Using Proxy Measures of the Survey Variables in Post-Survey Adjustments in a Transportation Survey

Ting Yan[1], Trivellore Raghunathan[2]
[1]NORC, 1155 East 60th Street, Chicago, IL, 60634
[2]Institute for Social Research, University of Michigan, Ann Arbor, MI, 48104

Abstract

Traditional post-survey adjustments tend to use demographic variables that are readily available. To reduce nonresponse error, the post-survey adjustments should use variables that are related not only to respondents?propensity to respond to a survey request but also to the survey variable of interest. This paper examines the use of proxy measures of survey variables in post-survey adjustments. We developed a new weighting scheme drawing on proxy measures of survey variables and compared the new weighting scheme to a traditional weighting procedure that employs only demographic variables to examine the effectiveness of the new weighting method in nonresponse error reduction.

KEY WORDS: post-survey adjustment, nonresponse error, record data

1. Introduction[1]

Household surveys in the United States have witnessed a decline in response rates for the past decades (Atrostic, Bates, Burt, & Silberstein, 2001; Curtin, Singer, & Presser, 2000; 2005; de Leeuw & de Heer, 2002). The declining response rates raise a concern about the accuracy and precision of the survey data among survey organizations and data users. The danger with a survey of a low response rate is the presence of nonresponse bias if sampled persons with low response propensities systematically differ from those with higher response propensities with regard to survey variables of interest. Therefore, survey organizations do their best to minimize potential nonresponse bias associated with survey statistics. On the one hand, survey organizations invest extensive resources on contacting sample persons and on persuading them to participate in the survey once contacted. Techniques such as advance letters, incentives, and customized call scheduling and calling rules and so on have been shown empirically to be effective in increasing response rates at the data collection stage. However, the extensive effort spent on contacting and recruiting sample persons with low response propensities are usually costly. Given limited resources and budget, survey organizations are restrained in the extent of their effort to pursue people of low response propensities.

On the other hand, post-survey adjustment is commonly used to address potential nonresponse bias after data is collected. Weighting is one such post-survey adjustment method. With nonresponse weighting (whether it is a weighting class adjustment method or a response propensity weighting method), survey respondents are given a weight to compensate for differential probability of response given selection. Weighting is aimed at reducing nonresponse bias, although it is often accompanied with an increase in variance. The success of nonresponse weighting relies on the auxiliary variables used in constructing nonresponse weights. Ideally, the adjustment variables should have two properties in order to reduce

[1] This paper, together with other papers in this session, is the result of a research seminar lead by Profs. Bob Groves and Trivellore Ragunathan, both at the University of Michigan.

nonresponse bias. The variables should 1) be predictive of survey variables of interest and 2) be predictive of sampled person's probability to respond to a survey request. The first property (high correlation with survey outcomes) is of the most importance in terms of nonresponse bias reduction. A recent simulation study by Little and Vartivarian (2005) demonstrated that weighting was only effective when the auxiliary variables used in constructing weights were highly correlated with both the survey variables of interest and also the response propensity. Auxiliary variables that are only related to response propensity but not survey outcome variables inflate the variance with no effect on bias reduction.

In practice, auxiliary variables that are strongly associated with both survey outcomes and response propensities are hard to find. Instead, practitioners often make use of variables that are readily available to do nonresponse adjustments. Demographic variables are commonly used because of their availability. However, the extent of their association with survey outcomes and with response propensity is open to discussion. When demographic variables are only weakly associated with survey variables of interest and/or response propensity, using them in weighting adjustment will not reduce nonresponse bias and might run into the danger of increasing variance.

The search for good auxiliary variables turns to rich sampling information, external administrative/record data, and paradata (such as interviewer observations and call record data). This paper examines the use of proxy measures of survey variables in nonresponse adjustment. The proxy measures are record data obtained from a different source and are highly correlated with survey variables of interest. We developed a new weighting scheme drawing on these measures and compared the new weighting scheme to a traditional method employing only demographic information to see whether the new method is more effective in reducing nonresponse bias and variance.

2. The Survey

The data used in this paper is from a survey conducted by the University of Michigan Transportation Research Institute (UMTRI). This is a survey of young adults who completed high school and have a valid Michigan driver license. A large sample of young adults were selected and interviewed when they were still in high school. A year later, they were followed up for a second interview by telephone. This paper only looks at this telephone interview.

The survey asks sampled young adults various driving behaviors and risk-taking behaviors. Questions such as seat belt wearing, traffic violation, and drink and driving are included in the survey.

