Journal of Official Statistics, Vol. 15, No. 2, 1999, pp. 283–303

Analysis of Nonresponse Effects in the 1995 Survey of Consumer Finances

Arthur B. Kennickell[1]

This article looks at systematic patterns in unit nonresponse in the 1995 Survey of Consumer Finances. A key contribution of the article is the application of a hazard model framework to examine the contributions of the characteristics of interviewers, respondents, and respondents' neighborhoods to nonresponse.

Key words: Unit nonresponse; hazard models; interviewer effects.

1. Introduction

Unit nonresponse is a serious problem in the Survey of Consumer Finances (SCF). In the area-probability (AP) part of the SCF sample, only about 70 percent of the selected respondents agree to participate in the survey; in the relatively wealthy SCF list sample, the cooperation rate is much lower.[2] Unsurprisingly in this light, the study of nonresponse has long been a core area of research for the project. The survey is fortunate in having extensive frame data on income and some other characteristics for the entire list sample, and this information has driven most of the project research on nonresponse. This work has contributed very substantially to our ability to measure the behavior of wealthy households. However, until the 1995 survey, the only systematic information available for the AP sample has been the identity of the primary sampling unit.

This article uses information available for the 1995 SCF to look more broadly at the causes of unit nonresponse and the efforts expended to obtain completed interviews. The new data used here include information about characteristics and attitudes of the interviewers, descriptive material about the first contact with the respondent, characteristics of the respondent's neighborhood, and the administrative logs that interviewers keep to track actions for each case.

[1] Board of Governors of the Federal Reserve System, Mail Stop 153, Washington, DC 20551, U.S.A. E-mail: Arthur.Kennickell@frb.gov

Acknowledgments: The views presented in this article are those of the author alone and do not necessarily reflect the views of the Board of Governors or the Federal Reserve System. The author wishes to thank Kevin Moore and Amy Stubbendick for outstanding research assistance, and Gerhard Fries and Brian Surette for help in assembling the data for the analysis. The author also thanks Fritz Scheuren for comments on this article. The author is particularly grateful to Mick Couper and Robert Groves who generously shared the results of their earlier research and offered critical advice in designing an important part of the data collected for this analysis, and to Louise Woodburn who also participated in the design of the data. This article is dedicated to the interviewers for the 1995 SCF and the other NORC staff, without whom there would be nothing to report here.

[2] See Kennickell and Woodburn (1997) for a description of the SCF sample and more detail on response rates. Other SCF information is available on the Federal Reserve webpages at http://www.bog.frb.fed.us/pubs/oss/oss2/scfindex.html. (see e.g., Woodburn 1991, and Kennickell and McManus 1993.)

Following the research of Groves and Couper (1996) on ''tailoring'' behavior by interviewers, the article develops a set of reduced form models describing the interaction of effects attributable to interviewers, respondents, and the contextual effects of neighborhoods. An innovation in the approach here is the use of a discrete time hazard model of the resolution of the sample cases into complete or refused final dispositions.

The article has three sections. The first section gives some background on the design of the SCF and provides a basic description of unit nonresponse in the survey. The second section develops a multivariate model of nonresponse. The final section summarizes the findings of the research and outlines future work.

2. Background

2.1. Description of the survey

The SCF is a triennial survey sponsored by the Board of Governors of the Federal Reserve System, with the cooperation of the Statistics of Income Division (SOI) at the Internal Revenue Service (IRS). Data for the 1995 survey, the basis of this article, were collected by the National Opinion Research Center at the University of Chicago (NORC) between the months of June and December using computer-assisted personal interviewing. There were 246 final interviewers for the cases released to the field. The median interview required approximately 90 minutes, but some took as long as three hours. The questionnaire focuses on households' assets, liabilities, and financial relationships. Data are also obtained on employment history, pension rights, marital history, demographic characteristics, and various attitudes and expectations. (Kennickell, Starr-McCluer, and Sundén 1997 provide an overview of the data.)

The SCF employs a dual-frame sample design, including an area-probability (AP) sample and a list sample (see Kennickell and Woodburn 1997 for details). The AP sample is a multistage design with equal probabilities of selection for each household included (see Tourangeau, Johnson, Qian, Shin, and Frankel 1993). The list sample is drawn from a special sample of tax returns selected and edited by SOI for research purposes, the Individual Tax File (ITF). These data are divided into seven strata for sampling. Empirically, the first three strata overlap strongly with the AP sample in terms of their wealth and the top four strata are generally substantially wealthier. Cases in higher strata are sampled at increasingly higher rates. List respondents are treated somewhat differently from AP respondents: by agreement with SOI, list sample respondents are initially sent a postcard offering them a chance to refuse participation in the survey. All list cases not returning a postcard and all AP cases are to be pursued with equal vigor. The AP sample provided about 2,800 of the survey participants in 1995, and the list sample about 1,500.

2.2. Unit nonresponse

The general experience over the history of the survey is that respondents feel that the survey is long and that it requests particularly sensitive information. Consequently, it is not surprising that response rates have been lower than those in many other U.S. government surveys. Table 1 provides information on nonresponse for different parts of the sample. For the AP sample, nonresponse is a particular problem in the northeast region

Table 1. Response rates as a percent of eligible respondents, 1995 SCF, for various parts of the sample

All AP sample cases	66.3
Northeast region	60.1
Northcentral region	70.9
Southern region	67.2
Western region	65.3
Largest urban areas	58.9
Other cities and towns	66.6
Non-urban areas	77.6
All list sample cases	30.4
Stratum 1	45.2
Stratum 2	39.5
Stratum 3	35.5
Stratum 4	35.0
Stratum 5	30.4
Stratum 6	23.9
Stratum 7	12.8
List sample participants as a % of those not refusing by postcard	
All list sample cases	38.7
Stratum 1	54.2
Stratum 2	54.7
Stratum 3	47.5
Stratum 4	45.4
Stratum 5	38.7
Stratum 6	29.9
Stratum 7	15.1

and in more urban areas. For the list sample, response rates decline from the bottom stratum to the top stratum. Even removing the postcard refusals from the calculation, the response rates in the lower strata are still substantially below those for the AP sample. Thus, it seems that there may be some factors affecting response for the list sample that are not as strong for the AP sample. Perhaps it is the effect of being contacted more times than AP cases or being contacted specifically by name, either of which might arouse suspicion. (Cartwright and Tucker 1967 discuss an example where advance contact has negative effects.)

Table 2. Reasons for noninterview, 1995 SCF, percent of eligible sample type

	AP	*List*
Postcard refusal	NA	30.7
No contact	2.1	0.0
Unlocatable	0.1	3.9
Unavailable	0.3	3.0
Language problem	3.3	0.7
Too ill	4.4	2.0
Refused by gatekeeper	2.8	3.6
Refused, too long	17.5	16.9
Refused, too personal	47.2	18.7
Refused, gov't involvement	7.6	2.8
Stopped work	5.4	15.2
Other incomplete	9.4	2.5

Based on the final case disposition codes, almost half of the final reasons entered for nonresponse in the AP sample indicate that the respondent thought the survey was ''too personal'' (Table 2). The length of the survey is also an important factor for the group. For the list sample, the length of the survey is about as important as for the AP sample; the lower proportion coded ''too personal'' and ''government involvement'' may be explained by the elimination of the group that refused by postcard.

