Models of Nonresponse and Their Relationship to Survey Estimates

John Dixon
Bureau of Labor Statistics, Room 1950
2 Massachusetts Ave., NE
Washington, DC 20212-0001

Abstract
The survey process as measured by paradata may be different for different respondents. Those initially reluctant to participate may be convinced by the interviewer to cooperate, or the reluctance may produce attrition. Difficulty to contact respondents may be related to interviewer effort and busy respondent schedules, or it may be a form of reluctance. This paper will use mixed models to attempt to identify subsets of respondents whose paradata relates differently to survey outcomes and measures of survey quality.

Key Words: Survey, paradata, interviewer

1. Introduction

The Consumer Expenditure Survey (quarterly) is a household survey which provides part of the "market basket" of consumer expenditures, which are the basis of the CPI as well as other indices. Sampled housing units in the Quarterly are interviewed for 5 consecutive quarters. These interviews are referred to as "time-in-sample" (TIS) 1 to 5.

A very useful feature was added in 2005 to collect detailed call history data (Bates, 2004). The interviewer records times and outcomes of attempted contacts, problems or concerns reported by reluctant households, and strategies used to gain contact or overcome reluctance. This provides a very rich source for studying the interview process, which is only lightly used in this study.

Dixon (2006) found that estimates of nonresponse bias weren't impacted much by the addition of call history variables. Those interviews which required a larger number of calls where the interviewer changed modes had lower expenditures (-39.4). This effect was partially offset by those interviews that required more calls but where respondents who reported no problems had higher expenditures (27.0). Those interviewers who reported "no strategy" for attempted contact ended up with lower expenditures (-66.2) and those who changed modes during the contact process ended up with higher expenditures.

2. Data Sources

The Call History Instrument used in conjunction with the Consumer Expenditure Survey (quarterly) will be combined from 2006 through 2008 to provide the paradata for this study. The interviews from the second interview were used for the bulk of the data, with refusals in that or subsequent interviews

used as an outcome variable. 97317 households were used for the analysis. Interviews which didn't provide CHI data were excluded.

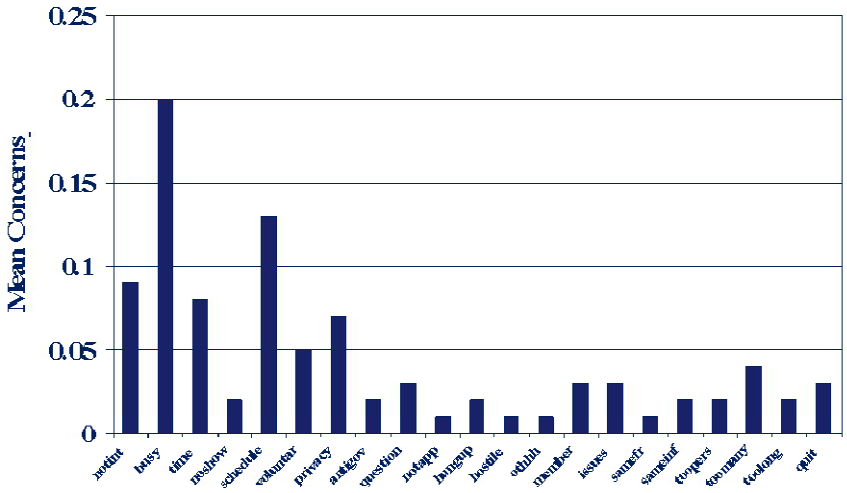
3. Methods

Logistic models (SAS version 9.1) were used to predict nonresponse using variables from CHI. A factor analysis (using MPlus) was be used to examine the inter-relationships among the concerns expressed in CHI.

