Asymmetric Information, Interviewer Behavior, and Unit Nonresponse

Arthur B. Kennickell
Senior Economist and Project Director
Survey of Consumer Finances
Mail Stop 153
Federal Reserve Board
Washington, DC 20551
Phone: (202) 452-2247
Fax: (202) 452-5295
Email: Arthur.Kennickell@frb.gov
SCF Web Site: http://www.bog.frb.fed.us/pubs/oss/oss2/scfindex.html

Abstract

Unit nonresponse in surveys is generally taken to be the result of an unsuccessful interaction, or series of interactions, between an interviewer and a respondent. However, in the case of in-person surveys, the intensity of interactions, and even the pursuit of a situations that would allow interactions are largely under the control of interviewers. Survey managers normally know only a very limited part of the characteristics of sample members—often the managers only know whether or not a case is completed. In contrast, interviewers directly observe many characteristics of the sample. In the absence of effective monitoring, interviewers will tend to arrange their work on cases to suit their own preferences in light of their expectations based on the information they observe and whatever constraints and compensation arrangements are set by the managers. If interviewers' preferences differ from those of the managers, they may apply effort to cases in a way that differs systematically from what would be desired by managers under an arrangement with more complete information unless this tendency is somehow offset through constraints or a compensation structure. The result of such a deviation would be a type of implicit field stratification—that is, participants may differ systematically from nonrespondents. This paper develops a very simple model of the information structure in a field survey, the nature of interviewer incentives, and the effects of differential field effort, and it provides illustrations based on the outcomes of the 1998 Survey of Consumer Finances. The paper argues for more attention to factors that determine the application of effort to individual cases in surveys.

Unit nonresponse in surveys is generally taken to be the result of an unsuccessful interaction, or series of interactions, between an interviewer and a respondent. However, in the case of in-person surveys, the subject of this paper, the intensity of interactions and the pursuit of situations that increase the likelihood of interactions are to a large degree under the control of interviewers. Normally, survey managers are able to target field resources only at a relatively gross level, and they observe only a very limited part of the set of characteristics of sample members that are known more fully to interviewers. Often the managers only know not much more than whether or not a case is completed and the total amount of time an interviewer worked on all cases over a given time period.

Because the structure of information usually makes it impossible to reward interviewers for the appropriateness of their efforts across all cases, their performance evaluations very often turn critically on the degree to which they complete the sample cases in their assignments. Thus, all other things being equal, interviewers have an incentive to pursue cases that have the highest likelihood of success   More generally, the absence of full monitoring implies that interviewers will arrange their work to suit their own preferences as far as such behavior is consistent with any overall constraints imposed by the field managers. To the extent that variations in effort are correlated with anything measured in a survey, the result will be a type of implicit field stratification of the original sample observations, where the dimensions of stratification are related to the level of difficulty or other features of the cases perceived by the interviewers.

In the economics literature, a great deal has been written on behavior under uncertainty in situations with asymmetric information, and contracts between various actors that may be written to reduce incentives for partially observed agents to act against the wishes of a principal.[1] Although the findings are often quite complex and particular, the essence of this work is that the presence of private information makes it difficult to align the incentives of the actors; consequently, attention must be paid to understanding the incentive effects of the contracts between the actors, and the cost and value of obtaining additional information. This paper attempts to bring some of the reasoning in this literature to bear on understanding the behavior of

---

[1]For example, see Ross (1973), Harris and Raviv (1979), Holmström (1979), and Grossman and Hart (1983).

interviewers and the consequences of their behavior for nonresponse and the distributions that are ultimately measured.

The first section of the paper discusses the potential consequences of the structure of interviewers' incentives in the presence of asymmetric information about the sample members, and it presents a very simple model of this relationship. The next section illustrates the problem with data from the 1998 Survey of Consumer Finances (SCF). The final section summarizes the findings in the paper and points toward related areas in need of additional research.

I. A Model of interviewer behavior

The response rate and the distribution of completed cases in a survey are functions of many factors: the number of interviewers, the amount time they are allowed to work, the degree to which interviewers choose to allocate effort to particular cases, the nature of interviewers' interactions with respondents, and the level and variation of resistance among respondents. Building on their earlier work and that of others, Groves and Couper (1996, 1998) have built a very clear and useful framework for characterizing the nature of interviewer-respondent interactions and respondent resistance. The more elementary question of the allocation of resources, either by management or by the interviewer, has received remarkably little attention.

Ideally, one would like effort to be applied to all cases in a sample such that the distribution of the completed cases (along dimensions relevant for the motivating research) is, at worst, a completely random subset of the original sample.[2] If resources were always evenly matched with expected resistance in the sample, it would be only surprises in the required level of effort for a given survey that would keep completion rates below 100 percent. It does appear that there is an important element of surprise in the response rates for most surveys, but there are also deeper complications. First, there are probably always respondents who will not agree to be interviewed no matter how much attention they are given, and respondents who would require so much effort to persuade that they far exceed the practical limits of effort. Second, for many reasons, interviewers are not often distributed in equal proportion to required effort, and it is

---

[2]This argument assumes that any treatment effects resulting from differential applications of effort are negligible or otherwise not important.

costly to move interviewers over long distances. Hiring and retention can be unusually difficult in some areas, and it may well be that labor market tightness is also related to the overall level of respondents' degree of resistance. Also, some areas with small numbers of sample cases may be relatively over-staffed even with just one interviewer; but sometimes the remoteness of the area makes it costly to use the interviewer elsewhere or, alternatively, to bring an interviewer into the area on a temporary basis. Third, interviewers may adjust the level of effort that they apply to particular cases in response to the incentives they face and given the set of information they have about the sample and individual cases. Because the information about the level of interviewers' efforts and the level of difficulty of cases is not the same for interviewers and managers, it is not possible for the managers to make clear decisions about whether redeploying resources might actually change outcomes. This section of the paper focuses on this third issue.

As an example of the general problem, consider the following very simple model. Suppose interviewers are rewarded R for completing cases, but there is no reward for the number of hours worked. A sample of respondents consists of two types: N^e easy respondents who cost interviewers E of effort, and N^d difficult ones who require D of effort, where $D>E$. The interviewers can observe the types of each of their own cases. The survey managers have no information to distinguish members of the two groups in the sample, but they want to obtain interviews with each group in proportion to its share of the sample. Assuming that interviewers' preferences are linear in the costs and benefits, the managers would have to set $R>D$ to have the difficult cases completed, and the response rate would be 100 percent. Interviewers would make $N^e(R-E)$ "profit" on the easy cases and the smaller amount $N^d(R-D)$ on the difficult ones. If there was even a remote possibility that interviewing would be halted before all cases were completed, interviewers would always complete the easy cases first. If easy cases differ in terms of analytically interesting variable from the hard cases, anything that terminated interviewing before all cases were completed would introduce bias.

