USING RANGE TECHNIQUES WITH CAPI IN THE 1995 SURVEY OF CONSUMER FINANCES

Arthur B. Kennickell
Senior Economist
Mail Stop 180
Board of Governors of the Federal Reserve System
Washington, DC 20551
Voice: 202/452-2247, FAX: 202/452-5295
Email: m1abk00@frb.gov

January 1997

The author wishes to thank Val Cook and Geoff Walker who wrote the CAPI program for the 1995 SCF, and without whose talent, dedication, and patience there would be no information to report in this paper. The author is also grateful to Gerhard Fries and Kevin Moore for invaluable help in preparing the information used in this paper, and to Steve Heeringa and Martha Starr-McCluer for comments. The author is particularly grateful to the respondents who gave their time for the interview and to the interviewers who collected the data. The views expressed in this paper are those of the author alone and do not necessarily reflect the official position of the Board of Governors.

The Survey of Consumer Finances (SCF) collects dollar amounts of a wide variety of assets, liabilities, payments, incomes, and other items.[1] Experience has shown that (1) item nonresponse on such items is a serious problems, and (2) range cards and decision trees posing sequences of bounding choices can provide important partial information.[2] Respondents may fail to provide complete information for a number of reasons. Some questions may appear overly intrusive to some respondents, and they may refuse to answer. There may also be respondents who could, in principle, know the exact answer to a question, but who do not remember the answer and cannot be persuaded either to take action to uncover the value or to make an estimate of the value. However, there may be cases where the exact answer is not clear even in principle. There may be ambiguity about the concept being probed or there may be a set of coherent responses to a single question: e.g., it may not be clear whether a question about house value is asking for the market value or the assessed value, or the value of the house depends on whether one wants to sell quickly or one wants the highest possible price. It may also be the case that the only way to determine the answer to a question is to take an extraordinary action: e.g., the value of a closely held business may not be knowable until one tried to sell it. All of these factors apply to respondents to the SCF.

The redesign of the 1995 SCF for computer-assisted personal interviewing (CAPI) provided an opportunity to integrate range data more formally into the data collection process than was feasible with a paper questionnaire. A computer subroutine was written to do three things for every potential dollar response: (1) provide a confirmation in words of the amounts reported, (2) provide a place to record ranges reported by respondents who are reporting items for which there is genuine uncertainty about the value, (3) confront every “don’t know” or “refuse” response with a request to use a range card or to go through a dollar decision tree to educe a bounding range.

Range data may provide information for imputation of missing data that cannot be

---

[1]In the 1995 SCF, there are 479 possible dollar questions, but it is not possible for a given respondent to be confronted with all of them.

[2]Kennickell [1991] provides information on the use of ranges in the 1989 SCF, and Heerings, Hill and Howell [1995] and Juster and Smith [1995] summarize the experience of the Health and Retirement Survey with an elaborate system for collecting range data.

obtained from conditioning on observed information—that is, nonresponse may be nonignorable. However, such information must be weighed against substantially higher respondent burden in many cases, and decreased interviewer flexibility in awkward situations. Unfortunately, we do not have the luxury of analyzing the results of a controlled experiment in the use of ranges. The evidence presented here suggests a complicated interaction between the range questions and the new mode of administration. The introduction of effective "automated probing" with CAPI could serve to lower the number of completely missing responses. However, if interviewers believe that ranges are equivalent to complete responses, or they find it convenient to "hide behind the computer" in probing initial nonresponses, one might expect the proportion of complete responses to decline as well.

In the next two sections of this paper, I provide some background information on the SCF, and outline the procedures used to collect partial information through ranges for amount questions. Next, I present some time series information on nonresponse patterns in the SCF. The data suggest that (1) item response rates have deteriorated over time, (2) the introduction of the elaborate system to collect ranges in 1995 appears to have reduced the proportion of "don't know" responses, (3) there may have been a "conversion" of responses that might have been complete responses in the old questionnaire to partial information. Some modeling also indicates that interviewer learning may have steered the outcomes toward a type of response described below as "volunteered ranges." The data suggest that there is substantial persistence in the use of ranges by individual respondents, and that respondents who use ranges tend to use the same type of range response. The fourth section assembles various pieces of information in an attempt to gauge whether the responses of range respondents are systematically different from those of respondents who provided complete information, and how much information is added to various types of estimates by the inclusion of such information. The evidence here is mixed. The data do suggest that the univariate distributions of outcomes for some types of range are different from those of complete reporters, but the differences are largely in the upper tails—a potentially important problem for the SCF, but not necessarily a critical one in most surveys. Unfortunately, it is difficult to create a universally appropriate significance test of the differences. Regressions using data imputed both with and without the reported range information suggest (not surprisingly) that ranges improve model fit, but any bias effects appear small in the case considered. The final section of the paper summarizes some important findings and suggests changes for the next SCF.

I. Background on the Survey of Consumer Finances

The SCF has been conducted on a triennial basis since 1983 by the Board of Governors of the Federal Reserve System in cooperation with the Statistics of Income (SOI) of the Internal Revenue Service.[3] The Survey Research Center at the University of Michigan collected the data for the survey from 1983 to 1989, and the National Opinion Research Center (NORC) at the University of Chicago has collected the data since that time.

