LONGITUDINAL SURVEY NONRESPONSE ADJUSTMENT BY WEIGHT CALIBRATION FOR ESTIMATION OF GROSS FLOWS

A.C. Singh, S. Wu, and R. Boyer, Statistics Canada
A.C. Singh, Survey and Analysis Methods Development, Statistics Canada, 16-A Coats Bdg, Ottawa K1A 0T6

KEY WORDS: Aggregate vs. unit level information; Asymptotic design consistency; Internal consistency.

1. INTRODUCTION

It is well known that the usual survey estimates based only on respondents in the case of unit nonresponse or complete items in the case of item nonresponse may be biased. There exist various adjustment methods to reduce bias due to nonresponse, see reviews by Platek, Singh and Tremblay (1978), Chapman, Bailey and Kasprzyk (1986) and Little (1986). In this paper, we restrict ourselves to adjustment methods for unit nonresponse and consider the problem of nonresponse adjustment for longitudinal surveys by respondent reweighting. In particular, we address two issues. First is the general question of ensuring asymptotic design consistency (ADC) of the estimates after nonresponse adjustment; the estimates of interest are either based on one-wave respondents for cross-sectional estimation or two-wave respondents for gross-flow estimation or multi-wave respondents for transition histories. Second is the question of ensuring internal consistency of estimates based on m-wave respondents ($m \geq 2$) with those based on m'-wave respondents ($1 \leq m' \leq m-1$). The second issue distinguishes the nonresponse problem for longitudinal surveys from that for cross-sectional surveys as this issue does not even arise for the latter case. The above issues are discussed in this paper in the context of only one- and two-wave respondents for convenience. Treatment of m-wave respondents is analogous.

First consider one-wave respondent reweighting for nonresponse. The weight adjustment methods for nonresponse correspond to some models for the response mechanism involving response predictors as auxiliary variables. Denoting by ψ_k the probability of response for the kth responding unit, its weight is adjusted by a factor $\hat{f}_k = 1/\hat{\psi}_k$ where $\hat{\psi}_k$ is an estimate of ψ_k (the factor f_k will also be termed as the f-weight.) The nonresponse model should include as many important predictors as possible. In addition, the following two properties for the method of model fitting seem desirable:

(a) the resulting adjustment factors $\hat{f}_k$ are not less than 1; and
(b) the resulting estimates are ADC.

Property (a) is desirable since ψ_k lies between 0 and 1, under the model. Moreover, a nonresponse adjustment factor should not be less than 1 because the weights of nonrespondents are in effect being distributed over respondents.

Now consider the case of two-wave respondent reweighting for nonresponse. For the kth unit, let $\psi_k(t-1)$, $\psi_k(t)$ denote the probability of response for the waves $t-1$ and t, respectively, and $\psi_k(t-1,t)$ denote the joint probability of response for both waves. These three response probabilities give rise to three types of adjustment factors: first for the respondent set for wave $t-1$, second for the respondent set for wave t, and third for the common respondent set for the two waves. These three types of adjustment factors lead to three sets of weights for expansion estimates involving the two waves. Therefore, an additional property seems desirable for the adjustment factors:

(c*) Estimates for each wave based on the common or two-wave respondents are *internally consistent*, i.e., are the same as the estimates based on corresponding single-wave respondents after reweighting for nonresponse.

Note that in the absence of nonresponse, the usual weights do have the property of internal consistency. Also note that since the number of two-wave respondents would usually be less than the number of one-wave respondents of each wave, the set of two-wave nonresponse adjusted weights would be different from each set of one-wave adjusted weights. Therefore, it may not be feasible to have the property (c*) satisfied for all the study variables. However, it may be feasible for a few selected variables deemed important. In this paper, the property (c*) will be replaced by a weaker property (c):

(c) For selected study variables, estimates for each wave based on the two-wave adjusted weights are *internally consistent* with the estimates based on single-wave adjusted weights.

