NONINTERVIEW ADJUSTMENT RESEARCH FOR THE CURRENT POPULATION SURVEY

Promod K. Chandhok
Bureau of the Census[1], Washington, DC 20233

KEY WORDS: Response propensity, nonresponse bias, accuracy

1. Introduction

This paper attempts to evaluate the presently used Current Population Survey (CPS) noninterview adjustment (NIA) procedure. It also compares the current NIA procedure with a NIA procedure that uses response propensities.

In CPS about five percent of the sampled housing units are Type A noninterviews. Hanson (1977) defines Type A noninterviews as "occupied units eligible for the survey for which the interviewer was unable to complete the interview." Noninterviewed households can be different from interviewed households in terms of survey characteristics. Estimates (based on respondents) of characteristics can thus be biased due to nonresponse.

The basis of this paper is to compare two procedures for NIA. Section 2 describes the current procedure. Background and a discussion of the response propensity procedure is in Section 3. In Section 4 we evaluate the current adjustment and compare it with the proposed adjustment. Research issues which include alternative approaches are presented in Section 5.

2. The Current Adjustment

The current NIA is described in detail in Jones (1984). Briefly, noninterview adjustment clusters are formed by grouping administrative PSUs (subsets of sample PSUs) which are similar with regard to MSA status, MSA size and urban versus rural population within PSU. The three variables used in grouping PSUs were thought to be correlated with the two principal survey variables viz., number unemployed and number in the civilian labor force. Noninterview clusters are designated MSA (belonging to or comprising a metropolitan statistical area) or non-MSA. Within MSA clusters, the cells are central city and not-central city. Within non-MSA clusters cells are urban and rural.

3. Background and the Response Propensity Approach

Ernst (1977) compared weighting with "hot deck imputation." He first determined a set of cells that minimized the variance of estimated mean total money income. Since variance was insignificant compared to bias, he decided to concentrate on reducing bias. A set of cells for which bias is minimum was determined. This weighting scheme was compared with "hot deck imputation."

Politz and Simmons (1949) introduced the idea of grouping units by their response probabilities. A discussion of this paper is in most textbooks on sampling theory (see for example, Raj, 1968).

In the context of CPS, Palmer (1968) reports results from a study done in September 1965. In the study nonresponse households from September 1965, CPS sample were followed up. Interviews were later obtained from about half of the nonresponse households. It was found that although there were demographic differences between the interviewed and the noninterviewed groups, the proportion of persons in the labor force was about the same in both groups. Palmer also states that overall employment and unemployment rates were about the same. In other words, the variables proportion in the civilian labor force, proportion employed and proportion unemployed had negligible nonresponse bias.

There are several approaches to reduce nonresponse bias. One approach is the response propensity approach.

In this approach we assume that response to the survey is a probability mechanism and each sampled household has a nonzero probability of responding to the survey. In other words, there are no never-ever respondents (NER). NERs are also called hard-core nonrespondents. The assumption of no NERs is indefensible. Nevertheless, this quasi-randomization approach (as called by Oh and Scheuren, 1983) may provide a good NIA.

We divide the sample into L response propensity groups. The

weighting class estimator (termed the sample-based adjustment cell weighting estimator by Kalton and Maligalig (1991)) of $\bar{Y}$ is

$$\bar{y}_{wc} = \sum_{h=1}^{L} \frac{n_h}{n \sum_{i=1}^{m_h} \frac{1}{P_{hi}}} \sum_{i=1}^{m_h} \frac{y_{hi}}{P_{hi}}$$

where y_{hi} is the y-value of the i-th sample unit in the h-th group ; P_{hi}, the probability attached to the i-th unit in the h-th group; n_h, the number of sample units in the h-th group; n, the sample size; and N, the population size. Under the assumption

$$E \ (n_h / n) = (N_h / N)$$

the approximate bias of the above weighting class estimator is derived by Kalton and Maligalig (1991). Here N_h is the number of units in the population that are in group h. If all units in the *h - th* group have the same response probability, then the weighting class estimator is approximately unbiased. If bias dominates the mean squared error (MSE) then this approach may provide a good NIA. For the univariate case Kalton and Maligalig (1991) suggest a procedure for collapsing cells. This procedure is based on the conditional MSE. A procedure that is optimal in terms of the conditional MSE may not be optimal in terms of the unconditional MSE after all adjustments. This is a reason why this collapsing procedure was not used by me.