To highlight the problem more formally, and somewhat more generally, let there be a set of cases $K=\{1,..., n\}$ in a sample, and suppose that there is a vector of characteristics z_j for all $j \in K$, and that this vector is fully observed by interviewers but it is not observed directly by the survey managers. The "difficulty" of a case is given by $D(z)$. Interviewers, taking D and a compensation function $C(w)$ as given, chose effort e_j^* for each case to maximize their utility

given by $U(-e_j, C(e_j), -D(z_j))$, where U has weakly positive first derivatives in all three arguments and it is taken to be additive over cases. Thus, interviewers get satisfaction from lower levels of effort, higher levels of compensation, and lower levels of difficulty. Field managers would like interviewers to apply effort, given by w_j, that is proportional to $D(z_j.)$, say $e_j^+ = \alpha D(z_j.)$, where α is a constant proportion, and they would like to minimize $\sum_{j \in K} C(e_j)$. Under complete information, the managers would chose a function C^+ such that

$$\sum_{j \in K} C(\alpha D(z_j)) \quad \text{is minimized where } \alpha D(z_j) \text{ is } \quad ARGMAX_{e_j} U(-e_j, C(\alpha D(z_j)), -D(z_j)) \;.$$

However, because the survey managers cannot observe z_j, they can only set C as a function of e_j. If C is chosen so that no observation has less than e_j^+ effort applied, say C^-, then under regularity conditions it is clear that C^- will be weakly greater than C^+. Furthermore, if C is chosen to that the minimum effort is less than e_j^+, then the most difficult cases will not receive the optimal level of effort.

The information model may be made much more complicated. For example, in a world where the difficulty of cases is uncertain and it is costly for interviewers to determine the true level, they may extrapolate from their experience to create a working model of the expected payoff for cases with common characteristics and adjust their effort accordingly. Sometimes managers may know things about cases that are not—or cannot be—told to interviewers, and this information might contradict the interviewer's mental model if it were known. The critical factor in generating distortions in effort relative to the ideal is that interviewers and managers conditioning on their available information have different expectations about the difficulty of cases.

The real world of field interviewing introduces many other complexities. Incentives for interviewers typically mix an hourly rate of compensation that may vary with experience, with a required number of completed cases to remain employed. There is almost surely great variation in the level of difficulty across cases, and some cases may so intimidate interviewers that the benefit in hourly pay is less than the psychic cost. Interviewers' preferences and abilities also vary, but because such variation is not directly observable to managers, there is uncertainty in comparing behavior across interviewers. Because the "types" of cases may not be completely obvious to interviewers, they may start cases that they expect will be easier than they turn out to be, and the opposite situation may also occur. Some information about the difficulty of cases is

revealed when interviewers classify a case as a nonrespondent, but it is virtually impossible to gauge whether this subjective classification indicates that an interview is impossible or that it is simply some order of magnitude more difficult than other cases. Competing opportunities for interviewers may arise during the field period, and if only more difficult cases remain, the minimum additional benefits needed to get the interviewer to switch jobs will be smaller. Numerous sources of uncertainty may raise costs and budgets are always ultimately limited; these facts make clear to everyone that the field period will almost surely not continue until every case is completed or even until every possibility of refusal conversion is exhausted. Finally, it is important to note that some interviewers clearly display an unusually high level of altruism in working on cases that are not necessarily in a narrow view of their interest.

An essential thread in many of the innumerable more complex models that could be built to describe behavior of interviewers in the field is the informational asymmetry between interviewers and survey managers. As long as managers and interviewers have different information, their conditional estimates of the level of difficulty involved in completing various cases will differ, and consequently their desired levels of effort on individual cases may differ from the effort managers would desire in the presence of full information. Where interviewers have information not known to the managers, the managers are unable to estimate sufficiently accurately the true level of difficulty of cases, and consequently they cannot make direct decisions about allocating time to individual cases. If the difficulty of obtaining cases is correlated with analytically interesting variables, this implicit field stratification will almost certainly lead to bias in the distribution of the final data.

There are several possibilities for dealing with the problems induced by the asymmetry of information. First, in some cases it may be that factors sufficiently correlated with difficulty can be observed for each observation during the field period, and that these factors can be used ex post to make nonresponse corrections through post-stratification of the analysis weights. Second, if such factors are observable sufficiently promptly during the field period, they might be used to develop explicit quotas. A third possibility is to design contracts with interviewers either to raise the level of their incentives and manipulate their constraints so that they are economically motivated to complete all their cases, or to divide the interviewing staff into units with different operational responsibilities. Each of these possibilities is addressed in turn below.

Whether explicitly or not, post-stratification is very often used with the aim of reducing nonresponse bias, including any component attributable to differential application of effort. Research more focused on this dimension of nonresponse could lead to the discovery of better ex-post adjustments.

Although there are many references to the use of quotas in market research, the subject has rarely been touched in a positive way in more formal survey research. Sudman (1966) develops the idea of using quotas within the overall constraints of a scientifically drawn sample. In essence, such an approach is analogous to post-stratification—perhaps a good name would be "ante-stratification." Research could be devoted to identifying reliably measurable indicators, and computer-assisted data collection in the field could make implementation practical. There may be a particular advantage to this approach over post-stratification if the strata are not homogeneous, as one would ideally like them to be. With either the ex-ante or ex-post approach, there may still be variations in the level of difficulty within strata. Pressing for additional cases in the ex-ante approach would have the advantage of obtaining more relatively difficult cases. However, to the degree that either approach failed fully to condition outcomes in terms of difficulty, bias would remain.

