

# Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors

Phillip S. Kott

---

Although originally developed as a method for reducing sampling errors while retaining randomization consistency, calibration weighting can also be used to adjust for unit nonresponse and/or coverage errors under appropriate quasi-randomization models. Linear and nonlinear calibration adjustments, which are asymptotically identical in a purely sampling context, can diverge asymptotically when used in this manner. In addition, defining instrumental variables makes it possible for nonresponse (say) to be a function of a set of characteristics other than the calibration vector and opens the possibility of using calibration to handle nonignorable nonresponse. A variant of the jackknife can remove the need for iteration in variance estimation when the calibration adjustment is nonlinear.

KEY WORDS: Prediction model; Quasi-random model; Quasi-randomization consistent; Instrumental variable; Generalized raking.

---

Phillip S. Kott is Chief Research Statistician, Research and Development Division, National Agricultural Statistics Service, 3251 Old Lee Highway, Room 305, Fairfax, Virginia, 22030. An earlier version for this paper was prepared for the Mini Meeting on Current Trends in Survey Sampling and Official Statistics, January 2004, Calcutta, India.

## I. INTRODUCTION

Calibration weighting was originally developed as a method for reducing sampling errors while retaining randomization consistency. Folsom and Singh (2000), among others, have shown that calibration weighting can also be used to adjust for coverage errors and/or unit nonresponse under appropriate quasi-randomization models.

Folsom and Singh is not in the refereed literature. The heart of this paper repeats their key results, adding to them the idea that the calibration variables need not be exactly the same as the explanatory variables in the function for the element response (or coverage) probabilities. In addition, a jackknife is proposed that computes replicate weights in one step even though the calibration weights themselves may be determined iteratively.

Section 2 provides a review of linear calibration in a purely sampling context. Section 3 describes Estevao and Sarndal's (2000) expansion of linear calibration to include "instrumental variables." Section 4 parallels Deville and Särndal (1992) treatment of nonlinear calibration to include the expansion of the previous section. Section 5 reviews variance/mean squared error estimation, proposing a new jackknife for certain designs. Section 6 describes how nonlinear calibration can be used to adjust for nonresponse and/or coverage errors, that is, frame under- or over-coverage. Section 7 contains a small empirical example supporting the new jackknife. Section 8 provides a discussion.

## 2. LINEAR CALIBRATION AND THE GREG ESTIMATOR

Suppose we knew the selection probability,  $\pi_k$ , for each sample element  $k$  in the sample  $S$ , then we could estimate any population total,  $T_y = \sum_U y_k$ , where  $U$  denotes the population, with the expansion estimator  $t_{y-E} = \sum_S y_k / \pi_k = \sum_U y_k I_k / \pi_k$ , where  $I_k = 1$  when  $k \in S$  and 0 otherwise. Treating the  $I_k$  as random variables, it is easy to see that  $t_{y-E}$  is an unbiased estimator for  $T_y$ . Properties arising when the  $I_k$  are treated as random variables are called *randomization-based*. We can also write  $t_{y-E} = \sum_U a_k y_k = \sum_S a_k y_k$ , where  $a_k = I_k / \pi_k$  is the *sampling weight* of element  $k$ .

Deville and Särndal (1992) coined the term “calibration estimator” to describe an estimator of the form  $t_{y-CAL} = \sum_S w_k y_k$ , where  $\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k = T_x$  for some row vector of auxiliary variables,  $\mathbf{x}_k = (x_{1k}, \dots, x_{pk})$ , about which  $T_x$  is known. Since there is generally a continuum of sets  $\{w_k \mid k \in S\}$  that satisfy the *calibration equation*:

$$\sum_{k \in S} w_k \mathbf{x}_k = T_x, \quad (1)$$

Deville and Särndal required that the difference between the set of weights,  $\{w_k \mid k \in S\}$ , satisfying equation (1) and  $\{a_k \mid k \in S\}$  minimize some loss function.

An alternative approach to survey sampling treats the  $y_k$  as random variables satisfying the linear *prediction* model:

$$y_k = \mathbf{x}_k \beta + \epsilon_k, \quad (2)$$

where  $E(\epsilon_k \mid \{\mathbf{x}_g, I_g \mid g \in U\}) = 0$  for all  $k \in U$ . By conditioning this expectation on the  $I_g$ , we are assuming the sampling mechanism can be ignored. This is a crucial, and

sometimes unreasonable, aspect of the (prediction) *model-based* framework.