A desirable property of the response propensity approach is that the weighting class estimator for any characteristic y is unbiased.

We should emphasize that besides unit nonresponse error and sampling error, there are various other errors in the data. Other errors include (1) response error, (2) coverage error and (3) item nonresponse error. As adjustments for some of the other errors are also made, we look at NIA in conjunction with other adjustments. What we desire is an optimal set of groups, the groups being evaluated at the ultimate stage, i.e., after all adjustments.

A requirement of the response propensity approach is that information on characteristics used to define groupings must be available for both the respondents and the nonrespondents. This requirement limits our choices. Numerous response propensity groupings for households can be considered. First, a parsimonious response propensity grouping is considered. The parsimonious response propensity grouping is the starting point. Results from this are evaluated. Based on this, other groupings can be considered.

The March 1990, CPS file was used in this research. For noninterviews only "control card" and master segment tape (MST) information is available. Variables that we thought could be good predictors of response propensity were selected. Logistic regression was then used to determine a parsimonious model to predict response propensity. This model gives 45 response propensity groups and uses two variables. The first variable misc is recoded from the variable Month-in-Sample, while the second variable stsz is recoded from variables MSA status and MSA size.

The variables misc and stsz in the response propensity model are stable over time. The model is robust in that sense.

The best response propensity model for March 1990 may not be the best response propensity model for other months. To see how this model performs in other months, this model was applied to March 1988 and April 1988 data. The results showed that in terms of predicting response propensity, the proposed response propensity model does not fit well either to March 1988 or to April 1988 CPS data. We could obtain the best response propensity models for March 1988 and April 1988 CPS data and see how they compare with the response propensity model for March 1990 CPS data.

The unconditional bias and the unconditional variance of an estimator after all adjustments can only be estimated. This is done in Section 4.

4. Evaluation of the Current NIA and its Comparison With the Proposed NIA.

A criterion for evaluation is the mean squared error (MSE) of key

estimates. Key estimates considered are:

1. Number unemployed,
2. Number employed,
3. Number not in the civilian labor force,
4. Number of whites unemployed,
5. Number of whites employed,
6. Number of whites not in the civilian labor force,
7. Number of blacks unemployed,
8. Number of blacks employed,
9. Number of blacks not in the civilian labor force,
10. Number of hispanics unemployed,
11. Number of hispanics employed, and
12. Number of hispanics not in the civilian labor force.

Bias and variance of an estimator are estimated separately. The estimation of bias requires knowledge of true values. Since true values are unknown, these would have to be estimated. A revised March 1990 CPS file is built to estimate "true values." Details of how this file is built, how "true values" are calculated, and how bias is estimated are given later. The procedure used for variance estimation is the method of generalized replication.

The labor force characteristics of March 1990 Type A noninterviewed households that responded either in February 1990 or in April 1990, were used to build a revised March 1990 CPS file. Six hundred and thirty - eight, March 1990 noninterview Type A households responded in February 1990. Also, there were 537 March 1990 noninterview Type A households that did not respond in February 1990 but responded in April 1990. The labor force characteristics of these 1,175 households were substituted from February 1990 / April 1990 files. Still there were 1,534 Type A noninterviewed households that did not respond either in February or in April 1990. The characteristics of these households can be (are) different from the characteristics of interviewed and imputed households. The impact of these households on "true values" should be assessed. In computing "true values" of the twelve characteristics we assume that imputes are the true values. This is a reasonable assumption for most households, as imputes are obtained from an adjacent month. However, this may not hold for all the noninterviewed households in the March 1990 CPS sample. The labor force characteristics of some households may have been different in February 1990 or in April 1990, but our imputation assumes that they did not change.

In determining "true values" I used no NIA. We note that all other adjustments viz., the first stage adjustment and the second stage adjustment were made. We also note that the second stage adjustment is population-based. It also compensates for noninterviews.

The final weights are obtained this way. First, the base weights (BWGT), the duplication control factors (DUPLCF) and the first stage adjustment factors (FSAF) are input. The Second Stage Adjustment program is then used to obtain the second stage adjustment factors (SSAF). Final weights (FNLWGT) are then determined using

FNLWGT = BWGT * DUPLCF * FSAF * SSAF.