The data also show that a significant fraction of apparently eligible observations cannot be classified as either complete or refused. About eight percent of AP cases and about 22 percent of list cases have final completion codes of ''no contact,'' ''unlocatable,'' ''unavailable,'' or ''stopped work.'' Moreover, it appears that even these figures understate the number of such ''censored'' cases. If we take the set of incomplete cases and reclassify them as censored if the last recorded action in the record of calls indicated that the case had not been contacted on that attempt, the proportion of such cases rises to about nine percent for the AP sample and 30 percent for the list sample. I suspect that the proportion of such observations in the SCF is high relative to what might be found in other surveys, but I know of no systematic investigation of such outcomes in other surveys.

2.3. Contacts

The project interviewers were diligent in pursuing the respondents. For each sample, Figure 1 shows an average shifted histogram (ASH) – a type of kernel density estimate – plot of the distribution of the number of contacts at the end of the field period.[3] The results for both are remarkably similar. The overall median number of contacts was only 3 (mean of 4.1), but ten percent of cases had eight or more contacts, and one case had 34 contacts.[4] As shown for the AP and list sample respondents respectively, the results differ surprisingly little when broken out by final disposition. The solid lines in the figures show the distribution for cases that were resolved as completed, the dashed lines those that were resolved as refused, and the dotted lines those that were unresolved at the end of the field period. For both samples, the distribution of contacts for the refusals is shifted to the right of that for the completed cases. This outcome is expected, since efforts are expected to be made to convert refused cases until it is judged that such efforts are no longer productive. It is also striking how much more alike the distributions are for completed and refused cases in the list sample than in the AP sample. For the cases that were unresolved at the end of the field period, it is interesting (the average of ''a surprise'' and ''a relief'') how similar the distributions are to those for the sample cases that had a final resolution: both the median and mean number of contacts are virtually the same, and the distribution is no less skewed. Thus, there is no indication at this level that there was any different effort expended on cases that were never resolved.

Although contacts were monitored by the field supervisory staff, it was impossible to enforce a strict protocol without more precise information than was available without great

[3] The number of contacts for the list sample excludes the initial postcard mailing. The interviewers reported that very many respondents had no recollection of having received any earlier survey materials.
[4] For a small number of cases, there were no recorded contacts: of the 385 cases with no recorded contacts, 13 were complete cases, two were refusals, 179 were censored cases, and 191 were ineligible.

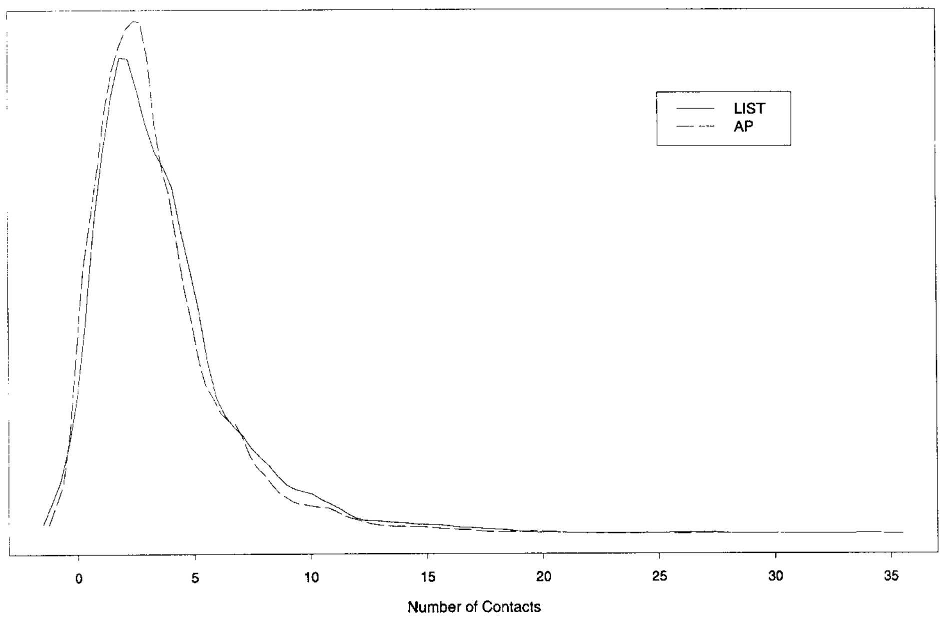

Fig. 1. ASH plot of number of contacts; AP sample and list sample, 1995 SCF

effort. Extreme numbers of calls to resolve a case have a clear monetary cost. It is also possible that some interviewers could also have ceased trying to make contact with cases that might have been particularly stressful; the potential cost of such behavior is harder to evaluate.

2.4. New data in the 1995 SCF

In the 1995 SCF, several new sources of information were added with the goal of furthering our understanding of unit nonresponse in the survey. First, new questions were added to the household enumeration folder (HEF), a paper document interviewers use to determine the respondent and record their actions on a case. The coded HEF data include a description of the first interaction with a person in the selected units, some characteristics of the informant for the initial household listing used to determine the eligible respondent, characteristics of the neighborhood surrounding the dwelling, and key items from the record of calls, a listing of all attempts to contact respondents. Second, interviewers completed a questionnaire about their own work and educational background and their attitudes. Third, ZIP code data were available for every case, and this information was used to link socio-demographic data derived from public files for the 1990 Census of Population.

There is no usable information on the record of calls for only 504 observations out of about 8,740.[5] The completion rate for the interviewer questionnaire was 100 percent, and missing information problems there are fairly small. There were minimal problems

[5] The totals exclude the 1,070 list sample cases that refused participation by returning the postcard. The cases with missing data are nearly equally divided between the two samples; only about 30 were incomplete cases, about 70 were ineligible, about 100 are in the censored group, and the remainder were refusals. There are also more minor problems with missing data within the record of calls.

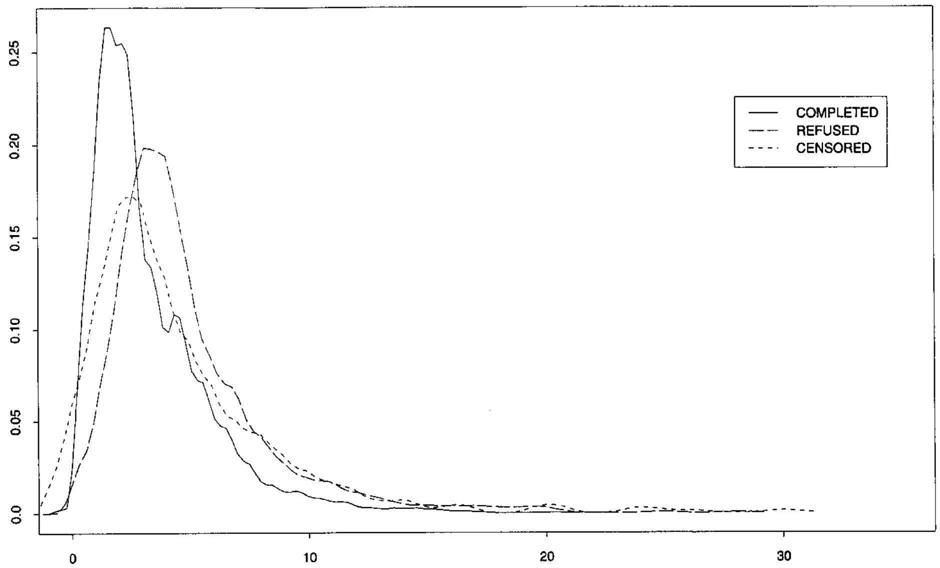

Fig. 2. ASH plot of number of contacts, by final disposition; AP sample, 1995 SCF

in matching the geographic variables by ZIP codes. Unfortunately, there are many cases with missing data among the variables in the HEF describing the structure of the sample household (about 4,100 cases) and those describing the first contact with the HEF respondent (about 1,500 cases).[6] Interestingly, these missing data problems were widely spread over the whole group of interviewers, rather than being concentrated in a smaller group. Logit modeling indicates interviews more likely to have substantial missing enumeration data were in ZIP codes with larger proportions of college educated adults, and less likely when the interviewer was older and more experienced and when the respondent was a list sample case selected from a stratum with high predicted wealth. Because of the severity of this missing data problem, the models reported in the next section that use the enumeration and contact variables should only be taken as suggestive.