A uniform rise in the wage structure for interviewers to ensure a high level of effort on all cases, though arguably highly desirable for other reasons, would be difficult to implement in the competitive world of survey research operations.[3] Potentially, a nonlinear compensation scheme with progressively larger reward for additional completed cases might be even more effective. However, both arrangements could be seen as handicapping subsequent surveys, particularly if there is substantial variation in difficulty across surveys, and developing a general formula that might apply across many surveys would probably be very complicated. Still, there appears to have been very little systematic exploration of the range of feasible compensation plans for interviewers, and it may well be that the large economic literature on contracting

---

[3]Approaches that lead to higher pay might have the additional benefit of leading to a more "professionalized" interviewer corps, the cost of which could be at least partially offset through reduced attrition and consequently lower costs for training, supervision, etc. A more permanent professional staff would presumably also improve data quality by developing and maintaining high quality data collection skills among interviewers.

developed in other areas could have significant payoffs. Another possibility would be to offer interviewers a menu of compensation plans. For example, interviewers might have an option of receiving a higher wage by agreeing to deal only with cases that other interviewers have classified as refusals, or receiving the base wage by agreeing to the normal conditions of completing a certain percentage of cases etc. Effectively, this approach would codify the use of "refusal converters," a common arrangement on many surveys. One negative effect of such a formal arrangement might be that the "regular" interviewers would tend to give up on moderately difficult cases that they would otherwise have completed. Other literature on asymmetric information suggests that sometimes incentives can be structured to make it in agents' own interest to reveal at least some part of their "private" information, as is commonly the case in the design of insurance contracts. A better understanding of the structure of respondents' propensity to be "difficult" would probably be needed to make significant progress in this direction.

II. Illustrative data from the SCF

A. Background on the survey

The data used here derive from the 1998 Survey of Consumer Finances (SCF), which was sponsored by the Federal Reserve Board (FRB) in cooperation with the Statistics of Income Division of the Internal Revenue Service (SOI). The SCF is designed to collect detailed information on households' assets, liabilities, pensions, work history and use of financial services. The survey also collects extensive demographic and other data useful in understanding the more narrowly economic data.[4] The sample is a dual-frame design. One part of the sample is selected using a standard multi-stage area-probability (AP) technique (see Tourangeau *et al.*, 1993). The other part is a list sample developed from statistical records created by SOI based on individual tax returns (see Kennickell, 1998a). This list sample oversamples wealthy families. Unlike the AP sample, respondents in the list sample were given an opportunity to refuse participation in the survey by returning a postcard; respondents who returned the postcard were

---

[4]For a more detailed description of the survey and the data, see Kennickell, Starr-McCluer, and Surette (2000).

not subjected to any type of refusal conversion. In 1998, about 21 percent of the list sample cases opted out of the survey by returning the postcard. Over both samples, 4,309 interviews were completed, with about a third of those observations coming from the list sample. In the AP sample, the response rate among eligible cases was about 66 percent. In the list sample overall, the rate was 29 percent, but this figure ranges from about 41 percent in the least wealthy stratum to about 8 percent in the wealthiest stratum. Data for the survey were collected between June and December of 1998 by the National Opinion Research Center (NORC) at the University of Chicago using CAPI.

B. The structure of information

In addition to the actual interview data for the survey participants, some things are known about the full sample either at the unit level or for groups of units. For each sample case assigned to them, interviewers were required to maintain records of all their actions in an electronic "record of calls" (ROC), and each entry records a working case disposition code as of the end of the action described. The ROC information also includes the date, time, mode and specific nature of each action (e.g., in-person call, message on answering machine, mailed literature, etc.). Each record of this file also includes an open-ended field that interviewers were instructed to use to contain any unstructured data about the respondent, the dwelling or any other factor that might be of use to another interviewer in the event that the case was transferred. Most likely because interviewers were told that the information in the ROC was accessible by supervisory staff, the information in the structured fields was much more complete than in previous years when the records were maintained on paper, and only the structured contact data were made machine-readable at the end of the field period.

The interviewers were also asked to provide some other types of information about each case. After the first visit to a sample unit, interviewers were instructed to complete a short questionnaire about the visit; the questions included a few characteristics of the dwelling and its relationship to others in the neighborhood, and a few characteristics of the neighborhood. After the initial contact with anyone at the selected address, the interviewer was required to complete a few questions about the person they spoke with and the comments that person may have made. In cases where the interviewer was at least able to complete the screener for a case, some characteristics of the person answering the screener questions may also have been obtained.

Although useful, these types of other information are not as complete in their coverage of the sample as the ROC data. Moreover, most of the information provided by the interviewer has a strong subjective component.

In order to administer the interview, the NORC central office staff and interviewers needed to have address information for every case; the FRB project staff knew only the nine-digit ZIP code for list sample cases and the five-digit ZIP code and census tract for area-probability cases.[5] In theory, the location data could have been linked to other geographically-based data during the field period in order to give some indication of the characteristics of the neighborhoods where completed cases were more or less common. In practice, such information was only used by the FRB project staff after the field period in evaluations of the distribution of nonrespondent cases (see Kennickell, 1999).

For each observation in the list sample, there is also important information in the original frame. Although the use of the SOI data for purposes other than pure sample selection is strictly controlled, FRB staff are permitted to use some key frame items (other than names and addresses, which were never available to FRB staff) for purposes of characterizing nonresponse. NORC staff were allowed access to a much more limited set of data for purposes of administering the survey: the supervisory staff, but not the interviewers, were allowed to know the original list sample stratum indicators (ranging from 1 to 7, where higher numbers are associated with higher likelihood of increasingly greater wealth), and both groups were given information on the age of the principal tax filer to use in cases where a distinction needed to be made between two people with the same name.[6]

_____

[5]One complication was that address information for list sample respondents was not always meaningful. By law, a taxpayers is required to provide his home address on the tax return. In fact, for a number of cases selected into the list sample, the address provided was either a work address, which was sometimes incomplete (e.g., a large office building without the specification of a suite number), or was the address of a law or accounting firm.

[6]Names alone were insufficiently precise to identify the correct respondent in two situations. First, a "senior" and "junior" may have lived at the address given to NORC. Second, in cases where the address information was not the home address of the respondent, sometimes a wider search needed to be conducted to locate the respondent, and in such situations there was a greater likelihood of finding multiple people with the same name as the respondent.

Finally, in working on a given case, interviewers observe idiosyncratic characteristics of the case that are probably often not effectively summarized in the formal data associated with the case. For example, the interviewer may see that the house has a sophisticated prowler detection system, the garden may be completely overgrown with weeds, or there may be a shiny new boat in the driveway. The factors the interviewer observes may have a meaning that varies depending on location. Even where such information is recorded somewhere in the records for a case, only the local interviewer sees the facts directly and in the context of the local meaning of the information.

C. Use of information during the field period

For practical reasons, attention to information about the full set of sample members was restricted to only a few key items during the field period. Obviously, NORC staff made extensive use of identifying information to locate respondents. Completion rates by PSU in the area-probability sample and completion rates by stratum in the list sample were important management variables, and these rates were tracked closely both by NORC and FRB staff. Although interviewers could clearly distinguish area-probability sample and list sample cases—the former had only an address, but the latter had a name and an address—they were given no information that would allow them to identify the strata to which the list sample cases belonged.