Currently there does not seem to exist a method of weight adjustment for nonresponse that satisfies all the properties (a)-(c). A method termed Generalized Weight Calibration in Sampling (GWCS) is proposed which satisfies (a)-(c). This method is based on the ideas of Weight Calibration in Sampling (WCS) of Deville and Särndal (1992). Although both GWCS and WCS are concerned with weight adjustment, their motivations are different; the former is for bias reduction, the latter is for variance reduction. The term generalized in GWCS is used to signify that it is not hinged on minimizing a distance function as is WCS. The idea behind GWCS may be useful, in general, for the problem of nonresponse adjustment for both cross-sectional and longitudinal surveys.

In Section 2, commonly used methods of weight adjustment for nonresponse are reviewed, with emphasis on the two-wave respondent weighting. This is useful for motivating the proposed GWCS method, which is presented in Section 3. For the commonly used logistic model for nonresponse, it is interesting to note that GWCS gives rise to a simple modification of the familiar raking-WCS. For all the methods, suitable conditions for ensuring the ADC property are discussed. In section 4, application of the GWCS is illustrated using data from the second round of the Labour Market Activity Survey (LMAS) -- a longitudinal annual survey conducted by Statistics Canada for the period 1988-1990. Finally, Section 5 contains concluding remarks.

2. REVIEW OF SAMPLING WEIGHT ADJUSTMENT METHODS FOR NONRESPONSE

Consider a probability sample $\{y_k: k \epsilon s\}$ from a finite population U and let π_k denote the sample inclusion probability for the k^{th} unit in U. Let n, N denote respectively the fixed sample and population sizes. The sampling (or design) weights are the inverse of the inclusion probabilities and will be denoted by $\{h_k\}$. Let s_R denote the subsample consisting of respondents and let f_k as before be the inverse of the response probability (ψ_k) for the k^{th} unit in the population. It is assumed that there exists a superpopulation model for nonresponse which assigns the value of 1 meaning "responds" to the k^{th} unit in the population with probability ψ_k, and the value of 0 meaning "does not respond" with probability $1-\psi_k$. Under the design-cum-nonresponse model, the usual weighted estimator based on respondents, $\sum_{k\epsilon s_R} y_k h_k$ is biased for $\tau_y (: = \sum_{k\epsilon U} y_k)$ but the nonresponse adjusted estimator, $\sum_{k\epsilon s_R} y_k h_k f_k$ is unbiased for τ_y where y denotes a study variable. Since the f-weights are usually unknown, they are estimated by fitting the nonresponse model to the available data on respondents and nonrespondents. Suppose the nonresponse model is given by $\psi_k = \psi(\boldsymbol{x}_k, \lambda)$, for $k \epsilon U$, where $\boldsymbol{x}_k$ is a p-vector of nonresponse predictors (or auxiliary variables) and λ is a p-vector of model parameters. To ensure that the estimated ψ_k is between 0 and 1, a logit model is often used. However, sometimes other models such as linear and log-linear are also used.

The above framework is applicable to one-wave, two-wave, or multi-wave respondents. The response set s_R will, in general, be non-increasing in size as the number of waves increases.

For the two consecutive waves $t-1$ and t, $\psi_k(t-1,t)$ is traditionally modelled in two stages; see, for example, Rizzo *et al.* (1994), Binder *et al.* (1994), and Folsom & Witt (1994). First, the unconditional probability of response to the first wave, $\psi_k(t-1)$ is modelled and then the conditional probability of response to the second wave (given response to the first wave) $\psi_k(t|t-1)$ is modelled. The joint response probability is calculated using the relation $\psi_k(t-1) = \psi_k(t-1)\, \psi_k(t|t-1)$. We call this approach *separate*. In this approach, when fitting a model for $\psi_k(t|t-1)$, wave $t-1$ nonrespondents are discarded. Thus, it is implicitly assumed that the nonresponse pattern is monotone (i.e., second wave respondents are also first wave respondents). However, nonresponse pattern is rarely monotone in practice. An alternative is to model the joint probability $\psi_k(t-1,t)$ directly in one stage. This will be termed the *combined* approach. In the next sub-section, we consider methods of model fitting that are applicable to both approaches.