It is easy to see that  $t_{y\_CAL}$  is an unbiased estimator for  $T_y$  under the model in the sense that  $E_\epsilon(t_{y\_CAL} - T_y) = 0$  (suppressing the conditioning for notational convenience); the subscript  $\epsilon$  refers to treating the  $\epsilon_k$  as random variables (and the  $I_k$  as fixed constants).

For our purposes, the general(ized) regression or GREG estimator has the form:

$$t_{y\_GREG} = t_{y\_E} + (T_x - \sum_{k \in S} a_k \mathbf{x}_k) (\sum_{k \in S} c_k a_k \mathbf{x}_k \mathbf{x}_k')^{-1} \sum_{k \in S} c_k a_k \mathbf{x}_k' y_k, \quad (3)$$

where  $c_k$  is an arbitrary constant which may or may not be a function of  $\mathbf{x}_k$ , and  $\lim_{N \rightarrow \infty} \sum_U c_k \mathbf{x}_k \mathbf{x}_k' / N = \Lambda$  is a positive definite matrix, where  $N$  is the size of  $U$ . This last condition means that  $\sum_S c_k a_k \mathbf{x}_k \mathbf{x}_k'$  will usually be invertible in practice. We will assume that it is always invertible for convenience.

The GREG estimator in equation (3) can be rewritten in calibration form as

$t_{y\_GREG} = \sum_S w_k y_k$ , where

$$w_k = a_k + (T_x - \sum_{j \in S} a_j \mathbf{x}_j) (\sum_{j \in S} c_j a_j \mathbf{x}_j \mathbf{x}_j')^{-1} c_k a_k \mathbf{x}_k'. \quad (4)$$

Strictly speaking, the  $w_k$  are functions of the realized sample,  $S$ , and the  $c_k a_k$ , but we suppress that in the notation for convenience. Observe that the calibration weights in equation (4) are linear functions of  $\mathbf{x}_k$ , hence the term “linear calibration.” The calibration equation itself (equation (2)) will always be linear here.

Let us assume that reasonable regularity conditions hold (see, for example, Kott 2004 for a more thorough treatment) and the sample plan is such that  $t_{y\_E} - T_y =$

$O_p(N/\downarrow n)$ , where  $n$  is the (expected) size of  $S$  (the actual size can be random),

$\sum_S a_k \mathbf{x}_k - T_x = O_p(N/\downarrow n)$ , and  $\sum_S c_k a_k \mathbf{x}_k' \mathbf{f}_k - \sum_U c_k \mathbf{x}_k' \mathbf{f}_k = O_p(N/\downarrow n)$ , where  $\mathbf{f}_k$  can be  $\mathbf{x}_k$  or  $y_k$ .

Let  $e_k = y_k - \mathbf{x}_k (\sum_U c_i \mathbf{x}_i' \mathbf{x}_i)^{-1} \sum_U c_i \mathbf{x}_i' y_i$ , so that  $\sum_U c_i \mathbf{x}_i' e_i = 0$ , and  $\sum_S c_k a_k \mathbf{x}_k' e_k = O_p(N/\downarrow n)$ .

We can express the error of  $t_{y\_GREG}$  as

$$\begin{aligned}
 t_{y\_GREG} - T_y &= \sum_{k \in S} w_k y_k - \sum_{k \in U} y_k \\
 &= \sum_{k \in S} w_k e_k - \sum_{k \in U} e_k \quad (\text{since } \sum_{k \in S} w_k \mathbf{x}_k - \sum_{k \in U} \mathbf{x}_k) \\
 &= \sum_{k \in S} a_k e_k + (T_x - \sum_{k \in S} a_k \mathbf{x}_k) (\sum_{k \in S} c_k a_k \mathbf{x}_k' \mathbf{x}_k)^{-1} \sum_{k \in S} c_k a_k \mathbf{x}_k' e_k - \sum_{k \in U} e_k \\
 &= \sum_{k \in S} a_k e_k - \sum_{k \in U} e_k + O_p(N/n). \tag{5}
 \end{aligned}$$

The GREG estimator is randomization consistent with a relative randomization bias and mean squared error of asymptotic order  $1/n$ .