The ROC was an important management tool for tracking the completion of cases and monitoring the progress of individual interviewers. However, the use of the information was not very sophisticated during the field period. The field managers directly above the interviewers were able to generate a few simple reports from the database (for example, counts of cases by disposition code, PSU, etc.) that were critical in managing their interviewers. The data were also used by the field managers in conferences with interviewers about specific cases. However, the data structure was too complex and its implementation too recent to expect that anyone could have used it during the field period to gain a broad grasp of the patterns in of each interviewer's efforts. The FRB staff only obtained the structured variables from the ROC data with a lag of about two months.

It would be hard to imagine, in general, that interviewers' own decisions to commit effort to various cases were not influenced by both the shared information and the less organized

information that only they had. Unfortunately, there is very little direct evidence on how interviewers used such information, and traces of their decision are embedded in the completion rates and co-mingled with the traces of many other factors.

D. Management decision that may affect differential application of effort

Over the course of the field period, survey managers made many decisions that had a very strong effect on the likelihood that various cases were acted upon or completed. Global decisions made by the central office staff in conjunction with the field managers determined the overall level of field resources and the initial geographic dispersion of interviewers. This information was fully observed. As the field period progressed, management decisions, constrained by contractual obligations and the desire to produce high-quality data, had a very strong effect on the degree to which resources were tailored to increasing response rates in different areas of the AP sample and in different areas and strata of the list sample, particularly through the use of traveling teams of interviewers. These allocation decisions were not recorded directly in a way that is analytically useful, though it is possible to extract some information about the use of traveling interviewers from the ROC and a database of information on individual interviewers. Interviewers had weekly case reviews with their field managers, and the instructions from the manager as a result of that conference may have shaded the application of work on an interviewer's cases in many subtle ways. None of these decisions are currently captured in any analytically usable way.

E. Illustrations

Because the signals generated by many types and levels of behavior by respondent, interviewers, and management personnel are superimposed in the observable data, it is very difficult to produce statistics that give an unambiguous picture of how limited information in particular may have real effects on the distribution of outcomes in the SCF. Nonetheless, there are some examples that may indicate a role for the effects of informational limitations. Three such examples are given here.

Illustration 1

As noted earlier, the field interviewers did not have any information that allowed them to discriminate directly between the strata of the list sample. Such information was available only

to the management and supervisory staff and to the FRB project staff. Although there were specific minimum completion targets set for each of the list sample strata, these targets in each survey were projected based on what was actually achievable in earlier surveys, and there was a very clear impression at the time of the survey that it was a struggle to reach these goals.

The completion rates for the bottom two strata of the list sample are particularly interesting. These two strata overlap in terms of wealth with the great majority of the interviewed part of the AP sample (see Kennickell 1998b). The cases interviewed in stratum 1 have a median that lies between the median and 25th percentile of the cases interviewed for the AP sample (table 1). By construction, the list sample excludes people who did not file a tax return—about 10 to 15 percent of all households—and almost all of that group could be expected to have wealth sufficiently low that they would have been included in stratum 1 if they had filed a tax return. Stratum 2 appears to be roughly equivalent to the wealthier end of the AP sample. As noted earlier, a key difference between the AP and list sample cases is the fact that before the interviewers were sent into the field, the list cases were given an opportunity to refuse participation by returning a postcard. Because the postcard refusals should have removed cases most determined not to be interviewed, one might expect the response rates in stratum 1 to be substantially above the rate for at least the tax filers in the AP sample. However, even if one assumes that all of the cases in the full AP sample who did not file tax return were actually interviewed, the implied response rate for the remainder of the AP sample falls to only about 60 percent—still far above the rates in strata 1 and 2 adjusted for the postcard refusals. It is possible that there were systematically greater problems in locating the list sample cases, but there is no sign of a sufficient level of such problems in the ROC data. It is also possible that the fact that list sample respondents were requested by name made some people more suspicious than would be the cases for AP cases, whose names were unknown a priori since they were selected only on the basis of their address.

Another interpretation is that the data show the effects of limited information. Because the samples for both stratum 1 and stratum 2 are relatively small—each comprises well under 10 percent of the list sample—most of the list sample cases interviewers see are much wealthier than the great majority of AP sample cases. If, as appears strongly to be the case, wealth is correlated with difficulty, interviewers may have attributed to stratum 1 and 2 cases a subjective estimate of the level of difficulty that was higher than the correct conditional level.

Table 1: Unweighted quantiles of net worth (1998 dollars) and response rate (percent) for AP sample and strata 1 and 2 of the list sample; 1992, 1995 and 1998 SCF.

	Survey year		
	1992	1995	1998
Full AP sample			
Net worth			
25th percentile	6,100	9,300	5,800
Median	47,800	57,600	58,800
75th percentile	158,000	179,000	202,000
Response rate	68.0	66.3	65.9
List sample			
Stratum 1			
Net worth			
25th percentile	3,100	4,200	2,700
Median	29,600	28,100	21,600
75th percentile	79,200	102,000	105,000
Response rate			
Overall	42.8	45.3	41.3
Given PC	53.0	54.2	40.9
Stratum 2			
Net worth			
25th percentile	103,000	140,000	117,000
Median	206,000	272,000	284,000
75th percentile	447,000	522,000	530,000
Response rate			
Overall	41.4	39.5	39.2
Given PC	54.3	49.6	34.3
All strata			
Net worth			
25th percentile	447,000	458,000	456,000
Median	1,889,000	1,887,000	2,182,000
75th percentile	8,505,000	7,943,000	10,430,000
Response rate			
Overall	31.1	30.4	28.6
Given PC	40.9	34.3	33.1

Illustration 2

The ROC allows one to construct a measure of the application of effort across different parts of the sample. These data show that the distribution of the number of attempts needed to obtain an interview in the AP sample and the bottom two strata of the list sample are similar (table 2 and figure 1).[7] For completed cases in list sample strata 3 and above, the data indicate that the distribution of the number of attempts needed to complete the case is shifted up by one relative to AP cases across most of the distribution. One would expect that even with the removal of many of the potentially most hostile cases via the postcard refusal, the wealthy families in these strata would be relatively more difficult to persuade to complete an interview.