2.1 Methods of Fitting Nonresponse Models

There are two commonly used methods of model fitting: the method of pseudo maximum likelihood (PML) and the method of moments (MOM). It is assumed that the predictor information is available for each responding unit. However, for nonrespondents, it may be available at the unit level for some predictors and at the aggregate level (i.e., for the nonrespondent population) for others. In this section, unit level information for nonrespondents for all the predictors is assumed to be available. For PML, suppose we use the Bernoulli distribution for the response variable r_k (with values 1 or 0). Under the superpopulation model, the estimating equations for λ are, for $i = 1, \ldots p$,

$$\sum_{k \epsilon s} h_k (\partial \psi_k / \partial \lambda_i) \psi_k^{-1} (1 - \psi_k)^{-1} (r_k - \psi_k) = 0, \qquad (2.1)$$

where $\psi_k = \psi_k (\boldsymbol{x}_k, \lambda)$.

Now consider the method of moments. Denote by "E" the expectation operator under the design-cum-nonresponse model and let T_{xi} be the expansion estimate of τ_{xi}, i.e., $\sum_{k \epsilon s} x_{ki} h_k$. We have for $i = 1, \ldots, p$, $E(\sum_{k \epsilon s_R} x_{ki} h_k f_k - T_{xi}) = 0$ and therefore, the corresponding estimating equations for λ are

$$\sum_{k \epsilon s_R} x_{ki} h_k f_k = T_{xi}, \; i = 1 \ldots p, \qquad (2.2)$$

where as before f_k is $\psi_k^{-1} (\boldsymbol{x}_k, \lambda)$.

Remark 2.1: For the simplest model in which x-variables form MEE (mutually exclusive and exhaustive) categories, the two methods (2.1) and (2.2) coincide and give rise to the familiar weighting cell adjustment method; the cells correspond to the MEE categories. Suppose, for convenience, the x-variables themselves correspond to the MEE categories. Then for the i^{th} category, $\hat{f}_k$ is given by $\hat{f}_k = T_{xi} / \sum_{k \epsilon s_R} x_{ki} h_k = \sum_{k \epsilon s(i)} h_k / \sum_{k \epsilon s_R(i)} h_k$, where $s(i)$ is the total sample in cell i and $s_R(i)$ is the respondent sample in cell i. In other words, $\hat{f}_k$ is constant for cell i and is simply inverse of the observed response rate based on weights. Clearly, $\hat{f}_k$ so defined is always greater than unity.

Remark 2.2: With the commonly used logistic regression model, we have

$$f_k = (1 + \exp(\boldsymbol{x}_k' \lambda)) / \exp(\boldsymbol{x}_k' \lambda) = 1 + \exp(-x_k' \lambda). \qquad (2.3)$$

The estimating equations for the pseudo-likelihood method are given by, for $i = 1, \ldots, p$,

$$\sum_{k \epsilon s} x_{ki} \left[r_k - \exp(\boldsymbol{x}_k' \lambda) / (1 + \exp(\boldsymbol{x}_k' \lambda)) \right] h_k = 0. \qquad (2.4)$$

For the method of moments, the estimating equations are $\sum_{k \epsilon s_R} x_{ki} h_k (1 + \exp(-\boldsymbol{x}_k' \lambda)) = \sum_{k \epsilon s} x_{ki} h_k$ or

$$\sum_{k \epsilon s_R} x_{ki} h_k \exp(-\boldsymbol{x}_k' \lambda) = \sum_{k \epsilon s - s_R} x_{ki} h_k. \qquad (2.5)$$

The right hand side of (2.5) is the estimate of the total of x_i-variable for the nonrespondent population. It can be obtained directly because unit level information for x-variables about non-respondents is assumed to be available. The p equations (2.5) can be solved for λ using the raking method of weight calibration in sampling (raking-WCS). The above method for logistic model fitting was proposed by Folsom (1991) but motivated in a somewhat different manner.

Remark 2.3: Remark 2.2 suggests a connection between MOM for model fitting and WCS for calibration estimation. If the model for f_k matches with the form of g-weight in WCS, then the corresponding WCS algorithm can be used for MOM. For example, a linear model for f_k leads to the regression-WCS and a log-linear model for f_k leads to the raking-WCS.

