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

Arthur B. Kennickell, Board of Governors of the Federal Reserve System
Federal Reserve, Mail Stop 180, Washington, DC 20551; email akennickell@frb.gov

Key words: Range data, nonresponse, CAPI, imputation, ignorability

Survey respondents may fail to provide complete information for a number of reasons. Questions may appear overly intrusive to some respondents, and they may refuse to answer. There may be respondents who could, in principle, know the exact answer to a question, but who do not know the answer and cannot be persuaded either to take action to uncover the value or to make an estimate of it. There may also be cases where the exact answer is not clear even in principle. The concept being probed may be ambiguous or multifaceted from some perspectives. It could also be that the only way to determine the answer to a question is to take an extraordinary action, such as selling an asset

Missing data raise two critical problems: diminished efficiency in estimates that depend on the data, and the possibility of bias through nonignorable nonresponse (Little [1984]). Collection of range data may provide important partial information to lessen these problems.[1] However, such information must be weighed against substantially higher respondent burden in many cases, and decreased interviewer flexibility in awkward situations. The introduction of probing formally structured to elicit ranges 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.

The Survey of Consumer Finances (SCF) collects dollar values of a wide variety of assets, liabilities, payments, incomes, and other items, and missing data rates are substantial for many variables. Range information has been an important part of the data collected since the survey began. The redesign of the 1995 SCF for 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 where there is genuine uncertainty about the value, (3) confront every "don't know" (DK) or "refuse" (REF) response with a request to use a range card or to go through a dollar decision tree to educe a bounding range.

I. Background on the SCF

The SCF has been conducted on a triennial basis since 1983 by the Board of Governors of the Federal Reserve System in cooperation with Statistics of Income (SOI) at the Internal Revenue Service. 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 (Kennickell and Starr-McCluer [1994]).

The SCF uses 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. There have been substantial variations over time in the size and composition of the sample. Unit response rates have varied little over the life of the survey. 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.

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, and a range card was available for the interviewers to use in probing. When the survey was restructured in 1989, the range card was revised, and a systematic effort was made to record ranges in a documented way for use in imputation. Interviewers were told to use the card with respondents who would otherwise be unwilling to give a response. In addition, a "decision tree"—a series of questions designed to bound a partial response—was added to for the key question on total income, traditionally a particularly sensitive question for respondents.

Judging from the number of data irregularities, it appears that some respondents had difficulty choosing ranges in 1989. The most serious practical problem seemed to be that respondents had difficulty with the number of zeroes 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 be more clear in their choice of ranges. In the event, the use of ranges actually declined in 1992, probably as a result of a seemingly innocuous decision about the arrangement of interviewers'

materials. In addition, it appeared that the reformatting of the card did little to lessen reporting errors.

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 we created a subroutine ("DKDOL") intended to capture a variety of types of partial information. For each of 479 dollar variables in the 1995 SCF, the interviewer and respondent had several options.[2] 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 DK or REF was asked to give a range from a reformatted range card. 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.[3] Respondents who refused at this point went on to the next question, while respondents who could not give a letter range or who answered DK were then confronted with a decision tree designed to select a range that contained the true value. Eight sets of ranges were developed using information from the 1992 SCF to cover the range of expected outcomes, with particular attention to the upper tail of the distributions. Respondents could refuse to continue at any point during this questioning. Whatever partial information the interviewer obtained in the decision tree was summarized in words and presented to the interviewer.

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. To use this option, 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 screen was set up to accept such responses as "more than a million dollars" as a lower bound with a missing upper bound.

Because of the nature of the findings in this paper, it is useful to comment on interviewer training. The DKDOL procedure was extensively demonstrated through mock interviews, and after a day and a half or training, I gave the interviewers a project overview talk in which I stressed (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. Interviewers expressed some initial resistance to the decision tree, but they appeared to become more comfortable by the end of training as they realized they could exit the question sequence by entering a refusal code. A minor problem appeared early in the field period with a small number of interviewers not understanding how to exit the range routine correctly. Once this problem was corrected, there were no other such problems during the field period. After the field period, we held a comprehensive project debriefing, where the collection of range data was discussed. There were two major complaints. Most importantly, the computers (386 machines) processed the range data very slowly, in part owing to a property of the version of Surveycraft in which the program was written. Some interviewers felt that the range questions pushed respondents too far.