3. Models of Unit Nonresponse

3.1. Background

The interactions between interviewers and respondents are at the heart of the survey process, but very many of the events that occur at that level either are unmeasurable without severe disruption of the interview, or are very difficult to define objectively. Most of the early research on these interactions examined behavior during an interview. Study in this area dates at least to Rice (1929) who studied the effects of interviewer beliefs (in his case, about prohibition) on the answers respondents give. Hanson and Marks (1958) focused on the relationship between interviewer characteristics and data quality in an experiment using the 1950 Census of Population. Cannel, Fowler, and Marquis (1968)

[6] The observations with missing enumeration data include about 1,300 completed cases, 1,700 refusals, and 1,100 cases that were neither complete nor refused at the end of the field period. Data on the first contact are missing in roughly equal numbers for completed, refused and censored cases.

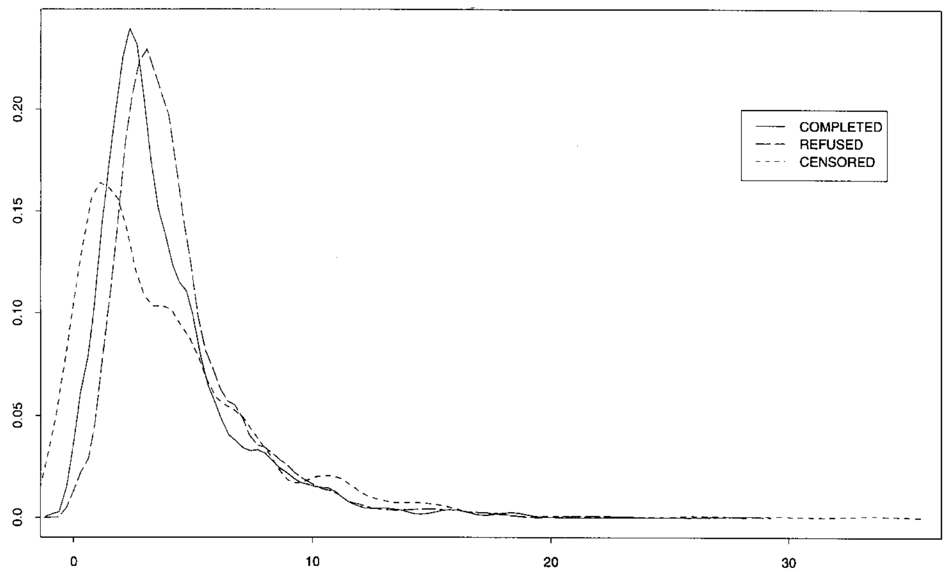

Fig. 3. ASH plot of number of contacts, by final disposition; list sample, 1995 SCF

devised a sophisticated study merging data from the Health Interview Survey, a set of observations of interviewer behavior, a questionnaire administered to respondents about the original interview, and an interview with the interviewers. Their results indicated that attitudinal variables had at best a minimal influence on accuracy, but behavioral variables had strong effects.

More recently, there has been much discussion of unit nonresponse in the literature (for extensive citations see Bogeström, Larsson, and Lyberg (1981) and Holt and Elliot (1991). However, most of this research has dealt with the effects of nonresponse in estimation and possible remedies through weight adjustments (see e.g., Little (1993)). Recent work by Groves and Couper (1993a, 1993b, 1996) has developed a theory of response and assembled a variety of types of information to test aspects of the theory. Because of the importance of their efforts for interpreting the research reported in paper, it is useful to describe their work briefly.

Groves and Couper hypothesize that two factors should figure prominently in interviewers' strategies as they approach a meeting with a respondent: keeping the respondents engaged, and tailoring their remarks throughout the interaction in a direction expected to avoid a permanent refusal. Interviewers may differ in their ability to generate responses to the subjects' reactions that lead in a positive direction, and in their ability to decode the cues respondents provide. The authors assembled several sources of data. As a part of the National Survey of Health and Stress, a long interview on physical and mental health, interviewers obtained information describing the person with whom the negotiation for the interview took place and the events that occurred during the interaction. The interviewers also maintained a record of calls for each sample element, and they filled out a questionnaire about their own backgrounds and attitudes. These data are brought together in a series of models describing the success of each of the first through fourth contacts with the sample households as well as an overall model of response.

In their models, they find some strong effects, and a general weakening of effects with repeated contacts with a respondent. For example, barriers to entry have a negative effect on completing a case, but the effect fades with additional contacts. They find that success is less likely with one-person households and male respondents, and more likely where the interviewer is confident. Interestingly, interviewer experience has no effect. Initial negative statements and time delay statements made by a respondent have a persistent negative effect over repeated contacts. A measure of the degree to which interviewers tailored their interactions from one contact to the next has no significant effect, perhaps because of the crude nature of the measure. There are problems with their models. As in the SCF study reported here, interviewer assignment was nonrandom, and many important variables are unobserved.

3.2. Prior SCF work on nonresponse

Most prior work on nonresponse in the SCF has focused on the list sample. Research reported in Woodburn (1991) investigated the effects of post-stratification for nonresponse adjustment in the list sample. Kennickell and McManus (1993) used more detailed information in the list sample frame to develop models of nonresponse for this group.[7] In these models, about three-quarters of the explanatory power came from a measure of financial income, with higher levels of this variable correlating with a lower response propensity.[8] Other important contributing factors were nontaxable income (positive effect on response); pension income (positive effect); real estate taxes (negative effect); wage and salary income (negative effect); estate, trust or royalty income (negative effect); age (negative effect); residence in the Western or Southern regions (positive effect); residence in California or any self-representing PSU (negative effect). The results of these models support the structure of the nonresponse adjustments applied to the SCF weights for the list sample. Unfortunately, because the variables available for modeling are so highly aggregated and abstract, it is difficult to extract much insight into the behavioral mechanisms that underlie the decision to participate in the survey.

3.3. Analysis using the 1995 SCF data

Respondents and interviewers come together usually with different information and perceptions about each other, and with very different incentives. Their interaction is a two-part behavioral game (or multi-part if we allow for the effects induced by supervisors, survey organizations, and principal investigators).