For all sampled young adults, UMTRI was able to get their driving record data from the State of Michigan. In this paper, we explore the use of five record variables for weighting purpose. They are the traffic points received in the past 12 months, number of traffic offenses, number of serious traffic offenses, number of crashes and the number of serious crashes in the past 12 months.

3. The New Weighting Scheme

As a base for comparison, we constructed one set of weights using a traditional weighting class adjustment method involving two demographic covariates (age and sex).

The new weighting scheme aimed to include the five record variables in nonresponse weights besides the two demographic covariates. We employed the response propensity score method. We first fitted a logistic regression of the response indicator on all five record variables and the demographic covariates. We started with a step-wise selection procedure and balance tested and fine tuned the propensity model. The final model included all five record variables besides age and race and two interaction terms (sex and age, and number of offenses and number of serious crashes). We estimated a response propensity score for each respondent based on the final propensity model and used the inverse of the estimated response propensity scores as the new weights.

We compared the two sets of weights on their effectiveness in reducing nonresponse bias.

4. Results

We first discuss some diagnostic measures of the new weighting method drawing on the five proxy measures of the survey variables before comparing the survey estimates weighted by the two different weighting methods.

4.1. Diagnostic measures

Since a strong association between auxiliary variables and survey outcomes is a desired condition for weighs to be effective in reducing nonresponse bias, we examined the correlations between the five record variables with four survey variables of interest. As displayed in Table 1, the record variables are significantly correlated with the survey outcome variables (except for two correlations), but the strength of correlations varies, ranging from .02 to .46.

We also examined the correlations between the proxy measures and the response indicator. Again, four of the five correlations are significant at .05 level, but the strength of the correlations is rather weak in overall. The absolute values of the correlations range from .02 to .08.

After we obtained the estimated response propensity scores for each respondent, we looked at the distribution of the estimated propensity scores by the response indicator, as displayed in Figure 1. The two distributions are quite similar, suggesting that the estimated propensity scores are balanced by respondents and nonrespondents. Table 2 further shows the correlations between the y variables and the estimated propensity scores. It seemed, from Table 2, that respondents with a lower response propensity were more likely to exhibit bad driving behaviors (e.g., never wearing a seat belt, having their license revoked, and more traffic points).

To further examine the relation between response propensities and survey outcomes, we divided respondents into five propensity classes, where respondents in propensity class 1 have the lowest average response propensities and respondents in class 5 the highest average response propensities. We displayed the means of the survey variables of interests by the five response propensity classes in Table 3. Again, we found that the means of all four survey variables do differ significantly by response propensity classes, pointing to the presence of possible nonresponse bias in unweighted survey estimates. The table also suggested a trend that respondents of lower propensities are more likely to exhibit bad driving behaviors and unweighted survey estimates run the risk of underestimating the prevalence of bad driving behaviors in young adults.

4.2. Effects of weights

We explored the effects of the new weights and old weights on survey estimates via two approaches. Survey variable y4 (number of traffic tickets) is a continuous variable and we examined how the two weights affected the distribution of this continuous variable. Table 4 presents the unweighted and weighted frequency distributions of the traffic ticket variable. It seemed that both the traditional and the new weights moved the distribution to the right (or to the upper end of the distribution). The shift to the right is more ostensible with the new weights including the proxy measures than with the old weights. Consistent with the previous tables, there seemed to exist a downward bias with the unweighted estimate.

Unweighted and weighted survey estimates are compared in Table 5. The weighted estimates are in general bigger than the unweighted ones, suggesting that the weights reduced the extent of the underestimation present in the unweighted estimates. In other words, the weighted estimates reduced the extent of bias caused by survey nonresponse. Furthermore, the new weights that incorporated the record variables shifted the means/proportions to the right to a bigger degree than the traditional weights created by age and sex. The degree of upward shift tends to be bigger for variables that have higher correlations with the record variables (e.g., the "traffic ticket" variable and the "license revoked" variable) than for the other two, suggesting that the association between the survey outcomes and auxiliary variables affect the extent of bias reduction.

5. Conclusions and Discussions

This paper explores the use of proxy measures of survey variables in post-survey adjustments in a transportation survey. The proxy measures are more highly correlated with survey outcomes, and less so with the response indicator. Including these proxy measures in the weight construction seemed to have led to a larger shift in the survey estimates to the right than when they are not included in the weight construction or when no weights are used. In addition, the extent of bias reduction (or the extent of the right shift) using the new weights is related to the strength of the associations between the survey outcomes and the proxy measures – larger shifts in estimates are observed with moderate correlations (between auxiliary and survey variables) and smaller effects with low correlations.