III. SCF Item Response Rates from 1983 to 1995

To provide context for the results on ranges in 1995, table 1 presents data on item nonresponse rates for a set of SCF variables for the period 1989 to 1995. To abstract from changes in sample composition over this

Table 1: Item response rates over time, AP Sample, Unweighted

	House	Chckng	Stocks	Income
Have item?				
Yes				
1989	55.2	80.5	16.6	100.0
1992	54.6	82.1	15.3	100.0
1995	56.8	84.2	14.9	100.0
Unknown				
1989	0.0	0.0	0.2	0.0
1992	0.0	0.2	0.3	0.0
1995	0.0	0.3	0.6	0.0
Final value				
Number				
1989	96.0	89.3	78.2	82.7
1992	93.9	86.9	71.7	78.2
1995	88.6	80.2	64.6	72.8
Dec. tree				
1989	NA	NA	NA	8.8
1992	NA	NA	NA	8.6
1995	1.1	1.9	3.6	1.4
Range card				
1989	0.6	3.3	7.4	3.7
1992	0.7	1.6	1.9	3.0
1995	7.5	10.5	14.2	15.7
$ range				
1989	NA	NA	NA	NA
1992	NA	NA	NA	NA
1995	0.8	0.5	3.1	0.1
DK				
1989	2.4	1.3	8.0	0.2
1992	4.5	3.7	20.3	2.0
1995	0.1	0.4	1.7	0.2
Oth. miss				
1989	1.0	6.1	6.4	4.5
1992	1.0	7.8	6.1	6.0
1995	1.9	6.4	12.8	9.7

period the table contains data for the area-probability sample only. Since 1989, item response rates have deteriorated sharply for most of the items shown. This decline may reflect a tradeoff between unit and item nonresponse: increased efforts to maintain approximately constant unit nonresponse rate 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 responses were a small, but important source of information for as much as a few percent of the respondents. The decision tree follow-up for total income provided information for nearly 9 percent of the area-probability cases in each year. 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 variables in 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 a relatively small amount of information on about the scale of 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 some part of the population that may have been complete reporters or DK respondents in 1992 were converted to range value reporters in 1995. Behaviorally, this outcome would 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 push 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 from past surveys 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, 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 doing so, 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, table 2 arrays the final types of responses for the variables in tables 1 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 are resolved into ranges, with those ranges about equally divided between range card responses and decision tree choices. This finding suggests that CAPI may have induced changes in interviewer behavior. The conversion rate for refusals is relatively low—overall, about 15 percent. The figures

Table 2: Final Incomplete Responses by Initial Response, 1995 AP sample, Unweighted

Init. resp. / Fin. resp.	House	Chckng	Stocks	Income
DK	26.1	16.6	33.3	19.2
Dec. tree	31.9	26.0	22.4	13.1
Card	36.2	32.5	20.4	38.6
Missing	31.9	41.6	57.1	48.3
All 100.0	100.0	100.0	100.0	
REF	10.6	14.5	30.2	30.6
Dec. tree	15.8	15.2	7.7	8.7
Card	5.3	7.9	7.7	1.7
Missing	78.9	77.0	84.6	89.6
All 100.0	100.0	100.0	100.0	
Vol. range	63.3	47.7	40.1	50.1
$ range	11.4	5.4	22.0	1.1
Card	88.6	94.6	78.0	98.9
All 100.0	100.0	100.0	100.0	
All 100.0	100.0	100.0	100.0	
Memo item: Percent with any inc. resp.	11.4	19.8	35.4	22.7

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 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. Using the available data, I estimated several models of range use. Not surprisingly, the use of ranges at all is positively related to the number of questions on which such 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.

Respondents varied widely in their use of ranges. 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 1 for the 73.2 percent of the full sample that reported at least one range.

Figure 2 shows a kernel density plot for those who gave at least one range response, of the fraction of applicable dollar questions elapsed until the first range response was given. There is an initial spike in the distribution, followed by a gradual decline. Thus, there appears there is a class of interviewers and respondents who quickly turn to range reporting. Otherwise there appears to be no universal trigger in the questionnaire that caused respondents to begin the use of ranges.

