WTADJX is Coming: Calibration Weighting in SUDAAN when Unit Nonrespondents Are Not Missing at Random and Other Applications

Phillip S. Kott[1]

Abstract

SUDAAN 10 (RTI 2008) introduced the WTADJUST procedure, which produces calibrated survey weights. By forcing the weighted totals for a set of "calibration variables" to equal benchmark totals computed from either the whole sample, a larger sample, or the entire population, calibration weights can reduce or eliminate the potential for bias from unit nonresponse under a reasonable response model. WTADJX, which will be available in SUDAAN 11 (RTI 2012) allows the set of model variables governing the response model to differ from the set of calibration variables. As a result, the new procedure may be used to assess and perhaps even correct for unit nonresponse that is not missing at random, that is to say, when unit response is a function of variables known only for the respondents. There are other potential uses of WTADJX. These include producing calibrated weights asymptotically identical to so-called "optimal" ones and simplifying the computation of replicate weights when using a nonlinear calibration routine.

Key Words: Response model; Nonresponse bias; Generalized raking; Variance estimation; Optimal calibration.

1. What is calibration weighting?

In the absence of nonresponse, calibration is a weight-adjustment method that creates a set of weights, $\{w_k\}$ that

1. Are asymptotically close to the original design weights: $d_k = 1/\pi_k$ (as the sample size grows arbitrarily large, w_k converges to d_k) and therefore *nearly* unbiased under probability-sampling theory.

2. Satisfy a set of calibration equations (one for each components of $\mathbf{x}_k$):

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

When estimating $T = \sum_U y_k$ with $t = \sum_S w_k y_k$ or $\bar{y}_U = T/N$ with $\sum_S w_k y_k / \sum_S w_k$, *calibration weighting* will tend to reduce mean squared error when y_k is correlated with components of $\mathbf{x}_k$ (but a real survey has many y_k's).

[1] RTI International, 6110 Executive Blvd., Suite 902, Rockville, MD 20852, U.S.A.
Email: pkott@rti.org

One way to compute calibration weights is linearly with the following formula:

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

$$= d_k\left[1+\mathbf{g}^T\mathbf{x}_k\right], \quad \text{where } \mathbf{g}=\left(\sum_S d_j\mathbf{x}_j\mathbf{x}_j{}^T\right)^{-1}\left(\sum_U \mathbf{x}_j - \sum_S d_j\mathbf{x}_j\right).$$

Observe that as the sample size grows arbitrarily large, $\mathbf{g}^T\mathbf{x}_k$ (which means **g**) converges to **0**.

This is the weighting scheme implied by the *generalized regression* (GREG**)** estimator since

$$\sum_S w_k y_k = \sum_S d_k y_k + \left(\sum_U \mathbf{x}_j - \sum_S d_j\mathbf{x}_j\right)^T\left(\sum_S d_j\mathbf{x}_j\mathbf{x}_j{}^T\right)^{-1}\sum_S d_k\mathbf{x}_k y_k$$

$$= \sum_S d_k y_k + \left(\sum_U \mathbf{x}_j - \sum_S d_j\mathbf{x}_j\right)^T\mathbf{b}.$$

2. **The GREG and unit nonresponse?**

How does the GREG handle unit nonresponse? The sample *S* is replaced by the respondent sample *R* in defining the GREG and **g:**

$$t_{GREG} = \sum_R w_k y_k = \sum_R d_k\left(1+\mathbf{g}^T\mathbf{x}_k\right)y_k$$

where

$$\mathbf{g} = \left(\sum_U \mathbf{x}_j - \sum_R d_j\mathbf{x}_j\right)^T\left(\sum_R d_j\mathbf{x}_j\mathbf{x}_j{}^T\right)^{-1} \quad \text{or}$$

$$= \left(\sum_S d_j\mathbf{x}_j - \sum_R d_j\mathbf{x}_j\right)^T\left(\sum_R d_j\mathbf{x}_j\mathbf{x}_j{}^T\right)^{-1},$$

depending on whether the respondent sample is calibrated to the population $\left(\sum_U \mathbf{x}_j\right)$ or to the original sample $\left(\sum_S d_j\mathbf{x}_j\right)$. Either way, the estimate 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:

