SOME METHODOLOGICAL ISSUES IN VALIDATION STUDIES*

Harold W. Guthrie, University of Illinois

Sample surveys designed to collect financial data from households have become an established part of the methodology of research on consumer behavior over the last twenty years. Yet the continued failure to obtain sample data which are consistent with data obtained from other sources has been a serious handicap to researchers advocating new approaches to analysis of consumer behavior -- approaches for which sample surveys are uniquely appropriate. This paper attempts to review briefly the present state of the arts in collecting financial data and to point out some methodological problems which we encounter in attempts to validate survey data. Since sample survey methods are shared by social scientists in several disciplines we have a common interest in research on methods of improving our survey techniques.

Analysts of household financial data benefit from a wealth of data for purposes of comparing estimates of aggregates derived from survey data with parameters measured by other means. On some items the discrepancies are not intolerable; on others, including some very important behavioral variables, the discrepancies are staggering. Since there are differences in coverage between the aggregate data and the survey data, comparisons require skillful adjustments. The resulting comparisons are approximations but they are useful signposts pointing toward weaknesses and strengths in the survey data.

On one important variable, income, the survey data have been reasonably good. The 1950 BLS Survey of Consumer Expenditures underestimated the aggregate for total income by only about 6 percent. The Surveys of Consumer Finances conducted by the Survey Research Center, University of Michigan, underestimated income from 3 percent to 13 percent over the period 1947-1955. While the estimates of total income can be regarded as within a tolerable range of error, the estimates of individual components of income are less encouraging. Data from the 1950 BLS study overestimated entrepreneurial income by 21 percent and underestimated interest income by 74 percent.

Another example of what seems to be a compensation effect in an aggregative measure is provided by saving, the flow of money income into financial and other assets held by households. Saving is one of the more important single dependent variables in analysis of household financial behavior, and our comparisons here are based on the pioneer work of the Survey Research Center in this area. Over a four-year period, 1947-50, the sum of the amounts of saving estimated by the surveys exceeds the comparable sum of amounts of personal saving reported by the Securities and Exchange Commission by only 4 percent, but the two major components of saving show gross discrepancies. The survey data estimate saving in the form of liquid assets at $-22.5 billion against the SEC aggregate of $+7.9 billion. This shocking underestimate is compensated in the measure of total saving by a survey estimate of non-liquid saving in the amount of $71.3 billion against the SEC aggregate of $39.4 billion.

In the early days of sample surveys, attempts to measure the flow of saving were based on the recall method. That is, respondents were asked, for example, the balance in their checking accounts as of a given date, and also the balance as of the same date one year prior to the given date. The recall method was largely abandoned in later surveys in favor of a reinterview procedure. Under this procedure respondents were interviewed in two successive years and they were asked for their balances for given dates, one year apart, in each interview. The flow of saving was therefore measured as the difference in the two reports of holdings in the given asset. This procedure focused attention on the wealth holdings of families, or taking into account their debts as well, on their net worth position.

Survey data on holdings of assets and debts, when compared with similar external data, suggest some of the methodological problems now under attack by many people and institutions concerned with surveys of household financial data. The survey data estimate the number of automobiles owned very well. Proceeding down the scale with other examples the survey data underestimate the number of checking accounts by 24 percent, the number of savings and loan accounts by 64 percent, the value of stock holdings by 75 percent.[1]

The problem of improving our sample data on household finances is now under attack in two major areas. The first area concerns sample design. We know that the size distributions of many of our important variables, such as holdings of common stock, are highly skewed to the right. The sample designs used by the Survey Research Center in the early years of household financial surveys stratified households by a measure of income and over-sampled high income families in an attempt to be able to describe this segment of the population. Since the results did not measure up to the expectations of many analysts, stronger medicine seemed to be called for. In 1963 the Board of Governors of the Federal Reserve System conducted a Survey of Financial Characteristics in which, with the cooperation of the Bureau of the Census, a much more drastic scheme of stratification was used. The top income stratum of this

*In preparing this paper I have used unpublished material prepared by my colleagues, Robert Ferber and E. Scott Maynes. I am grateful for their permission to do so.