Simple examination of the data suggests persistence in a given respondent's use of ranges, and probit modeling confirms this finding, 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 also show signs of an increase in the propensity to use ranges as the interview progresses:

Fig. 1: Distribution of fraction of elegible dollar questions answered with ranges, for those giving at least one range response

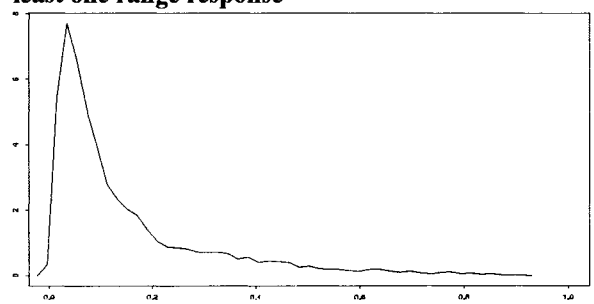
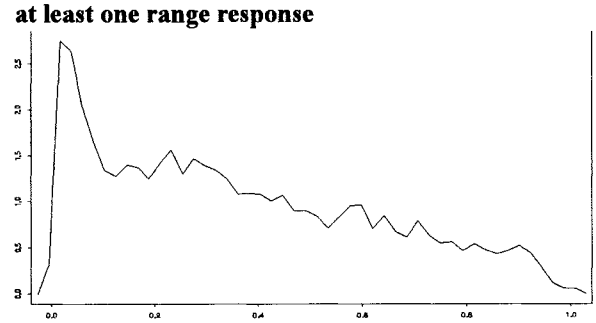


Fig. 2: Distribution of fraction of dollar questions elapsed until first range response, for those giving at least one range response



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. Overall, the largest change in range use is a tendency to migrate to volunteering a range from the card from the other types of range responses. 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.[4]

IV. Effects of Range Data on Data Quality

Ultimately, the most important statistical question here is whether the information gained by using ranges significantly reduces variance and bias. The variance gain is obvious, but the second issue is more subtle.

One simple, though 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. Aggregating over a number of variables, it appears for the area-probability sample that the distribution of values for range respondents who initially answered DK or REF lies below the distribution of complete respondents. The distribution of volunteered ranges appears substantially higher. Looking at the full sample, the whole distribution of range responses appears higher than that for the full respondents.

A better way of evaluating the important distributional differences is to control for systematic observable differences between the different types of respondents. A straightforward, though complex, way of doing this is to impute the data both with and without using the range information, and compare the two distributions. Since the 1989 SCF, missing data have been imputed using an iterative process (FRITZ) to produce multiple imputations (see Kennickell [1991]). During the process of imputation, the range data are used to truncate the conditional distributions from which the imputations are randomly drawn. At the time this paper was written, the 1995 data were still actively being processed, and it was not possible not possible to create comparable final imputations for the complete dataset. For this paper, I ran the part of the first iteration of FRITZ that imputes financial assets and total income. This yielded only a single imputation.

Quantile-quantile (Q-Q) plots are a useful device for gauging the distortions induced by ignoring the range data in imputation. Figures 3 through 6 are unweighted Q-Q plots for a selection of variables imputed under the two methods. Generally, the plots differ most at the top of the distribution, with a tendency for the range distribution to be more top-heavy. One exception is certificates of deposits, for which the distribution of the imputations made without the range data is above the distribution of the imputations made using the range dataa until the top three observations.

Fig. 3: CDs

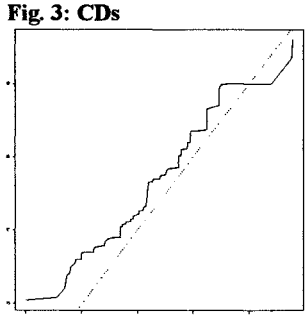


Fig. 4: Savings account

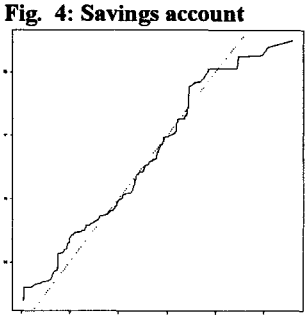


Fig. 5: Pub.-traded stock

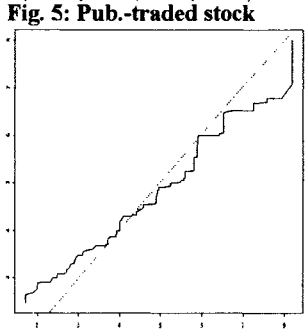
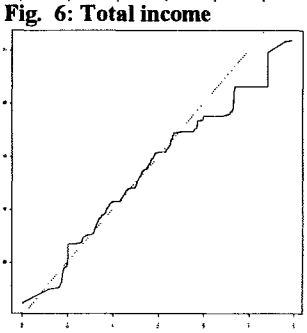


