#/195

Testing A New Attrition Nonresponse Adjustment Method For SIPP

Ralph E. Folsom and Michael B. Witt, Research Triangle Institute
P. O. Box 12194, Research Triangle Park, NC 27709-2194

KEY WORDS: Response Propensity, Constrained Logistic Regression, Longitudinal Nonresponse

1. Introduction

While unit or questionnaire nonresponse can seriously degrade the quality of any survey, nonparticipation is particularly threatening to a longitudinal survey like the U.S. Census Bureau's Survey of Income and Program Participation (SIPP). To minimize the potential biasing effects of second and subsequent wave attrition from SIPP panels, staff at the Census Bureau perform weight adjustments. These adjustments are designed to make the self-selected subsample of longitudinal respondents more representative of the initial wave one sample. The Bureau's SIPP weight adjustments take the form of post-stratum or weighting class specific multipliers applied to the wave one base sample weight. The associated weighting classes are defined by collapsing cells in a multi-way cross classification of categorical variables until each resulting cell satisfies two conditions: cells are collapsed until the sample size and the estimated response propensity are greater than predefined thresholds.

The Bureau funded research project reported here (Folsom and Witt, 1994) tests a new nonresponse adjustment methodology recently developed at the Research Triangle Institute (RTI) (Folsom, 1991). This new method uses weight adjustment multipliers defined at the person level. These multipliers are created by modeling a sample person's response propensity using constrained forms of either a logistic or exponential model. This nonresponse adjustment methodology was tested on data from the 1987 SIPP panel. The goal of this project was to develop a new SIPP nonresponse adjustment that would reduce the attrition bias in cross sectional and longitudinal estimates derived from the survey.

2. Overview of the 1987 SIPP

The 1987 SIPP was a longitudinal household, panel survey designed to collect demographic and economic data on all household occupants over a 28 month data collection period. The initial sample of households was divided into four *rotation groups*, and data was collected from respondents in each rotation group every four months. Each four month data collection period is referred to as a sample *wave*; consequently, the 1987 SIPP had seven waves of data collection. During each interview, demographic and economic data was collected by month, for the four previous months. By carefully matching waves of data collection, rotation groups and reference months one can see that monthly data is available for all four rotation groups for 25 continuous reference months and monthly data for subsets of rotation groups are available for an additional 6 months.

The SIPP was designed to be a ongoing survey with new, seven or eight wave samples (or panels) introduced each year. In our evaluation of the new weight adjustment, we take advantage of this panel overlap noting that the reference period for the first wave of the 1989 SIPP is equivalent to the reference period in the last wave of the 1987 SIPP. To increase the length and sample size of individual panels without increasing survey costs, the Bureau plans to change the design of future SIPP surveys to nonoverlapping panels with 48 months of data collection. The impetus for this research project was based on the likelihood that the increase in panel length and associated respondent burden would increase the attrition rate and the associated nonresponse bias.

At the first wave of data collection for the 1987 SIPP, a sample of roughly 11,700 responding households was established. For the months in which they remained survey eligible, person-level data was sought on residents of these households at each subsequent wave of the SIPP panel. Generally, people can become ineligible for the survey if they died or move out of the country. Within this sample of first wave responding households, a person sample of roughly 33,100 was identified.

While SIPP person nonresponse occurs at all waves of data collection, the wave specific attrition rates decline monotonically through the sample waves. At wave one, 30,767 persons were interviewed and of these, 24,429 responded for each month in which they were eligible during the seven waves of data collection.

For this project, we were interested in developing a nonresponse adjustment for the SIPP longitudinal respondent weights. Longitudinal respondents are wave one participants that provide data for all subsequent waves in which they were eligible. Consequently, this project began with an initial sample of 30,767 people; 24,429 longitudinal respondents and 6,338 nonrespondents for a 20.6% attrition rate. The initial sample weight (or base weight) for the 30,767 people used for this analysis retained the Bureau's adjustment for wave one household and person nonresponse.