The intended role of the interviewer during the negotiation stage is to communicate information to the respondent that will lead to an agreement to complete an interview. Interviewer behavior is influenced by a number of factors. As with other workers, it is important that they perform sufficiently well to keep their jobs. There is monitoring of interviewers' performance along several axes, including the proportion of cases they complete, and some indications of the quality of the data collected. However, it seems likely that interviewers are driven by other, less traditional incentives as well. It is striking

[7] The variables constructed from the ITF include a number of income, tax, and other dollar values, the age of the filer, geographic information, and other variables related to the SCF sample design.
[8] Financial income includes all types of interest and dividend income.

how frequently the SCF interviewers talk about the importance of the research that gets done with the data they collect, the interest they have in the lives of other people, the adventure they find in visiting strangers in unusual places, and the appreciation they have of the independence of their work. While it is clear that they find most respondents enjoyable, there are sometimes very stressful and unpleasant interviews. Potential SCF interviewers are made aware of the nature of the survey, and they are selected based on their past performance and credentials, and at least implicitly on their ability to deal with strangers with a reasonable lack of fear. Because there is generally other work that competes in the same salary range as interviewing, experience is likely to weed out people who do not fit the desired profile. SCF interviewers are also extensively trained in order to minimize variations in technique. Nonetheless, many important variations likely remain in this group.

Randomization in the SCF sample designs virtually guarantees that respondents are more varied than interviewers. Respondents are taken to have a set of preconceptions and an internal structure that determines their responses to stimuli. Prominent among the factors that might influence respondents in their willingness to participate in an interview are: a desire for attention or company, a sense of the competing uses or value of their time, their past experience with surveys, their sense of social integration and the value of public service, their faith in government, their sense of their physical security, and their feelings about privacy. Respondents' reactions to an interview may also be shaped by their education or sophistication. It may also be that respondents who understand a survey and who feel themselves to be particularly interesting in the context of the survey might also be made particularly suspicious. No doubt there are many other psychological and demographic considerations that also enter into a decision to cooperate in an interview.

Although it would be interesting to model separately the interviewers' efforts and the respondents' receptivity, our ability to monitor what actually happens during the negotiations between interviewers and respondents is very limited. The work reported here takes a reduced form approach, focusing on the factors influencing the resolution of cases into ''completes'' and ''refusals.'' These resolutions are taken to be indicators for a latent variable reflecting something one might call the respondents' ''enthusiasm'' – denoted $\mathbf{E}$, where this variable is a function of respondents' pre-existing attributes and their cumulative reactions to the interviewer. If $\mathbf{E}$ rises above a certain upper level $\mathbf{E}^+$, the respondent completes the interview, and if it falls below a certain lower level $\mathbf{E}^-$, the respondent refuses ''permanently.'' Until a respondent passes up to or beyond either $\mathbf{E}^+$ or $\mathbf{E}^-$, the respondent remains ''at risk'': for respondents at risk, all we know is that their level of $\mathbf{E}$ lies between $\mathbf{E}^+$ and $\mathbf{E}^-$.

We might approach modeling the outcomes in several ways. One might simply model overall response versus all other outcomes, as has been common in most of the literature, or the probability of response at a given contact, as in Groves and Couper (1996). This article adopts a different approach.[9]

A respondent's decision at each contact to participate, refuse participation, or to stop

[9] Estimates comparable to those of Groves and Couper are given in the working paper version of this article on the Internet at http://www.bog.frb.fed.us/pubs/oss/oss2/method./html.

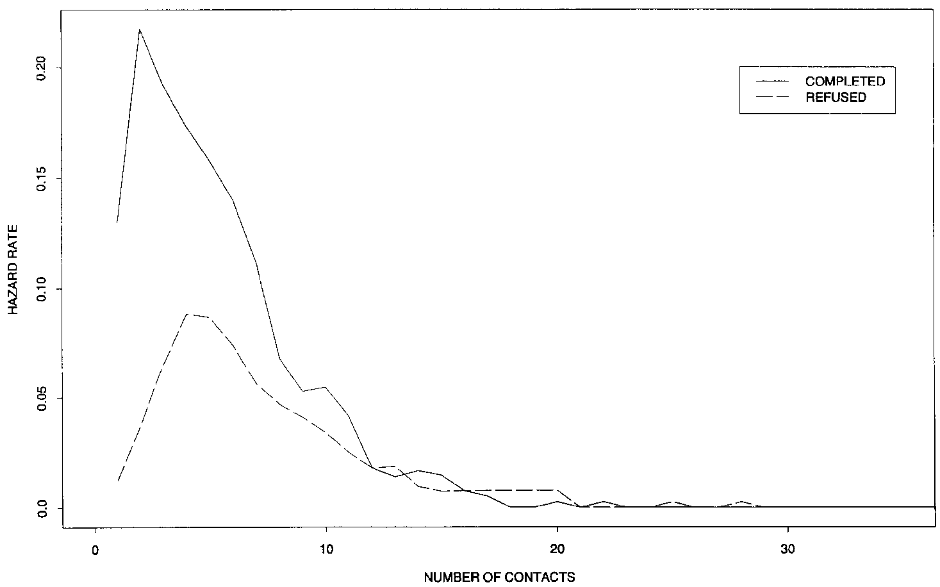

Fig. 4. Hazard rate over contacts, by final disposition; AP sample, 1995 SCF

short of either fits within the framework of a discrete time hazard model.[10] A classic example of the application of a hazard model is in the biometric literature where there is the possibility that a person under study might die of a number of causes over the period of observation or might continue to live at risk of dying through the period of observation. For the model considered here, the exit states are completed status and refused status, and the population at risk at each contact consists of the cases that have not received a final disposition as of the previous contact. Cases that cease to be contacted before they achieve a final resolution into complete or refused cases are treated as censored. The time dimension is taken to be indexed by contacts with the survey respondent.[11] The general form assumed for the model is a form of logit:

$$\log \frac{P_{ijt}}{P_{i0t}} = \beta_j X_{it}$$

where P_{ijt} is the probability of outcome j for case i at time t, P_{i0t} is the probability that case i remains at risk after period t, X_{it} is a vector of possibly time-varying covariates for case i at time t, and β_i is a vector of parameters conformable with X_{it}. Because the likelihood function is the product of the probabilities at each period observed and β_i is not time-varying, the model can be estimated using a standard multinomial logit procedure

[10] Another possibility might be to model the process incorporating the ordering implied by **E** using, for example, a version of the ordered probit for repeated events. Although a simple two-state hazard model should be less efficient than a correctly specified model incorporating the ordering, it is also a more flexible form than the ordered probit: the ordered probit estimates one set of coefficients with an event-specific shift parameter, while the hazard model allows a full separate set of parameters for each outcome. Investigations not reported here suggest that the simple ordered probit model is insufficiently flexible to capture the asymmetric effects captured by the hazard model.

[11] There are other possible choices for the time dimension in the model, most notably attempts on a case. The general distribution of contacts and attempts is similar in shape. However, because the coding of case actions is insufficiently strong to distinguish trivial actions from serious actions, the variable attempts appears to be too noisy an indicator to use in modeling.

