Providing Double Protection for Unit Nonresponse with a Nonlinear Calibration Routine

Darryl V. Creel[1]
Phillip S. Kott[2]

[1,2]RTI International, 6110 Executive Blvd., Suite 902, Rockville, MD 20852

Abstract

There are at least two reasons to calibrate survey weights: force estimators to be unbiased under a prediction model and adjust for the bias caused by unit nonresponse. Although a prediction-model justification is possible, Lundströrm and Särndal (1999) argued that a unit's weight adjustment under calibration estimates the inverse of the unit's response probability. The functional form of the response model in their linear calibration adjustment is awkward and unlikely. We describe a nonlinear calibration procedure available in SUDAAN that includes a logistic response model, generalized raking, and bounds the weight adjustments limiting their inflationary impact on mean squared errors. Using this procedure provides double protection against nonresponse bias. If the linear prediction model or implied unit response model holds, the resulting estimator is asymptotically unbiased.

Key Words: Nonresponse, Calibration, SUDAAN

1. What is calibration?

In the absence of nonresponse, calibration is a weight adjustment method that creates a set of weights, $\{w_k | k \in S\}$, where S is the set of selected sampling units, that

1. Are close the original design weights, $d_k = 1/\pi_k$ where π_k is the probability of selection for the k^{th} selected sampling unit. Since the w_k weights are close to the original design weights, we assume that they will produce nearly unbiased estimates under the randomization distribution.
2. Satisfy a set of calibration equations with one equation for each component of $\boldsymbol{x}_k$ which is the vector of auxiliary variables for the k^{th} selected sampling unit. That is, the sum of the weighted auxiliary information from the selected sample units equals the sum of the auxiliary information from the population

$$\sum_S w_k \boldsymbol{x}_k = \sum_U \boldsymbol{x}_k$$

When estimating a total, $T = \sum_U y_k$, from the weighted sample total, $\hat{T} = \sum_S w_k y_k$, or estimating a mean, $\frac{T}{N} = \frac{\sum_U y_k}{N}$, from the weighted sample mean, $\frac{\hat{T}}{N} = \frac{\sum_S w_k y_k}{\sum_S w_k}$, calibration will tend to reduce mean squared error when y_k is correlated with the components of $\boldsymbol{x}_k$.

More formally, $\hat{T}$ is an unbiased estimator for T under the prediction model:

$$y_k = \boldsymbol{x}^T \boldsymbol{\beta} + \varepsilon_k,$$

where $E(\varepsilon_k | \boldsymbol{x}_k) = 0$ whether or not k is in the sample, i.e., the sample design is ignorable. This means $E(\hat{T} - T) = 0$.

The simplest way to compute calibration weights is linearly,

$$w_k = d_k \left[1 + \left(\sum_U \boldsymbol{x}_j - \sum_S d_j \boldsymbol{x}_j\right)^T \left(\sum_S d_j \boldsymbol{x}_j \boldsymbol{x}_j^T\right)^{-1} \boldsymbol{x}_k\right] = d_k [1 + \boldsymbol{g}^T \boldsymbol{x}_k],$$

which is the generalized regression (GREG) estimator. There are nonlinear calibration routines where $w_k = d_k f(\boldsymbol{g}^T \boldsymbol{x}_k)$, but they are asymptotically equivalent to the GREG because $f(0) = f'(0) = 1$ and $\boldsymbol{g}^T \boldsymbol{x}_k$ converges to zero as the sample size grows.

2. What is double protection for unit nonresponse?

Most surveys experience some level on nonresponse, which is usually beyond our control. We are forced to assume, either explicitly or implicitly, some type of model to adjust for nonresponse. A prediction model on the survey variable usually assumes the response/nonresponsne mechanism, like the sampling design, is ignorable. A response model assumes the response mechanism behaves like a round of Poisson sub-sampling. Double protection means that if *either* the prediction or response model is specified correctly, the estimator is nearly unbiased in some sense.

