Calibration Weighting for Nonresponse that is Not Missing at Random: Allowing More Calibration than Response-Model Variables

Phillip S. Kott and Dan Liao[1]

Abstract

Calibration weighting can be used to remove bias when unit nonresponse is a function of one or more survey variables. This is done by allowing the model variables in the weight-adjustment function to differ from the variables in the calibration equation. An extension of calibration weighting allows there to be more calibration variables than model variables. Rather than equating the two sides of a calibration equation, the difference between the sides is minimized in some sense. This paper discusses some ways of doing that. A promising solution results instead from an alternative version of the calibration equation. A helpful insight into choosing calibration variables for given model variables follows.

Key words: Weight-adjustment function, Shadow variables, Calibration equation. Double protection.

1. Introduction

The standard approach to applying calibration weighting when adjusting for nonresponse (Fuller, Loughin, and Baker 1994; Folsom and Singh 2000) can provide double protection against bias due to unit nonresponse (Kott 2006; Kim and Park 2006). That is to say, if *either* the expected value of the survey variable is the same linear function of the calibration variables for both respondents and nonrespondent *or* the probability of unit response is a function of the calibration variables identical in form to the inverse of the weight-adjustment function used in calibration weighting, then a calibration-weighted estimator will be nearly unbiased (i.e., its relative bias will be asymptotically ignorable) in some sense. A set of assumptions about the distribution of the survey variable has been called an "outcome" or "prediction model" (because the survey variable is predicted by the model; Royall 1976) while a set of assumptions about which units respond and which do not is usually called a "selection" or "response model."

In this standard calibration-weighting approach to nonresponse adjustment, unit nonrespondents are assumed to be missing at random, that is, unit nonresponse does not depend of variable values known only for respondents. Deville (2000), however, showed that calibration weighting can be used to remove unit nonresponse bias even when nonrespondents are not missing at random by letting the (response-) model variables in the weight-adjustment function differ from the calibration variables. The number of model and calibration variables needed to be the same in Deville's setup, and many of the model and calibration variables could coincide. We call thse coinciding variables "dual variables" here, while model variables that that are not calibration variables are "model-only variables," and calibration variables that are not model variables are "shadow variables." The popular term "instrumental variable" is avoided because model-only variables have the form of instrumental variables (as in Brewer 1995) while shadow variables share their function (as in Wang, Shao, and Kim 2014).

Kott and Chang (2010) extended the notion of double protection to cover Deville's approach to calibration weighting, but as in Chang and Kott (2008), this extension allowed there to be more calibration than model variables. That required expanding what was meant by calibration weighting. Rather than forcing the weighted mean among respondents for a vector of calibration variables to equal a mean estimated from the

[1] RTI International, 6110 Executive Blvd #902, Rockville, MD 20852

full sample or provided by outside source, the difference between the two means, call it **s**, needed to be minimized in some sense.

In Chang and Kott's expanded formulation of calibration weighting, one chooses a symmetric, positive definite matrix **Q** and then finds the calibration weights that minimize $\mathbf{s}^T\mathbf{Q}\mathbf{s}$. Any valid choice for **Q** results in calibration weights that produced nearly unbiased estimators in some sense.

Chang and Kott suggested a methodology for choosing **Q**. We will propose a different approach which we argue will likely lead to a more efficient calibration-weighted estimator. In making the proposal, a revised version of the calibration equation (also noted by Chang and Kott) emerges and with it a simplified variation of our proposal that abandons Chang and Kott's original formulation of calibration weighting (i.e., $\mathbf{s}^T\mathbf{Q}\mathbf{s}$ is not longer minimized for some **Q**). Our two proposals are based on the simple prediction-modeling idea: shadows variables should be chosen that predict the model-only variables. Nevertheless, the resulting calibration weights produce nearly unbiased estimators when the response model holds but the prediction model does not.

Section 2 reviews the background theory in more detail. For simplicity we only treat calibration weighting to the original sample here. In practice, calibration weighting can also be targeted to known population totals, to estimated totals from a different source, or a vector whose components reflect a combination of sample, total and outside sources. A rigorous treatment of the background theory can be found in Chang and Kott (from a response-model viewpoint) and Kott and Chang (from a prediction-model viewpoint).

Section 3 discusses our two new proposals for calibration weighting when there are more calibration than model variables. Section 4 describes a modest simulation experiment demonstrating the increased efficiency from using one of our proposals rather than a potential competitor. Section 5 offers some concluding remarks.

