries [2002] "The 2001 SCF Editing Process," Paper presented at the Annual Meetings of the ASA, New York, NY.

Loosveldt, G., A. Carlton, and J. Billiet [1999] "Evaluation of the relationship between data quality and interviewer behavior during doorstep contact in face-to-face interviews," http://www.jpsm.umd.edu/icsn/papers/loosveldt2.htm.

Japec, L. [2002] "What we Know and Don't Know About Process Steps in Interview Surveys," Paper presented at the Annual Meetings of the ASA, New York, NY.

Kennickell, A.B. [2000] "Wealth Measurement in the Survey of Consumer Finances: Methodology and Directions for Future Research," paper presented at the Annual Meetings of the AAPOR, Portland.

Kennickell, A.B. [1999] "Measuring Data Quality in the 1998 Survey of Consumer Finances," *Proceedings of the Section on Survey Research Methods*, Annual Meetings of the ASA, Baltimore, MD.

Kennickell, A.B., M. Starr-McCluer, and B.J. Surette [2000] "Recent Changes in Family Finances: Results from the 1998 Survey of Consumer Finances," *Federal Reserve Bulletin*, pp. 1-29.

Safir, A., F. Scheuren, and K. Wang [2001] "National Survey of America's Families: Survey Methods and Data Reliability," Chapter 1 in 1999 NSAF Collection of Papers, Report No. 7, Urban Institute.

Endnotes

1. The author is grateful to Ana Aizcorbe, Ryan Bledsoe, Gerhard Fries, Kevin Moore, and Brooke Wells at the Federal Reserve who made it possible to have edited data to analyze in this paper, and NORC staff who made it possible to have data to edit. All opinions expressed in this paper are those of the author alone and do not necessarily reflect those of the Federal Reserve Board. An unabridged version of this paper may be found at www.federalreserve.gov/pubs/oss/oss2/method.html.
2. See Kennickell, Starr-McCluer and Surette (2000) for an overview of the data.
3. See Kennickell (2000) for an overview of the general methodology of the survey.
4. See Bledsoe and Fries (2002) for discussion of the SCF editing procedures.
5. After the initiation of imputation for missing values, graphical techniques were also used to examine the data.
6. See Japec (2002) for an overview of the literature on the role of interviewers in data collection.
7. The "eligible set" of variables is defined for each observation as the set of instances in either the baseline or comparison data sets where responses were given by the respondent, or where it is not known whether or not a question should have been answered.
8. The expected number of instances of the dependent variable under the Poisson model is given by $E[Q^i] = \{\#\text{ eligible questions}\} * \exp(X^r\alpha + Y^i\beta + \nu^i)$; thus, the exponentiated term has the interpretation of a rate.