DISCUSSION

Kent H. Marquis, U.S. Bureau of the Census

We've heard two excellent papers today on the pitfalls of survey measurement, a subject on which Robert Ferber did pioneering work.

Pearl and Schuman have told us what's wrong with some of our questionnaire measures. Pearl offers some potential solutions and Schuman offers an underlying theory that might guide future questionnaire designers.

As discussant today, my goals are to highlight what I feel are especially key points in each paper and then to comment on some general methods issues that I feel will be important in extending this work.

Bob Pearl has nicely summarized some of the main measurement problems in net worth surveys. His main point is that surveys underestimate net worth; he discusses various kinds of response effects and what might be done about them.

He got our attention right-off-the-bat by showing that, in survey after survey, estimates of most asset holdings have been consistently and substantially below estimates produced from independent data. He follows up by citing record check studies that imply large omission response biases. I'd like to return to these points in a few minutes since aggregate comparisons and record checks are becoming widely used tools in survey methods research and we need to do some careful thinking about what we can learn from them.

Pearl gives first place in his presentation to problems of nonresponse. I certainly agree about the potential importance of this problem, not only for the balance sheet issues but also for household measures of flow in all surveys of consumer expenditures.

But the dimensions of the problem aren't all that clear. We do observe high nonresponse rates but there is very little information about how they translate into estimation biases. We're unsure of what causes missing data or the best ways to neutralize missing data effects on the subject matter estimates. Pearl summarizes some of the techniques available to address nonresponse and some reasons why nonresponse occurs but he is correct in stopping short of making cause-and-effect statements.

As to possible causes of nonresponse and other response errors, Pearl gives us a nice taxonomy including misunderstandings, lack of knowledge, memory failures, lack of general motivation to cooperate, and specific motives to lie.

He offers a range of design features to overcome reporting errors that includes using repeated interviews, detailed questioning, probes, questions about the attributes of the asset rather than its value, self rather than proxy responses, diaries, consulting other people, randomized response, and sealed envelope techniques. To overcome nonresponse, we have other design features such as advance letters, commitment contracts, feedback strategies, gifts and cash incentives.

I agree with Pearl that we really don't know which of these design features is best for overcoming response problems but I view the presented information as a good start: What we have is a nice taxonomy of sources of response effects and a range of practical ways to alter the data collection design to overcome them.

But it is probably never going to be possible to design a net worth survey that is problem free so statistical procedures are needed to deal with the remaining errors. In this area Pearl focuses on two procedures for handling item nonresponse, hot-deck and regression techniques. Neither is offered as a best or even good solution. The lack of known remedies poses serious problems for net worth surveys: As Pearl points out some subject matter mean estimates change by 50% or more as a result of imputation so it's important that we get our imputation procedures right.

Turning now to the Schuman and Ludwig work, Schuman described what I consider to be a very nice attempt to induce a theory to explain one part of the context effect.

Over the past few years, through Schuman and Presser's book[1] and the work of the National Academy of Sciences[2], we've become aware of the possibility that answers to questions about subjective phenomena can be influenced by the content of earlier questions. If the hypothesis is true, then we can't draw meaningful conclusions, say, about time trends in answers to a particular question if the question was imbedded in a different context each time it was asked.

Schuman reports a series of split-panel experiments that lead to a theory about one kind of context effect: The norm of evenhandedness. He treats us to the interesting detective story about how an initial plausible hypothesis about a psychological need for consistency got refined to the more specific present version. As of today, the hypothesis is that context effects operate in questions asking about the rights to act of parties in competition with each other. If we ask "Should B be allowed to do X?" we will get different answers depending on whether just the B question is asked or whether we ask about A's rights first. When both questions are asked (in AB order), the answers about B are more evenhanded in that the mean of the B answer distribution moves closer to that of the A distribution.

Schuman and Ludwig do raise questions about the practical utility of the theory. In the doctor-lawyer example, the a priori decision

that A and B are in conflict proved wrong. Future question writers will also face this ambiguity.

The recent experiment with the Japanese-American trade question raises the possibility that evenhandedness norms influence answers in the absence of context; and when this is true, the evenhandedness context effect may not operate. The possibility that a respondent's education level can predict how evenhandedness norms operate does not receive as much support as the authors would like.

The mid-level theory is promising enough to warrant further refinements. The important point here, though, is that we are taking positive steps toward a taxonomy of errors for questions about subjective phenomena and someday that taxonomy may be as extensive as the one Pearl proposes for fact questions about net worth.

Next I'd like to make some comments about methodology for future studies of survey response errors. My thinking is heavily influenced by a different context effect, namely the the climate of austerity that statistical activities and social science research now operate in. I'm also thinking of the negative technical climate surrounding survey measurement generated by National Academy of Science panels evaluating the crime survey[3] and measures of subjective phenomena[2].