"True values" of key characteristics, eg., number unemployed (NUM-UE) are determined using

$$NUM\text{-}UE = \sum (FNLWGT)_i * (UE)_i$$

where $(UE)_i$ is the unemployment status of the i-th interviewed adult and

$(UE)_i$ = 1 if the i-th interviewed adult is unemployed
= 0 otherwise.

"True values" for other characteristics are obtained in a similar way. "True values" are presented in Table 1.

The current NIA procedure is then applied to the March 1990 CPS file to obtain estimates of key characteristics. This is done in a way similar to the procedure described earlier for obtaining "true values." The only difference is that NIA factors are input before the Second Stage Adjustment program is run. Here the final weights (FNLWGT) are determined using

FNLWGT =
BWGT * DUPLCF * NIAF * FSAF * SSAF.

Estimates using the current NIA procedure are also presented in Table 1.

The proposed NIA procedure is then used to obtain estimates of key characteristics. Estimates are constructed like constructing estimates using the current NIA procedure. Here we would have different noninterview adjustment factors and thus different second stage adjustment factors. Estimates using this procedure are also presented in Table 1.

Since variables considered are at different levels, we consider the relative bias. Relative bias of an estimator is determined by dividing the bias by the "true value" of the characteristic. Relative biases are given in Table 2.

Variance after all adjustments was obtained using the generalized replication method for variance estimation. We used 48 replicates. The mean squared error (MSE) of estimators using both the current NIA and the proposed NIA was then calculated. As MSE depends on the level, relative MSE is also considered. The relative MSE of an estimator is obtained by dividing the MSE by the square of the "true value" of the characteristic. Relative MSEs are presented in Table 3.

The average absolute relative bias using the current adjustment is 0.6060%, while for the proposed adjustment it is 0.5879%. Thus, the proposed adjustment is better in terms of the average absolute relative bias. The average relative MSE using the proposed adjustment is 0.0820%, while the average relative MSE using the current adjustment is 0.0830%. In this sense, the proposed adjustment is better. We also look at the minimax NIA adjustment. The maximum relative MSE under the current NIA adjustment is 0.6015%, while for the proposed adjustment this quantity is 0.5934%. Hence, under the minimax rule the proposed NIA adjustment is preferred.

The proposed adjustment is better than the current adjustment under each of the three criteria considered. We cannot say whether the proposed adjustment is significantly better.

When looking at the average absolute relative bias and the average relative MSE, we assume that the twelve characteristics are of equal importance. If characteristics are not of equal importance, other linear combinations of characteristics can be considered. We note that the results of comparisons between adjustments depend on the characteristics considered and the weight given to each characteristic.

5. Research Issues

We have compared the NIA proposed by the best response propensity model using March 1990 CPS data to the current NIA. The comparison was done using March 1990 CPS data. We may apply the proposed NIA and the current NIA to other months to see how the two NIAs compare in other months.

Some alternative approaches that need consideration are listed below. These approaches may provide a better adjustment.

1. Group 'similar' PSUs into noninterview clusters. Within clusters form cells.

2. Further subdivide response propensity groups into cells. Groups that have a large contribution to the Relative MSE can be considered candidates for subdivision.

3. Do NIA by type of nonresponse. We can have a separate NIA for each type of noninterview.

4. Determine the best response propensity models for months January through June of 1990. A response propensity model that uses all variables and levels of variables present in the six models can be chosen for the whole decade.

5. Use longitudinal data to obtain a better model for predicting response propensity. Information from previous months, on variables such as number of persons in the household, can be used for a different NIA for households that responded in a previous month.

6. Look at other noninterview

adjustment factors. Another noninterview adjustment factor that can be used is the reciprocal of the response propensity. An advantage of using this is that NIA factors are fixed for the decade.

Acknowledgements

I would like to thank Lynn Weidman for helpful discussions on this topic, Patrick Cantwell for reading an earlier draft and Lawrence Ernst for giving me a draft of his paper and helpful discussions.

References

Ernst, L. (1977). Weighting to Adjust for Partial Nonresponse. Proceedings of the Survey Research Methods Section, American Statistical Association, 468-473.

Hanson, R. H. (1977). The Current Population Survey: Design and Methodology, Technical Paper No. 40, United States Bureau of the Census, U.S. Government Printing Office, Washington, D.C.