The SCF is intended to provide detailed information on the distribution of a large number of financial characteristics of U.S. households. The 1995 questionnaire took, on average, about 90 minutes to complete, but for some households with very complex finances, the interview lasted over three hours. The most detailed data are collected on assets (including checking, money market, and savings accounts, IRAs and Keogh accounts, savings bonds, other types of bonds, mutual funds, publicly-traded stocks, trust accounts, annuities, businesses, the principal residence, other real estate, vehicles, loans made to others, and other assets) and liabilities (including credit card debt, principal residence mortgage debt, other mortgage debt, lines of credit, automobile loans, education loans, other installment loans, margin loans, loans against pensions and insurance policies, and other loans). Information is also collected on employment history, pension rights, inheritances, marital history, attitudes, and numerous other items.[4]

Some of these variables are relatively broadly-distributed in the population (e.g., credit

---

[3]There is an earlier series of surveys of the same name associated with the University of Michigan and the Federal Reserve, but these surveys were largely geared toward collecting information on purchase intentions. The earliest comparable U.S. survey was the Survey of Financial Characteristics of Consumer, conducted by the Bureau of the Census for the Federal Reserve in 1963 with a follow-up in 1964 (see Projector and Weiss [1966]).

[4] See Kennickell and Starr-McCluer [1994] for a description of the data in the 1992 survey.

card debt, principal residence mortgages, and automobile ownership), while many other variables have a highly skewed distribution (e.g., most financial assets, investment real estate, and businesses). A standard area-probability sample would provide sufficient coverage for the first type of variables, but such a sample would provide very inefficient estimates for the distribution of the second type. Moreover, as noted in Kennickell and McManus [1994], there is strong evidence that unit nonresponse is much more likely among wealthy households. Thus, analysis that fails to account for the nonrandomness of the response mechanism would yield biased estimates of many characteristics of the distributions of assets and liabilities that are held disproportionately by wealthy households (e.g., the mean of stock holdings). The SCF addresses these problems by using a dual-frame sample incorporating both an area-probability sample, and a special list sample developed from a sample of tax records that strongly oversamples wealthy households (see Kennickell, McManus and Woodburn [1996]). The great majority of wealthy households in the survey derive from the list sample. Because the frame data allows us to identify some key systematic aspects of unit nonresponse, we have a reasonable hope that the SCF provides a reasonably reliable basis for wealth estimation for the entire population.

There have been substantial variations in the size and composition of the list sample over time. In the 1983 survey, the list sample was selected using an income-based definition, and that group was sent a letter describing the survey and a postcard to be returned if they were willing to participate. This need for active agreement to participate was a condition for using tax data to select the sample. Not surprisingly, the response rate in that part of the sample was quite low—about 10 percent.[5] For the 1989 survey, a more systematic effort was

[5]In 1983, 159 area-probability sample cases were deleted because the information provided was unusably incomplete or was insincere. Following past analysis of the 1983 data, I have excluded these cases from the analysis of missing data. A large proportion of the incomplete cases would now be treated as "partial interviews," or "breakoffs." Under current practices, such cases must provide information (missing data or otherwise) on a core of critical items to be included in the final set of cases. This effect of this treatment has been to include a much higher fraction of such cases in the final datasets of later surveys.

undertaken to stratify the list sample (see Heeringa, Connor and Woodburn [1994]).[6] In an important revision of the approach to obtaining consent from this group, the selected units were mailed and letter with a postcard to be return only if they refused to participate. Response rates improved dramatically. Subsequent years of the survey have built on this model. The size of the list sample was expanded again in 1992 (see Kennickell, McManus and Woodburn [1996]). In 1995, a new model-based approach to stratification was developed to improve the efficiency of the sample (see Frankel and Kennickell [1995]).

Other than the large increase in response rates for the list sample between 1983 and 1989, the rates for the area-probability sample and for comparable strata of the list sample have not varied much over time.[7] However, rather than being a reflection of a general population stasis, this outcome is the conscious result of a decision to devote ever-increasing resources to maintain acceptable response rates. Table 1 provides information on the number of completed cases in the area-probability and list samples of the surveys discussed in this paper.

Table 1: Number of completed cases in the 1983, 1989, 1992, and 1995 SCF, by sample type.

	Full sample	AP sample	List sample
1983 SCF	4103	3665	438
1989 SCF	3143	2278	865
1992 SCF	3906	2456	1450
1995 SCF	4299	2781	1518

II. Use of Range Data in the SCF

As noted in more detail below, missing data rates for cases that complete the survey are substantial for many of the key dollar variables. This appears to be the common experience of most surveys, though SCF response rates may look somewhat worse for some

[6] The complete 1989 SCF was a very complex sample, an overlapping panel-cross-section design. The part of the survey discussed here is the combination of a cross-section sample constructed using the 1983 design and an independent cross-section sample selected in 1989.

[7] The area-probability sample response rate has been about 70 percent. It is much harder to state the list sample response rate simply because of the complex nature of the stratification and the strong variations in response rates over the strata. In 1992, the stratum corresponding to the wealthiest households had a response rate of about 14 percent, while that corresponding to the lowest wealth group had a response rate of about 43 percent.

variables because the survey contains a disproportionate number of wealthy respondents who are much more likely to be asked many of the dollar questions than are other respondents. Given an initial expectation of non-negligible missing data rates, beginning with the 1983 survey the survey has attempted with increasing sophistication to incorporate the possibility of reporting partial information.[8] In the 1983 survey, respondents were allowed to report dollar ranges, which were later translated into a single value by coders using a set of rules (e.g., "just over a million" might be coded as "1111111"), and a range card was available for the interviewers to use in probing. Unfortunately, none of this information is systematically recoverable from the final coded data. However, the impression gathered from the coders and editors was that range information was an important element of the information collected.