4. Results

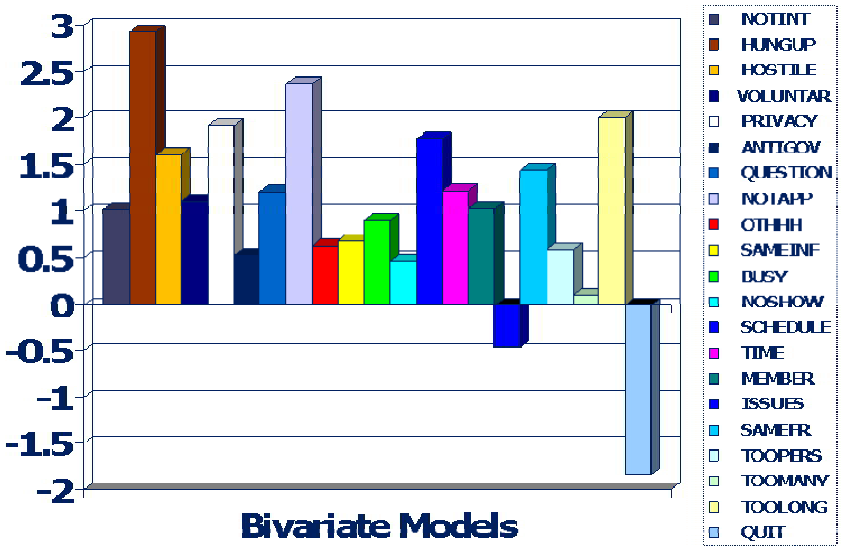
The most common concern expressed by respondents was "busy" (Figure 1), followed by "schedule difficulties", and "not interested", which was also most predictive of a refusal outcome. Other notable concerns were "time the interview takes" and "privacy concerns".

Figure 1: Mean Rates of CHI concerns.



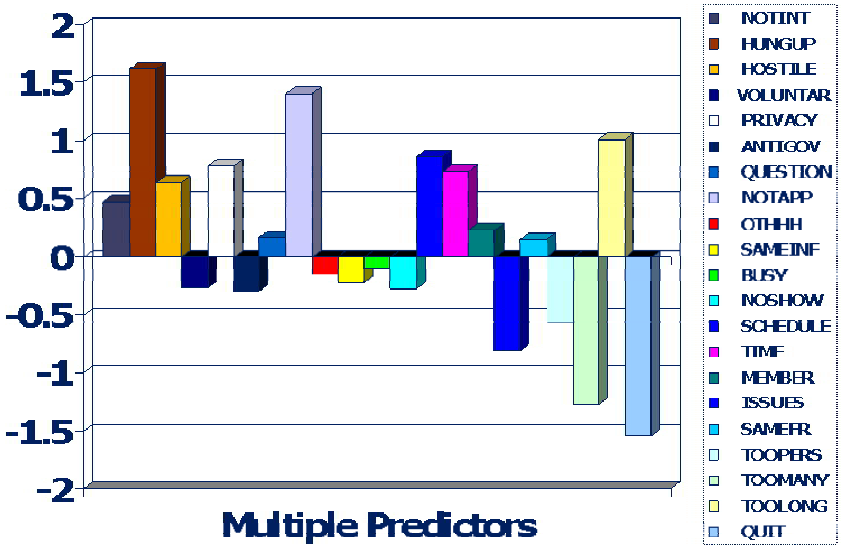
The univariate logistic models (Figure 2) showed positive relationships between most of the concerns and refusal during some of the interviews. "Family issues" (issues, which was not significantly related to refusal) and "intends to quit" are the two related to not refusing.

Figure 2: Predicting nonresponse with bivariate logistic models.



The multivariate model showed some coefficients which reversed direction or became non-significant when adjusting for the other variables.