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

and **g** is a consistent estimator for **γ.** Put another way: $t_{GREG} = \sum_R w_k \mathbf{x}_k = \sum_R d_k \hat{p}_k^{-1} \mathbf{x}_k$. Notice that with nonresponse neither $(\sum_U \mathbf{x}_j - \sum_R d_j \mathbf{x}_j)^T$ nor $(\sum_S d_j \mathbf{x}_j - \sum_R d_j \mathbf{x}_j)^T$ converges to $\mathbf{0}^T$, and so neither does $\mathbf{g}^T$. This, at the time surprising, use of calibration weighting was proposed by Fuller *et al.* (1994).

3. **Nonlinear Calibration Weighting**

The problem with the probability-or-response function in equation (1) is that it can fall below unity and even be negative. A useful nonlinear form of calibration weighting finds a **g** (through repeated linearization) such that

$$\begin{aligned} \sum_R w_k \mathbf{x}_k &= \sum_R d_k \alpha(\mathbf{g}^T \mathbf{x}_k)\mathbf{x}_k = \sum_U \mathbf{x}_k \quad \text{or} \\ \sum_R w_k \mathbf{x}_k &= \sum_R d_k \alpha(\mathbf{g}^T \mathbf{x}_k)\mathbf{x}_k = \sum_S w_k \mathbf{x}_k, \end{aligned} \tag{2}$$

where $$\alpha(\mathbf{g}^T \mathbf{x}_k) = \frac{\ell(u-c) + u(c-\ell)\exp(A\mathbf{g}^T \mathbf{x}_k)}{(u-c) + (c-\ell)\exp(A\mathbf{g}^T \mathbf{x}_k)}, \tag{3}$$

and $A = (u-\ell)/[(u-c)(c-\ell)]$ (which makes taking the derivate of $\alpha(\mathbf{g}^T \mathbf{x}_k)$ easier).

The *weight adjustment* $\alpha(\mathbf{g}^T \mathbf{x}_k)$ is centered at *c* (i.e., α(0) = 1) with a lower bound $\ell \geq 0$ and an upper bound $u > c > \ell$, which can be infinite. The user sets these *centering* and *bounding* parameters. Equation (3) is a generalization of both raking, where $\ell = 0$, $c = 1$, $u = \infty$, and the implicit estimation of a logistic-regression response model, where $\ell = 1$, $c = 2$, $u = \infty$.

When *c* = 1, equation (3) is the generalized-raking adjustment introduced by Deville and Särndal (1992) to bound the range of the $\alpha(\mathbf{g}^T \mathbf{x}_k)$. Centering at 1 was a requirement of calibration weighting in that landmark paper (α'(0) = 1 was required as well), but setting c > 1 with $\ell = 1$ is more sensible when adjusting for unit nonresponse.

SUDAAN's WTADJUST allows separate weights functions for each *k*:

$$\alpha_k(\mathbf{g}^T \mathbf{x}_k) = \frac{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp(A_k \mathbf{g}^T \mathbf{x}_k)}{(u_k - c_k) + (c_k - \ell_k)\exp(A_k \mathbf{g}^T \mathbf{x}_k)} \tag{4}$$

but with a common **g** chosen to satisfy one of the two versions of the calibration equation (the population or original-sample version). When adjusting for nonresponse (or coverage), it makes sense to center at the inverse of the overall response (coverage) rate.

Although WTADJUST allows $\alpha_k(\mathbf{g}^T \mathbf{x}_k)$ to be *k*-specific, when adjusting for nonresponse (or coverage), it is sensible to select a single value for the c_k and a very limited number of ℓ_k and u_k values.