[1] Federal Reserve Bulletin, September, 1958, pp. 1041-51.

study includes households in which income is $150,000 or more. The data from this study should tell us much about the potential gains in reliability to be expected from a high degree of stratification of highly skewed populations.

The second major effort to improve the reliability of household financial data is a growing number of validation surveys. The remainder of this paper summarizes some of the recent developments in this area and points out some methodological difficulties in conducting validation surveys.

The prime mover in several important studies of errors of response and nonresponse in household financial data has been the Inter-University Committee for Research on Consumer Behavior, a committee of scholars who have an active interest in using survey data.[2] This Committee, through its Consumer Savings Project, has sponsored three major studies. The first of these studies was conducted by the Survey Research Center under the direction of John B. Lansing. The report of this study states the urgency of the need for validation research: "When the problem is one of obtaining the most accurate possible estimate of a parameter, and the interviewing technique being used tends to result in an underestimate of that statistic by 25 to 50 percent, it is not an efficient use of resources to take a large sample. Money spent on a large number of interviews could much better be spent on a smaller number of interviews plus a study of response error."[3] The studies of response error should be aimed at measuring the errors and developing analytical techniques by which biases might be taken into account in analysis of the data.

A second major phase in the studies sponsored by the Inter-University Committee was a set of panel studies directed by Robert Ferber of the University of Illinois. In the panel studies in three midwestern cities and a farm area several split-sample experiments were designed to discover survey methods with which errors of response and nonresponse were correlated. The samples for these separate surveys were drawn from lists of families who were known to hold savings accounts, checking accounts, consumer debt, life insurance, or farm debt. A third major study under the Consumer Savings Project is directed by Ferber and conducted with the assistance of the Response Research Branch, Bureau of the Census. This study focuses attention on two of the items of household finances which have large biases: savings accounts and holdings of common stock. No built-in experiments on methods were included in this study. Rather, an attempt was made to approximate as closely as possible the field operations and data processing methods used in the Federal Reserve Board's 1963 Survey of Financial Characteristics.

The findings of these studies have been, or will be, reported elsewhere. I shall draw upon these studies to illustrate some methodological problems encountered in validation studies. I shall also attempt to point out some analogues of these problems in validation studies conducted in substantive areas other than household finances.

The first issue that we must cope with in financial validation studies is the greater complexity of our variables as compared with the variables of primary concern in studies of medical conditions or voting behavior. In the latter two kinds of studies, the variable of primary concern is essentially an attribute. Does the element have an illness? Is the element a registered voter? Did the element vote in a given election? Our primary variables are of a similar nature but we also need to measure errors in the amounts. First, does the element have a savings account? Second (and like which kind of illness) is it in a commercial bank, a savings and loan association, a building and loan association, or some other kind of institution? Finally, how much was the balance as of a given date? The bias observed in the comparisons of aggregates can be attributed, in part, to errors at any one of these three stages of specification of the variable. Validation studies aimed at explaining the bias must therefore break down total nonsampling error into at least three components: errors of nonresponse, reporting error, or error in reporting the attribute, and response error, or error in stating the amount of the variable.

Two findings from recent validation studies are related to the nature of the variable and the reduction of response errors in the observation of values of the variable. First, many respondents are sufficiently motivated to give correct values, but they simply do not remember the information required. We have found that it is feasible to ask respondents to check their records so that they can give accurate data. A second useful finding is that a second interview frequently can correct errors resulting from misunderstandings existing at the time of the first interview. A similar finding was noted for the study of data on hospitalization in the National Health Survey.[4]

Since repeated interviews have become a standard procedure in studies of financial data we have also extended our validation research into panel studies involving as many as 5 waves of interviews at intervals of 3-6 months. Although panel studies suffer from declining response rates, Ferber's work has shown that there is a substantial reduction of reporting error in

[2] Members of the Inter-University Committee are: Guy Orcutt, Chairman, Lincoln Clark, Robert Ferber, George Katona, Theodore Newcomb, Howard Raiffa, and James Tobin. Raymond Goldsmith was a member of the Committee until his departure to join OECD in Paris in the summer of 1963.

[3] John B. Lansing, Gerald P. Ginsberg, Kaisa Braaten, An Investigation of Response Error. Studies in Consumer Savings, No. 2 (Urbana, Ill.: 1961), p. 201.