My recommendation is not that we "cover up" survey errors. Such a strategy would only degrade the profession and ensure its ultimate demise in any context. What I propose is that we give more careful attention to the methods we use to evaluate surveys. And, in presenting our work, we try not to give the impression that the few isolated errors we have studied somehow pervade all kinds of survey answers.

Let me briefly mention four points about survey error detection: The rarity of general (as opposed to local) context effects; the pitfalls of aggregate comparisons; the misleading inferences that can result from record checks; and a suggestion that we use psychometric methods to minimize effects of single item response errors in measures of subjective phenomena.

Some people, who've had a casual exposure to the report on survey measures of subjective phenomena[2], are left with the impression that general context effects are pervasive in attitude surveys: For example that answers to a given question are influenced strongly by all the questions that preceeded it. I'd like to mention some work previously published by Schuman and Presser[1] on this: They describe their 1971 Detroit Area Study that used the split panel to conduct experiments with question wording. This survey also contained 113 attitude items that didn't vary and were placed after the experimental questions. On only 8 of these items were between sample differences significant at the .05 level.

The important point here is that differences weren't found on 113 items, or 100 items, or even 50 items, and there is a high probability that most or even all of the few significant differences were a chance phenomenon. I appreciate Schuman and Presser's willingness to develop this kind of analysis and to feature it prominently in their recent book. I would encourage future researchers to do likewise.

To understand the next two points about aggregate comparisons and record checks, let me just remind you that there are several sources of errors in surveys. In addition to response errors, there are potential nonobservation errors arising from inadequate frames or sample designs and from imputation and adjustment strategies.

What I want to say about aggregate comparisons, such as the one Pearl made, is that they reflect all these survey errors and all the errors in the independent estimate of the phenomenon and also the definitional differences between the two sources. It seems as if everyone who presents such comparisons concludes that discrepancies are due to survey response errors but this is a major leap of faith.

If we really want to know about characteristics of response errors, the methodology of choice is the well-designed and appropriately interpreted record check, a procedure that makes case-by-case comparisons of survey and criterion values and thus excludes the effects of many of the other errors. But I need all those qualifiers because it is very easy to be misled by record checks especially when it comes to estimating the systematic response biases that have been a principal theme in session.

I was reminded of this issue by Pearl's citation of a record check study that seemed to show that between 1/4 and 1/3 of known savings accounts were unreported in a survey. The implication is that there was a large survey response bias that caused a large downward bias of the estimated dollars in personal savings accounts. But this isn't necessarily so because errors associated with the match, but not with the reported balance in the savings account, get interpreted as omission errors in reporting the balance.

For example, Mrs. Smith may not have reported her recorded $1,000 balance at the First National Bank but she may have reported $1,000 in the Second National Bank or she may have reported it under Mr. Smith's name, or the interviewer may be talking to the wrong Smith household or the matching clerk could have paired up documents from the wrong Smith family. Any of these errors will be interpreted as an unmentioned $1,000 savings account yet Mrs. Smith did not make an error in reporting the account balance.

There are ways of designing record checks to avoid an overestimate of response bias and these have been discussed by myself and others

elsewhere[4]. The point here is the recommendation to conduct and interpret future record checks so as not to dramatically overstate the size of the response bias of interest.

The final point is a suggestion that, in surveys of subjective phenomena, designers and analysts adopt already existing solutions when they suspect that answers to single items have several systematic determinants such as a true preference and the evenhandedness norm. These strategies, based on factor analysis and structural equation models, use multiple questions in a survey and scaling techniques to obtain purer measures of the underlying trait of interest[5]. The general point is that when we have good statistical ways of controlling or correcting for survey errors, let's use them and remind our readers of them.

May I say I enjoyed hearing the papers today and I hope these lines of enquiry will be pursued in the future but with careful evaluations of response errors and with an emphasis on developing design and statistical methods of bringing the real errors under control.

Footnotes

[1]Schuman, Howard and Stanley Presser, Questions and Answers in Attitude Surveys: Experiments on Question Forms, Wording and Context, Academic Press, New York, 1981.

[2]E.g., Turner, Charles F. and Elizabeth Martin (eds.), Surveys of Subjective Phenomena: Summary Report, National Academy Press, Washington, D.C., 1981.

[3]Eidson Penick, Bettye K. (ed.), Surveying Crime, National Academy of Sciences, Washington, D.C., 1976.

[4]E.g., Marquis, Kent H., Record Check Validity of Survey Responses: A Reassessment of Bias in Reports of Hospitalizations, R-2319-HEW, The Rand Corporation, 1978.

[5]E.g., Nunnally, Jim, Psychometric Theory, McGraw-Hill, New York, 1967 or David J. Jackson and Edgar F. Borgatta (eds.), Factor Analysis and Measurement in Sociological Research, Sage Publications, Beverly Hills, 1981.