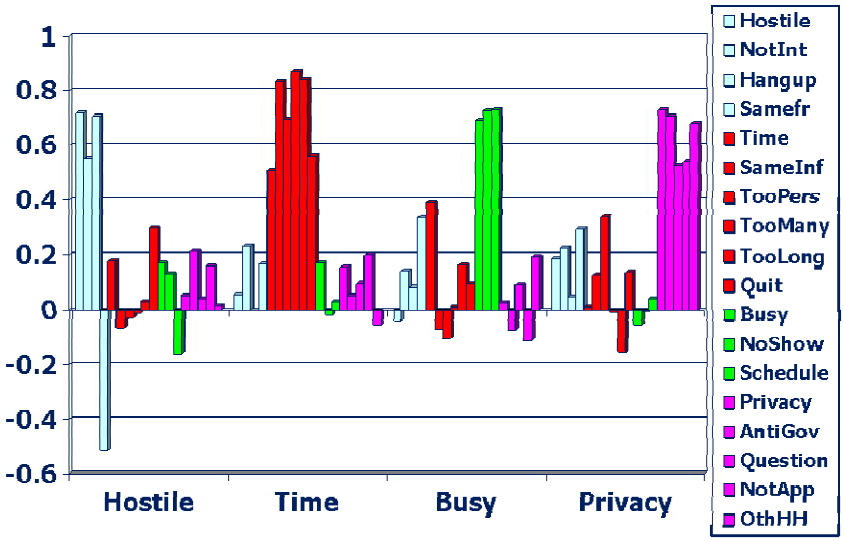
Figure 3: Predicting nonresponse with a multivariate logistic model.



The univariate estimates could be interpreted as the relationship between those concerns and refusal, while the multivariate estimates could be interpreted as the unique relationship of those concerns beyond the other concerns. The combination gives a more complete picture of the relationship between concerns and refusal. The most common concern "busy", showed a strong relationship with refusal, but didn't contribute anything beyond the other variables. "Not interested", which also was a frequent concern (almost 10% of cases) had a strong relationship with refusal even after adjusting for the other variables. Counter to expectations, the concern "planning to quit" (quit) showed a strong negative relationship to refusal. This might have been related to increased efforts by the interviewer to persuade the respondent to stay with the survey.

To examine the inter-relationships between concerns a factor analysis was used.

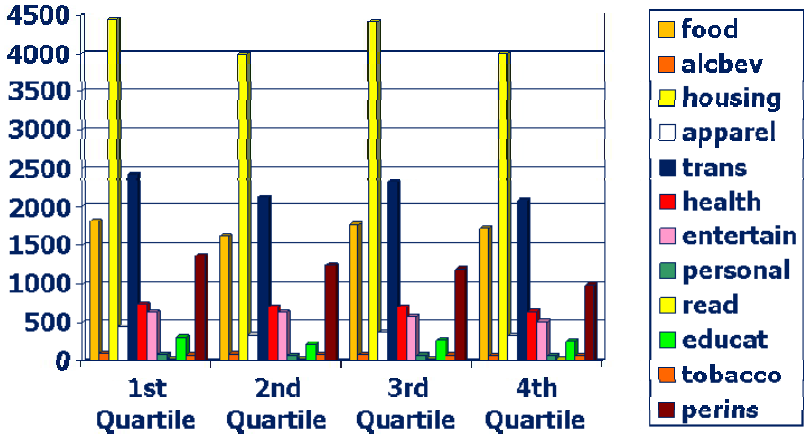
Figure 4: Factor pattern for CHI concerns.



Four factors were identified based on concerns expressed in the contact history instrument. The "Hostile" factor included hostile behavior, hangups, "not interested", and in a negative relationship, wanting the same FR from the previous interview. They "Privacy" factor included concerns about privacy, expression of anti-government sentiment, not understanding the survey, not thinking the survey was applicable to them, and other household members advising the respondent not to participate. Although the "Time" and "Busy" factors may seem similar, the "Time" factor was negatively related to refusal and seems more related to being overwhelmed or over burdened while the "Busy" one seems to relate to the respondent not being able to find the time. The prediction of refusal based on the 4 scales gave a pseudo r-square of .19 compared to a model using all the items which went into the scales pseudo r-square of .20.