Kalton, G. and Maligalig, D.S. (1991). A Comparison of Methods for Weighting Adjustment for Nonresponse. Seventh Annual Research Conference, Washington, D.C.

Little, R. J. A. (1986). Survey Nonresponse Adjustments. Int. Statist. Rev., 54, 139-157.

Oh, H.L. and Scheuren, F.J. (1983). Weighting Adjustments for Unit Nonresponse. Incomplete Data in Sample Surveys, Vol. 2. Theory & Bibliographies, Madow, Olkin and Rubin (eds.) New York: Academic Press, pp. 143-184.

Palmer, S. (1968). On the Character and Influence of Nonresponse in the Current Population Survey. Proceedings of the Social Statistics Section, American Statistical Association, 73-80.

Politz, A.N. and W.R. Simmons (1949). An Attempt to Get the "Not at Homes" into the Sample Without Call-backs. J. Amer. Stat. Assoc., 44, 9-31.

Politz, A.N. and W.R. Simmons (1950). An Attempt to Get the "Not at Homes" into the Sample Without Call-backs. J. Amer. Stat. Assoc., 45, 136-137.

Raj, D. (1968). Sampling Theory. McGraw-Hill, Inc., New York, N.Y.

U.S. Department of Commerce, Bureau of the Census. Internal Memorandum from Jones to Walsh, "1980 CPS Redesign: Specifications for Computing Noninterview Adjustment Factors Beginning January 1985," November 2, 1984. ID # K-23.

TABLE 1. "True Values" (in millions) of key characteristics and their estimates using the current NIA procedure and the proposed NIA procedure.

Characteristics	"True Value"	Estimate using current NIA	Estimate using proposed NIA
Number Employed	116.910595	117.039090	117.037820
Number Unemployed	6.761442	6.697402	6.698170
Number not in CLF	63.856625	63.792524	63.793025
Number of Whites Employed	101.178440	101.236800	101.239040
Number of Whites Unemployed	5.065067	5.032723	5.032140
Number of Whites not in CLF	53.832340	53.806317	53.804664
Number of Blacks Employed	11.881552	11.930843	11.926362
Number of Blacks Unemployed	1.464410	1.431347	1.432715
Number of Blacks not in CLF	7.865320	7.849092	7.852206
Number of Hispanics Employed	8.766612	8.751627	8.749391
Number of Hispanics Unemployed	0.735201	0.722509	0.723294
Number of Hispanics not in CLF	4.657128	4.684805	4.686256

TABLE 2. Relative Bias (in percent) in estimating key characteristics using the current NIA procedure and the proposed NIA procedure.

Characteristics	Relative Bias using current NIA	Relative Bias using proposed NIA
Number Employed	0.1099	0.1088
Number Unemployed	-0.9471	-0.9358
Number not in CLF	-0.1004	-0.0996
Number of Whites Employed	0.0577	0.0599
Number of Whites Unemployed	-0.6386	-0.6501
Number of Whites not in CLF	-0.0483	-0.0514
Number of Blacks Employed	0.4149	0.3771
Number of Blacks Unemployed	-2.2578	-2.1644
Number of Blacks not in CLF	-0.2063	-0.1667
Number of Hispanics Employed	-0.1709	-0.1964
Number of Hispanics Unemployed	-1.7263	-1.6195
Number of Hispanics not in CLF	0.5943	0.6254

TABLE 3. Relative MSE (in percent) of estimates using the current NIA and the proposed NIA.

Characteristics	Relative MSE current NIA	Relative MSE proposed NIA
Number Employed	0.0010	0.0010
Number Unemployed	0.0485	0.0482
Number not in CLF	0.0029	0.0029
Number of Whites Employed	0.0008	0.0008
Number of Whites Unemployed	0.0567	0.0567
Number of Whites not in CLF	0.0026	0.0026
Number of Blacks Employed	0.0135	0.0132
Number of Blacks Unemployed	0.1908	0.1874
Number of Blacks not in CLF	0.0235	0.0230
Number of Hispanics Employed	0.0120	0.0120
Number of Hispanics Unemployed	0.6015	0.5934
Number of Hispanics not in CLF	0.0428	0.0429

[1] This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the author and do not necessarily reflect those of the Census Bureau.