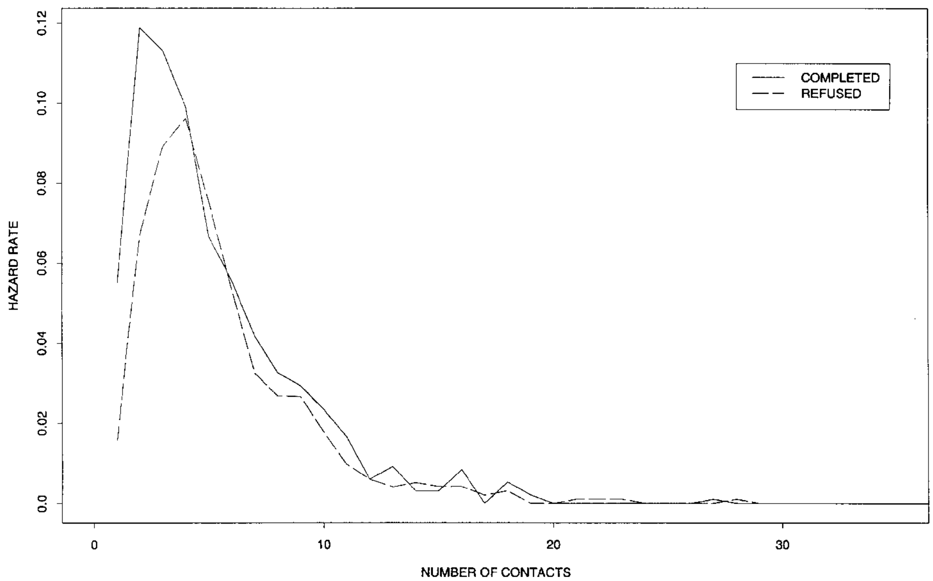

Fig. 5. Hazard rate over contacts, by final disposition; list sample, 1995 SCF

taking each period for each survey case that is still at risk at that point as a separate observation (see Allison 1984, 1995 for a discussion of the estimation of discrete time hazard models. The SAS procedure CATMOD was used for estimation.)

Plots of the unconditional discrete time hazards of resolving a case as a complete interview or a final refusal are shown in Figures 4 and 5 for the AP and list samples respectively.[12] The general shape of each of the plots is very similar: there is a sharp initial rise in the hazard, followed by a sharp decline and a trailing off of the rate. The fact that the hazard first rises and then declines most likely reflects two factors. First, many respondents express a desire to read the study materials, confirm the authenticity of the study, or simply think over the decision to participate. Second, reluctant respondents (even quite reluctant ones) are given additional information in subsequent attempts or exposed to different interviewers until the respondent unequivocally refuses to participate. In practice, the two effects are entangled. As expected, the refusal hazard for the list sample cases is initially much greater than that for the AP cases. Overall, the shape of these plots reflects the simple intuition that ''the easy cases resolve first.'' The important question is, what are the systematic components of this process?

Although there are interesting new data available for modeling the hazards, the information is still limited relative to the task. It is particularly problematic that to understand unit nonresponse more fully, we need information on the characteristics of the respondent, which are very likely to be unavailable from the respondent in cases where that person wishes strongly to avoid giving the interviewer information. The first column of Table 3 shows the simplest model incorporating variables constructed by matching sample

[12] The hazard for refusal is computed for each contact as the number of cases that resolve at that point as refusals, divided by the number of cases still ''at risk'' just before that contact less the number of observations censored at that point. This calculation and those that follow exclude the list sample cases that refused participation by returning the postcard.

Table 3. Discrete time hazard models of completion and refusal, 1995 SCF

INTRCPT	1.53+	1.41+	1.59*	3.06+	ITALK1	.	.	0.01	−0.01
	0.51	*0.51*	*0.72*	*1.13*		.	.	*0.02*	*0.02*
	−5.21+	−3.82+	−0.55	−7.86*		.	.	0.01	0.10
	0.77	*0.78*	*1.15*	*3.92*		.	.	*0.03*	*0.08*
CCCMSA	0.05	0.04	−0.05	0.21	BARR	.	.	.	0.04
	0.06	*0.06*	*0.07*	*0.11*		.	.	.	*0.11*
	0.37+	0.23+	0.22*	0.17		.	.	.	0.25
	0.08	*0.08*	*0.09*	*0.33*		.	.	.	*0.31*
OCMSA	−0.09	−0.09	−0.06	−0.02	RHRES	.	.	.	−0.08
	0.05	*0.05*	*0.06*	*0.09*		.	.	.	*0.11*
	0.06	0.04	0.04	−0.03		.	.	.	0.05
	0.06	*0.06*	*0.07*	*0.24*		.	.	.	*0.40*
MSA	−0.40+	−0.42+	−0.38+	−0.48+	POOR	.	.	.	0.18
	0.05	*0.05*	*0.06*	*0.08*		.	.	.	*0.10*
	0.04	−0.26*	−0.16	−0.37		.	.	.	0.00
	0.10	*0.11*	*0.12*	*0.30*		.	.	.	*0.35*
PWHITE	0.37+	0.43+	0.33+	0.37	RICH	.	.	.	−0.12
	0.12	*0.12*	*0.13*	*0.20*		.	.	.	*0.08*
	0.54+	0.56+	0.80+	0.84		.	.	.	−0.57*
	0.17	*0.19*	*0.21*	*0.66*		.	.	.	*0.25*
PGT65	−2.39+	−2.54+	−2.44+	−2.14*	MALE	.	.	.	0.02
	0.54	*0.54*	*0.61*	*1.00*		.	.	.	*0.06*
	−0.23	−1.21	−0.85	0.70		.	.	.	−0.43*
	0.76	*0.78*	*0.88*	*3.44*		.	.	.	*0.18*
AHHSZ	0.13	0.14	0.19*	0.24	ALE30	.	.	.	−0.33+
	0.08	*0.08*	*0.08*	*0.14*		.	.	.	*0.09*
	0.11	0.02	0.07	0.35		.	.	.	−0.85+
	0.10	*0.11*	*0.12*	*0.41*		.	.	.	*0.33*
PCOLL	0.53*	0.51*	0.62*	1.22+	A31_40	.	.	.	−0.31+
	0.25	*0.25*	*0.27*	*0.41*		.	.	.	*0.08*
	−0.12	0.14	0.03	−0.67		.	.	.	−0.25
	0.34	*0.35*	*0.38*	*1.24*		.	.	.	*0.25*

Table 3. *Discrete time hazard models of completion and refusal, 1995 SCF (continued)*