3. How does the GREG handle unit nonresponse?

The sample S is replaced by the respondent sample R in defining the GREG and $\boldsymbol{g}$ by

$$\widehat{T_{GREG}} = \sum_R w_k y_k = \sum_R d_k (1 + \boldsymbol{g}^T \boldsymbol{x}_k)\, \boldsymbol{x}_k,$$

where

$$\boldsymbol{g} = \left(\sum\nolimits_U \boldsymbol{x}_j - \sum\nolimits_R d_j \boldsymbol{x}_j\right)^T \left(\sum\nolimits_R d_j \boldsymbol{x}_j \boldsymbol{x}_j^T\right)^{-1}$$

or

$$\boldsymbol{g} = \left(\sum\nolimits_S d_j \boldsymbol{x}_j - \sum\nolimits_R d_j \boldsymbol{x}_j\right)^T \left(\sum\nolimits_R d_j \boldsymbol{x}_j \boldsymbol{x}_j^T\right)^{-1}$$

depending on whether the respondent sample is calibrated to the population $(\sum_U \boldsymbol{x}_j)$ or to the original sample $(\sum_S d_j \boldsymbol{x}_j)$. When calibrating to the population, the estimator is unbiased under the prediction model, $y_k = \boldsymbol{x}^T \boldsymbol{\beta} + \varepsilon_k$ and $E(\varepsilon_k | \boldsymbol{x}_k) = 0$, whether or not k is in the respondent sample. When calibrating to the sample, the estimator is nearly unbiased under a combination of the prediction model above and the original sampling design. Either way the estimator is also nearly unbiased under the quasi-sample design that treats response as a second phase of random sampling as long as each unit's probability of response has the form

$$p_k = \frac{1}{1+\boldsymbol{\gamma}^T \boldsymbol{x}_k},$$

and $\boldsymbol{g}$ is a consistent estimator for $\boldsymbol{\gamma}$. Put another way

$$\widehat{T_{GREG}} = \sum_S w_k \boldsymbol{x}_k = \sum_S d_k \frac{1}{\widehat{p_k}} \boldsymbol{x}_k.$$

Notice that with nonresponse $\boldsymbol{g}^T \boldsymbol{x}_k$ no longer converges to 0.

4. What about raking?

Raking is a form of nonlinear calibration in which effectively the weights for the raking adjustment have the form

$$w_k = d_k \exp\,(\boldsymbol{g}^T \boldsymbol{x}_k).$$

Traditionally, the components of $\boldsymbol{x}_k$ are indicator variables, i.e., they are 0/1 variables with the value of 1 indicating the observation has the characteristic, and an iterative proportional fitting routing is used to solve the calibration equations. The components do not have to be binary, but an iterative search using Newton's method is still needed to find the g that satisfies one of the two versions of the calibration equations, i.e., calibration to the population or the sample.

Using raking to adjust the weights results in a calibration estimator (nearly) unbiased under the same linear prediction model as the GREG. The quasi-random response model under which the raking estimator is nearly unbiased has a more reasonable form than the GREG. It is

$$p_k = \exp\,(-\boldsymbol{\gamma}^T \boldsymbol{x}_k).$$

Raking is asymptotically equivalent to the GREG when there is no nonresponse or when every unit is equally likely to respond.

5. What about fitting a logistic regression model?

Search for a $\boldsymbol{g}$ that forces the w_k to satisfy a version of the calibration equations the weights using the logistic adjustment have the from

$$w_k = d_k \exp\,(1 + \boldsymbol{g}^T \boldsymbol{x}_k),$$

which produces a calibration estimator that is nearly unbiased under the same linear prediction model as the GREG. This estimator is also nearly unbiased under the logistic response model

$$p_k = [1 + exp(-\boldsymbol{\gamma}^T \boldsymbol{x}_k)]^{-1} = \frac{exp(\boldsymbol{\gamma}^T \boldsymbol{x}_k)}{1+exp(\boldsymbol{\gamma}^T \boldsymbol{x}_k)}.$$

The weight adjustments $f(\boldsymbol{g}^T \boldsymbol{x}_k)$ are centered at 2, when $\boldsymbol{g}$ = 0. By contrast, raking and GREG adjustments are centered at 1.