[4] U.S. Public Health Service, Health Statistics from the U.S. National Health Survey, Series D, No. 4. (Washington, D.C.: 1961), pp. 39-52.

successive waves.[5] Apparently those who remain in the panel become increasingly cooperative and more willing to report the existence of assets and debts. Whether the gain in reliability is worth the price of panel mortality is still to be resolved.

A second methodological issue concerns the unit of analysis and the interviewing unit. The obviously relevant unit of analysis in a study of election behavior is the individual person. Although the analyst might be interested in interactions between persons to find out how behavior is determined he needs to be concerned only with the individual for validation purposes. Similarly in studies of health conditions the primary variable is unique to the individual person. Given reasonably good identification data the matching of an interview report with a validation record is not too difficult.

The relevant unit of analysis in studies of household finances, both for substantive analysis and validation studies, is a group of persons, either the spending unit or the family. The necessity of observing the larger unit arises, in part, from the complexity of ownership of assets. Several persons in the same family may own a given type of asset, and joint ownership of savings accounts and corporate stock, not only by husband and wife, but by many different combinations of family members, is very common. While the legal status of the ownership of the asset is clearly defined by the records of the financial institution, the respondent's perception of ownership is frequently much less clear. Wives, aged parents, and children are often not aware of assets which they own singly or jointly and the ownership patterns frequently extend outside the immediate family.

We find that nonsampling errors can be measured meaningfully only if we consider the pattern of ownership of multiple units of the given asset among all members of the reporting unit. The several reported units of the asset must frequently be matched with an unequal number of units reported by the financial institution. The discrepancies between the two reports with respect to ownership require a complex matching of validation reports with respondents' reports.

Given that any of several members of a family may be holders of assets, which one should be designated as the respondent? The increased response error resulting from using a single respondent as a proxy for other family members was noted for the National Health Survey.[6] We face the same hazard in financial surveys. The general practice in financial surveys has been to choose the husband as the respondent in families containing married persons, and the wife or other family members are allowed to be present during the interview. The presence of other persons in the family during the interview improves the accuracy of response in some cases, and reduces accuracy in others.

Two recent developments in this area are worth noting. In the Federal Reserve Board's Survey of Financial Characteristics respondents with extensive and complex wealth holdings were encouraged to ask their financial advisers to fill out a self-enumeration form. For many wealthy families this seems to be the only feasible way to obtain accurate data. The second development is an attempt in a validation study to correlate response errors with the degree of participation of various family members in the interview. The results of this attempt are not yet available.

A third methodological issue is the link between the individual interviewing unit and the aggregate which is the parameter to be estimated. Again, a contrast between validation studies of household finances and studies of voting behavior and health studies illustrates the difficulty. Registration for voting provides ideal conditions for a validation study. Here the validation records are open to the public and local. The aggregate number of registrations to be estimated for a county can easily be related directly to a validation study of individual persons and their interview reports. Validation of voting behavior in terms of how votes were cast is, of course, made difficult by the lack of a validation record. Health studies can also be locally oriented because the market for health care is generally a local market.

One very important consequence of the local and open character of validation records is that both primary and secondary validation can be performed in a single study. That is, the study can determine both the errors of underreporting (respondent reports that he is not registered when, in fact, he is) and overreporting (respondent reports he is registered when, in fact, he is not).

Only rarely do we find very favorable conditions for validation of household financial data. First, the records necessary for validation are never public records, and, indeed, the sources of the financial records are generally business firms who take seriously the ethic of guarding confidential personal data. Moreover, the firms are not easily convinced that validation research is a legitimate justification for their spending the necessary effort to cooperate in a validation study. Second, for many kinds of financial variables, the market is national rather than local. An ideal validation study of life insurance holdings, for example, would require access by the researcher to the records of hundreds of insurance companies. Perhaps at the other extreme are certain kinds of consumer credit for which the risk element requires a local market rather than a national market.

Up to the present time our validation studies have not attempted to cover completely a single local area. We have conducted primary validation

[5] Robert Ferber, "Does a Panel Operation Increase the Reliability of Survey Data: The Case of Consumer Savings," 1964 Social Statistics Proceedings of the American Statistical Association, pp. 210-216.