PMWK	−1.22+	−1.41+	−1.83+	−1.31	A41_50	.	.	.	−0.20+
	0.43	*0.43*	*0.47*	*0.73*		.	.	.	*0.07*
	0.19	−0.84	−1.54*	−1.07		.	.	.	−0.23
	0.65	*0.66*	*0.72*	*2.51*		.	.	.	*0.23*
PFWK	0.66	0.76	0.92	−0.01	ONEP	.	.	.	−0.77+
	0.43	*0.44*	*0.47*	*0.70*		.	.	.	*0.15*
	0.30	1.00	1.50*	1.45		.	.	.	1.14+
	0.61	*0.64*	*0.69*	*2.34*		.	.	.	*0.23*
ATRAV	−0.21+	−0.22+	−0.19+	−0.02	INFOQ	.	.	.	0.09
	0.04	*0.04*	*0.05*	*0.07*		.	.	.	*0.06*
	0.09	0.01	0.07	0.11		.	.	.	0.35
	0.06	*0.06*	*0.07*	*0.24*		.	.	.	*0.20*
MHVAL	−0.22+	−0.21+	−0.22+	−0.26+	TIMEQ	.	.	.	−0.10
	0.04	*0.04*	*0.05*	*0.08*		.	.	.	*0.06*
	0.11	0.00	−0.02	0.26		.	.	.	−0.56+
	0.06	*0.06*	*0.07*	*0.23*		.	.	.	*0.19*
IEXP	.	.	−0.01	−0.01	INCENQ	.	.	.	0.12
	.	.	*0.01*	*0.02*		.	.	.	*0.16*
	.	.	0.04+	0.05		.	.	.	−1.22
	.	.	*0.01*	*0.05*		.	.	.	*0.75*
ICOMEX	.	.	0.09*	0.14*	RNEG	.	.	.	−0.47+
	.	.	*0.05*	*0.07*		.	.	.	*0.07*
	.	.	−0.05	0.00		.	.	.	0.51+
	.	.	*0.07*	*0.23*		.	.	.	*0.18*
ICOLL	.	.	−0.02	0.02	RDELAY	.	.	.	−0.50+
	.	.	*0.06*	*0.08*		.	.	.	*0.06*
	.	.	0.09	−0.30		.	.	.	−0.05
	.	.	*0.09*	*0.26*		.	.	.	*0.18*

Table 3. Discrete time hazard models of completion and refusal, 1995 SCF (continued)

IAGE	.	.	−0.01	−0.24	DAYS	.	−0.03+	−0.03+	−0.01
	.	.	*0.10*	*0.16*		.	*0.01*	*0.01*	*0.01*
	.	.	−0.82+	−0.40		.	0.14+	0.14+	0.14+
	.	.	*0.15*	*0.54*		.	*0.01*	*0.01*	*0.02*
ICONV	.	.	−0.05*	−0.03	NATT	.	0.03+	0.03+	0.05+
	.	.	*0.02*	*0.03*		.	*0.00*	*0.01*	*0.01*
	.	.	−0.28+	−0.20*		.	0.02+	0.02*	0.04
	.	.	*0.03*	*0.10*		.	*0.01*	*0.01*	*0.03*
IOUTGO	.	.	0.07*	−0.03	NCON	.	0.01	0.01	0.00
	.	.	*0.03*	*0.04*		.	*0.01*	*0.01*	*0.01*
	.	.	0.12+	0.24		.	−0.06+	−0.04+	−0.01
	.	.	*0.04*	*0.14*		.	*0.01*	*0.01*	*0.04*
ICURIO	.	.	−0.09+	−0.08*	NEW1	.	0.18+	0.09	0.18*
	.	.	*0.02*	*0.03*		.	*0.04*	*0.05*	*0.07*
	.	.	−0.02	−0.02		.	1.09+	1.02+	0.86+
	.	.	*0.03*	*0.09*		.	*0.06*	*0.06*	*0.20*
INEIGH	.	.	0.03	0.08	LSSTGE4	−0.31+	−0.32+	−0.33+	−0.11
	.	.	*0.03*	*0.04*		*0.03*	*0.03*	*0.03*	*0.06*
	.	.	−0.03	0.19		0.10+	0.09*	0.11+	0.08
	.	.	*0.04*	*0.13*		*0.04*	*0.04*	*0.04*	*0.06*
IRES	.	.	0.03	0.07	LSSTLT4	0.04	0.10*	0.14+	0.31
	.	.	*0.04*	*0.06*		*0.04*	*0.04*	*0.05*	*0.17*
	.	.	0.22+	−0.05		−0.02	0.10	0.05	0.08
	.	.	*0.07*	*0.28*		*0.05*	*0.05*	*0.06*	*0.18*
IHAM	.	.	−0.03	0.02					
	.	.	*0.02*	*0.03*					
	.	.	−0.23+	−0.19	N_EVENTS	32434	32434	27564	10037
	.	.	*0.03*	*0.10*	N_CASES	7524	7524	6443	2111

Variable definitions for Table 3

INTRCPT: Model intercept.
CCCMSA: Dummy variable: R's residence in center city of a CMSA (1 = included).
OCMSA: Dummy variable: R's residence is in a non-center-city CMSA (1 = included).
MSA: Dummy variable: R's residence is in an MSA (excluding CMSAs) (1 = included).
PWHITE: Fraction of residents of R's ZIP code who are white.
PGT65: Fraction of residents of R's ZIP code who are age 65 and older.
AHHSZ: Average number of people in households in R's ZIP code.
PCOLL: Fraction of adults in R's ZIP code with at least some college education.
PMWK: Fraction of adult males in R's ZIP code who are in the labor force.
PFWK: Fraction of adult females in R's ZIP code who are in the labor force.
ATRAV: Average number of minutes workers in R's ZIP code travel to get to work divided by 10.
MHVAL: Logarithm of the median dwelling value in R's ZIP code.
IEXP: Logarithm of years of interviewer's experience.
ICOMEX: Dummy variable: interviewer experience with computers (1 = experienced).
ICOLL: Dummy variable: interviewer has at least some college education (1 = college).
IAGE: Logarithm of the age of the interviewer.
ICONV: Scale variable: interviewer believes every R can be converted with enough effort (1 = strongly disagree, 5 = strongly agree).
IOUTGO: Scale variable: interviewer considers self outgoing (1 = strongly disagree, 5 = strongly agree).
ICURIO: Scale variable: interviewer curious about other people and what they do (1 = strongly disagree, 5 = strongly agree).
INEIGH: Scale variable: interviewer enjoys challenge of unfamiliar neighborhoods (1 = strongly disagree, 5 = strongly agree).
IRES: Scale variable: interviewer likes being part of a research project (1 = strongly disagree, 5 = strongly agree).
IHAM: Scale variable: interviewer thinks of self as a bit of an Actor (1 = strongly disagree, 5 = strongly agree).
ITALK1: Scale variable: interviewer believes it is better on the first contact to keep a conversation going rather than press for a quick decision (1 = strongly disagree, 5 = strongly agree).
BARR: Dummy variable: barriers (including physical barriers and gatekeepers) to contacting R (1 = barriers).
RHRES: Dummy variable: by observation, R's neighborhood mostly residential (1 = residential).
POOR: Dummy variable: by observation, R's neighborhood is poor (1 = poor).
RICH: Dummy variable: by observation, R's neighborhood is rich (1 = rich).
MALE: Dummy variable: R for listing was male (1 = male).
ALE30: Dummy variable: R for listing was aged 30 or younger (1 =<= 30).
A31_40: Dummy variable: R for listing was aged 31 to 40 (1 = 31 to 40).
A41_50: Dummy variable: R for listing was aged 41 to 50 (1 = 41 to 50).
ONEP: Dummy variable: R for lives alone (1 = alone).
INFOQ: Dummy variable: R for asked for information about the survey at the first contact (1 = asked).
TIMEQ: Dummy variable: R asked about the length of the interview at the first contact (1 = asked).
INCENQ: Dummy variable: at the first contact, R asked about the possibility of monetary incentives (1 = asked).
RNEG: Dummy variable: at the first contact, R made negative comments about the survey (1 = made comments).
RDELAY: Dummy variable: at the first contact, R made comments to delay interview (1 = made comments).
NOREF: Dummy variable: at the first contact, R did not refuse to do interview on first contact (1 = did not refuse).
DAYS: Number of days elapsed since first attempt on case, divided by 10.
NATT: Number of attempts made on case including current contact.
NCON: Number of contacts made on case including current contact.
NEWI: Dummy variable: interviewer changed since case originally fielded (1 = changed).
LSSTGE4: Dummy variable: case in list sample strata 4 or higher (1 = included).
LSSTLT4: Dummy variable: case in list sample strata less than 4 (1 = included).