One appealing feature of this two step (wave one plus subsequent attrition) weight adjustment is that the wave one data can be used as explanatory information in our models of longitudinal response propensity. Since our response propensity models can effectively incorporate more of these wave one variables than the weighting class method, the potential clearly exists for a significant reduction in attrition bias.

Before launching into a description of how our new weights were created, it will be useful to consider how the Census Bureau currently creates their longitudinal nonresponse adjustment. The Bureau refers to their weighting class adjustment for longitudinal nonresponse as the *first stage* adjustment. Weighting classes are formed using such wave one auxiliary information as race, education, welfare and unemployment benefits indicators, a labor force status indicator, a bonds indicator and categorized average monthly household income (126 total classes). The nonresponse adjustment is defined as the inverse of the base weighted longitudinal response rate observed within each weighting class. In order to minimize the effect of unequal weights on the variance of estimates, weighting classes are collapsed to avoid cells with less than 30 longitudinal respondents or cells with inverse response rates (nonresponse adjustments) exceeding 2.00.

After performing the weighting class adjustments, Census Bureau statisticians performed a *second stage* adjustment (also commonly referred to as a post-stratification adjustment) to January 1987, person level control totals derived from the Current Population Survey (CPS). The second stage adjustment was applied using standard raking methods in order to preserve CPS counts for the cross classifications defined by:

- Age Group x Race x Gender x Hispanic Indicator. Several cells in this cross-classification were collapsed; 98 population totals initially controlled for.
- Complex Interaction Of Householder Status, Living With Relative Indicator, Spouse and/or Child Present Indicator; 19 population totals initially controlled for.

As with the first stage adjustment procedure, marginal cells in the second stage adjustment are collapsed when the total number of respondents is less than 30 or the marginal adjustment is greater than 2.00.

3. Exponential and Logistic Model

Extending a constrained exponential model suggested by Deville and Särndal (Deville and Särndal, 1992), our constrained exponential and logistic weight adjustment multipliers may be written as:

$$\lambda_i = \begin{cases} \alpha_o^{-1}\alpha_i & \text{\{Exponential Model\}} \\ (1+\alpha_o^{-1}\alpha_i) & \text{\{Logistic Model\}} \end{cases} \quad (1)$$

Where:

$$\alpha_i = \left\{ \frac{\hat{L}(\hat{U}-1) + \hat{U}(1-\hat{L})\exp(-AX_i\beta)}{(\hat{U}-1) + (1-\hat{L})\exp(-AX_i\beta)} \right\} \quad (2)$$

i = Indexes the sample units (wave one respondents),

X_i = Vector of explanatory variables known for all i,

λ_i = the inverse response propensity (ρ_i^{-1}) weight adjustment,

A = $(\hat{U}-\hat{L})/(1-\hat{L})(\hat{U}-1)$. This is simply a scale factor that helps minimize the effects of the constraints on the shape of the exponential function.

The scale factors α_o are derived from the base sample weighted overall response rate, ρ_o; specifically,

$$\alpha_o = \begin{cases} \rho_o & \text{\{Exponential Model\}} \\ \rho_o \div (1-\rho_o) & \text{\{Logistic Model\}} \end{cases} \quad (3)$$

$\rho_o = (\Sigma w_i r_i) \div (\Sigma w_i)$,

r_i = 0/1 Sample response indicator, and

w_i = Sample base weight.

And the bounds $\hat{L}$ and $\hat{U}$ in α_i [see equation (2)] are set as follows:

$$\begin{array}{lll} \hat{L} = \alpha_o L, & \hat{U} = \alpha_o U & \text{\{Exponential Model\}} \\ \hat{L} = \alpha_o(L-1), & \hat{U} = \alpha_o(U-1) & \text{\{Logistic Model\}} \end{array} \quad (4)$$

Our estimates for β are found by solving the following generalized raking equation using a Newton-Raphson algorithm:

$$\sum w_i r_i \lambda_i X_i = \sum w_i X_i \tag{5}$$

Notice, the constants U and L in (4) are chosen to bound (or constrain) the resulting adjustment factor λ_i:

As $X_i\beta \Rightarrow +\infty$ then $\alpha_i \Rightarrow \hat{L}$ in (2) and $\lambda_i \Rightarrow L$ in (1),

As $X_i\beta \Rightarrow -\infty$ then $\alpha_i \Rightarrow \hat{U}$ in (2) and $\lambda_i \Rightarrow U$ in (1).