### 3. REDEFINING CALIBRATION WEIGHTS

In their original definition of calibration weights, Deville and Särndal (1992) required that the set of calibration weights,  $\{w_k \mid k \in S\}$  minimize some distance function between the members of the set and the original sampling weights, the  $a_k$ , subject to satisfying the calibration equation. As a result, the calibration estimator,  $t_{y\_CAL} = \sum_S w_k y_k$ , was both unbiased under the model in equation (2) and usually randomization consistent.

Estevao and Särndal (2002) suggested removing the requirement that the

calibration weights minimize a distance function. Instead, they essentially proposed that the  $w_k$  need only satisfy the calibration equation and be of the “functional form.”

$$w_k = a_k(1 + \mathbf{h}_k \mathbf{q}), \quad (6)$$

where  $\mathbf{h}_k$  is a row vector with the same dimension as  $\mathbf{x}_k$  such that  $\sum_S a_k \mathbf{h}_k' \mathbf{x}_k$  is invertible, and  $\mathbf{q}$  is a column vector of that same dimension. It is a generalization of the GREG where  $\mathbf{h}_k$  effectively replaces  $c_k \mathbf{x}_k$

It is not hard to see that  $\mathbf{q} = (\sum_S a_j \mathbf{h}_j' \mathbf{x}_j)^{-1} (\mathbf{T}_x - \sum_S a_j \mathbf{x}_j)'$ . Moreover, under mild conditions we assume to hold,  $t_{y\_CAL} = \sum_S w_k y_k = \sum_S a_k y_k + (\mathbf{T}_x - \sum_S a_j \mathbf{x}_j) (\sum_S a_j \mathbf{h}_j' \mathbf{x}_j)^{-1} \sum_S a_k \mathbf{h}_k' y_k$  is randomization consistent whenever  $t_{y\_E}$  is. It is unbiased under the linear prediction model in equation (2) when  $E(\epsilon_k | \{\mathbf{x}_g, \mathbf{h}_g | g \in S\}, \{l_g | g \in U\}) = 0$  for all  $k \in U$ .

This suggests another alternative definition of calibration weights: a set of weights,  $\{w_k | k \in S\}$ , such that, *i*, the  $w_k$  satisfy the calibration equation for  $\{\mathbf{x}_k | k \in U\}$  and, *ii*,  $t_{y\_CAL} = \sum_S w_k y_k$  is randomization consistent whenever  $t_{y\_E}$  is under mild conditions. That is the definition we will use.

It follows that Estevao and Särndal’s functional-form calibration is indeed a form a calibration weighting. Borrowing from econometric theory, the components of  $\mathbf{h}_k$  that are not linear combinations of components of  $\mathbf{x}_k$  are called “instrumental variables.”

## 4. NONLINEAR CALIBRATION

Building on ideas in Deville and Särndal (1992), we can generalize the linear form for the calibration weights in equation (6) to

$$w_{k\_GEN} = a_k f(\mathbf{h}_k \mathbf{q}^*), \quad (7)$$

where  $f$  is a monotonic, twice-differentiable function with  $f(0) = 1$ ,  $f'(0) = 1$  ( $f'(0)$  is the first derivative of  $f$  evaluated at 0), and  $\mathbf{q}^*$  is chosen so that the calibration equation holds.

Strictly speaking, there should be an additional symbol on  $w_{k\_GEN}$  (and later on  $w_{k\_LIN}$ ) to denote the particular choice of  $\mathbf{h}_k$ . It has been dropped for convenience.

A solution,  $\mathbf{q}^*$ , to equation (7) can be approached iteratively. One can start with  $\mathbf{q}^{(0)} = \mathbf{0}$ ; that is,  $\sum_S w_k^{(0)} y_k$ , where  $w_k^{(0)} = a_k f(0)$ . For  $r = 1, 2, \dots$ , one then sets  $\mathbf{q}^{(r)} = \mathbf{q}^{(r-1)} + [\sum_S f'(\mathbf{h}_k \mathbf{q}^{(r-1)}) a_k \mathbf{h}_k' \mathbf{x}_k]^{-1} (T_x - \sum_S w_k^{(r-1)} \mathbf{x}_k)$ , and  $w_k^{(r)} = a_k f(\mathbf{h}_k \mathbf{q}^{(r)})$ . Iteration stops at  $r^*$  when  $T_x = \sum_S w_k^{(r^*)} \mathbf{x}_k$  for all practical purposes. One should be aware, however, that *there may not be a set of weights that can be expressed in the form of equation (7) while satisfying the calibration equation.*