Fig. 6: Total income



Given the very small number of cases in upper tails, it is hard to gauge the importance of the differences between the two distributions in each of these plots.

The differences become much less pronounced in the context of the entire weighted distribution of real and imputed values. Aggregated to the level of total financial assets (figure 7), the differences become even smaller, probably because of offsetting errors in the component imputations. It appears that there is some tendency to overstate the amount of financial assets until about the top 25 cases in the data. The plot for income (figure 8) looks even closer to the 45 degree line. With multiple imputation, it would be possible to put confidence bands on these results.

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 informational gains from range data for this purposes, I ran 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 each asset, and the log of the age of the

Fig 7: Total fin. assets

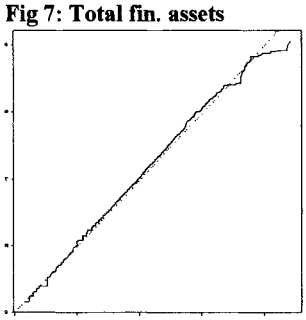
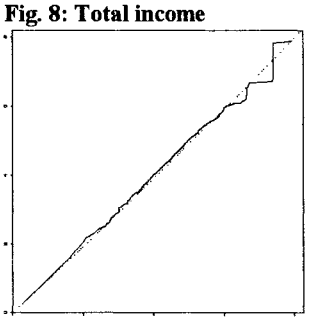


Fig. 8: Total income



household reference person. This model was 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. 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. For the variables judged significant by the customary 95 percent confidence standard (ignorinb variance attributable to sampling and imputation), 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. In almost all cases, the pairs of coefficients lie within the regression confidence interval.

VI. Future Research

Close examination of the imputations has shown some relatively weak points in the underlying modeling. I hope to examine this in more detail. As the imputation of the 1995 SCF proceeds, multiple imputation will allow me to compute estimates of the observed variablity of the two sets of imputations reported in the last section of this paper. For future interviewer training, it appears important to explain more fully the role we expect ranges to play in overall data collection. Finally, the results reported here differ substantially from those of Juster and Smith [1996] with the HRS, suggesting to me that there may be strong cognitive effects that need to be explored before we can more fully understand the collection of range data.

BIBLIOGRAPHY

Juster, F.T. and J.P. Smith [1996] "Improving the Quality of Economic Data: Lessons from the HRS and AHEAD," mimeo SRC U of Michigan.

Kennickell, A.B. [1991] "Imputation of the 1989 Survey of Consumer Finances," 1991 Proc. Sect. on Survey Research Methods.

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

_____, D.A. McManus, and R.L. Woodburn [1996] "Weighting Design for the 1992 Survey of Consumer Finances," mimeo, Federal Reserve Board.

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

ENDNOTES

A longer version of the paper is available from the author. The author thanks Val Cook and Geoff Walker who wrote the 1995 SCF CAPI program, and the many other NORC staff members who collected that data and helped to make the project a success. The author is also grateful to Gerhard Fries and Kevin Moore for help in preparing the data used here, and to Steve Heeringa and Martha Starr-McCluer for comments. 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.

1. The use of ranges to collect partial information has an interesting history. The earliest evidence I have found 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 give 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. 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 an extensive battery of decision trees (Juster and Smith [1996]).

2. Because of the questionnaire skip sequences, no respondent was asked all of these questions.

3. In the case of these letter ranges and the other letter ranges discussed below, the computer did not return a confirmation screen. Such screens were used 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.

4. The volunteered card range may have been the fastest route in some cases. The routine that translated dollar amounts into words for the confirmation screen was slow, but as noted earlier, responses from the range card bypassed the confirmation screen.