4. Missing Not at Random

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

$$p_k = \left[\alpha_k(\boldsymbol{\gamma}^T\mathbf{x}_k)\right]^{-1} = \frac{(u_k - c_k) + (c_k - \ell_k)\exp(A_k\boldsymbol{\gamma}^T\mathbf{x}_k)}{\ell_k(u_k - c_k) + u_k(c_k - \ell_k)\exp(A_k\boldsymbol{\gamma}^T\mathbf{x}_k)},$$

where some components of the *model* vector $\mathbf{x}_k$ are known only for respondents (such as y_k itself) but fit calibration equations using a *benchmark* $\mathbf{z}$-vector containing values know for respondents and nonrespondents. In other words, replace equation (2) by

$$\begin{aligned} &\textstyle\sum_R w_k\mathbf{z}_k = \sum_R d_k\alpha_k(\mathbf{g}^T\mathbf{x}_k)\mathbf{z}_k = \sum_U \mathbf{z}_k \quad \text{or} \\ &\textstyle\sum_R w_k\mathbf{z}_k = \sum_R d_k\alpha_k(\mathbf{g}^T\mathbf{x}_k)\mathbf{z}_k = \sum_S w_k\mathbf{z}_k, \end{aligned} \tag{5}$$

so that $\mathbf{g}$ again estimates $\boldsymbol{\gamma}$.

Mathematically, finding a $\mathbf{g}$ that satisfies either the first or second line of (5) can often be done as long as the number of calibration (benchmark) variables in $\mathbf{z}_k$ is at least as great of the number of model variables in $\mathbf{x}_k$. A routine to do that is coming in SUDAAN 11 (RTI 2012): WTADJX. It will work easiest when the numbers of model and calibration variables coincide so that one of the sets of calibration equations in (5) holds. Otherwise, there are more unknowns than equations, and the equations in (5) cannot hold exactly. See Chang and Kott (2008) for a discussion of minimizing the difference between, say, $\sum_R d_k\alpha_k(\mathbf{g}^T\mathbf{x}_k)\mathbf{z}_k$ and $\sum_U \mathbf{z}_k$ as a means for estimating $\boldsymbol{\gamma}$.

5. An Example: Nonresponse in the 2002 Census of Agriculture

The USDA used traditional poststratification within each county to adjust for nonresponse in the 2002 Census of Agriculture. Five groups were formed using frame information on historic sales. Sales increases with Group number, except that Group 5 had same sales as Group 3 or 4 but no survey responses in the previous five years. Call these group identifiers the calibration variables ($\mathbf{z}_k$).

In the following table, we contrast using a poststratification scheme at the state level with using WTADJX where model variables ($\mathbf{x}_k$) are formed based on *actual* 2002 sales. Since the model variables are identifiers of mutually exclusive groups, the choice of $\alpha_k(.)$ doesn't matter in this case except when the linear solution is out of bounds for an alternative choice of $\alpha_k(.)$.

Summary Statistics on the Two Methods of Reweighting for Nonresponse

Model group	Census counts	*Using poststratification*			*Using survey values*		
		Adjusted counts	Response rate	Standard error	Adjusted counts	Response rate	Standard error
1	408,243	465,059	88	219	446,322	91	531
2	288,040	330,097	87	197	329,755	87	1,401
3	256,280	297,589	86	187	309,230	83	7,321
4	216,022	257,410	84	150	268,045	81	327
5	53,107	66,327	80	110	96,127	55	344
Total	1,221,692	1,416,482	86	256	1,449,479	84	721

It appears that poststratification, which assumes nonresponse to the census is a function of a farm's expected size is biased downward if nonresponse is actually a function of a farm's realized size. Kott (2005) provides a thorough discussion of these results.

Observe that the standard errors are higher when a farm's realized sales is used in nonresponse adjustment. Moreover, the possibility exists that there is no satisfying solution to the calibration equation. When that happens, one may need to drop a model variable or add another. See Chang and Kott (2008).

6. "Optimal" Calibration

In the absence of nonresponse and coverage errors, a linear estimator often better than the GREG calibrated on $\mathbf{z}_k$ also calibrates on $\mathbf{z}_k$ but sets $\mathbf{x}_k = (d_k - 1)\mathbf{z}_k$. This produces the nearly unbiased linear estimator with the smallest asymptotic mean squared error under Poisson sampling and similarly under stratified simple random sampling with large stratum samples sizes. As a result, it has been called the "optimal estimator" under Poisson sampling (Rao, 1994) and "pseudo-optimal estimator" more broadly (Banker, 2002).