By comparing those most like refusers on their contact history characteristics we can guess about potential bias.

Figure 5: Potential bias based on CHI predicting refusal.



Here, the 4th quartile of the propensity scores from a logistic model predicting refusal will respresent the refusers. In general, they reported lower expenditures, but the more convincing pattern is if the trend over the quartiles was consistently lower. Alcoholic beverages, Health, Entertainment, Personal care, Tobacco, and Personal insurance were all lower. The others were usually lower in the 4th quartile, but often low in the 2nd as well. The largest expenditures (Housing, Transportation, and Food) showed little change, indicating the bias is low overall. A similar model predicting noncontact showed less consistent patterns, probably because prediction was poorer and there was so little noncontact. A model based on contact effort, which was found to work well with the Current Population Survey didn't fit this data very well, although the patterns of bias were similar.

5. Discussion

Other studies have found a relationship between CHI data and refusal on both the National Health Interview Survey and the Consumer Expenditure surveys (Bates 2004, Bates et. al. 2008). This study extended their research to examine differences in the interview experience and subsequent refusal based on the concerns expressed by the respondents.

The studies by Bates (2004), and Henley and Bates (2006) found that the number of concerns was a more important predictor of refusal than particular concerns for the NHIS. They found privacy concerns, the voluntary nature of the survey, "not interested", and "Survey takes too long" to be the primary concerns for refusers. This study found a similar overall pattern, but added "schedule difficulties" to the list. Some unexpected effects were found with negative relationships to refusal for respondents which had "family issues" (issues) or "intends to quit survey" (quit). The family issues were thought to make the difficult process of reporting expenditures more difficult, but although non-significant, that didn't seem to be a problem for respondents. The respondent "intending to quit" probably triggered a strong anti-refusal response from the

interviewer (which isn't captured by the CHI instrument), resulting in a decrease in the likelihood of refusal.

Limitations and future research

The rarity of some of the concerns pose a problem in modeling, but since they are so rare they don't present much of a problem for possible new approaches for avoiding refusal.

The effect of reluctance on data quality may also be of interest, since one of the groups used fewer records in reporting expenditures.

References

Bates, N., Contact Histories: A Tool for Understanding Attrition in Panel Surveys., A paper presented at the 59th Annual AAPOR Conference, May 11-13, 2004

Bates, N., Dahlhamer, J., and Singer, E., "Privacy Concerns, Too Busy, or Just Not Interested? Exploring the Relationship Between Doorstep Concerns and Nonresponse Bias", Journal of Official Statistics, Vol. 24.4, December, 2008.

Chopova, B., Edgar, J., Gonzalez, J., King, S., McGrath, D., and Tan, L., "Assessing nonresponse bias in the CE Interview Survey: A summary of four studies", presented at the Federal Committee for Statistical Methodology Workshop on How to Do Nonresponse Bias Analyses in Household and Establishment Surveys, June 10, 2009.

Dahlhamer, J., Simile, C., and Taylor, B., "Exploring the Impact of Participant Reluctance on Data Quality in the National Health Interview Survey (NHIS)", paper presented at the Statistics Canada Symposium; Methodological Issues in Measuring Health, 2006.

Dixon, J., "The Effects of Item and Unit Nonresponse on Estimates of Labor Force Participation", Paper presented at the Joint Statistical Meetings, N.Y., N.Y., 2002.

Dixon, J., "Models of Nonresponse Bias, Survey Quality, Paradata, and Household Characteristics in the Consumer Expenditure Survey.", A paper presented at the International Survey Nonresponse Workshop, Omaha, Nebraska, 2006.

Henley, M., and Bates, N., "Using Call Records to Understand Response in Longitudinal Surveys", paper presented at the American Association for Public Opinion Research, 2006.

Tucker, C., and Dixon, J., "Predicting Interviewer Nonresponse Rates from Household and Regional Characterstics", Paper presented at AAPOR, 2000.

Formatted: Bullets and Numbering