Thus, when the survey was revised in 1989, a more systematic effort was made to record ranges and to use that information in imputation.[9] A range card was created that contained a list of letters associated with a set of ranges, and interviewers were told to use the card with respondents who would otherwise be unwilling to give a response.[10] The card was a separate piece of card stock in the interviewers' materials. In addition, a "decision tree"—a series of questions designed to bound a partial response—was added to the questionnaire for

---

[8]The use of ranges to collect partial information has an interesting history. The earliest evidence I have been able to find is in the 1967 Survey of Consumer Finances conducted by the Survey Research Center at the University of Michigan. In that survey a "yellow card" with ranges was used for respondents who did not want to or not able to give dollar responses for asset values. In the 1977 Survey of Consumer Credit, also conducted by SRC, all dollar values were collected as ranges, reportedly in the belief that response rates would be raised if only ranges were asked. There is also a history of the use of bounding questions in earlier SRC surveys. The 1984 Panel Study on Income Dynamics introduced a decision tree for key asset and income variables. The 1992 Health and Retirement Survey (HRS) and the Asset and 1994 Health Dynamics Survey (AHEAD) employed a more extensive battery of decision trees.

[9]The 1986 survey was only a limited telephone reinterview of a set of the 1983 SCF respondents. Because it is so different in character from the other waves of the survey discussed here, I will not address the data in that survey. For a discussion of the use of ranges in imputation of the SCF, see Kennickell [1991].

[10]There were 10 ranges identified by the letters from "A" to "J." The corresponding ranges went from the interval "zero to $500," to the open interval "more than $100,000,000." A conventional was also created to handle negative range responses.

the key question on total income, traditionally a particularly sensitive question for respondents.

One problem that emerged from the 1989 data was that some respondents had difficulty choosing the correct range. Sometimes there was direct confirmation of this problem from marginal notes, but more often the problem was detected during mechanical screening of the data for inconsistent or otherwise very anomalous values. The most serious practical problem seemed to be that respondents had difficulty understanding how many zeroes were associated with different orders of magnitude. In addition, it appeared upon further analysis that the ranges allowed were also too broad relative to the distribution of many of the variables for which they were used. For the 1992 survey, the number of ranges was expanded and the card was organized in a way intended to help respondents avoid errors in their choice of ranges.[11]

In the event, the use of ranges actually declined in 1992. My suspicion is that a seemingly innocuous decision to bind the card that contained the ranges with the other interviewer showcards made the interviewers less likely to use the card. In addition, it appeared that the reformatting of the card had done nothing noticeable to the prevalence of seemingly incorrect range responses.

Because of the complexity of the SCF interview, it had long been apparent that the survey should migrate to CAPI as soon as the software became adequate for such a large survey. In 1995, we decided to make this transition, and encouraged by the efforts of the Health and Retirement Survey to elicit ranges in a systematic way (see Juster and Smith [1996]), we designed a special routine ("DKDOL") intended to capture a variety of types of partial information. Because this procedure is so important to the analysis that follows, I will devote some time to describing it here in detail.

For each of 479 dollar variables in the SCF, the interviewer and respondent had

---

[11]In 1992, the range intervals were redesigned to be more choices in ranges appropriate for more variables. As a result, 20 ranges were created going from "zero to $100″ to "more than $100,000,000."

several options (see figure 1).[12] The ideal response was a complete dollar response. In this case, the interviewer typed in a string of numbers, and the laptop computer returned a screen with the amount written out in words, along with a request to the interviewer to confirm that this is what she meant to enter. A respondent who answered either “don't know” or “refuse” (hereafter noted as DK and REF, respectively) was asked to give a range from a range card (see figure 2).[13] If the respondent agreed to use the card, the interviewer was presented with a screen on which to enter the letter selected from the card.[14] Respondents who refused at this point went to the next question, while respondents who could not give a letter range or who answered DK were then confronted with a decision tree (see figure 3) designed to put respondents into a range. Eight sets of ranges were developed using information from the 1992 SCF to cover the expected outcomes, with particular attention to the upper tail of the distributions. Generally, the entry point in the range sequence was selected to contain the median value from the previous survey. Respondents could refuse to continue at any point in the decision tree. As a check, whatever partial information the interviewer obtained in the decision tree was summarized in words and presented to the interviewer before the program proceeded.

Finally, to allow for respondents who preferred to provide their own ranges, and for those who had used the range card for earlier questions and preferred to continue to do so, the program incorporated a section for reporting “volunteered” ranges. As noted earlier, volunteered ranges were believe to have been important in earlier SCFs. To use this option,

---

[12]There are are a few other other dollar response fields in the survey for which ranges were not collected. Because of the questionnaire skip sequences, no respondent could possibly answer all of the dollar questions.

[13]The range card was also reformatted with values larger than $999,999 being written with the word “million,” rather than writing out the appropriate number of zeroes. The hope (dashed again, alas) was that this would make it more difficult for respondents to choose incorrect ranges.

[14]In the case of these letter ranges and the other letter ranges discussed below, the computer did not return a confirmation screen. We included such screens in the survey pretest, but the interviewers protested that respondents used the range card because the values seemed somewhat more “confidential,” and when they noticed that the interviewers got a translation on the screen, they reportedly felt betrayed.

the interviewer pressed a special function key. This action generated a screen which offered a choice between entering upper and lower bound dollar figures, or a letter from the range card. The dollar bounding was set up to accept both closed-interval ranges and such open-ended responses as "more than a million dollars." The screen that accepted the range card value was also set up to accommodate verbatim responses, most of which it turned out could have been entered as upper and lower bound data.[15]

[15]This verbatim field was also intended to handle negative ranges.

Figure 3: Example of a decision tree in the 1995 SCF.