2. Background Theory and Notation

Suppose we have a probability sample S subject to unit nonresponse. Let
y_k be the outcome variable of interest in a population of size N,
I_k be a sample indicator (1 if $k \in S$, 0 otherwise),
$d_k = 1/E(I_k)$ be the original sampling weight,
R_k be a response indicator (1 if k responds, 0 otherwise),
$w_k = R_k d_k \alpha_k$ be a nonresponse-adjusted weight, defined to be 0 for nonrespondents,
p_k be a possibly incorrect implicit guess at $E(R_k)$, so that $\alpha_k = 1/p_k$ and $w_k = d_k(R_k/p_k)$, and
$\mathbf{z}_k$ be vector of calibration variables, which usually includes unity or the equivalent (i.e., some linear combination of components of $\mathbf{z}_k$ is unity).
We will treat y_k, I_k, R_k, and $\mathbf{z}_k$ as random variables here. With the exception of I_k, however, their distributions are unknown.

2.1 Missing at Random

When calibration is to the full sample before unit nonresponse, it is not hard to show that the bias in a calibration-weighted estimator $t_c = \sum_S w_k y_k = \sum_S d_k(R_k/p_k) y_k$ satisfying the calibration equation $\sum_S w_k \mathbf{z}_k = \sum_S d_k(R_k/p_k) \mathbf{z}_k = \sum_S d_k \mathbf{z}_k$ is

$$E\left[\sum_{k\in S} d_k y_k \frac{R_k}{p_k} - \sum_{k\in S} d_k y_k\right] = E\left[\sum_{k\in S} \frac{d_k\left(y_k - \mathbf{z}_k{}^T\mathbf{q}\right)\left(R_k - p_k\right)}{p_k}\right] \quad (1)$$

for *any* vector $\mathbf{q}$, but a more useful choice might be the full-population regression coefficient of y_k on $\mathbf{z}_k$ (Kott 2014).

If either the response model

$$E(R_k - p_k | y_j, \mathbf{z}_j, I_j) = 0 \quad (2)$$

or the prediction model

$$E(y_k - \mathbf{z}_k^T \boldsymbol{\beta}_\mathbf{z} | \mathbf{z}_j, I_j, R_j) = 0 \quad (3)$$

holds (for every j given each k), then the equality in equation (1) can easily be used to show that the nonresponse bias vanishes asymptotically. This is double protection against nonresponse bias.

In an ignorable prediction model, $y_k|\mathbf{z}_k$ is the same regardless of which units are sampled and respond (Little and Rubin 2002); that is, both the sampling and response mechanisms are ignorable under the prediction model, as they are in expectation under the model in equation (3). Notice that equation (2) also assumes that the sampling mechanism is ignorable under the response model. In practice p_k is usually assumed to be a function of $\mathbf{z}_k$, but not y_k or any other y-value. In other words, nonrespondents are missing at random (Rubin 1976).

2.2 Missing Not at Random

What if $y_k|\mathbf{z}_k$ is correlated with R_k? Deville (2000) supplied a quasi-probability-theory solution – "quasi" because response is treated as an additional phase of random sampling. Suppose $E(R_k)$ can be described by a known function:

$$E(R_k) = p(\mathbf{x}_k^T \boldsymbol{\gamma}) \quad (4)$$

with unknown parameter values in $\boldsymbol{\gamma}$. The vector $\mathbf{x}_k$ in Deville's formulation has the same number of components as $\mathbf{z}_k$ but y_k can replace one of the latter's components and other survey variables can be in $\mathbf{x}_k$ as well. Using our terminology, equation (4) allows model-only variables to replace shadow variables in the response model. When calibration-weighting for unit nonresponse, it is sensible to assume that one component of $\mathbf{x}_k$ is unity (or the equivalent). This allows the possibility that every unit is equally likely to respond.

Suppose equation (4) correctly specifies the unit response mechanism. Finding a consistent estimate for $\boldsymbol{\gamma}$, call it $\mathbf{g}$, that satisfies the calibration equation:

$$\sum_{k \in S} w_k \mathbf{z}_k = \sum_{k \in S} d_k \alpha_k \mathbf{z}_k = \sum_{k \in S} d_k \mathbf{z}_k \quad (5)$$

where $\alpha_k = \alpha(\mathbf{x}_k^T \mathbf{g}) = 1/p(\mathbf{x}_k^T \mathbf{g}) = 1/p_k$, $\alpha(.)$ being the weight-adjustment function, results in a nearly unbiased estimator $t_c = \sum_S w_k y_k$ for $T = \sum_U y_k$ under mild conditions. Thus, calibration weights can be used when nonresponse is not missing at random. Among those mild conditions is that the matrix $\mathbf{M} = N^{-1} \sum_S d_k R_k \alpha'(\mathbf{x}_k^T \mathbf{g}) \mathbf{x}_k \mathbf{z}_k^T$ is invertible for $\mathbf{g}$ near $\boldsymbol{\gamma}$ and that $\mathbf{M}$ converges to a finite matrix as the sample size grows arbitrarily large.