Note that  $\mathbf{q}^{(1)}$  above equals the  $\mathbf{q}$  in  $w_{k\_LIN} = a_k(1 + \mathbf{h}_k \mathbf{q})$ . A Taylor expansion around zero reveals  $f(\mathbf{h}_k \mathbf{q}^{(1)}) = \mathbf{h}_k \mathbf{q}^{(1)} + O_p(1/n)$  under mild conditions, so  $\sum_S w_k^{(1)} y_k = \sum_S w_{k\_LIN} y_k + O_p(N/n) = T_y[1 + O_p(1/n)]$ . Furthermore, it is not difficult to see that  $w_{k\_GEN} = w_{k\_LIN}[1 + O_p(1/n)]$ , an equality that proves helpful in variance estimation.

The most common example in practice of a nonlinear  $f$  is  $f(\mathbf{h}_k \mathbf{q}) = \exp(\mathbf{x}_k \mathbf{q})$ ,

where the values of each of the components of  $\mathbf{x}_k$ , denoted  $x_{1k}, \dots, x_{pk}$ , are either 0 or 1. That is effectively the form of Deming and Stephan's (1940) raking weights computed via iterative proportional fitting. Many have observed that the iterative routine described above can be used even when the components of  $\mathbf{x}_k$  are not binary as they are in Deming and Stephan. Note that the *generalized raking* calibration weights that result are always nonnegative.

## 5. VARIANCE ESTIMATION

Särndal, Swensson, and Wretman (1989) proposed this *plug-in* variance/mean-squared-error estimator for  $t_{y\_GREG}$  under an arbitrary sampling plan:

$$v_{SSW} = \sum_{k \in S} \sum_{j \in S} [(\pi_{kj} - \pi_k \pi_j) / \pi_{kj}] (w_k r_k) (w_j r_j). \quad (8)$$

The term derives from  $r_k$  being “plugged into”  $v_{SSW}$  in place of the unknown  $e_k$  (for randomization-mean-squared-error estimation) or  $\epsilon_k$  (for prediction-model-variance estimation). Following the reasoning of Deville and Särndal (1992),  $v_{SSW}$  also applies to  $t_{y\_CAL}$  with calibration weights defined by equation (7), and  $r_k = y_k - \mathbf{x}_k (\sum_S \mathbf{h}_j \mathbf{x}_j)^{-1} \sum_S \mathbf{h}_j y_j$ .

Developing asymptotic properties for  $v_{SSW}$  can be elusive when it contains  $n(n-1)/2$  distinct terms. That is not a problem under stratified simple random sampling, where

$$v_{ST1} = \sum_{\alpha=1}^A (n_\alpha / [n_\alpha - 1]) \sum_{k \in S_\alpha} (1 - n_\alpha / N_\alpha) (w_k r_k - \sum_{j \in S_\alpha} w_j r_j / n_\alpha)^2, \quad (9)$$

$S_\alpha$  denotes the sample of  $n_\alpha$  units in stratum  $\alpha$  ( $\alpha = 1, \dots, A$ ), and  $U_\alpha$  the stratum population containing  $N_\alpha$  elements.

For a multi-stage sample it makes sense to allow the possibility that  $e_k$  and  $e_j$  in the prediction model are correlated when  $k$  and  $j$  are elements in the same PSU, but not otherwise. When finite-population correction can be ignored, the model variance of a calibration estimator is approximately  $V_m = \sum_{i \in S'} E_e [ (\sum_{k \in S(i)} w_k e_k)^2 ]$  under mild conditions, where  $S(i)$  is the set of sampled elements in PSU  $i$ , and  $S'$  is the set of PSUs selected in the first stage of sampling.

The following variance estimator often has good randomization and model-based properties (when the first-stage selection probabilities are all small):

$$v_{ST2} = \sum_{\alpha=1}^A (n_{1\alpha}/[n_{1\alpha} - 1]) \left\{ \sum_{j \in S_{1\alpha}} \left( \sum_{k \in S_{\alpha j}} w_k r_k \right)^2 - \left( \sum_{j \in S_{1\alpha}} \sum_{k \in S_{\alpha j}} w_k r_k \right)^2 / n_\alpha \right\}, \quad (10)$$

where  $\alpha$  denotes a first-stage stratum of PSU's,  $n_{1\alpha}$  the number of sampled PSU's in stratum  $\alpha$ ,  $S_{1\alpha}$  the set of sampled PSU's in  $\alpha$ , and  $S_{\alpha j}$  the set of subsampled elements from PSU  $j$  of stratum  $\alpha$ . There can be many stages of sampling involved.