R1. Was it $250,000 or more?			
YES	DK	REF	NO
▼	▼	*out*	▼

R2. Was it $500,000 or more?				R5. Was it $50,000 or more?			
YES	DK	REF	NO	YES	DK	REF	NO
▼	*out*	*out*	*out*	▼	*out*	*out*	*out*
R3. Was it $1,000,000 or more?				R6. Was it $100,000 or more?			
YES	DK	REF	NO	YES	DK	REF	NO
▼	*out*	*out*	*out*	▼	*out*	*out*	*out*
R4. Was it $5,000,000 or more?				R7. Was it $150,000 or more?			
YES	DK	REF	NO	YES	DK	REF	NO
out	*out*	*out*	*out*	*out*	*out*	*out*	*out*

Ranges yielded:

>$50,000	$50,000 to $100,000
>$150,000	$50,000 to $250,000
>$250,000	$100,000 to $150,000
>$500,000	$100,000 to $250,000
>$1,000,000	$150,000 to $250,000
>$5,000,000	$250,000 to $500,000
<$50,000	$500,000 to $1,000,000
<$250,000	$1,000,000 to $5,000,000

Actual set of ranges:

R1	R2	R3	R4	R5	R6	R7
$500K	$1M	$5M	$10M	$10K	$100K	$250K
$11M	$5M	$10M	$25M	$50K	$100K	$500K
$250K	$500K	$1M	$5M	$50K	$100K	$150K
$100K	$250	$500K	$1M	$5K	$25K	$50K
$50K	$100K	$250K	$750K	$5K	$10K	$25K
$500	$1K	$5K	$10K	$25K	$75K	$250K
$1K	$2K	$10K	$50K	$100	$250	$500
$50	$100	$250	$500	$1K	$5K	$10K

Since the 1989 SCF, we have tried to retain as much information as possible about the original state of variables and transformations applied to them. This information is contained in a set of "shadow variables" that parallel all of the main survey variables. For the 1995 SCF, we defined a large number of codes for each such varaible to track the initial response given by respondents (complete response, volunteered ranges of the two types, DK, and REF), and the large array of secondary responses (paths through the eight classes of decision trees, letter responses from the range card, and upper and lower bound data). This information is retained primarily for use during imputation, but such careful record keeping also imposes a very useful rigor and clarity on data processing. These shadow variables form the basis of a large part of the analysis reported here.

Because of the nature of the findings in this research, it is useful to say a little about the information that interviewers were given about the collection and use of range information. I participated very actively in the design and execution of the interviewer training for the 1995 survey. During a project overview talk I gave to interviewers after they had been through a day and a half of practical exercises with the survey including experience with range information, I made a particular point of focusing on the collection of range data. An appendix contains a copy of the written outline of that part of the talk, which was given to the interviewers with their training materials. The points stressed in the presentation were: (1) complete responses are preferred to range responses, (2) range responses may be legitimate answers for items that vary in value over time or where there is no ready market, (3) range information is strongly preferred to no information when the respondent is unwilling to provide complete information.

During training, interviewers expressed some initial resistance to the decision tree, but they appeared to become more comfortable as they realized they could exit the question sequence by entering a refusal code. During the field period, I queried interviewers and field managers several times on how the process was working, but detected no problems. However, once we began to receive the preliminary data (beginning in August 1995, about a month into the field period) we noticed a suspicious use of the number "1" in some dollar fields by a group of interviewers. Having identified this problem, we contacted the interviewers. Apparently, there had been a misunderstanding about how to enter data in the case of strong respondent refusals. A few interviewers were entering a "1" in dollar fields to indicate a strong refusal under the assumption that the value entered would be "obviously" incorrect to us in processing. After these errors were resolved, subsequent tracking of the process indicated no other obvious systematic problems. After the completion of the field work, we held a comprehensive project debriefing, one important agenda item was the performance of the part of the CAPI program that collected range data. There were two major complaints. Most importantly, the computers used for interviewing (386 machines) processed the range data very slowly owing to a quirk of the version of Surveycraft in which the program was written. It also appeared that some interviewers felt that the range questions pushed some respondents too far.

III. SCF Item Response Rates from 1983 to 1995

Tables 2 through 4 present information on item nonresponse rates for a set of variables from the SCF for the period 1983 to 1995. The variables are intended to cover a wide range of the types of data collected in the survey for which the data in the 1983 survey are roughly comparable with the later data. The items are shown in the order the underlying questions were asked in the surveys from 1989 forward. The tables differ in their treatment of the set of completed cases. Table 2 displays unweighted data for the full sample in each survey year, table 3 looks at only the unweighted area-probability samples, and table 4 looks at the weighted full sample. Table 2 is included as basic documentation of the SCF data, table 3 is included for ease of comparison with other surveys, and table 4 is given for comparisons over time that are unaffected by the large changes in the size of the list sample.

Overall, the same story emerges from all three of these tables. Item response rates moved inconsistently between 1983 and 1989. Generally, the questions where there were large improvements were ones where the question was rewritten in 1989 in light of serious data problems (e.g., business values). Since 1989, item response rates have deteriorated sharply for most of the items shown.[16] This decline in item response may reflect a tradeoff between unit and item nonresponse. Although there has been a continuing rise in interviewer efforts to maintain approximately constant unit nonresponse rate, the additional effort may yield respondents who on the margin are less cooperative. Evidence from other surveys would be useful on this point.

The record of range data in 1989 and 1992 suggests that such information was a small,

[16]The improvement in the reporting of the cash value of whole life insurance probably reflects a change in the question wording to improve the ability of respondents to choose between reporting term and whole life insurance.

but important source of information for as much as a few percent of the respondents who had certain items. For total income, the questionnaire incorporated a decision tree follow-up for respondents who answered DK or REF and who would not agree to use the range card. The decision tree sequence provided range information for 7.9 percent of the sample in 1989 and 9.8 percent in 1992. Both the DK and other missing data (REF and a small number of other types of missing data including mainly interviewer errors) were also lowered for this question. Although the rate of complete responses also went down, in light of other movements in response rates, it was not obvious that this movement had anything to do with the introduction of the follow-up questions.