Kott and Chang (2010) gave a prediction-modeling double-protection justification for Deville's calibration by suggested the following two-equation prediction model could hold even when the response model in equation (4) fails:

$$y_k = \mathbf{x}_k^T\boldsymbol{\beta}_\mathbf{x} + \varepsilon_k, \text{ and}$$
$$\mathbf{z}_k^T = \mathbf{x}_k^T\boldsymbol{\Gamma} + \boldsymbol{\eta}_k^T, \tag{6}$$

where $\boldsymbol{\Gamma}$ is of full rank (although square here, it will be generalized soon), $E(\varepsilon_k|\mathbf{x}_j, I_j, R_j) = 0$ and $E(\boldsymbol{\eta}_k|\mathbf{x}_j, I_j, R_j) = \mathbf{0}$.

Observe that $y_k = \mathbf{x}_k^T\boldsymbol{\beta}_\mathbf{x} + \varepsilon_k$ is degenerate when y_k is a component of $\mathbf{x}_k$. The two-equation prediction model in equation (6) does not necessarily assume it is. Under that model, $(y_k - \mathbf{z}_k^T\boldsymbol{\beta}_\mathbf{z})\,|\,\mathbf{x}_k = (\varepsilon_k - \boldsymbol{\eta}_k^T\boldsymbol{\Gamma}^{-1}\boldsymbol{\beta}_\mathbf{x})\,|\,\mathbf{x}_k$, where $\boldsymbol{\beta}_\mathbf{z} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\beta}_\mathbf{x}$.

Let $\mathbf{b}_\mathbf{z}$* be the asymptotic limit of

$$\mathbf{b}_\mathbf{z} = \left(\sum_{k\in S} d_k R_k \alpha'(\mathbf{x}_k^T\mathbf{g})\mathbf{x}_k\mathbf{z}_k^T\right)^{-1}\sum_{k\in S} d_k R_k \alpha'(\mathbf{x}_k^T\mathbf{g})\mathbf{x}_k y_k,$$

which, like $\mathbf{b}_\mathbf{z}$, is assumed to exist even when the prediction model fails. It would be equal to $\boldsymbol{\beta}_\mathbf{z}$ otherwise so long as $\mathbf{g}$ converged to a finite $\mathbf{g}$* as the sample grew arbitrarily large.

We can write

$$\sum_{k\in S} d_k y_k \frac{R_k}{p_k} - \sum_{k\in S} d_k y_k = \sum_{k\in S} \frac{d_k\left(y_k - \mathbf{z}_k^T\mathbf{b}_\mathbf{z}^*\right)\left(R_k - p_k\right)}{p_k}$$
$$= \sum_{k\in S} d_k\left(y_k - \mathbf{z}_k^T\mathbf{b}_\mathbf{z}^*\right)\left[R_k\alpha(\mathbf{x}_k^T\mathbf{g}) - 1\right], \tag{7}$$

from which the near unbiasedness of t_C can be inferred if either $E[R_k - p(\mathbf{x}_k^T\boldsymbol{\gamma})|\mathbf{x}_j, \mathbf{z}_j, I_j] = 0$ or $E(y_k - \mathbf{z}_k^T\boldsymbol{\beta}_\mathbf{z}|\,\mathbf{x}_j, I_j, R_j) = 0$. Kott (2006) and others have shown that the insertion of $\alpha'(.)$ into $\mathbf{b}_\mathbf{z}$, which otherwise looks like an instrumental-variable regression coefficient, removes the contribution to large-sample variance under the response model from estimating $\alpha(\mathbf{x}_k^T\boldsymbol{\gamma})$ with $\alpha(\mathbf{x}_k^T\mathbf{g})$ (because $\alpha(\mathbf{x}_k^T\mathbf{g}) - \alpha(\mathbf{x}_k^T\boldsymbol{\gamma}) \approx \alpha'(\mathbf{x}_k^T\mathbf{g})\,\mathbf{x}_k^T(\mathbf{g} - \boldsymbol{\gamma})$ under mild conditions).

2.3 When There are More Calibration than Model Variables

In a strictly quasi-probability framework, Chang and Kott (2008) allowed more calibration than model variables. Their extension of Deville's weighting approach replaced finding the $\mathbf{g}$ in this reformulation of the calibration equation

$$\mathbf{s} = N^{-1}\left|\sum_{k\in S} d_k R_k \alpha(\mathbf{x}_k^T\mathbf{g})\mathbf{z}_k - \sum_{k\in S} d_k\mathbf{z}_k\right| = \mathbf{0},$$

with finding the $\mathbf{g}$ that minimized $\mathbf{s}^T\mathbf{Q}\mathbf{s}$ for some symmetric and positive definite $\mathbf{Q}$.

Observe that if a $\mathbf{g}$ could be found that solved $\mathbf{s} = \mathbf{0}$, then $\mathbf{s}^T\mathbf{Q}\mathbf{s}$ would automatically be minimized. Otherwise, under mild conditions, $\mathbf{s}^T\mathbf{Q}\mathbf{s}$ is minimized when its derivative is set to zero: $N^{-1}\sum_S d_k R_k \alpha'(\mathbf{x}_k^T\mathbf{g})\mathbf{x}_k\mathbf{z}_k^T\mathbf{Q}\mathbf{s} = \mathbf{0}$.