[6] U.S. Public Health Service, op. cit., p. 8.

studies from known holders of given assets in a given institution. One important cost of using this approach is that we can say very little about the incidence of overreporting. We know that overreporting is typical for some behavioral characteristics as shown by the Survey Research Center election studies which measure voting turnout.[7] A related error is also indicated in financial validation studies which show that some respondents overstate the amount of their asset holdings.

Although overreporting of an attribute may be an important phenomenon in other areas of inquiry, we feel that it is not a serious source of error in financial data. We are sure that our net biases are negative because of the results of comparisons of aggregates. Also, in two surveys under the Consumer Savings Project in which secondary validation was attempted, no clear cases of overreporting occurred. Additional evidence that overreporting of amounts, if not attributes, can be reduced by reinterview is provided by Lansing's finding that gross overreporting was almost eliminated by a second interview.[8]

A fourth methodological issue arises from the fact that one important objective of validation studies is to develop interviewing procedures which will reduce nonsampling errors. Most validation studies of household finances have collected data on many different characteristics of the interview situation. Some attempts to correlate nonsampling errors with these characteristics have suggested possible minor improvements. Many of us who conduct surveys, however, look to the interviewer-respondent interaction as the most plausible explanation of errors. We tend to think that if we could just get perfect interviewers and field operations our error problem would be much smaller. So far in our validation studies of household finances we have not made much progress in this area of research. Our validation studies have been conducted within the context of normal procedures for interviewer selection and field control. We have included in our survey design neither controlled experiments nor random assignments of sample addresses to interviewers. The development of special purpose studies which combine validation, experimental controls, and random assignment of interviewers to respondents is one task which we, and perhaps researchers in other fields as well, might undertake in the future.

Although I advocate making an attack on the problem for the same reason that men climb mountains, I do not expect any significant results. There is undoubtedly a large number of possible types of respondent-interviewer interactions, and advance data about the respondent to guide optimal assignments of interviewers to respondents is usually scanty. Thus we would have to find a very strong and strategically useful relationship between nonsampling errors and type of respondent-interviewer interaction before these research efforts would pay off in substantial gains.

The last issue to be considered is partly methodological and partly theoretical. We noted earlier that comparisons of survey based aggregates with external aggregates produced widely variant results. Examples illustrated a range from an underestimate of about 75 percent for stock holdings to an underestimate of about 6 percent for total income. The data suggest a continuum extending into overestimation. Our ultimate goal in studies of nonsampling error might well be an explanation of differences in degree of error along the continuum and among different items of information.

For example, the underreporting of mental and nervous disorders in the National Health Survey[9] and the underreporting of financial data may have a common cause, a desire to suppress secret information. One of Lansing's results throws some light on this issue. He found that respondents who grew up in families in which financial information was concealed from the children tended to conceal information from the interviewer.[10] Perhaps mental and nervous disorders and holdings of common stock are not very far apart on a continuum of sensitivity and, if so, it may be that similar techniques to reduce response errors will be successful for both kinds of information.

To summarize this brief review of methodological issues, we in the sample survey field have made substantial progress in measuring nonsampling errors but we must go much farther. We have learned how to conduct useful but imperfect validation studies by paying careful attention to some methodological pitfalls. We have persuasive results to indicate that motivation of the respondent is important, but that we must also take into account errors due to lack of information, deliberate concealment, and other causes. These concepts form the basic structure of a theory of nonsampling errors. Now we must search out and identify specific reasons for differences in errors among different items of information in order to develop feasible remedies. Thus we must take advantage of every opportunity to incorporate validation research into surveys conducted primarily to obtain substantive data. At the same time, some methodological issues, such as the interaction between respondent and interviewer, seem to require specifically controlled validation studies.

[7] Angus Campbell, Philip E. Converse, Warren E. Miller, Donald E. Stokes, The American Voter. New York: 1960. pp. 93-96.

[8] John B. Lansing, et al., op. cit., p. 186.

[9] U.S. Public Health Service, op. cit., p. 54.

[10] J. B. Lansing, et al., op. cit., p. 184.