2.2 Asymptotic Design Consistency (ADC)

The ADC property of the estimator T_y of τ_y means that the difference goes to zero in probability as n, $N \rightarrow \infty$ in the sense of Isaki and Fuller (1982). Both methods, PML and MOM, give rise to ADC estimates under the assumed nonresponse model because the difference $\hat{\lambda} - \lambda$ goes to zero in probability. However, to ensure ADC for PML, the sampling weights h_k in Equation (2.1) should not be omitted because the design, in general, may not be ignorable for the nonresponse model, see for example Pfeffermann (1993). In the case of MOM, the familiar WCS methods such as regression and raking may not lead to ADC estimates unless the g-weights correspond to the model for f_k in view of Remark 2.3; see Fuller *et al.* (1994) for the use of regression-WCS in the presence of nonresponse. It should also be noted that enlarging the nonresponse model to include important predictors as they become available will not disturb ADC under the original model as long as the "true" model is believed to be nested in the current model.

2.3 Internal Consistency

Suppose single-wave estimates for selected y-variables at t-1 and t are available. Now both existing methods (PML and MOM) have the limitation that internal consistency is not satisfied under both separate and combined modelling approaches. One may perform calibration as an additional step, after the model is fitted such that the estimates become internally consistent. However, the resulting estimates will no longer be ADC. In principle, it may be possible to use the constrained PML method to satisfy internal consistency, but this may be difficult.

For MOM, observe that although the existing method relies on unit level predictor information about nonrespondents, for solving Equation (2.2), it is sufficient to have aggregate level information in the form of T_{xi}'s. Thus, for selected y-variables for which single wave estimates are available at t-1 and t, we could easily enlarge the nonresponse model to include these y-variables as predictors. The resulting estimates based on model fitting by MOM will be internally consistent because single wave estimates are used as (calibration) controls. This is the motivation behind the GWCS method (a version of MOM) proposed in the next section.

3. PROPOSED METHOD OF GENERALIZED WCS

Consider the MOM of fitting a nonresponse model to two-wave data using the separate or combined approach. It follows from Remark 2.3 and Section 2.2 that a known WCS-method will give ADC estimates if the corresponding form of g-weights matches with the form of f-weights obtained under the nonresponse model. Now the g-function for WCS is based on the underlying distance function between old weights and new weights. Since the concept of distance function is secondary for the nonresponse problem, we can generalize the WCS method by choosing the g-function according to the nonresponse model, i.e., by setting g_k equal to f_k (inverse of the response probability) and then solving for the λ-parameters from the usual calibration equations. Although it may not be easy in general to find the corresponding distance function, it is no longer important. In particular, there does not exist a known WCS-method which corresponds to the logistic model, i.e., for which the g-weights have the form $g_k = 1 + \exp(-\boldsymbol{x}_k'\lambda)$.

We thus propose a generalized WCS method for nonresponse adjustment defined by the following calibration equations:

$$\sum\nolimits_{k \in s_r} x_{ki} h_k f_k = \tau_{xi},\ i=1,\ldots p. \tag{3.1}$$

where f_k is a function of $\boldsymbol{x}_k$ and λ, as given earlier in Section 2, i.e., equal to $(\psi(\boldsymbol{x}_k, \lambda))^{-1}$ and τ_{xi} may be replaced by its estimate T_{xi} if necessary. The ψ-function is governed by the underlying nonresponse model. The system (3.1) of p-equations will generally require an iterative method such as Newton-Raphson for solving for λ if the solution exists.

We can apply the above method for the logistic nonresponse model given by (2.3). The calibration equations are for $i = 1, \ldots p$,

$$\sum\nolimits_{k \in s_r} x_{ki} h_k (1 + \exp(-\boldsymbol{x}_k'\lambda)) = \tau_{xi}$$

or

$$\sum\nolimits_{k \in s_r} x_{ki} h_k \exp(-\boldsymbol{x}_k'\lambda) = \tau_{xi} - \sum\nolimits_{k \in s_r} x_{ki} h_k. \tag{3.2}$$

The right hand side of (3.2) is simply the deficiency for the i^{th} control total. If control totals are replaced by deficiencies in controls due to nonrespondents, then (3.2) simply reduces to the calibration equations for raking-WCS. Note that the f-weights for GWCS are obtained from the g-weights for raking-WCS after adding unity. It is interesting to note that the GWCS method, in spite of being motivated quite differently, turns out to be a natural generalization of the method of Folsom (1991) (given earlier by (2.5)) when only aggregate level information about nonrespondents for some or all of the x-variables is available. Clearly, the two methods coincide if calibration constraints are based on only unit level information about nonrespondents.