With WTADJX centered at 1, we can bound the weights and retain the asymptotic properties of the optimal estimator by setting $\mathbf{x}_k = (d_k - 1)\mathbf{z}_k$. In particular, when $d_k > 1$, we can set $\ell_k = 1/d_k$ to assure that all w_k are at least unity, which many find desirable. If some $d_k = 1$, we can simply set $w_k = 1$ and remove k from U and S before applying equation (4). See Kott (2011).

7. Easier Replication Weights

We know how to estimate the mean squared errors of WTADJUST-calibration estimators using linearization. Unfortunately, that knowledge has has yet been incorporated into SUDAAN 10. SUDAAN 11 will include proper asymptotic mean-squared-error estimation for a single-step calibration. Often, however, calibration is done is separate nonresponse adjustment and mean-squared-error reducing steps. In fact, there can be

multiple nonresponse-adjustment steps as well as a coverage-adjustment step. When there are multiple calibration steps, one can use replication to estimate mean squared errors in SUDAAN as long as the first stage sample is drawn with replacement (or we can reasonable treat the sample as if it were).

The problem with replication when using WTADJUST is that it is possible that the one can find calibration weights can satisfy calibration equations in (2) with a particular set of centering and bounding parameters but not replicate weights. The following alternative method of computing replicate weights based on an idea in Kott (2006) may prove helpful. Rather than reproducing the calibration with the same centering and bounding parameters, run WTADJX centered at 1 (and any bounds) on $\alpha_k \times$ the pre-adjusted replicate weight, $d_{k(r)}$. Set $\mathbf{x}_k = (\phi_k/\alpha_k)\mathbf{z}_k$, where

$$\phi_k = \frac{(\alpha_k - \ell_k)(u_k - \alpha_k)}{(c_k - \ell_k)(u_k - c_k)} \quad \text{is the derivate of } \alpha_k(.).$$

This replication approach can also be used with WTADJX, but special handling is necessary when there are more calibration variables than model variables. Sadly, this special handling will be easily accomplished with SUDAAN 11.

8. Concluding Questions

There remains much we don't know about calibration weighting. In particular,

How many calibration variables are too many?

Is using calibration weighting to account for nonrespondents not missing at random practical, and, if so, when?

Empirical research is needed on these questions. WTADJX will makes this research easier.

References

Bankier, M. (2002). Regression Estimators for the 2001 Canadian Census, presented at the International Conference in Recent Advances in Survey Sampling.

Chang, T. and P. S. Kott (2008). Using Calibration Weighting to adjust for Nonresponse Under a Plausible Model, *Biometrika*, 95, 557-571.

Deville, J-C. and Sarndal, C-E. (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.

Fuller, W.A., Loughin, M.M., and Baker, H.D. (1994). Regression Weighting for the 1987-88 National Food Consumption Survey, *Survey Methodology*, 20, 75-85.

Kott, P.S. (2011). A Nearly Pseudo-optimal Method for Keeping Calibration Weights from Falling Below Unity in the Absence of Nonresponse or Frame Errors, *Pakistan Journal of Statistics*, forthcoming.

Kott, P.S. (2009). Calibration Weighting: Combining Probability Samples and Linear Prediction Models, in *Handbook of Statistics_29B: Sample Surveys: Inference and Analysis*, D. Pfeffermann, & C.R. Rao (Eds.), New York: Elsevier.

Kott, P.S. (2006). Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors, *Survey Methodology*, 32, 133-142.

Kott, P.S. (2005). "No" Is the Easiest Answer: Using Calibration to Assess Nonignorable Nonresponse in the 2002 Census of Agriculture, *Proceedings of the ASA Survey Research Methods Section* (available online).

Rao, J.N.K. (1994). Estimating Totals and Distribution Functions Using Auxiliary Information at the Estimations Stage, *Journal of Official Statistics* 25, 1-21.

RTI (2008). *SUDAAN Language Manual, Release 10.0.* Research Triangle Park, NC: RTI International.

RTI (2012). *SUDAAN Language Manual, Release 11.0.* Research Triangle Park, NC: RTI International.