Although there is some variability in the use of ranges over the different variables shown for the 1995 survey, some patterns seem clear. First, complete responses declined sharply in some cases. Second, as might be expected, the proportion of DK responses also declined substantially. Third, other types of missing values moved inconsistently, with some large declines, some large increases, and some rates nearly unchanged. Fourth, the use of the range card went up, generally by a very substantial amount. Fifth, the decision tree ranges provided information on about the same scale as the range card data in 1992. Finally, respondent-provided dollar ranges generally appear to be little used except in the case of business and stock values.

Overall, the 1995 patterns suggest that a part of the population that may have been complete reporters or DK respondents in 1992 became range value reporters in 1995. The decline in DK responses suggests that genuine uncertainty, and possibly some privacy concerns, were well addressed though ranges. Behaviorally, the conversion of complete reporters to range reporters would also be easy to understand. Interviewers are faced with the very difficult problem of extracting information on the value of sensitive items, and it is well-known that some respondents may become hostile when interviewers probe for dollar values. Furthermore, although SCF interviewers are generally highly motivated, in 1995 they faced a compensation system that gave positive rewards for completed cases, some limited punishment for very high rates of missing data, but gave no differential disincentive for collecting high fractions of value information as ranges.

Historically, the SCF has trained interviewers to probe for single dollar amounts rather than accept a DK or REF. There is ample evidence from margin notes in past surveys done on paper that interviewers probed for respondents' best guesses for items where they were unsure of an amount. Some evidence also exists for a comparable treatment of refusals, though this information is largely from conversations with interviewers and from following behavior during training. The 1995 SCF CAPI program made a fundamental change in the nature of the interviewers' engagement with the questionnaire and the respondent. The program forced the interviewers to ask every applicable question (though interviewers have found paths around all manner of a priori seemingly impassable barriers), and interviewers were very much aware that the program also enforced a form of structured probing for item nonresponse on value questions. From an interviewer's perspective this routine could have a mixture of effects. An interviewer who might otherwise have probed could be assured that even by acting passively, the computer would automatically generate at least the first level of probes that an interviewer would have been expected to do in the past. In relying on computer-generated probes, the interviewer could have deflected the stress of the questioning to the necessity of asking the questions the computer presented—and we have often encouraged interviewers in training to "blame it on us" when an interview gets difficult.

To get more deeply at the behavior that underlies the response patterns in 1995, tables 5, 6, and 7 array the final types of responses for the variables in tables 2, 3, and 4 by the respondents' initial responses. Here the data show a very much higher rate of DK responses than in 1992. Of these DK responses, about half were resolved into ranges, with those ranges about equally divided between range card responses and decision tree choices. This finding lends strong support to the hypothesis about the CAPI-induced changes in interviewer behavior. The conversion rate for refusals is relatively low—overall, about 15 percent. The figures also show a very high use of volunteered ranges, with the largest proportion attributable to the use of the range card. Respondents would not automatically be aware of the existence of the range card, so interviewers must have used it at least initially as a type of probing instrument. If this is the case, then interviewers would also have resolved fewer probes into single values.

There is some limited information to be brought to bear to analyze these patterns further. We track the sequence of completed cases for each interviewer. In addition, we collected some information from interviewers on their attitudes and characteristics before they began work, as a part of another research project we are conducting on interviewer behavior. In table 8, I present some models of reporting of different types of information, using the following set of explanatory variables: the log of the number of dollar questions the respondent was asked; the age of the household head; a set of indicators equal to one if the case was the interviewer's first case, second through fifth case, sixth through tenth case, or greater than the tenth case, and otherwise equal to zero; a variable on a scale of one to six indicating the strength with which the interviewer agreed that respondents in general were unlikely to answer financial questions; and a variable on a scale of one to six indicating the strength of the interviewer's discomfort with the idea of asking financial questions. A second model is presented for each dependent variable including the log of total household income.

Not surprisingly, the use of ranges at all is positively associated with the number of questions on which such responses could be given. Ranges were less likely to be used later in interviewers' production, though this could reflect the performance of a relatively small number of interviewers who had very high production, and who were often assigned the most difficult cases. Interviewers who either experienced personal discomfort in asking financial questions, or who expected discomfort in the respondent were significantly more likely to accept ranges at all, though their proportion of range responses appears no different than that of other interviewers. However, interviewers who were themselves uncomfortable tended to accept a higher proportion of completely missing data.

Given that interviewers accepted ranges in a particular interview, the data suggest that they were more likely to record a type of voluntary range (recall that these are overwhelmingly entries from the range card) in their later interviews, or if they were uncomfortable about asking financial questions. Even more interestingly, this result also holds for the first range response a respondent gave. The results make sense in light of the fact that interviewers who offered the range card directly were able to bypass the computer-directed offering of the range card and the decision tree, a move that could save both time and stress. The immediate offering of the range card suggests that interviewers viewed the range card as a replacement for more detailed probing to "negotiate" a single value with the respondent, an action that would tend to lower the proportion of complete responses.