It is not hard to show that  $v_{ST2}$  is asymptotically indistinguishable from the jackknife variance estimator:

$$v_J = \sum_{\alpha=1}^A ([n_\alpha - 1]/n_\alpha) \left\{ \sum_{j \in S_{1\alpha}} (t_{y\_CAL(\alpha j)} - t_{y\_CAL})^2 \right\}, \quad (11)$$

where  $t_{y\_CAL(\alpha j)} = \sum_{k \in S} w_{k(\alpha j)} y_k$ , and the *jackknife replicate calibration weights* are

$w_{k(\alpha j)} = w_k a_{k(\alpha j)} / a_k + \left( \sum_{m \in U} \mathbf{x}_m - \sum_{m \in S} w_m [a_{m(\alpha j)} / a_m] \mathbf{x}_m \right) \left( \sum_{m \in S} a_{m(\alpha j)} \mathbf{c}_m \mathbf{x}_m' \mathbf{x}_m \right)^{-1} a_{k(\alpha j)} \mathbf{c}_k \mathbf{x}_k'$ , where  $a_{k(\alpha j)} = 0$  when  $k$  is in PSU  $j$  of stratum  $\alpha$ ,  $a_{k(\alpha j)} = a_k$  when  $k$  is not in stratum  $\alpha$  at all, and

$\mathbf{a}_{k(\alpha_j)} = (n_\alpha / [n_\alpha - 1])\mathbf{a}_k$  otherwise. The  $w_{k(\alpha_j)}$  are constrained so that  $\sum_{k \in S} w_{k(\alpha_j)} \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k$  for all  $\alpha_j$ . Now, under mild conditions we assume to hold,

$$\sum_{m \in U} \mathbf{x}_m - \sum_{m \in S} w_m [\mathbf{a}_{m(\alpha_j)} / a_m] \mathbf{x}_m = (n_\alpha / [n_\alpha - 1]) (\sum_{k \in S(\alpha_j)} w_k \mathbf{x}_k - \sum_{k \in S(\alpha)} w_k \mathbf{x}_k / n_\alpha) = \mathbf{O}_P(N/n),$$

$$(\sum_{m \in S} c_m \mathbf{a}_{m(\alpha_j)} \mathbf{x}_m' \mathbf{x}_m) = \mathbf{O}_P(N), \text{ and}$$

$$\sum_{m \in S} c_m \mathbf{a}_{m(\alpha_j)} \mathbf{x}_m' \mathbf{e}_m = \mathbf{O}_P(N/n),$$

where  $S(\alpha)$  is the set of elements in stratum  $\alpha$ , and  $S(\alpha_j)$  is the set of elements in PSU  $j$  of stratum  $\alpha$ . As a result,

$$\begin{aligned} \mathbf{t}_{y\_CAL(\alpha_j)} - \mathbf{t}_{y\_CAL} &= \sum_{k \in S} w_{k(\alpha_j)} \mathbf{e}_k - \sum_{k \in S} w_k \mathbf{e}_k = \\ &= (n_\alpha / [n_\alpha - 1]) (\sum_{k \in S(\alpha_j)} w_k \mathbf{e}_k / n_\alpha - \sum_{k \in S(\alpha_j)} w_k \mathbf{e}_k) + \mathbf{O}_P(N/n^{3/2}), \end{aligned}$$

and  $v_j = v_{ST2} [1 + \mathbf{O}_P(1/n)]$  when  $\text{plim}_{n \rightarrow \infty} (nv_{ST2} / N^2) > 0$ .