Chang and Kott pointed out that this implies the following reformulated calibration equation:

$$\sum_{k\in S} w_k\tilde{\mathbf{z}}_k = \sum_{k\in S} d_k\alpha_k\tilde{\mathbf{z}}_k = \sum_{k\in S} d_k\tilde{\mathbf{z}}_k \tag{8}$$

where $\tilde{\mathbf{z}}_k = [N^{-1}\sum_S d_j R_j \alpha'(\mathbf{x}_j{}^T\mathbf{g})\mathbf{x}_j\mathbf{z}_j{}^T\mathbf{Q}]\mathbf{z}_k = \mathbf{A}\mathbf{z}_k$. Note that if **A** were square and invertible, then satisfying the calibration equation in (8) would be equivalent to satisfying the calibration equation in (5).

Suppose **A** is not square but is of full rank and converges in probability as the sample size grows arbitrarily large to **A***. The reformulated calibration equation allows all results derived when $\mathbf{z}_k$ and $\mathbf{x}_k$ have the same number of components to hold with $\mathbf{z}_k$ replaced by either $\tilde{\mathbf{z}}_k$ or $\tilde{\mathbf{z}}^*_k = \mathbf{A}^*\mathbf{z}_k$. Unlike $\tilde{\mathbf{z}}_k$, $\tilde{\mathbf{z}}^*_k$ is not a function of the other $\tilde{\mathbf{z}}_j$. This is very helpful ssary for deriving large-sample response-model and prediction-model results. Under the prediction model in equation (6), $(y_k - \tilde{\mathbf{z}}^*_k{}^T\boldsymbol{\beta}_{\tilde{\mathbf{z}}^*}) \mid \mathbf{x}_k = (\varepsilon_k - \boldsymbol{\eta}_k{}^T[\boldsymbol{\Gamma}\mathbf{A}^*]^{-1}\boldsymbol{\beta}_{\mathbf{x}}) \mid \mathbf{x}_k$, where $\boldsymbol{\beta}_{\tilde{\mathbf{z}}^*} = [\boldsymbol{\Gamma}\mathbf{A}^*]^{-1}\boldsymbol{\beta}_{\mathbf{x}}$. Note that equation (8) need not hold exactly when the $\tilde{\mathbf{z}}_k$ are replaced by the $\tilde{\mathbf{z}}^*_k$.

One still to needs a **Q** when applying $\mathbf{A}\mathbf{z}_k$ to reduces the number of components in $\tilde{\mathbf{z}}_k$. Any choice leads to a consistent calibration-weighted estimator under mild conditions when the response model in equation (4) holds. With that in mind, an obvious choice for **Q** is the identity matrix. A slightly better one removes the scales from the components of $\mathbf{z}_k$ (i.e., $\mathbf{Q}^{-1} = DIAG[(N^{-1}\sum_S d_k\mathbf{z}_k)(N^{-1}\sum_S d_k\mathbf{z}_k{}^T)]$) so that one doesn't get a different calibration-weighted estimator if, say, a component of $\mathbf{z}_k$ is measured in pounds rather than kilograms. This is what the default of the SUDAAN procedure WTADJX uses for **Q** (Research Triangle Institute 2012).

Chang and Kott suggested finding a **Q** that comes as close as possible to being *N* times the matrix inverse for the variance of the estimated mean of the calibration vector:

$$\boldsymbol{\tau} = N^{-1}\left(\sum_{k\in S} d_k[R_k / p(\mathbf{x}_k{}^T\boldsymbol{\gamma})]\mathbf{z}_k\right).$$

Iteration would be necessary to find such a **Q** because, until convergence, an estimator for $\boldsymbol{\gamma}$ found by solving equation (6) for **g** given changes the estimate for the matrix inverse of the variance of $\boldsymbol{\tau}$ (found by replacing $\boldsymbol{\gamma}$ by **g**) and so **Q**, which then changes the estimator **g**, and so forth.

One can start the iteration by replacing $p(\mathbf{x}_k{}^T\boldsymbol{\gamma})$ in $\boldsymbol{\tau}$ by the overall response rate and computing the first iteration of **Q** accordingly. Before that, however, one needs to decide what variance to minimize: the variance of τ as an estimator for the population mean $\sum_U \mathbf{z}_k/N$ or the conditional variance of $\boldsymbol{\tau}$ as an estimator for the full-sample-estimated mean $\sum_S d_k\mathbf{z}_k/N$.