Fig. 4: Distribution of fraction of elegible dollar questions answered with ranges, for those giving at least one range response

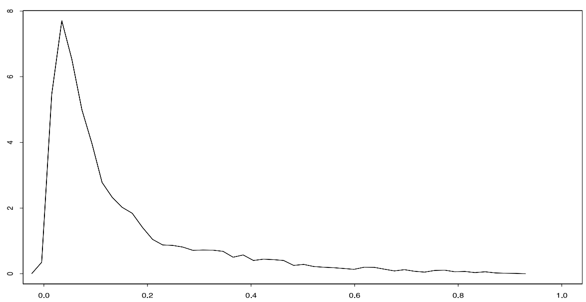
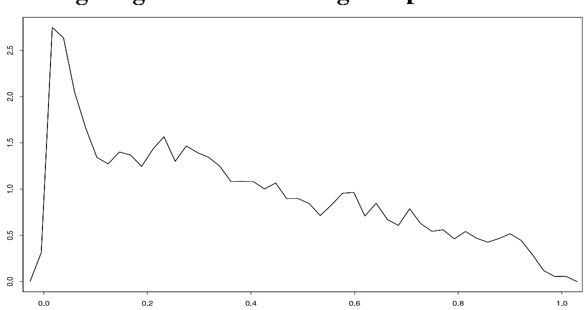


Fig. 5: Distribution of fraction of dollar questions elapsed until first range response, for those giving at least one range response



Respondents varied widely in their use of ranges. As shown in table 9, the median respondent in the full sample (unweighted) gave almost 17 percent of their applicable dollar responses as ranges; the figure for the area-probability sample was about 5 percent. However, 10 percent of the full sample (unweighted) gave over 69 percent of such responses as ranges. The skewness of the distribution is obvious from the kernel density plot of this distribution given in figure 4 for the 73.2 percent of the full sample that reported at least one range.

For respondents who gave at least one range responses, figure 5 shows a kernel density plot of the fraction of applicable dollar questions elapsed until the first range response is given. There is an initial spike in the distribution, followed by a gradual decline. Thus, there appears to be no universal trigger in the questionnaire that caused respondents to begin the use of ranges.

Table 10 gives an indication of the patterns of range use over the questionnaire. At this simple level, the data indicate a moderate degree of persistence of range use. Probit modeling (see table 11) confirms this persistence, even when I control for the number of questions asked in the section and the number of questions asked in the entire interview. Interestingly, the data in table 10 also show signs of an increase in the propensity to use ranges as the interview progresses: in the full sample, 27.6 percent of respondents used at least one range in the first part of the interview (credit cards, housing, and lines of credit), and the proportion rises monotonically to 54.5 percent who used ranges in the last section (employment, pensions, income, and inheritances). Although this trend could be subject matter driven (traditionally, respondents have had only weak knowledge about their pension, and it is well-known that income is among the most sensitive of questions), or it could reflect growing respondent suspicion or fatigue as the interview progresses.

There is also persistence in respondents' use of a given type of range response. For the unweighted full sample, the area-probability sample unweighted, and the weighted full sample, tables 12 through 14 give the distribution of responses for the second and third range responses arrayed by the first type of range response used. Overall, the largest change seems to be a migration to volunteering a range from the card from the other types of ranges. This result reinforces the earlier results suggesting that respondents learn that volunteering a range from the card is the easiest outcome short of giving a complete response.[17]

IV. Effects of Range Data on Overall Data Quality

To this point, I have tried to provide a picture of the mechanism that may lie behind the use of ranges in the 1995 SCF. Ultimately, the most important statistical question in most cases is whether the information gained by using ranges adds significantly to the ability to use the survey data for its intended purposes. Two key questions in this line are (1) whether the variance of key estimates is substantially reduced by the introduction of ranges, and (2) whether there is a nonignorable response process (see Little [1983]) that would induce bias in imputation and other estimates in the absence of the true data. If the mechanism introduced to collect the range data may is not neutral, as I believe the data suggest for the 1995 SCF, there may also be a trade-off between the gains from respondents who provided ranges who would otherwise not have given any information, and the loss of efficiency from converting complete responses to range responses.

One simple, and possibly misleading indicator, of the differences between full reporters and range reporters is a comparison of the univariate distribution of the values of the survey variables. For the unweighted full and area-probability samples respectively, tables 15 and 16 report the median, 75th percentile and mean of the variables in tables 2-4 by the type

---

[17]The volunteered card range may have been the fastest route in some cases. The routine that translated dollar amounts of single numbers, decision tree responses, or volunteered dollar ranges into words for the confirmation screen was surprisingly computationally intensive. However, as noted earlier, responses from the range card bypassed the confirmation screen.

of final type of non-missing response. Cells with 3 or fewer observations are surpressed. There is tremendous variability in the relative shapes of these distributions. One crude way of summarizing the data is to compute the weighted distribution of outcomes relative to the complete reporters, where the weights are the number of cases giving a particular type of response. The bottom of the table gives the median, mean and standard deviation of this relative measure. The results are similar in form for the full and area-probability samples, but they are stronger for the former. Initial DK and REF cases yield outcomes that are not very different in terms of mean or median outcomes. For the full sample, but not the area-probability sample, the mean values are larger than the value for full reporters, but given the standard deviation of this estimate, it is unlikely that the difference is significant. The difference in the two samples very likely reflects the greater overall likelihood of list cases having a larger number of applicable dollar questions and larger underlying values of the items. There are other ways of aggregating the data—e.g., on the basis of the contribution to an estimate of aggregate net worth—that might give a different impression.

A better way of evaluating the important distributional differences is to control for systematic differences between the different types of respondents. One straightforward, though complex, way of doing this is to impute the data both with and without the use of ranges and compare the two distributions. Beginning with the 1989 SCF, missing data have been imputed using the FRITZ model, an iterative process employing techniques of Gibbs sampling and multiple imputation (see Kennickell [1991]). Normally, range information is used in the imputations to truncate the conditional distributions from which the imputations are drawn.