The replicate weights described above are nonstandard. More common is  $w_{k(\alpha_j)} = \mathbf{a}_{k(\alpha_j)} + (\sum_{m \in U} \mathbf{x}_m - \sum_{m \in S} \mathbf{a}_{m(\alpha_j)} \mathbf{x}_m) (\sum_{m \in S} c_m \mathbf{a}_{m(\alpha_j)} \mathbf{x}_m' \mathbf{x}_m)^{-1} c_k \mathbf{a}_{k(\alpha_j)} \mathbf{x}_k'$ , which “look like” the original calibration weights. Our version generates a  $v_j$  with a model expectation closer to  $\sum_{i \in S} E_e [(\sum_{k \in S(i)} w_k \mathbf{e}_k)^2]$ . Replacing  $\mathbf{e}_k$  in the arguments above by  $\mathbf{e}_k$ , it is not hard to show that  $E_e(v_j) = V_m [1 + \mathbf{O}_P(1/n)]$  under mild conditions.

## 6. UNIT NONRESPONSE AND COVERAGE ADJUSTMENT

In this section we explore handling unit (whole-element) nonresponse as an additional phase of Poisson sampling. That is the essence of a *quasi-randomization* model. Each element  $k$  in the original sample, now denoted  $F$ , is assumed to have a probability of response,  $p_k$ . The probability of elements  $k$  and  $j$  jointly responding is  $p_k p_j$ , and whether element  $k$  responds (given a vector of covariates) is independent of whether it is chosen for the original sample.

It is often possible to construct a set of weights so that the calibration estimator is randomization consistent under the quasi-randomization model. We are interested here in a particular way of constructing those weights. To this end, we assume that the quasi-randomization model is correct. Each element has attached to it a row vector of auxiliary variables,  $\mathbf{x}_k$ , for which  $T_x = \sum_U \mathbf{x}_j$  is known. Finally, each  $p_k$  is assumed to have the form:

$$p_k = 1/f(\mathbf{h}_k \phi), \quad (12)$$

where  $\phi$  is an unknown column vector,  $\mathbf{h}_k$  is a row vector with the same dimension as  $\mathbf{x}_k$ , and  $\sum_S \mathbf{h}_k \mathbf{x}_k / N$ , where  $S$  now denotes the “subsample” of respondents, is invertible both for the realized  $N$  and in the probability limit. The function  $f$  is assumed to be monotonic and twice differentiable. Its functional form is known, but the value of the governing parameter,  $\phi$ , is not. Neither  $f(0)$  nor  $f'(0)$  need be 1.

The most obvious choice for  $\mathbf{h}_k$  when postulating the response model in equation

(12) is  $\mathbf{x}_k$  itself. In some applications, however, some component(s) of  $\mathbf{x}_k$  may have been chosen because it was the best measures we had for a variable *before* sampling. An example of such a variable in a survey of farms is the total land area of an operation. After collecting survey values, it may be possible to replace a component of  $\mathbf{x}_k$  (in  $\mathbf{h}_k$ ) with a better measure of the variable in question. In our example, response is more likely a function of the actual land area of a farm than a predetermined proxy for that value. As a result, replacing the corresponding proxy value with the survey value is tempting. A theoretical problem with this procedure is discussed below.

Using the iterative method described in the last section to find  $\mathbf{q}^*$ , we will often be able to uncover a row vector,  $\mathbf{q}$ , such that  $T_x = \sum_S a_k f(\mathbf{h}_k \mathbf{q}) \mathbf{x}_k$ . As a result, estimating  $T_y$  with  $t_{y\_CAL} = \sum_S w_k y_k$ , where the adjusted calibration weights have the form,  $w_k = a_k f(\mathbf{h}_k \mathbf{q})$ , may have good properties under the linear prediction model:  $y_k = \mathbf{x}_k \beta + \epsilon_k$ , where  $E(\epsilon_k | \{\mathbf{x}_g, \mathbf{h}_g, I_g | g \in U\}) = 0$  for all  $k \in U$ ,  $I_k = 1$  if element  $k$  is both in the original sample and responds, 0 otherwise.

Prediction-model unbiasedness is simply a result of the weights satisfying the calibration equation ( the prefix “prediction” is needed to distinguish this model from the quasi-random one). Note, however, that if some components of  $\mathbf{h}_k$  come from the survey rather than  $\mathbf{x}_k$ , the prediction-model assumption that  $E(\epsilon_k | \mathbf{h}_k) = 0$  can be problematic. At the extreme, consider the case where one such component is  $y_k$  itself. Obviously,  $E(\epsilon_k | y_k)$  is not usually 0. In the example described above,  $y_k$  may be the total land area on farm operation  $k$ . Putting total land area in  $\mathbf{h}_k$  makes the associated calibration estimator prediction-model biased.