Chang and Kott were calibrating to the population, so they chose the former. In our context, calibration is to the full sample, so the latter seems more appropriate. When the response model is Poisson (i.e., independent across units), the conditional variance of $\boldsymbol{\tau}$ is simply $\mathbf{V}_{\boldsymbol{\tau}} = N^{-2}(\sum_S d_k{}^2[\alpha(\mathbf{x}_j{}^T\mathbf{g}) - 1]\mathbf{z}_k\mathbf{z}_k{}^T)$, a value that can be estimated provisionally at every iteration of **g** (and then inverted) by letting $\boldsymbol{\gamma}$ equal that **g**.

3. Our Proposals

We suggest using an iterative process to find a **Q** and **g** such that

$$\mathbf{Q} = \left(N^{-1}\sum_{j\in S} d_j R_j \alpha'(\mathbf{x}_j{}^T\mathbf{g})\mathbf{z}_j\mathbf{z}_j{}^T\right)^{-1},$$

and $\mathbf{g}$ satisfies $N^{-1}\sum_S d_k R_k \alpha'(\mathbf{x}_k{}^T\mathbf{g})\mathbf{x}_k\mathbf{z}_k{}^T\mathbf{Q}\mathbf{s} = \mathbf{0}$. Under these conditions and assuming $\mathbf{Q}$ is of full rank,

$$\begin{aligned}\tilde{\mathbf{z}}_k{}^T &= N^{-1}\mathbf{Q}\sum_{j\in S} d_j R_j \alpha'(\mathbf{x}_j{}^T\mathbf{g})\mathbf{z}_j\mathbf{x}_j{}^T \\ &= \mathbf{z}_k{}^T\left(\sum_{j\in S} d_j R_j \alpha'(\mathbf{x}_j{}^T\mathbf{g})\mathbf{z}_j\mathbf{z}_j{}^T\right)^{-1}\sum_{j\in S} d_j R_j \alpha'(\mathbf{x}_j{}^T\mathbf{g})\mathbf{z}_j\mathbf{x}_j{}^T \\ &= \mathbf{z}_k{}^T\mathbf{B}_{\mathbf{z}}, \end{aligned} \tag{9}$$

where $\mathbf{B}_\mathbf{z}$ is the weighted linear regression of the model vector onto the calibration vector in the respondent sample using $d_j\alpha'(\mathbf{x}_j{}^T\mathbf{g})$ as the weights. Each column of $\mathbf{B}_\mathbf{z}$ is the weighted linear regression of the corresponding component of $\mathbf{x}_k$ onto $\mathbf{z}_k$. Each component of $\tilde{\mathbf{z}}_k$ is a prediction of the corresponding component of $\mathbf{x}_k$, although not necessarily a nearly unbiased one for both respondents and nonrespondents when nonresponse is not missing at random.

If y_k is a component of $\mathbf{x}_k$, and it could be expressed exactly as a linear combination of the components of $\mathbf{z}_k$, then no additional variance would come from unit nonresponse because

$$\sum_{k\in S} d_k y_k \frac{R_k}{p_k} - \sum_{k\in S} d_k y_k = \sum_{k\in S} \frac{d_k\left(y_k - \tilde{z}_{(y)k}\right)\left(R_k - p_k\right)}{p_k} = 0,$$

where $\tilde{z}_{(y)k}$ is the component of $\tilde{\mathbf{z}}_k$ that predicts y_k, in this case perfectly. Usually, however $\tilde{z}_{(y)k}$ is not a perfect prediction of y_k.

If the response model in equation (4) is Poisson and correct, a nearly unbiased estimator for the added variance in the calibration-weighted estimator t_c due to unit nonresponse using the $\tilde{\mathbf{z}}_k$ in equation (9) has this large-sample approximation:

$$addVar_R(t_c) = \sum_{k\in S} d_k{}^2\left(y_k - \tilde{\mathbf{z}}^*{}_k{}^T\,\mathbf{b}_{\tilde{\mathbf{z}}}{}^*\right)^2\left[\alpha(\mathbf{x}_k{}^T\boldsymbol{\gamma}) - 1\right],$$

where $\mathbf{b}_{\tilde{\mathbf{z}}}{}^*$ is the probability limit of $\mathbf{b}_{\tilde{\mathbf{z}}} = (\sum_S d_k R_k \alpha'(\mathbf{x}_k{}^T\mathbf{g})\mathbf{x}_k\tilde{\mathbf{z}}_k{}^T)^{-1}\sum_S d_k R_k \alpha'(\mathbf{x}_k{}^T\mathbf{g})\mathbf{x}_k y_k$ under mild conditions. Now $\tilde{\mathbf{z}}_k{}^T\mathbf{b}_{\tilde{\mathbf{z}}}$ is a nearly unbiased predictor of y_k under the two-equation prediction model in (6) given $\mathbf{x}_j$, I_j, R_j (in the sense that the ratio of its bias to y_k is asymptotically 0). Using $\tilde{\mathbf{z}}_k{}^T\mathbf{b}_{\tilde{\mathbf{z}}}$ corrects for any bias $\tilde{z}_{(y)k}$ as a predictor for y_k under this prediction model.