Remark 3.1 (ADC) It follows from Section 2.2 that the proposed method ensures ADC of the resulting estimates. In addition, it can be used to enhance any nonresponse model (by including additional predictors for which only aggregate level information about nonrespondents is available) and thus reducing further the bias in estimates due to nonresponse, i.e., the estimates will get closer to being ADC.

Remark 3.2 (Internal Consistency) It follows from Section 2.3 that the proposed method satisfies internal consistency for selected y-variables. Thus, GWCS meets all the three properties (a)-(c) as set out in the introduction. Note however that in practice only key y-variables should be used in calibration because use

Table 1: Differences between the existing methods and GWCS in Gross Flow Estimates.

(Each pair of entries in the table indicates (D_{ij}^{M}, $ARD_{ij}^{M}\times 100$).)

Flow Categories for Work Duration (1989 → 1990)	Existing Method: Weighting cell		PML		Raking	
0 → 1-26	(2789,	5.88)	(2590,	5.46)	(-964,	2.03)
0 → 27-48	(1726,	5.75)	(1792,	5.97)	(-746,	2.49)
0 → 49-52	(1959,	6.30)	(1914,	6.16)	(-1333,	4.29)
1-26 → 0	(-7773,	6.90)	(-7940,	7.05)	(-224,	0.20)
27-48 → 0	(-7082,	8.82)	(-6595,	8.21)	(-897,	1.12)
49-52 → 0	(-8984,	9.89)	(-8395,	9.24)	(-1923,	2.12)

of many predictors may lead to instability in $\hat{\lambda}$.

4. APPLICATION TO THE LMAS DATA

As an illustration, we apply the proposed GWCS method to the Statistics Canada's Labour Market Activity Survey (LMAS) data consisting of a three year panel, 1988-90. The population consists of residents of Ontario who are in the labour force in the year 1988 and suppose the parameters of interest are the gross flows corresponding to the number of people who experienced change in work duration from 1989 to 1990. The variable "work duration" is in number of weeks, grouped into four categories: 0, 1-26, 27-48 and 49-52. The nonresponse rate for 1989 and 1990 waves is respectively 7.6% and 2.4%. The nonresponse rate for both waves is 9.5%. Note that here the nonresponse pattern is not monotone. The possible predictors used for the nonresponse model are age in 1988, work duration in 1989 and work duration in 1990. The age variable has four categories: 0-24, 25-34, 35-54 and 55+.

The usual weighting cell method for nonresponse adjustment with age as predictor is used to obtain single-wave estimates for work duration in 1989 and 1990. The GWCS method provides estimates of gross flows for the variable work duration in the form of a 4 by 4 table such that the margins coincide with the single-wave estimates, i.e., internal consistency is preserved.

It would be of interest to compare the gross flow estimates obtained under GWCS and some existing methods. Three existing methods are applied to the same data set. These are: 1) the conventional weighting cell method under the separate approach; 2) the PML method with a logistic regression again under the separate approach; and 3) the naive raking method under the combined approach. Suppose, the possible predictor variables are age, work duration in 1989 and in 1990. The first two methods use age as predictor for the first stage, and age and work duration in 1989 for the second stage. The naive raking method uses work duration for 1989 and 1990 as predictors. For the proposed GWCS method, the combined approach with predictors as age and work duration for 1989 and 1990 is used.

The naive raking method is chosen because it is widely used and it does preserve internal consistency. Note that it corresponds to a log-linear model for nonresponse. The other two existing methods correspond to logit models for nonresponse: the weighting cell method to a saturated model and PML to an additive model. Apart from the difference in models, fitting procedures are also different. Given the set of possible predictors, the two existing methods (other than raking) do use all the available predictors except for the work duration in 1990, as it is unknown for nonrespondents at the unit level. The GWCS does use the additional predictor because it only needs information at the aggregate level for nonrespondents.