observations by ZIP code with characteristics measured in the 1990 Census, and some terms describing the sample design. The matched census variables are available for almost all cases.[13] The variables selected for modeling here include the percent of nonwhites in the neighborhood, the percent of residents older than age 65, the percent of adults who have at least some college education, the percent of adult males working, the percent of adult females working, the average household size, the median house value, the average commuting time, and the degree of urbanicity of the neighborhood.[14] These variables reflect three effects: (1) the pure effects of neighborhood context, (2) indirect characteristics of respondents who choose to live in such areas, and (3) other unobserved characteristics of the respondent that may happen to be correlated with the variables. Some of these variables are included to allow for obvious demographic variation. The percent of workers, household size, and commuting time are intended largely to reflect characteristics related to the value of time. Such effects are also likely captured by the income and house value variables. To allow for some differences in the two samples, all the models shown also include dummy variables indicating whether an observation derived from the bottom three strata of the list sample or the higher strata of that sample.

In this model and those that follow, the cases included are those that had at least one contact and for which the variables in the models contain no missing data. The first line for each variable in the table shows the estimated marginal effect on the propensity to complete an interview compared to remaining unresolved, and the third line shows the effect on the propensity to refuse compared to remaining unresolved. The second and fourth lines give standard errors for the coefficients above them. A ''+'' indicates that an estimate is significant at the 1 percent level, and a ''*'' indicates that it is significant at the five percent level.

The pure geographic effects are limited, but interesting. Respondents living in central areas of the largest cities are more likely to refuse than people living in nonurban areas (on average subjected to higher levels of stimuli?), but they are not different in their response propensity. Those outside the central areas of the largest cities are not significantly different from those in nonurban areas. Respondents in other cities are less likely than those in nonurban areas to give a complete interview (smaller populations may raise questions of privacy?), but are no different in their propensity to refuse.

Most of the neighborhood variables have strong effects: cases in neighborhoods that are disproportionately white in their racial composition are more likely to be resolved overall, but refusals are the more likely outcome. Neighborhoods with higher concentrations of people over the age of 65 are less likely to give an interview (security issues or suspicion?), but are no different in their refusal propensities. Neighborhoods with higher proportions of college graduates are more likely to complete an interview, suggesting that respondents who are more educated may be more likely to understand and appreciate the purpose of the survey. Two of the variables expected to proxy for the value of the respondents'

[13] The ZIP code information for AP cases is based on the actual sample address, but the code for list sample cases is taken from the original address from which their tax return was filed. Although tax filers are required by law to use their home address on their tax return, it is clear from interviewers' remarks that many list respondents were, in fact, interviewed elsewhere. Thus, neighborhood characteristics may be measured with error for such observations.

[14] Other variables, such as the median household income, are also available, but failed to account for significant additional variation in other exploratory modeling.

time have significant effects: neighborhoods with higher proportions of working males and neighborhoods where people have longer commuting times to work are less likely to complete interviews, though they are no different in terms of their refusal propensities. Consistent with earlier SCF findings of a wealth effect in nonresponse, cases in neighborhoods with higher housing values were significantly less likely to complete an interview. As expected, relative to AP cases the observations from the higher strata of the list sample are more likely to refuse and less likely to complete an interview; the cases from the lower strata are not significantly different from other cases in terms of their estimated hazards. Generally, these effects persist in the more complicated models below.

The hazard model offers a convenient way of including contact-varying characteristics. The model in the second column of Table 3 adds a variable indicating whether the interviewer at a given contact is a different one from the one who started the interview, and variables intended to capture time effects, including the number of days elapsed from the first contact to the current contact, the total elapsed number of attempts (including contacts), and the elapsed number of contacts. Cases that have been taken over by a new interview are strongly more likely to be resolved overall, but such events are significantly more likely to be refusals; this outcome undoubtedly reflects the fact that most changes of interviewer take place when a case has already given a refusal just short of a final one and it is believed that a different interviewer might ''convert'' the case. Unsurprisingly, the larger the number of days a case has been ''in play,'' the more likely it is to exit as a refusal and less likely as a complete case. The effect of ''persistence'' is shown in the coefficients on number of attempts: more attempts correlate with higher probability of exit in both states. Increasing numbers of contacts lessen the likelihood of exiting as a refusal; this result could be taken to suggest that the personalization of the process over repeated contacts makes it harder for a respondent to make a firm refusal. Alternatively, the result may simply reflect unobserved heterogeneity in the population modeled (see Allison 1995).

The model in the third column of the table adds variables obtained from the questionnaire administered to the project interviewers. The values entered into the model are based on the responses of the interviewer who was working on each case at each contact. The variables included are selected from a much larger number through initial modeling with simpler estimation methods (e.g., probit models of overall response, or response given that a case was still at risk at a given contact.)

Cases assigned to more experienced interviewers are more likely to resolve as refusals; this result likely reflects a tendency to assign more difficult cases to more experienced interviewers. Previous computer experience is associated with a higher completion propensity; perhaps such interviewers appear more ''professional'' to respondents. Cases administered by college educated interviewers do not differ significantly in their response hazards. Older interviewers are less likely to have refusals; this result accords with survey ''folklore'' that respondents find it harder to say ''no'' to older interviewers. However, the propensity for completing an interview is not significantly different for cases approached by older interviewers. Interviewers who are relatively confident that they can persuade reluctant respondents are actually less likely to obtain either final resolution, but refusals are relatively less likely than completions. Outgoing interviewers are more likely overall to resolve their cases. Interviewers who think of themselves as being a little like actors are significantly less likely to have refusals; this group may be particularly good at tailoring their remarks to deal with respondents' reservations. Those who favor a strategy to emphasize engagement with the respondent on the first contact do not have notably different outcomes. Interviewers who are relatively curious about other people are likely to have lower completion rates. Curiously, interviewers who have relatively greater interest in the research are significantly more likely to have their cases resolve as refusals.

The fourth model adds variables based on data interviewers recorded about the respondents on the first contact and about respondents' neighborhoods on the first in-person attempt. Because, as noted earlier, the missing data rate is very high for these variables, the model estimates should be taken as merely suggestive. Cases with barriers (either physical ones or gatekeepers) are not significantly different from other cases; perhaps barriers are more important in determining the possibility of contact at all. There is a counterintuitive lower propensity for cases in ''rich'' neighborhoods to refuse; because of the presence of the other economic controls, this may indirectly reflect characteristics of neighborhoods that have changed since the 1990 Census. Otherwise, the interviewers' perceptions of the relative prosperity of neighborhoods have little effect. Contrary to the customary presumption, male respondents appear less likely to refuse, though they are no different in their propensity to complete a case. Younger respondents tend to be less likely to achieve a final resolution of their interviews. Not surprisingly, single-person households were both more likely to refuse and less likely to complete an interview; security concerns are likely to be important for such cases. Respondents who asked informational questions or questions about possible incentives to participate do not appear to differ from other respondents. However, those who made negative comments at the time of the first contact were more likely to resolve as refusals and less likely to resolve as completed cases. Respondents who asked questions about the length of the interview were less likely to refuse, but those who made comments indicating that they wanted to delay the interview were less likely to resolve as completed cases.