Although the evidence for the completed cases could be taken to suggest that an overall higher level of effort was devoted to the list sample cases, a similar analysis for the final refusals and the "censored" cases (eligible cases that were classified as neither completed nor refused at the end of the field period) suggests the contrary (table 2 and figures 2, 3, and 4).[8] For both of the subgroups of the list sample, the distribution of the number of attempts lies distinctly below that for the AP sample alone. Even so, it could be that refused and censored list sample cases were so much more difficult than the completed cases that the expected payoff to addition effort was negligible. Perhaps surprisingly, the available data do not support this hypothesis (table 3). For the list sample refusals and for the censored cases, there are far fewer references to a hard refusal.[9] Moreover, there were relatively fewer attempts with the list sample refusals to convert

---

[7]An "attempt" is defined broadly to include any effort devoted to securing a complete interview with a case. In principle, duplicate entries for the same action, and simple comment entries in the ROC are excluded from the calculation.

[8]Across the censored cases, only about 13 percent of the working disposition codes for the ROC entry before the one indicating the suspension of work indicated a "temporary hard refusal." Other codes at this point suggest that additional work might well have been productive.

[9]In cases where an initial informant agreed to complete a screener interview designed to identify the appropriate respondent, but the person refused to complete the interview, the interviewer was asked to record whether the refusal was "hard," that is, one that appeared to be relatively poor prospect for refusal conversion. In the record of calls for cases, interviewers also were able to record such information at any stage of negotiation through the working case disposition codes. If the interviewer recorded a hard refusal either in the screening or at any point in working the case, it is counted here as a hard refusal.

the refusal than was the case for the AP sample.[10] As noted earlier, although the ROC data were used for some types of interviewer supervision during the field period, there is no sign that these data were used in a systematic way to regulate the efforts on individual cases. Thus, there was an effective asymmetry of information between interviewers and managers, and one interpretation of the data is that interviewers may have shaded their efforts away from the list sample.

Further evidence of possible aversion of interviewers to the list sample cases may also be gleaned from looking at the application of effort at the beginning of the field period. During the first three weeks of the field period, all cases were supposed to be "touched" in some way. The list cases were released a few weeks after the AP cases to ensure that interviewers had sufficient experience to be able to deal with the list sample cases, a practice recommended by earlier experience. Of the AP cases, there is a record of some attempt to reach about 3/4 of the cases during the first 28 days of the field period, and about half had at least some sort of contact. In contrast, 28 days after work had begun on the list sample cases, there had been attempts on only about a fifth, and only 10 percent had any sort of contact. The interviewers' direct supervisors were supposed to have enforced work on all cases, but it is very clear from the data that effort was more focused on the AP cases. Like the interviewers, the supervisors were under pressure to produce completed cases.

---

[10]Active refusal conversion was indicated by working case disposition codes in the ROC.

Table 2: Quantiles of distribution of number of attempts to interview cases, AP sample, list sample, strata 1 and 2 of list sample, and strata 3-7 of list sample, by final case disposition, 1998 SCF.

| | Distribution of number of attempts | | |
	25th %ile	Median	75th %ile
AP cases			
Completes	3	5	9
Refusals	6	9	13
Censored	6	9	14
List sample cases			
Completes	4	6	10
Refusals *	5	8	11
Censored	4	7	10
List strata 1 and 2			
Completes	3	5	8
Refusals *	5	7	10
Censored	4	6	9
List strata >2			
Completes	4	6	10
Refusals *	5	8	11
Censored	4	7	10

* Excludes postcard refusals.

Table 3: Percent of cases classified as giving "hard" refusals, percent where refusal conversion was attempted, and percent in group, by final refusal status and final censored status, and by AP sample, and refusals other than postcard refusals in the list sample, strata 1 and 2 of list sample, and strata 3-7 of list sample, 1998 SCF.

| | Sample group | | | |
	AP	All list	Strata 1+2	Strata 3-7
Refusals *				
Hard refusal	55.0	42.6	39.8	42.9
Refusal conversion attempted	80.1	63.3	62.7	63.3
Refusals as percent of elig. cases in sample group	24.9	24.8	16.9	27.9
Censored				
Hard refusal	39.1	28.4	27.4	28.5
Refusal conversion attempted	36.5	19.8	18.8	20.0
Censored as percent of elig. cases in sample group	2.3	27.1	49.2	29.1

* Excludes postcard refusals

Figure 1: Density of number of attempts to complete a case, by sample group, for cases that were ultimately completed, 1998 SCF.
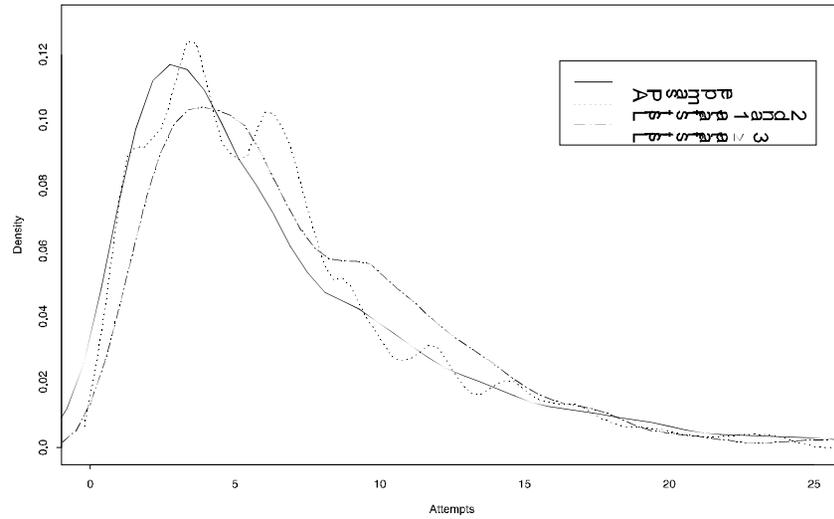

Figure 2: Density of number of attempts to complete a case, by sample group, for cases that were ultimately classified as refused, 1998 SCF.
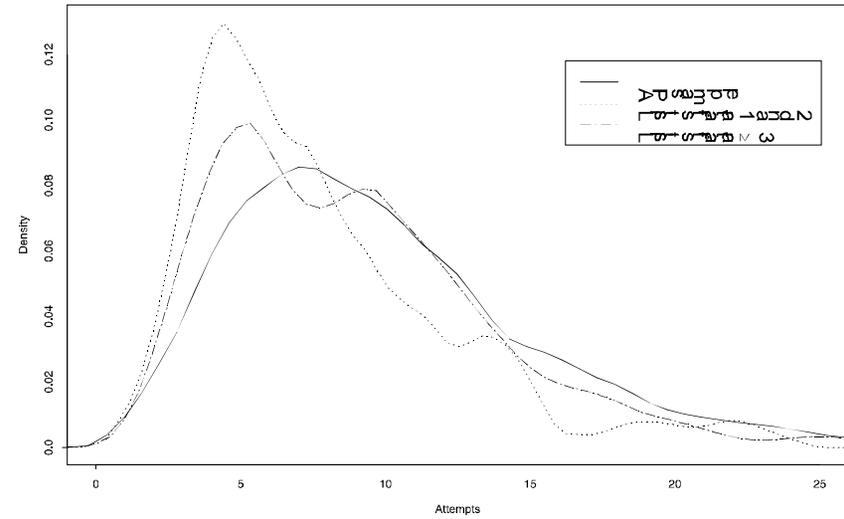