Whether or not  $t_{y\_CAL}$  can reasonably be called prediction-model unbiased has no effect on its quasi-randomization-based properties. Since  $T_x = \sum_S a_k [1 + f(\mathbf{h}_k \mathbf{q})] \mathbf{x}_k$ , our assumptions and the mean value theorem reveal

$$T_x - \sum_{k \in S} a_k f(\mathbf{h}_k \phi) \mathbf{x}_k = \sum_{k \in S} a_k [f'(\mathbf{h}_k \mathbf{q}^0) \mathbf{h}_k (\mathbf{q} - \phi)] \mathbf{x}_k = \mathbf{O}_p(N/n)$$

for some  $\mathbf{h}_k \mathbf{q}^0$  between  $\mathbf{h}_k \mathbf{q}$  and  $\mathbf{h}_k \phi$ . From this we see that if  $\sum_S a_j f'(\mathbf{h}_j \phi) \mathbf{h}_j' \mathbf{x}_j / N$  is invertible both for the realized  $N$  and at the probability limit (recall that  $f$  is monotonic so  $f'$  is never zero), then

$$\begin{aligned} \mathbf{q} - \phi &= \left\{ \sum_{k \in S} a_j f'(\mathbf{h}_j \mathbf{q}^0) \mathbf{h}_j' \mathbf{x}_j \right\}^{-1} \left\{ T_x - \sum_{i \in S} a_i f(\mathbf{h}_i \phi) \mathbf{x}_i \right\} = \mathbf{O}_p(1/n) \\ &= \left\{ \sum_{j \in S} a_j f'(\mathbf{h}_j \phi) \mathbf{h}_j' \mathbf{x}_j \right\}^{-1} \left\{ T_x - \sum_{i \in S} a_i f(\mathbf{h}_i \phi) \mathbf{x}_i \right\} + \mathbf{O}_p(1/n). \end{aligned}$$

The estimator  $t_{y\_CAL}$  has an error of

$$\begin{aligned} t_{y\_CAL} - T_y &= \sum_{k \in S} a_k f(\mathbf{h}_k \mathbf{q}) y_k - \sum_{k \in U} y_k \\ &= \sum_{k \in S} a_k f(\mathbf{h}_k \mathbf{q}) \mathbf{e}_k - \sum_{k \in U} \mathbf{e}_k, \end{aligned}$$

where  $\mathbf{e}_k = y_k - \mathbf{x}_k (\sum_U f'(\mathbf{h}_j \phi) \rho_j \mathbf{h}_j' \mathbf{x}_j)^{-1} \sum_U f'(\mathbf{h}_j \phi) \rho_j \mathbf{h}_j' y_j$ , and  $\rho_j = 1/[1 + f(\mathbf{h}_j \phi)]$  so

$\sum_S a_k f'(\mathbf{h}_k \phi) \mathbf{h}_k' \mathbf{e}_k = \mathbf{O}_p(N/n)$ . Continuing:

$$\begin{aligned} t_{y\_CAL} - T_y &= \sum_{k \in S} a_k f(\mathbf{h}_k \phi) \mathbf{e}_k - \sum_{k \in U} \mathbf{e}_k + \sum_{k \in S} a_k \{ f(\mathbf{h}_k \mathbf{q}) - f(\mathbf{h}_k \phi) \} \mathbf{e}_k \\ &= \sum_{k \in S} a_k f(\mathbf{h}_k \phi) \mathbf{e}_k - \sum_{k \in U} \mathbf{e}_k + \sum_{k \in S} a_k f'(\mathbf{h}_k \phi) \mathbf{h}_k (\mathbf{q} - \phi) \mathbf{e}_k + \mathbf{O}_p(N/n) \\ &= \sum_{k \in S} a_k f(\mathbf{h}_k \phi) \mathbf{e}_k - \sum_{k \in U} \mathbf{e}_k + (\mathbf{q} - \phi)' \sum_{k \in S} a_k f'(\mathbf{h}_k \phi) \mathbf{h}_k' \mathbf{e}_k + \mathbf{O}_p(N/n) \\ &= \sum_{k \in S} a_k f(\mathbf{h}_k \phi) \mathbf{e}_k - \sum_{k \in U} \mathbf{e}_k + \mathbf{O}_p(N/n) \end{aligned} \tag{13}$$