A large-sample approximation of the added prediction-model variance on t_c due to unit nonresponse when the $y_k - \tilde{\mathbf{z}}^*{}_k{}^T\boldsymbol{\beta}_{\tilde{\mathbf{z}}^*} = \varepsilon_k - \boldsymbol{\eta}_k{}^T[\boldsymbol{\Gamma}\mathbf{A}^*]^{-1}\boldsymbol{\beta}_\mathbf{x}$ are uncorrelated is

$$addVar_P(t_c) = \sum_{k\in S} d_k{}^2 E\left[\left(y_k - \tilde{\mathbf{z}}^*{}_k{}^T\,\boldsymbol{\beta}_{\tilde{\mathbf{z}}^*}\right)^2 \mid \mathbf{x}_j,\, I_j, R_j\right]\left\{R_k\alpha(\mathbf{x}_k{}^T\mathbf{g}^*)\left[\alpha(\mathbf{x}_k{}^T\mathbf{g}^*) - 2\right] + 1\right\},$$

where $\mathbf{g}^*$ (again) is the assumed-to-exist asymptotic limit of $\mathbf{g}$ even when the response model in equation (4) fails.

These two *addVar* approximations suggests that the choice for $\tilde{\mathbf{z}}_k$ in equation (9) should do a good job at limiting the added error of limit the added variance of t_c due to nonresponse. By contrast, the Chang-Kott proposal of the last section has the form:

$$\tilde{\mathbf{z}}_k^T = \mathbf{z}_k^T \left(\sum_{j \in S} d_j^2 \left[\alpha(\mathbf{x}_j^T \mathbf{g}) - 1 \right] \mathbf{z}_j \mathbf{z}_j^T \right)^{-1} \sum_{j \in S} d_j R_j \alpha'(\mathbf{x}_j^T \mathbf{g}) \mathbf{z}_j \mathbf{x}_j^T . \quad (10)$$

That is to say, $\mathbf{B_z}$ in equation (8) is replaced by the awkward weighted regression in the full sample of $(R_k / d_k)[\alpha_k'/(\alpha_k - 1)]$ $\mathbf{x}_k$ onto $\mathbf{z}_k$ using the $d_k^2(\alpha_k - 1)$ as weights. When the response model is logistic $\alpha_k'/(\alpha_k - 1) = 1$, which is a bit less awkward but still not as likely to limit the size of the errors in t_C due to nonresponse as $\mathbf{B_z}$ in $\tilde{\mathbf{z}}_k^T \mathbf{b}_{\tilde{\mathbf{z}}} = \mathbf{z}_k^T \mathbf{B_z} \mathbf{b}_{\tilde{\mathbf{z}}}$.

The use of equation (8) relies on the subsequent adjustment from calibration weighting to remove the potential bias from $\tilde{z}_{(y)k}$ as a predictor of y_k. A simpler variant that does not require iteration or even rely on finding a $\mathbf{Q}$. Instead, it defines calibration weighting as satisfying equation (8) and lets

$$\tilde{\mathbf{z}}_k^T = \mathbf{z}_k^T \mathbf{A}_0^T , \quad (11)$$

where $\mathbf{A}_0^T = (\sum_S R_j \mathbf{z}_j \mathbf{z}_j^T)^{-1} \sum_S R_j \mathbf{z}_j \mathbf{x}_j^T$. This is a variant of the component-reduction technique used in Andridge and Little (2011), which treated not-at-random unit nonresponse from a purely prediction-model point of view. Andridge and Little allowed uncertainty as to whether unit response was a function of y_k, $\tilde{z}_{(y)k}$, or some affine combination of the two. In our notation, their $\mathbf{x}_k = (1, y_k)^T$ at one extreme.

Whether $\mathbf{A}$ in $\tilde{\mathbf{z}}_k = \mathbf{A}\mathbf{z}_k^T$ equals $\mathbf{B_z}^T$ as in equation (9) or $\mathbf{A}_0$, it is not hard to show that dual variables in $\mathbf{x}_k$ will also be in $\tilde{\mathbf{z}}_k$. It is the shadow variables $\mathbf{z}_k$ that get reduced to the shadow variables in $\tilde{\mathbf{z}}_k$, although the latter can be linear combination of shadow and dual variables from $\mathbf{z}_k$ so long as no shadow variable is exactly equal to a linear combination of other components of $\tilde{\mathbf{z}}_k$. Each model-only component of $\mathbf{x}_k$ has a corresponding shadow variable in $\tilde{\mathbf{z}}_k$.