Figure 3: Density of number of attempts to complete a case, by sample group, for cases that were ultimately classified as censored, 1998 SCF.
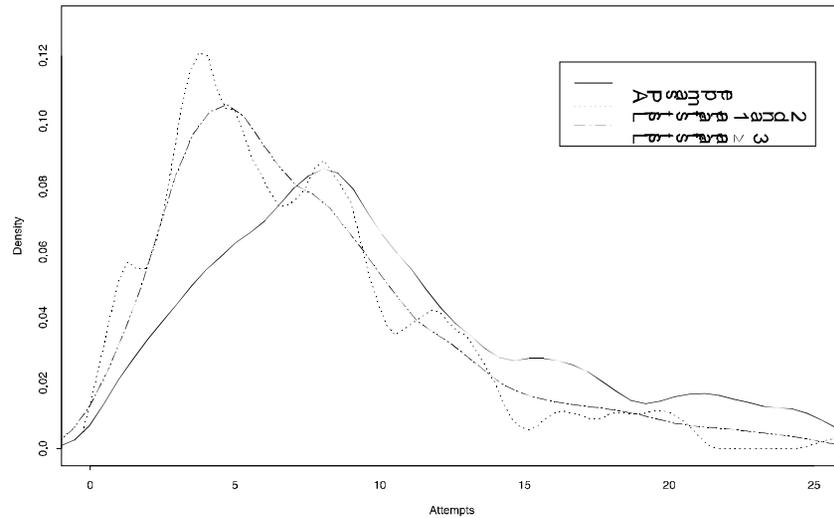

Figure 4: Density of number of attempts to complete a case, by sample group, for cases that were ultimately classified as refused or censored, 1998 SCF.
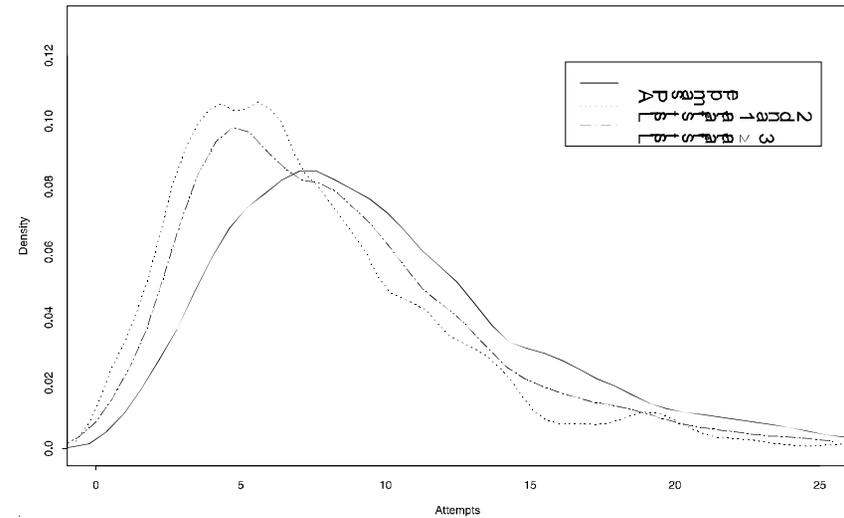
Illustration 3

For almost all AP cases and for about 80 percent of list sample cases, interviewers recorded their perceptions of the wealth of the neighborhood surrounding each respondent's residence.[11] Although this information was entered into the computer database that interviewers maintained for each case, the information was not accessible by the field managers, and it was not monitored by the central office staff. If neighborhoods are relatively homogeneous economically, and if wealthy cases tend to be more difficult to interview (either in fact or in expectation), then lack of effective monitoring might tend to lead to a tendency to under-work cases in wealthy areas.

The data (table 4 and figures 5-8) suggest that among completed cases, final refusals, and censored cases, the distribution of the number of attempts made on cases does not differ much between observations in unusually rich neighborhoods and those that were not in neighborhoods at either economic extreme. For cases in unusually poor neighborhoods, the data indicate that less effort was required for the cases that were actually completed, but that there was a small tendency to apply relatively more effort to this group in both the final refusal and censored cases.

A somewhat different impression may be had from looking at data on hard refusals and refusal conversion. Among final refusals the data show a bit lower rate of hard refusals for the cases in unusually wealthy neighborhoods than for those in the middle-wealth neighborhoods, and among the censored cases the gap is wider. The rate of attempted refusal conversion is high among the final refusals, but it is notably lower among the cases in unusually poor neighborhoods and those for whom the neighborhood was not observed. Among the censored cases, there is a notable shortfall relative to the cases in middle-wealth neighborhoods of both hard refusals and conversion attempts in all other groups. Thus, the data suggest that additional effort may have been relatively fruitful among the cases in places other than the middle-wealth neighborhoods.

---

[11]For most of the list cases where the neighborhood observation is not available, the address available for locating the respondent was most likely something other than a home address. In a fairly small number of cases, the neighborhood observations were simply not completed.

Table 4: Distribution of number of attempts to complete cases, percent of cases where refusal conversion attempted, and percent hard refusals, by observed wealth of neighborhood around sample dwellings and by final case disposition, 1998 SCF.

Final case disp.	Description of neighborhood around sample dwelling			
	Unusually rich	Unusually poor	Neither rich/poor	Not observed
Completed				
%-ile of dist. of attempts				
25	3	2	3	5
Median	6	4	5	8
75	9	8	9	12
% refusal conversion	4.1	3.0	5.0	4.8
% hard refusal	3.0	3.0	4.4	2.6
Completed cases as percent of cases with neighborhood type	40.8	80.6	58.7	27.9
Refused [*]				
%-ile of dist. of attempts				
25	6	6	6	5
Median	8	9	9	7
75	12	13	12	11
% refusal conversion	78.7	69.1	77.9	50.6
% hard refusal	51.1	54.6	52.6	38.7
Refused cases as percent of cases with neighborhood type	32.4	14.6	28.7	35.4
Censored				
%ile of dist. of attempts				
25	5	4	5	4
Median	7	8	7	6
75	11	12	11	9
% refusal conversion	20.7	22.2	29.1	18.4
% hard refusal	28.8	22.2	37.8	26.7
Censored cases as percent of cases with neighborhood type	26.7	4.8	12.6	36.6

* Excludes postcard refusals.