At the time this paper was written, the 1995 data were still being actively processed. Because the complete imputation process is very time-consuming, it was not possible not possible to create comparable final imputations for the complete dataset. For purposes of this paper, I ran the part of the first iteration of the part of the FRITZ model that imputes financial assets and total income. Because the first iteration is intended only to provide starting values for the iterative model, it entails only single imputation. For the full sample, table 17 shows characteristics of the distribution of the imputations using the range data and comparable figures for the imputations made with the range data completely suppressed.

Table 18 shows the same information for the area-probability sample alone[18]

As one possible summary of these relationships, I computed the weighted mean, median, and standard deviation of the values for the imputations without range constraints relative to those that used the ranges. For the full sample, ignoring the range information barely alters the mean outcome overall. However, as the median estimate of the relative mean (0.33) suggests, this distribution has some odd tail behavior. In fact, the mean of the savings bond imputations without ranges is 9 times that without ranges. Deleting the savings bonds from the aggregation suggests that ignoring the ranges hurts relatively little at the bottom of the distribution, but may be more of a problem at the top. Results for the area-probability sample alone suggest that the differences there are less strong. It would be useful to be able to compute some sort of significance test for the differences incorporating a measure of imputation variance. This is not possible at this time, but I will return to it later as time permits. One should keep in mind that the aggregate I use here is entirely arbitrary. It may not be either the most natural or meaningful one for all purposes.

Quantile-quantile (Q-Q) plots may be a more transparent device for gauging the distortions induced by ignoring the range data in imputation.[19] Figures 6 through 16 are unweighted quantile-quantile plots for the full sample of all the variables in table 17. Generally, the plots differ most at the top of the distribution, with a tendency for the distribution using ranges to be more top-heavy. Two exceptions are certificates of deposits and trusts and annuities. For certificates of deposit, the distribution of the imputations without ranges is generally above those made using the range data until the top three observations. For trusts and annuities, the imputations that do not incorporate range data generally understate the imputations made using the ranges. Given the very small number of cases in upper tails, it is hard to gauge the importance of the differences between the two

---

[18]The results in table 18 for the area-probability sample are not independent of the results in table 17. In both cases, the underlying moment matrices that underlie the imputations were computed using data for the full sample. For table 18, I merely subsetted the area-probability cases.

[19]Q-Q plots are graphs of the quantiles of distributions plotted against each other. If the figure lies on the 45 degree line, they are the same distribution.

distributions in each of these plots.

The differences become much less pronounced in the context of the entire distribution of real and imputed values for these variables, as shown in figures 17 through 28. Aggregated to the level of total financial assets (figure 27), the differences become even smaller, probably because of offsetting errors in the component imputations. For total financial assets, it appears that there is a slight tendency to overstate the amount of financial assets until about the top 25 cases in the data, where range responses appear to be particularly important.

Much of the research done using the SCF leans heavily on the sort of partial correlations obtained from regressions and related modeling. To address the importance of the informational gains from range data for this purposes, table 19 presents the results of a set of regressions of the log of total household income on a set of dummy variables for ownership of various financial assets, the log of the maximum of one and the value of the asset, and the log of the age of the household reference person. This model is selected only as an example, and it has no particular importance for any economic theory. I estimated the model on both sets of imputations using OLS, and following the common current practice in economics, I also ran it using a robust regression routine available in Stata. Overall, one would expect the fit on the unbounded data to be noisier, and this is confirmed by the R^2 of the OLS regression, which is three percentage points lower with the unbounded data. For the variables judged significant by the customary 95 percent confidence standard, there were no changes of sign between the different datasets, though a couple of variables were judged significant with the range data, but not with the unbounded data (the dummy variable for ownership of "other bonds" in the OLS model, and the log of savings accounts holdings in the robust model). In almost all cases, the pairs of coefficients lie within the regression confidence interval (estimated without accounting for design effects).

VI. Summary and Conclusion

If the alternative to collecting partial information on value variables as range data is to collect no information at all, collecting range data is the statistically dominant strategy. However, as indicated by data reported in this paper, the tradeoff is more subtle. It seems clear that many responses that would otherwise have been coded DK were resolved as ranges.

However, the data also suggest that some interviewers may have collected range information when it might have been possible to probe a respondent for a single value—tough even that value may have been an estimate made by the respondent. The data also suggest that there are complex interactions effects that determine the types of ranges that interviewers and respondents negotiate. It may be possible to improve data collection by making changes in interviewer recruiting and training.

One justification for obtaining range data that is frequently heard is that ranges ought to mitigate the effects of nonignorable item nonresponse. Although the results presented here tend to support the collection of range data as an efficiency-improving measure, there is not strong evidence that imputations would otherwise be biased.

Collection of range data comes at a price in terms of respondent burden, and possibly higher unit nonresponse (or higher costs per interview). Because the SCF contains so many dollar questions, it may be seen as an extreme case. However, many surveys operate near the margin of respondents' tolerance and systematic probing on sensitive items, such as income, may have disproportionately deleterious effects. Evidence in this paper suggests that one could do nearly as well using a simple range card as a probing tool as using the complex range apparatus of the SCF. An important qualification is that interviewers must see the use of the card as important.

The results presented in this paper differ from the experiences of the Health and Retirement Survey, particularly in the assessment of the ability of range data systematically to reduce bias. Clearly, more analysis of both the SCF and HRS data is warranted. Work should also be started on the interaction of interviewers and respondents to gain a better understanding of the underlying cognitive processes that generate the range data.

Bibliography

Frankel, M.A. and A.B. Kennickell [1995] “Toward an Optimal Stratification Paradigm for the Survey of Consumer Finances,” Proceedings of the Section on Survey Research Methods.

Heeringa, S.G., J.H. Connor, and R.L. Woodburn [1994] “The 1989 Surveys of Consumer Finances Sample Design and Weighting Documentation” mimeo, Survey Research Center, University of Michigan.