Due to the weak correlations between proxy measures and the response indicator, we were not able to observe any apparent changes in the variance of the estimates. In fact, both sets of weights increased the variance of the estimates slightly (the changes happened at the forth decimal place). Thus, for the four survey variables examined in this paper, weights including the record variables reduced nonresponse bias at no expense of variance.

The challenge of weighting adjustment, for survey researchers and practitioners, lies in the search for the right auxiliary variables that are predictive of both response propensities and survey variables of interest. We encourage survey researchers to actively engage in finding the right auxiliary variables in developing nonresponse weights.

This paper didn't consider the measurement error properties of the record variables. We made a simplistic assumption that there is no measurement error in these record variables. Of course, this assumption is debatable in the real world. Future research is needed to examine the effect of the measurement error in auxiliary variables on survey estimates and on the bias variance trade-offs.

References

Atrostic, B.K., Bates, N., Burt, G., and Silberstein, A. (2001). Nonresponse in U.S. Government Household Surveys: Consistent Measures, Recent Trends, and New Insights. *Journal of Official Statistics*, 17, 209-226.

Curtin, R., Presser, S., and Singer, E. (2000). The Effects of Response Rate Changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, 64, 413-428

Curtin, R., Singer, E., and Presser, S. (2005). Changes in Telephone Survey Nonresponse Over the Past Quarter Century. *Public Opinion Quarterly*, 69, 87-98.

de Leeuw, E. and de Heer, W. (2002). Trends in Household Survey Nonresponse: A Longitudinal and International Comparison. In Groves et al. (Eds.). *Survey Nonresponse*. (pp. 41-54). Wiley.

Little, R. J, and Vartivarian, S. (2005). Does Weighting for Nonresponse Increase the Variance of Survey Means? *Survey Methodology*, 31, 161-168.

Table 1. Correlations between Record Variables and Survey Outcomes

	Y1 (Never wearing seat belt at local travel)	Y2 (Never wearing seat belt at long travel)	Y3 (License revoked)	Y4 (Number of traffic tickets)
Traffic points received	.13	.11	.28	.41
Number of traffic offenses	.14	.12	.30	.46
Number of serious offenses	.12	.12	.26	.35
Number of crashes	.04	.04	.02	.05
Number of serious crashes	.02	.03	.04	.04

Table 2. Correlations between survey outcomes and estimated response propensities

	Y1 (Never wearing seat belt at local travel)	Y2 (Never wearing seat belt at long travel)	Y3 (License revoked)	Y4 (Number of traffic tickets)
Estimated Response Propensity scores	-.06	-.05	-.16	-.09

Table 3. Means of Survey Outcomes by Response Propensity Classes

Propensity Classes (Number of Nonrespondents /Respondents)	Y1 (Never wearing seat belt at local travel)	Y2 (Never wearing seat belt at long travel)	Y3 (License revoked)	Y4 (Number of traffic tickets)
1(1778/760)	.079	.068	.301	.475
2(1592/948)	.067	.053	.248	.463
3(1563/1028)	.038	.033	.170	.288
4(1422/1060)	.046	.036	.154	.277
5(1326/1217)	.037	.033	.123	.317

Table 4. Unweighted and Weighted Frequency Distribution of Y4

Value (n)	Unweighted	Traditional weights (weighting class by age and sex)	New weights (Including proxy measures)
0 Ticket (3735)	75.03%	74.68%	74.27%
1 Ticket (920)	18.48%	18.70%	18.93%
2 Tickets (223)	4.48%	4.54%	4.64%
3+ Tickets (100)	2.01%	2.08%	2.16%

Table 5. Comparisons of Survey Estimates

Survey Outcomes	Unweighted	Traditional weights (weighting class by age and sex)	New weights (Including proxy measures)
Mean # of Traffic tickets	.354 (.012)	.360 (.012)	.367 (.012)
Proportion of ppl with license revoked	.190 (.006)	.196 (.006)	.201 (.006)
Proportion never wearing seatbelt on local travels	.051 (.003)	.052 (.003)	.054 (.003)
Proportion never wearing seatbelt on long travels	.043 (.003)	.044 (.003)	.045 (.003)

Figure 1. Histograms of Estimated Response Propensities