Figure 5: Density of number of attempts to complete a case, by wealth of neighborhood, for cases that were ultimately completed, 1998 SCF.
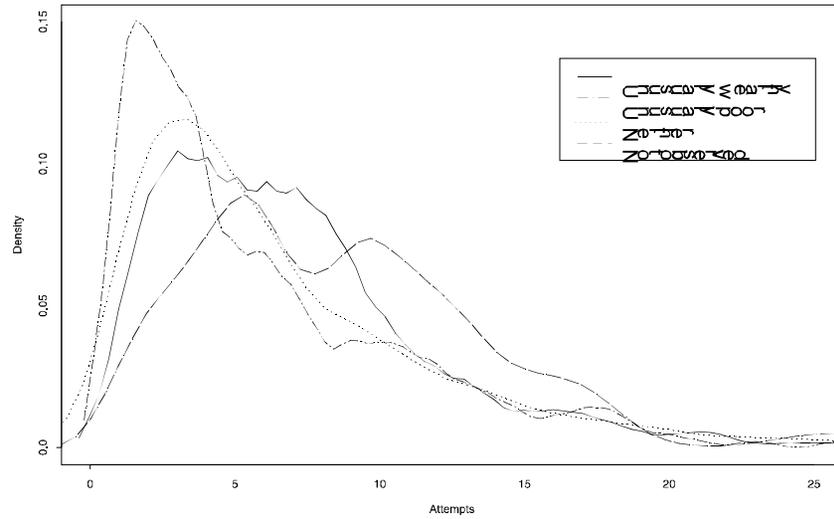

Figure 6: Density of number of attempts to complete a case, by wealth of neighborhood, for cases that were ultimately classified as refused, 1998 SCF.
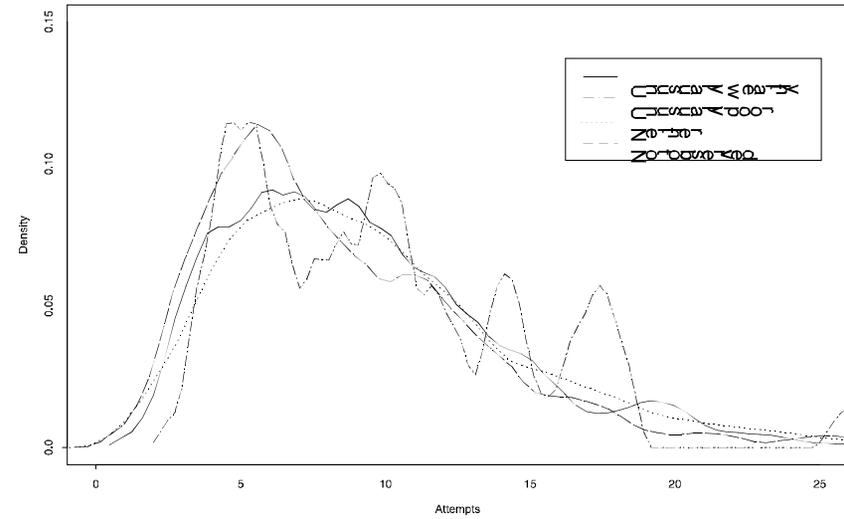

Figure 7: Density of number of attempts to complete a case, by wealth of neighborhood, for cases that were ultimately classified as censored, 1998 SCF.
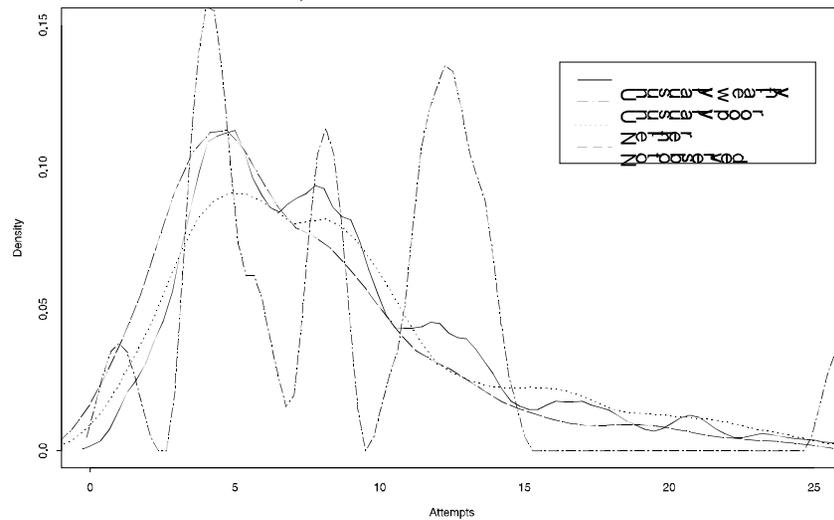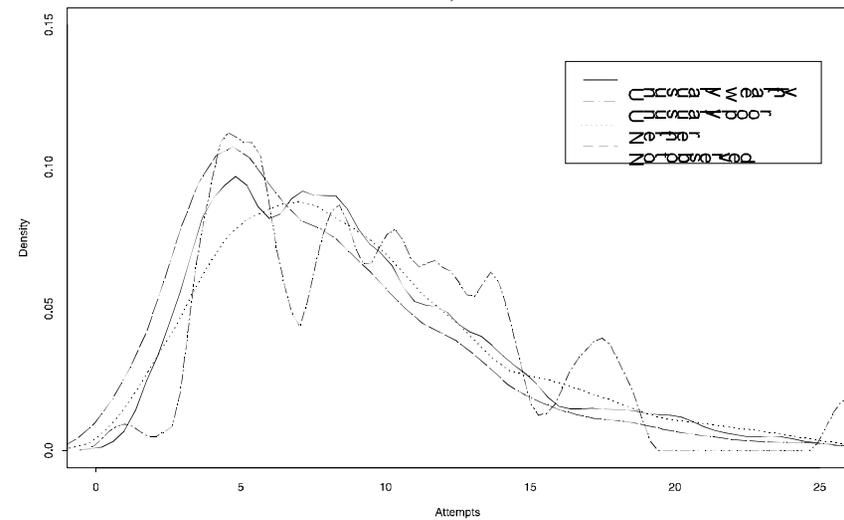

Figure 8: Density of number of attempts to complete a case, by wealth of neighborhood, for cases that were ultimately classified as refused or censored, 1998 SCF.