The gross flow estimates by the four methods are compared. Define measures of difference and absolute relative difference with respect to the GWCS method as $D_{ij}^{M}=y_{ij}^{M}-y_{ij}^{G}$ and $ARD_{ij}^{M}=|D_{ij}^{M}/y_{ij}^{G}|$, where y_{ij}^{G} is the estimate of the (i,j)th cell in the 4 by 4 gross flow table produced by the proposed method and y_{ij}^{M} is the estimate produced by an existing method M. In terms of the stock (or margin) estimates, there is little difference between the methods; results are not shown here. Except for the cell category "0", the ARD is at most 1.5%; for "0", the ARD is between 1.1% and 2.7%. However, for flow estimates, the difference could be appreciable although most of the ARDs (not shown here) are well below 3%. In particular, for flows representing transitions from employment to unemployment and vice-versa, there are significant differences, especially for the weighting cell and the PML methods. The values of the corresponding D_{ij}^{M}'s and ARD_{ij}^{M}'s are listed in Table 1. These differences indicate that the nonresponse model should be carefully chosen when estimating gross flows. Since the "true" nonresponse model is rarely known in practice, choice of an adjustment procedure should be based on a principled approach. In light of the discussion in the introduction about the desirable properties, we recommend the use of GWCS for estimating gross flows.

5. SUMMARY AND CONCLUDING REMARKS

We discussed an interesting finding that for longitudinal surveys, the commonly used methods for respondent reweighting for nonresponse do not preserve the internal consistency property for estimation of gross flows. To overcome this problem, we proposed a method of generalized weight calibration in sampling for nonresponse adjustment for multi-wave longitudinal surveys. Its properties are summarized below:

- It is easy to implement in practice.
- It does not require monotone nonresponse pattern.
- It ensures the weight adjustment factor ≥ 1, which is clearly desirable.
- It produces design-consistent estimates when the underlying nonresponse model holds.
- It preserves internal consistency, i.e., it maintains additivity with respect to the margins which are the single-wave estimates.
- It allows for enhancing the nonresponse model by adding important predictors which may have only aggregate level information about nonrespondents; this phenomenon arises naturally in the context of longitudinal surveys.

Finally, we remark that the proposed method's capability of being able to use predictors with only aggregate level information about nonrespondents may have implications, in general, for enhancing the nonresponse adjustment for both cross-sectional and longitudinal surveys.

Acknowledgement The authors are grateful to M.P. Singh for general discussion and encouragement, to Stephan Roller for providing the LMAS data set and to S. Michaud and C. Le Petit for useful discussions and programs for the LMAS application. The first author's research was supported in part by a grant from NSERC of Canada held at Carleton University, Ottawa.

REFERENCES

Binder, D.A., Michaud, S. and Poirier, C. (1994). Model-based reweighting for nonresponse adjustment. *Proceedings of the American Statistical Association, Survey Research Methods Section*, 989-994.

Chapman, D.W., Bailey, L. and Kasprzyk, D. (1986). Nonresponse adjustment procedures at the U.S. Bureau of the Census. *Survey Methodology*, 12, 161-180.

Deville, J.C. and Särndal, C.-E. (1992). Calibration estimators in survey sampling. *JASA*, 87, 376-382.

Folsom, R.E. (1991). Exponential and logistic weight adjustments for sampling and nonresponse error reduction. *Proceedings of the American Statistical Association, Social Statistics Section*, 197-202.

Folsom, R.E. and Witt, M.B. (1994). Testing a new attrition nonresponse adjustment method for SIPP. *Proceedings of the American Statistical Association, Survey Research Methods Section*.

Fuller, W.A., Loughin, M.M. and Baker, H.D. (1994). Regression weighting in the presence of nonresponse with application to the 1987-88 nationwide food consumption survey. *Survey Methodology*, 20, 75-85.

Isaki, C.T., and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 137-139.

Pfeffermann, D. (1993). The role of sampling weights when modelling survey data. *International Statistical Review*, 61, 317-337.

Platek, R., Singh, M.P. and Tremblay, V. (1978). Adjustment for nonresponse in surveys. In *Survey Sampling and Measurement*, (N.K. Namboodiri, ed.) Academic Press, 157-174.

Rizzo, L., Kalton, G., Brick, M. and Petroni, R. (1994). Adjusting for panel nonresponse in the survey of income and program participation. *Proceedings of the American Statistical Association, Survey Research Methods Section*.