4. A Simulation Experiment

We conducted a simulation experiment using mostly public-use files (PUFs) from the National Survey on Drug Use and Health (NSDUH; http://www.icpsr.umich.edu/icpsrweb/ ICPSR/ series/64), an annual national survey that collect data on substance use, mental health, and other health outcomes among members of the noninstitutionalized U.S. civilian population aged 12 or older. Using 2006-2010 data from those PUFs we restricted our attention to children 12-17 who sought counselling for mental-health problems. Our goal was to estimate for such children both the average number of visits to a specialty mental health facility, denoted SMHVST, and the prevalence for not making any visits, which we created and denoted NONE. We used the actual data in the PUFs, removing records with the missing values for SMHVST and reweighting the remainder to compensate. This gave us a complete data set of 2,454 records that looked very much like real data. (The original PUFs contained no other missing variables of interest to us.)

Unfortunately, after doing the computations, we learned that a variable used in reweighting the data was not on the public use files. As a result, we will need to recalculate our empirical finding using a revised reweighting scheme and cannot present them here.

In the PUFs, SMHVST has six numerical categories, but we treated it as continuous, reassigning category 6, no visits, to SMHVST = 0 and leaving the rest of the ordered categories as they were (categories 3, 4, and 5 had 3 to 6 visits, 4 had 7 visits, and 25 or more visits, respectively). We set NONE = 1 when SMHVST = 0 (originally 6) with NONE = 0 otherwise.

For our calibration variables, we used the binary variables MALE, WHITE (non-Hispanic Caucasian), and YOUNG (12 or 13 years old), and the categorical variable YOTMTHLP, how much did counseling help, which ranged from 0, not at all, to 5, extremely. We treated YOTMHLP as continuous.

We compared five methods of creating calibration weighting to account for the nonresponse. The first treated the calibration variables and unity as if they were also the (response) model variables: $\mathbf{x}_k = \mathbf{z}_k =$ (1 MALE WHITE YOUNG YOTMTHLP)^T. The second through fifth methods treated SMHVST and 1 as the model variables, $\mathbf{x}_k =$ (1 SMHVST)^T, and the same calibration-variable vector as method 1, but used different techniques to reduce the dimension of the latter, that is, create $\tilde{\mathbf{z}}_k$.

Method 2 used the default in SUDAAN's WTADJX (ADJUST = NONRESPONSE; BESTIM = REDUCED). Method 3 used the Chang-Kott method expressed in equation (10). Method 4 was our first proposed method (equation (9)). It required iteration. Method 5 was our second proposed method (equation (11)), which did not.

Finally, Method 6 added the three binary variables from vector of calibration variables to vector of model variables: $\mathbf{x}_k =$ (1 MALE WHITE YOUNG SMHVST)^T. As a result, the dimensions of the model and calibration vectors were the same and reduction of the size of the calibration vector became unnecessary. We investigated this method because it is relatively simple to implement with available software. It has SMHVST as a model variable while employing all the calibration variables. The extra model variables (MALE, WHITE, and YOUNG) should have coefficients asymptotically equal to zero. If the efficiency loss from using it was not too great, this method would be very appealing in practice.

We created and simulated probabilities of response with these three logistic models:
Model 1: $1/[1 + \exp(-2 + .3\ \text{SMHVST})]$
Model 2: $1/[1 + \exp(-2 + .75\ \text{SMHVST})]$
Model 3: $1/[1 + \exp(-5 - .3\ \text{SMHVST})]$.
Response decreased with SMHVST in the first two models, but decreased with SMHVST in the third. The nonresponse rate, which varied across simulations, under the first and third models was roughly 25%. It was roughly 50% under the second model.

5. Some Concluding Remarks

We have seen that if one knows what survey variables cause units to respond or fail to respond, then a prudent strategy would be to choose shadow variables that can predict them within the respondent sample using linear regression, which, if our modest simulations are any indication, may not need to be weighted. Weighting for both the sampling (through the d_k) and the response mechanism (through the $\alpha(.)$) can perhaps wait until after the shadow variables have been selected (say with equation (11)) and occur in the calibration-weighting process itself (equation (8)). In addition, there appears to be no need to chose a $\mathbf{Q}$ matrix as claimed in Chang and Kott (2008) for calibration-weighting to have desirable properties.

To be honest, this result surprised us. We had hoped by using real data not generated by a prediction model in our simulations while simulating nonresponse with an exact response model our first proposal based on a chosen $\mathbf{Q}$ would produce clearly smaller mean squared errors. It did, but not in all cases. This finding may be analogous to the $\mathbf{b}$ in the general regression estimator for $\sum_U y_k$ in the absence of unit nonresponse, $t_{\text{GREG}} = \sum_S d_k y_k + [\sum_U \mathbf{x}_k - \sum_S d_k \mathbf{x}_k)]^T \mathbf{b}$, not itself having to be a weighted estimator for t_{GREG} to be both unbiased under the linear prediction model and consistent under probability sampling theory.