Bounding the adjustments constrains the associated variance inflation which results from the increase in adjusted weight variability compared to the original base weights.

In order to obtain a solution to (2), $\hat{L}$ and $\hat{U}$ must satisfy: $0 < \hat{L} < 1 < \hat{U}$. Notice in (4) that the scale factors α_o are introduced to simply shift the limits on $\hat{L}$ and $\hat{U}$. This shifting allows us to obtain feasible inverse response propensities from the exponential and logistic models as follows:

Exponential Model:

$$\hat{L} \in (0,1) \text{ shifts to } L \in (0, \rho_o^{-1})$$
$$\hat{U} \in (1,\infty) \text{ shifts to } U \in (\rho_o^{-1}, \infty).$$

Without this shift, we note from (1) that a lower bound could not be set that would force the λ_i's to be greater than one since the lower bound on α_i could not be set to one without forcing all α_i to equal one uniformly.

Logistic Model:

$$\hat{L} \in (0,1) \text{ shifts to } L \in (1, \rho_o^{-1}) \text{ and}$$
$$\hat{U} \in (1,\infty) \text{ shifts to } U \in (\rho_o^{-1}, \infty).$$

Without this shift, equation (1) implies that an upper bound could not be chosen that would force λ_i's to be less than or equal to 2 since the upper bound on α_i could not be set to one without again forcing uniformity on all α_i.

Thus the scale factor is introduced in order to allow one to use either the logistic or exponential response propensity model and achieve the desirable property: $1 \leq L \leq \lambda_i \leq U$ where U and L are predetermined constants.

For the unconstrained cases, (i.e. when $\alpha_o = 1$, $\hat{U} = +\infty$, and $\hat{L} = 0$), the terms $A \rightarrow 1$, and $(\hat{U}-1)^{-1} \rightarrow 0$ in equation **(2)**. In these cases, the model reduces to the familiar form:

$$\rho_i = \begin{cases} [\exp(-X_i\beta]^{-1} & \{\text{Exponential Model}\} \\ [1+\exp(-X_i\beta)]^{-1} & \{\text{Logistic Model}\} \end{cases}$$

Folsom (1991) proposed the unconstrained logistic model for nonresponse adjustment and the unconstrained exponential model for sampling error adjustments akin to post-stratification or double sampling ratio estimation. In particular, Folsom (1991) discusses the variance and bias reduction properties which result from an equivalence between logistic re-weighted respondent means and regression imputation-based estimates, and between exponential reweighted sample means and a survey regression estimator.

In practice, we most often use the exponential model for second stage, post-stratification type adjustments since the required adjustment factors are not logically bound below by one as they are with adjustments resulting from the logistic response propensity model. Further, to facilitate obtaining a solution to the raking equations **(5)**, we tend to use the scaled versions of the exponential model for nonresponse adjustment when the over-all response rate ρ_o is close to one. For our application to the 1987 SIPP data, we used the scaled logistic model to predict longitudinal response propensity. With the overall response rate $\rho_o = 0.8$, we were able to obtain convergent solutions for each of our subpopulation models with $U=2$, the Bureau's weighting class adjustment bound.