The clearest problem in these models is the fact that cases are not randomly assigned to interviewers. Almost surely, there are also important dimensions of unobserved heterogeneity across the sample cases, though the expected effect of such omissions should be to bias coefficients toward zero.[15] However, there are also several noteworthy potential problems that entail the condition of ''informative censoring,'' which occurs when censored cases would have been more likely to have exited in one state or another had they been contacted a sufficient number of additional times.

There are at least three mechanisms that might induce informative censoring in the SCF sample. First, unless a respondent refuses very strongly, he or she should be pursued until he or she does so or agrees to complete an interview. Given the tremendous pressures on interviewers to produce completed cases, it would be very surprising if they attempted to contact every case with equal vigor, particularly those they might have believed to have been more likely refusals. Second, during the field period, a concerted effort is made to avoid (to the degree feasible) large disparities in completion rates across PSUs, and there are fairly hard targets for numbers of cases in the various list sample strata. Although this

[15] Estimated standard errors and significance tests are not affected by unobserved heterogeneity bias (Allison 1995).

balancing has some desirable effects (particularly on estimated variances), it may induce differences in effort since it is clear that cases are not equally difficult in all areas. Third, some respondents may make themselves hard to contact rather than have to deal with an interviewer, and such people may be more likely to have refused had they been contacted further.

Informative censoring may lead to complex biases, and there are no simple tests for bias. However, sensitivity tests excluding the censored cases from the modeling altogether suggest that informative censoring may not be a big problem. Moreover, the distribution of effort in Figure 4 showing very similar patterns of effort across censored and fully resolved cases also offers some comfort.

4. Future Research

The discrete time hazard model developed in this article suggests that there are previously undetected dimensions of differential nonresponse in the SCF. At the least, the distinct response patterns at the level of characteristics from census data suggest that the design of nonresponse adjustments should consider variation across sampled areas in factors such as house values, commuting time, and the proportion of older people.

The results confirm the intuitive proposition that more effort leads to a greater likelihood that a case is resolved. Perhaps more importantly, the data suggest that the increased personalization of the relationship that comes with repeated contacts between respondents and interviewers lessens the probability that a case will refuse.

Older interviewers are less likely to obtain outright refusals, though they are no different in their propensity to gain complete interviews. Similarly, interviewers who view themselves as being somewhat like actors are less likely to obtain refusals. Some other interviewer effects are paradoxical. For example, experience as an interviewer seems to increase the likelihood that a case will resolve as a refusal. This result may reflect the assignment of more difficult cases to such interviewers.

Although there are substantial problems of missing data at the level of information about the initial contact with the respondent, there are some interesting findings. Barriers to entry, such as doormen or locked gates, do not appear to have a direct effect on the resolution of a case, perhaps because these obstacles make contact of any sort difficult. Contrary to normal survey folklore, male respondents are less likely to refuse participation. Among statements made by respondents on the initial contact, two sorts seem to have a persistent effect over future contacts. Those who made negative comments were, in fact, more likely to refuse. Those who made comments to delay the interview were less likely to resolve as either a refusal or a complete.

Partially in response to the data problems encountered in this article, the collection of the ancillary data has been redesigned in electronic form, and the use of parts of the information collected for administrative purposes has dramatically raised the incentives to record correct and complete data. Because of the growing importance of nonresponse, I hope to continue this line of research with the new data. I also hope that others will conduct similar work with other surveys to explore the generality of the findings.

A point outside the general discussion of this article deserves emphasis. I believe strongly that to improve response on surveys (or even to maintain current levels), we must account for the humanity of both respondents and interviewers. Respondents are not filing cabinets to be rifled at will, but people who face conflicting demands on their time. It is generous of respondents to share their time with survey takers, and this fact should never be forgotten or taken for granted. Interviewers are paid for their work. Nonetheless, in almost every area of work, other factors than money appear to be important determinants of superior performance. It is a wasted opportunity when survey managers fail to engage interviewers' interest beyond the level of pure production. If interviewers fail to communicate a compelling vision of a survey and a deep respect for respondents' generosity, response rates will suffer.

5. References

Allison, P.D. (1984). Event History Analysis. Sage Publications, Beverly Hills, CA, U.S.A.

Allison, P.D. (1995). Survival Analysis Using the SAS System: A Practical Guide. SAS Institute, Cary, NC, U.S.A.

Bogeström, B., Larsson, M., and Lyberg, L. (1981). Bibliography on Nonresponse and Related Topics. Mimeo, National Central Bureau of Statistics, Stockholm, Sweden.

Cartwright, A. and Tucker, W. (1967). An Attempt to Reduce the Number of Calls on an Interview Inquiry. Public Opinion Quarterly, 31, 2, 299–302.

Canell, C.F., Fowler, F.J. Jr., and Marquis, K.H. (1968). The Influence of Interviewer and Respondent Psychological and Behavioral Variables on the Reporting in Household Interviews. National Center for Health Statistics, Series 2, Number 26.

Groves, R.M. and Couper, M.P. (1996). Contact-Level Influences in Face-to-Face Surveys. Journal of Official Statistics, 12, 63–83.

Groves, R.M. and Couper, M.P. (1993a). Multivariate Analysis of Nonresponse in Personal Visit Surveys. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Groves, R.M. and Couper, M.P. (1993b). Households and Interviewers: The Anatomy of Pre-Interview Interactions. SMP Working Paper No. 11, Survey Research Center, University of Michigan, Ann Arbor, MI, U.S.A.

Hanson, R.H. and Marks, E.S. (1958). Influence of the Interviewer on the Accuracy of Survey Results. Journal of the American Statistical Association, 53, No. 283, 635–655.

Kennickell, A.B. (1991). Imputation of the 1989 Survey of Consumer Finances. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Kennickell, A.B. (1996). Using Range Techniques with CAPI in the 1995 Survey of Consumer Finances. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Kennickell, A.B. and McManus, D.A. (1993). Sampling for Household Financial Characteristics Using Frame Information on Past Income. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Kennickell, A.B., Starr-McCluer, M., and Sundén, A. (1997). Family Finances in the U.S.: Recent Evidence from the Survey of Consumer Finances. Federal Reserve Bulletin, 83 (January), 1–24.

Kennickell, A.B. and Woodburn, R.L. (1997). Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth. Working paper, Board of Governors of the Federal Reserve System, Washington, DC, U.S.A.

Little, R.J.A. (1993). Post-Stratification: A Modeler's Perspective. Journal of the American Statistical Association, 88, No. 423, 1001–1012.

Rice, S.A. (1929). Contagious Bias in the Interview. American Journal of Sociology, 35, 420–423.

Tourangeau, R., Johnson, R.A., Quian, J., Shin, H.C., and Frankel, M.R. (1993). Selection of NORC's 1990 National Sample. Working paper, National Opinion Research Center at the University of Chicago, Chicago, IL, U.S.A.

Woodburn, R.L. (1991). Using Auxiliary Information to Investigate Nonresponse Bias. Proceedings of the Section on Survey Research Methods, American Statistical Association.

Received January 1998
Revised December 1998