6. A Useful Generalization of Raking and Logistic Weighting

In general, logistic weight adjustments cannot be less than 1; raking weight adjustments cannot be less than 0; and GREG weights can be negative. None of the three weight adjustments have an upper bound. A useful generalized weight adjustment that contains bounds proposed by Deville and Sarndal (1992) is

$$f(\boldsymbol{g}^T \boldsymbol{x}_k) = \frac{L(U-1)+U(1-L)\exp(A\boldsymbol{g}^T \boldsymbol{x}_k)}{(U-1)+(1-L)\exp(A\boldsymbol{g}^T \boldsymbol{x}_k)},$$

where

$$A = \frac{U-L}{(1-L)(U-1)}$$

is centered at C with lower bound $L \geq 0$ and upper bound $U > C > L$. The user sets these parameters.

7. SUDAAN's WTADJSUT Procedure

To extend the bounded function further, SUDAAN's WTADJUST procedure allows for separate weights for each sample respondent *k* as proposed by Folsom and Singh (2000) with the function defined as

$$f_k(\boldsymbol{g}^T \boldsymbol{x}_k) = \frac{L_k(U_k - C_k) + U_k(C_k - L_k)\exp(A_k \boldsymbol{g}^T \boldsymbol{x}_k)}{(U_k - C_k) + (C_k - L_k)\exp(A_k \boldsymbol{g}^T \boldsymbol{x}_k)}$$

where

$$A_k = \frac{U_k - L_k}{(U_k - C_k)(C_k - L_k)}$$

but with common $\boldsymbol{g}$ chosen to satisfy one of the calibration equations. Some d_k may also be scaled by a trimming factor, but these factors should be used sparingly, because trimming perturbs the sample design weights which could induce bias into the estimates.

We can set common bounds and centers for all respondent sample units indexed by *k*. Alternatively, we may want to bound the weights themselves or bound the weighted total. This can be done with SUDAAN's WTADJUST procedure. For example, setting $L_k = \frac{1}{d_k}$ forces all weights to be at least 1, and, setting $U_k = \frac{U}{w_k y_k}$ keeps the weighed totals no greater than *U*.

When adjusting for nonresposne, it makes sense to center at the inverse of the overall response rate (Folsom and Witt 1994). When using a logistic distribution to model nonresponse, not being able to do set the value of the center is a limitation.

When the population totals for the components of the $\boldsymbol{x}$ vector of auxiliary variables in known, WTADJUST calibration can also be used to adjust for coverage errors in the frame. In this quasi-randomization context, calibration implicitly estimates the expected number of times the k^{th} respondent sample unit is on the frame. Here, a reasonable center would be the overall coverage rate.

When calibrating for consistency with outside sources or for mean squared error reduction in the absence of nonresponse and coverage errors, we can center at 1, like the GREG, and make use of the bounding properties of WTADJUST.

7. Generalizing Further

Although we know how to estimate the mean squared errors of WTADJUST-calibration estimators using linearization, it has not yet been programmed in SUDAAN. The problem is the multiple calibration steps. One could use replication to estimate those mean squared errors in SUDAAN as long as the first-stage sample is selected with replacement, or we can treat the sample as if it were.

Why not allow for the possibility that nonrespondents are not missing at random? In particular, what if assumed response model

$$p_k = f_k(\boldsymbol{g}^T\boldsymbol{z}_k)^{-1} = \frac{L(U-C)+(C-L)\exp\,(A_k\boldsymbol{g}^T\boldsymbol{z}_k)}{(U-C)+U(C-L)\exp\,(A_k\boldsymbol{g}^T\boldsymbol{z}_k)},$$

where some components of $\boldsymbol{z}_k$ are known only for respondents, but fit calibration equations as an $\mathbf{x}$ vector containing values for respondents and nonrespondents? We often can when the dimension of $\boldsymbol{x}_k$ is greater than that of $\boldsymbol{z}_k$. This is coming in SUDAAN 11 and will be called WTADJSTX.

References

Deville, Jean-Claude and Sarndal, Carl-Erik (1992), "Calibration Estimators in Survey Sampling," *Journal of the American Statistical Association*, Vol. 87, No. 418, pp. 376-382.

Folsom, R. E. and Singh, A. C. (2000), "The Generalized Exponential Model for Sampling Weight Calibration for Extreme Values, Nonresponse, and Poststratification, *Proceedings of the American Statistical Association, Survey Research Methods Section*, pp. 598-603.

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*, pp. 428-433.