Heeringa, S.G., D.H. Hill and D.A. Howell [1995] “Unfolding Brackets for Reducing Item Nonresponse in Economic Surveys,” Health and Retirement Study Working Paper Series, Paper No. 94-029.

Juster, F.T. and J.P. Smith [1996] “Improving the Quality of Economic Data: Lessons from the HRS and AHEAD,” mimeo, Survey Research Center, University of Michigan.

Kennickell, A.B. [1991] “Imputation of the 1989 Survey of Consumer Finances," Proceedings of the Section on Survey Research Methods.

_____ and D.A. McManus [1993] “Sampling for Households Financial Characteristics Using Frame Information on Past Income, Proceedings of the Section on Survey Research Methods.

_____, _____, and R.L. Woodburn [1996] “Weighting Design for the 1992 Survey of Consumer Finances,” mimeo, Bpard of Governors of the Federal Reserve System.

_____ & M. Starr-McCluer [1994] "Changes in Family Finances from 1989 to 1992," *Federal Reserve Bulletin* (October), pp. 861-882.

Little, R.J.A. [1983] “The Nonignorable Case” in *Incomplete Data in Sample Surveys*, Academic Press.

Projector, D.S. and G.S. Weiss [1966] “Survey of Financial Characteristics of Consumers,” Board of Governors of the Federal Reserve System.

Appendix: Interviewer Training Material on DKDOL

a. Dollar amounts are a very important part of the SCF: we have designed a new routine ("DKDOL" will appear in the upper left-hand corner of the screen) to make it easier (we hope) to record figures correctly.
   i. Every time you enter a dollar answer, the CAPI program will present you with a screen with the amount you have entered written out in words.
      (1) If you enter a negative number on a screen where this is allowed, the program will confirm that number as a negative number (e.g., "NEGATIVE THREE THOUSAND").
   ii. Helps you to catch entry errors, particularly when typing very large numbers, and will help you catch cases where the R reports only a part of a number (e.g., "my house is worth 200", but R means $200,000).

b. We always prefer to have the R give the exact answer to every dollar question, but this is not always possible because the R may genuinely not know the answer, or the R may be hesitant to answer.
   i. However, experience shows that very many Rs who cannot be persuaded to give a direct dollar response are actually able and willing to give some other sorts of information. Some Rs don't realize that even partial information or range data can be very helpful to us when we analyze the data.
      (1) However, we need to be careful that we don't wind up encouraging the R to give imprecise answers everywhere!
   ii. Because Rs differ in the type of additional information they are able or willing to give, we have built in several options for recording data that we hope will fit as naturally as possible with the types of answers Rs give ("DKDOL" again will appear in the upper left-hand corner of the screen). (See figure)
   iii. Sometimes when the R doesn't know the exact answer or the exact answer may not be easily knowable (for example, the value of R's house might not be known without a formal appraisal), we know from past experience that it is possible to get some information. But the best thing is always to record an exact dollar figure.
   iv. Sometimes the R does not know or does not want to share an exact value with you, but the R may be willing to give a range answer. For this reason, we have built two types of range responses into the CAPI program. To access these ranges directly, you enter [F9] in a dollar field, and you will get a screen asking you whether you want to enter one of the two types of ranges.
      (1) One response is that the R is willing to give a range from the range card that is in the booklet with the showcards. (See figure)

(a) Past experience tells us that a large number of Rs can provide at least this much information.

(b) Use of a letter also allows Rs who are embarrassed about the size of the true number (too large or too small) to avoid saying the number.

(2) Some Rs will actually volunteer (or you may easily elicit) a dollar range, such as "150,000 to 200,000" or "in the low 10 thousands." Such responses may be particularly appropriate in the SCF for assets that vary in value.

(a) For the "150,000 to 200,000" response you would enter: LOW END OF RANGE=$150,000 and UPPER END OF RANGE=$200,000.

(b) The "in the low 10 thousands" response could be entered as LOW END OF RANGE=$10,000 and UPPER END OF RANGE=$50,000. It is appropriate to confirm such decisions with the R.

v. Alternatively, the R might respond DK/Ref to a dollar question and not be persuaded to give an exact answer and not give a range at that point.

(1) The program will present a screen asking whether R can give an answer from a range card using a letter to identify the range. This is the same range field that you can access more directly with [F9].

(a) This is a chance to prompt Rs to think about trying the range card.

(b) If R refuses at this point, the program will skip to the next question.

(2) If R says that he/she cannot give a letter from the card, or does not know which letter would apply, the program will go through a "decision tree." (See figure)

(a) For example: "Is the amount more than $30,000?" If NO "Is it more than $5,000?" Etc.

(b) If R refuses at any point in the decision tree, the program will skip to the next question so that you do not have to badger the R.

(c) The amounts are tailored to each question using data from earlier surveys.

(i) These ranges are pretty broad, but they still provide very valuable information because they tell us where the R fits into the overall distribution of an item.

vi. Again, the idea of these tools in CAPI is to help you record the sorts of responses past interviewers have told us we can expect.

(1) This sequence may seem complicated when you first see it, but we have tried to make the sequences natural for you (with the

help of feedback from other interviewers), so we expect you will get accustomed to it quickly.

vii. As in the case in which the R gives a single dollar figure, when the R volunteers a dollar range or answers questions in the decision you will be given a screen to confirm the data.

(1) The difference from the confirmation screen for single dollar responses is the number confirmed will be a figure somewhere in the range that the R gave.

(2) You will not be given a confirmation screen if the R gives a letter from the card because you will usually not be able to look at the card to see if the figure makes sense. Some Rs may view the actual number as private and get upset if you try to look.