4. Estimating Model Parameters

We began the development of the alternate nonresponse adjustments for the 1987 SIPP by examining marginal response rates across several main effect and low order interaction terms in order to specify an initial set of response propensity predictor variables. During this search we found that testing the statistical significance of the univariate response rate differentials across the levels of the categorical variables was not a useful indication of a variable's predictive power. The combined wave one sample size of 30,767 guaranteed that most of the associated variable specific chi-squared tests would exhibit highly significant results. This power to declare negligible response rate differentials statistically significant persisted when proper account was taken of the SIPP design induced clustering effects. Therefore, for this initial screening, we chose to disregard variables with response rate differentials that fail to exceed a subjective threshold of 10 percentage points. We observed that age group, race/ethnicity, relationship to reference person, living quarters owned indicator, race/ethnicity of the head of household, household type (levels), wage dollars imputed indicator, the source dollars imputed indicator, covered by Medicare indicator, and covered by General Assistance indicator all exhibited a response rate differentials approaching or exceeding 10 percent.

This initial look at response rate differentials, coupled with some exploratory modeling to estimate the significance of various lower-order interactions, continuous variables and linear spline functions of several variables, led us to eventually build separate models for seven subsamples (nonresponse classes) defined in terms of household income, race/ethnicity, marital status, and Census Region. Within each class, we began with models containing a large number of explanatory variables, and eliminated statistically nonsignificant parameters using a backwards elimination process. The statistical test to determine the significance of the parameters was based on students-T type statistics derived from sample design-based (cluster sampling) variances estimated using the Taylor Series method. The level of significance used for these tests was set at $\alpha=.10$. The variables and β coefficients retained in our final models are presented in Table 1.

In order to minimize the effect of these adjustments on the coefficient of variation of the resulting respondent sample weights, a lower bound of 1.0 and an upper bound of 2.0 was set in the scaled logistic models for each nonresponse class. Further, generalized Wald statistics adjusted for design effects were created to test the overall significance of each model. These Wald statistics test the null hypothesis that all slope parameters are zero. The significance probabilities of the Wald statistics are presented at the bottom of Table 1.

While we tested for several two-way interactions among the significant main effects in our models, very few significant interactions were found. One should note however that by virtue of fitting separate models within the seven classes, we have implicitly interacted the nonresponse class variable with all the fitted main effects.

5. Evaluation Of Alternate Weights

Our new longitudinal sample weights were evaluated in terms of estimated relative bias, relative standard error and the error in 90% confidence interval (CI) coverages. The coverage error in the estimated 90% CIs approximates the probability that a population parameter lies outside the bounds of a 90% CI, assuming our bias and variance estimates are error free and the sampling distribution is normal. These evaluation statistics were computed using the new "RTI-Revised" weight and using the Bureau's "Current" nonresponse-adjusted longitudinal sample weight. Standard errors in this analysis were based on the linearization variance for cluster sample ratio means and proportions which treat the response propensities and second stage, post-stratification adjustments as known without error. Though this assumption may lead to bias in the variance estimates, our experience with variance estimators based on the proper generalized raking variances suggests that these biases are likely to be modest.

To evaluate the RTI-Revised weights, two comparative evaluations were conducted (Table 2):

- 1987 SIPP wave one estimates derived from longitudinal respondent data using the RTI-Revised and the Current weight were compared against benchmark statistics created using the original, 1987 SIPP wave one sample (30,767) with their household and nonresponse adjusted base sample weight. For this analysis, a second stage adjustment was applied to each weight using the scaled exponential model specified in Section 3 with explanatory variables representing all the effects currently used in the Bureau's second stage adjustment as specified in Section 2. No attempt was made to omit nonsignificant terms in the second stage adjustment model.

- 1987 SIPP wave seven estimates using the RTI-Revised weight and the Current weight were compared against benchmark statistics created using wave one data from the 1989 SIPP. Recall from Section 2 that the seventh wave of the 1987 SIPP overlaps the first wave of the independently selected 1989 SIPP. Further, except for births and immigrations added to the 1989 panel, there is no difference in the population coverage of the two panels. To minimize this population coverage difference, those 1989 SIPP respondents who were 0-2 years old in wave one were omitted from the analysis since they had not been born when the 1987 panel was established at wave one.

Intuitively, in our first analysis of the weights, the bias calculation derived from the 1987 panel wave one data should maximize the performance of our RTI-Revised weights since our generalized raking solution [equation (5)] forces mean equality within