III. Summary and future research

This paper focuses on the role of asymmetric information between interviewers and survey managers in determining the distribution of completed cases in a survey. When interviewers and survey managers do not have the same information about the sample cases, and there is a fixed incentive scheme for interviewers, there will be a tendency for the interviewers to arrange their work to suit their own preferences. Those arrangements may well differ from the behavior desired by managers with complete information. As noted in the paper, such behavior does not deny the clear fact that many interviewers exhibit dedicated behavior that goes well beyond what would normally be seen as being in their narrow self interest. Indeed, I believe it is very likely that the managers involved in data collection do not normally give sufficiently clear information about the sampling objectives of a project for the interviewers even to develop a notion of how suiting themselves might lead to suboptimal outcomes in the distribution of completed interviews.

A simple model was presented to characterize the basic problem, and three suggestive illustrations were presented. In all three of the illustrations, the information that is used to extract possible signs of distortions in the application of effort was, in principle, knowable during the field period. Thus, one might argue that there was not meaningful asymmetry in the knowledge of this information. However, as noted in each instance, the information was not used at that time in a way that would have been informative about systematic differentials in the application of effort across cases. A particular benefit of ignoring such information then is that it is available for this analysis without being contaminated by feedback to interviewers' behavior.

Because many complex factors determine the overall propensity of a case to be completed, it is not possible to draw unambiguous conclusions from the data presented, but the evidence does support the hypothesis that in the absence of direct monitoring, the incentives interviewers face may drive them to apply effort to cases in ways that differs from the ideal. At the very least, it gives impetus to thinking about how the structure of incentives for interviewers and selective monitoring might be useful in enforcing the objectives of survey managers

In the absence of monitoring, structuring a compensation plan that disproportionately rewards interviewers for completing difficult cases could induce the desired behavior, but such plans may also induce counterproductive behavior. For example, defining difficulty in terms of

the number of attempts needed to complete a case could have the effect of driving interviewers to apply effort in an inefficient way. A compensation plan that paid off at a progressively higher rate with the number of cases completed would tend to encourage interviewers to undertake the additional work to obtain the more difficult cases. A possible problem with this plan is that in the absence of other factors, it would tend to cause completed cases to be roughly ordered during the field period in terms of difficulty; if timing is important in the analytical goals of a survey, such ordering would be problematic. Such a scheme might also be viewed as "unfair" if interviewers in geographic areas that are traditionally easier face the same rewards, though perhaps existing pay regional differentials could be extended in a way to yield an outcome that would be viewed as fair. An potentially fatal problem with such a plan is that allowing such escalation of compensation could lead to very high field costs.

Introduction of even limited monitoring raises the possibility of being able obtain similar results as under a pure compensation scheme, but at a lower cost. Information might come from teams of interviewers (such as the traveling interviewers commonly used near the end of projects, or special refusal converters), data provided by the interviewers, or external data (including frame data and data that can be linked with frame data). The first type of information could be used to monitor the behavior of individual interviewers, but it might be difficult to use in a systematic way to affect the distribution of outcomes. The other types of information might be used to impose what would effectively be *quotas* on the field staff, a term which often makes people who work on scientifically designed surveys quite nervous. However, very many surveys already use quotas of some sort, though the label may be more euphemistic. For example, survey researchers frequently refer to "maintaining response rates" and "equalizing response rates across PSUs"; both of these tactics taken in a hardened form are equivalent to quotas. In any case, the issue at stake is the effect of additional effort on the distribution of important characteristics in the survey.

At the close of a survey, many statisticians compute post-stratification weights to account for nonresponse (usually hoping to improve efficiency simultaneously). However, it is also widely recognized that even at the end of a field period, additional effort yields completed cases, but the overall rate of completion is judged too low to justify the administrative structure needed to support a field staff. If factors used as post-strata can be monitored during a survey, and if

those factors can be used to steer the application of effort, one needs to ask whether it is "better" to post-stratify or to "ante-stratify" (that is, set field quotas). If the additional cases obtained by using quotas turned out to be drawn from the same distribution as the other cases, there is a cost/statistical-efficiency tradeoff; if the additional cases represent a different distribution, then there is a cost/bias tradeoff. Often people are willing to pay more to reduce bias than to increase efficiency. Further research should explore these tradeoffs.

Bibliography

Grossman, Sanford J. and Oliver D. Hart [1983] "An Analysis of the Principal-Agent Problem," *Econometrica*, v. 51, No. 1, pp. 7-46.

Groves, Robert M. and Mick P. and Couper [1996] "Contact-Level Influences in Face-to-Face Surveys," *Journal of Official Statistics*, Vol. 12, No. 1, pp. 63-83.

Groves, Robert M. and Mick P. and Couper [1998] *Nonresponse in Household Interview Surveys*, John Wiley and Sons, New York.

Harris, Milton and Artur Raviv [1979] "Optimal Incentive Contracts with Imperfect Information," *Journal of Economic Theory*, v 20, pp. 231-259.

Holmström, Bengt [1979] "Moral Hazard and Observability," *Bell Journal of Economics*, pp. 74-91.

Kennickell, Arthur B.[1998a] "List Sample Design for the 1998 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

Kennickell, Arthur B. [1998b] "Using Income Data to Predict Wealth," paper presented at the Annual Meetings of the Allied Social Science Associations, New York, 1999a.

Kennickell, Arthur B.[1999c] "How Different Are the "Late" Cases? Evidence from the 1998 Survey of Consumer Finances," working paper, Board of Governors of the Federal Reserve Board, http://www.federalreserve.gov/pubs/oss/oss2/method.html.

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

Ross, Stephen A. [1973] "The Economic Theory of Agency: The Principal's Problem," *American Economic Review, Papers and Proceedings of the Eighty-fifty Annual Meeting of the American Economic Association*, (May) pp. 134-139.

Sudman, Seymour [1966] "Probability Sampling With Quotas," *Journal of the American Statistical Association*, v.61 (315), September, pp. 749-771.

Tourangeau, Roger, Robert A. Johnson, Jiahe Qian, Hee-Choon Shin, and Martin R. Frankel [1993] "Selection of NORC's 1990 National Sample," working paper, National Opinion Research Center at the University of Chicago, Chicago, IL.