The calibration-weighting approach using equation (11) differs from the pure prediction-modeling approach to unit response in Andridge and Little (2011) in including a calibration-weighting step. A pure prediction-modeling approach when combined with a non-ignorable original sampling design might produce implicit weights like

$$w_k = d_k \left[1 + \sum_{j \in S} d_j (1 - R_j) \tilde{\mathbf{z}}_j{}^T \left(\sum_{j \in S} R_j d_j \mathbf{x}_j \tilde{\mathbf{z}}_j{}^T \right)^{-1} \mathbf{x}_k \right] \tag{12}$$

$$= \sum_{j \in S} d_j \tilde{\mathbf{z}}_j{}^T \left(\sum_{j \in S} R_j d_j \mathbf{x}_j \tilde{\mathbf{z}}_j{}^T \right)^{-1} d_k \mathbf{x}_k \quad \text{if } \mathbf{x}_k \text{contains 1 as a component (or the equivalent),}$$

where the $\tilde{\mathbf{z}}_k$ come from equation (11).

The theory under which using the weights in equation (12) result in nearly unbiased estimators only covers survey variables for which the two-equation model in (6) applies, unlike NONE in our simulations (although, to be fair, ideally each survey variable could be modeled separately using a plausible model for each). Ironically, the response model corresponding to the weights in equation (12) is $E(R_k) = 1/(1 + \mathbf{x}_k{}^T \boldsymbol{\gamma})$ featured in the purely quasi-randomization approach of Särndal and Lundström (2005).

In our framework, calibration variables as well as survey variables can be in the weight-adjustment function, which is ideally the inverse of the response model in equation (4). The problem, which is still unresolved, is how best to determine what variables belong in the weight-adjustment function. As our simulations suggests, there is a variance penalty from allowing survey variables to be in the weight-adjustment function. This suggests to us that weight adjustment made to remove potential biases due to nonresponse be separated from weighting adjustments to increase efficiency or control for coverage errors.

Finally, although we used the SUDAAN routines WTADJUST (for Method 1) and WTADJX (for all other methods) in ours simulation, the R package "Sampling" (Tille and Matei 2013) and other routines in R can also be used after the vector of calibration variables is reduced to a vector with the same number of components as the model vector.

Acknowledgements

Much of this work was supported by a grant from the National Science Foundation, award number SES-1424492. The authors sincerely thank Dai Lanting and Peter Frechtel for their help is setting up the data files and Neung Soo Ha for his support with R code.

References

Andridge, R.H. and Little, R.J. (2011), "Proxy Pattern-Mixture Analysis for Survey Nonresponse," *Journal of Official Statistics*, 27, 153-180.

Brewer, K.R.W. (1995), "Combining Design-Based and Model-Based Inference" *Business Survey Methods*, ed. by B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott, New York: Wiley, 589-606.

Chang, T. and Kott, P.S. (2008), " Using Calibration Weighting to Adjust for Nonresponse Under a Plausible Model, *Biometrika*, 95, 557-571. Version for full appendices online at http://www.stat.virginia.edu/documents/NASS_tech_reportwCover.pdf

Deville, J. C. (2000), "Generalized Calibration and Application to Weighting for Non-response," *COMPSTAT: Proceedings in Computational Statistics, 14th Symposium, Utrecht, The Netherlands,* J.G. Bethlehem and P.G.M. van der Heijden, (Eds.), New York: Springer-Verlag.

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*, 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.

Kim, J.K., and Park, H. (2006), "Imputation Using Response Probability, *Canadian Journal of Statistics,* 34, 1-12.

Kott, P. (2014), " On Voluntary and Volunteer Government Surveys in the United States," *Statistics Journal of the IAOS*, 30, 249-253.

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

Kott, P.S. and Chang, T. (2010), "Using Calibration Weighting to Adjust for Nonignorable Unit Nonresponse," *Journal of the American Statistical Association*, 105, 1265-1275.

Little, R.J. and Rubin, D.B. (2002), *Statistical Analysis with Missing Data* (2nd ed.), New York: Wiley.

Research Triangle Institute (2012), *SUDAAN Language Manual*, Volumes 1 and 2, Release 11. Research Triangle Park, NC: Research Triangle Institute.

Royall, R. M. (1976). The linear least squares prediction approach to two-stage sampling. *Journal of the American Statistical Association,* 71, 657-664.

Rubin, D.B. (1976). Inference and Missing Data, *Biometrika* 63, 581-592.

Särndal,C-E. and Lundström, S. (2005), *Estimation in Surveys with Nonresponse*. Chichester: Wiley.

Tille, Y. and Matei, A., (2013), *Package* '*Sampling.*' A software routine available online at http://cran.r-project.org/web/packages/sampling/sampling.pdf.

Wang, S., Shao, J. and Kim, J.K. (2014). "An Instrument Variable Approach for Identification and Estimation with Nonignorable Nonresponse," *Statistica Sinica*, 24, 1097-1116.