Thus,  $t_{y\_CAL}$  is quasi-randomization consistent under mild conditions whenever

$$t = \sum_S a_k f(\mathbf{h}_k \phi) y_k \text{ is.}$$

To estimate the quasi-randomization mean squared error of  $t_{y\_CAL}$  (i.e., the estimator's randomization mean squared error under the quasi-randomization model), we first note that the probability that elements  $k$  and  $j$ ,  $k \neq j$ , are both in the respondent subsample is  $\pi_{kj}^* = \pi_{kj} p_k p_j$ . Let  $\pi_k^* = \pi_k p_k$ , and recall that  $a_k = 1/\pi_k$  and  $1/p_k = f(\mathbf{h}_k \phi)$ . From equation (13), we see that the randomization mean squared error of  $t_{y\_CAL}$  is approximately

$$\begin{aligned} E_{-1}[(t_{y\_CAL} - T_y)^2] &\approx \sum_{k \in U} \sum_{j \in U} (\pi_{kj}^* - \pi_k^* \pi_j^*) (e_k / \pi_k^*) (e_j / \pi_j^*) \\ &= \sum_{k \in U} (1 - \pi_k^*) e_k^2 / \pi_k^* + \sum_{\substack{k \in U \\ j \in U \\ k \neq j}} (\pi_{kj} - \pi_k \pi_j) (e_k / \pi_k) (e_j / \pi_j) \end{aligned} \quad (14)$$

If the original sample is Poisson, then  $v_m = \sum_S (w_k^2 - w_k) r_k^2$  with

$$r_k = y_k - \mathbf{x}_k \left[ \sum_{j \in S} a_j f(\mathbf{h}_j \mathbf{q}) \mathbf{h}_j' \mathbf{x}_j \right]^{-1} \sum_{j \in S} a_j f(\mathbf{h}_j \mathbf{q}) \mathbf{h}_j' y_j, \quad (15)$$

serves as both a reasonable estimator for prediction-model variance and quasi-randomization mean squared error under mild conditions, since  $w_k \approx 1/\pi_k^*$  and  $r_k \approx e_k$ . A close relative of the non-intuitive sample residual in equation (15) can be found in Folsom and Singh (2000). See Kott (2004) for a further discussion of  $v_m$  in a purely sampling context.

For a general design, we can get close to the a good variance/mean-squared-error estimator by starting with  $v_{SSW}$  in equation (8), where  $r_k$  is again defined by

equation (15). We need to add a term like

$$v_{\text{add}} = \sum_{k \in S} (w_k^2 \pi_k - w_k) r_k^2,$$

so that  $\sum_U (1 - \pi_k^*) e_k^2 / \pi_k^*$  in equation (14) is estimated by  $\sum_S (w_k^2 - w_k) r_k^2$  rather than  $\sum_S w_k^2 (1 - \pi_k) r_k^2$ . This correction to  $v_{\text{SSW}}$  in equation (8) has good prediction-model-based properties when the  $e_k$  are uncorrelated, and  $\sigma_k^2 = \mathbf{x}_k \zeta$ , for some  $\zeta$ . It can be made even in the absence of nonresponse.

When the actual sample is multistage, and the first stage selection probabilities are ignorably small,  $v_{\text{ST2}}$  in equation (10) can be used as the variance/mean-squared-error estimator with  $r_k$  defined once more by equation (15).

Observe that when there is no nonresponse,  $\phi = 0$ , so that  $f(\mathbf{h}_j, \mathbf{q}) = f(0) + f'(0) \mathbf{h}_j \mathbf{q} + O_p(1/n) = f(0) + O_p(1/n)$ . As a result, the  $f$ -terms in equation (15) are all asymptotically identical and can be removed from the definition of  $r_k$  without altering the asymptotics of the variance/mean-squared-error estimators.

When  $f$  is linear,  $f(\theta) = 1$ , and the  $r_k$  in equation (15) are computed as if there were no nonresponse. The same holds true for the variance/mean-squared-error estimator  $v_{\text{ST2}}$ . Unfortunately, this  $f$  corresponds to an awkward response-probability function:  $p_k = 1/(1 + \mathbf{h}_k \phi)$ . Fuller, Loughin, and Baker (1994) made these observations for the case where  $\mathbf{h}_k = c_k \mathbf{x}_k$ .

The jackknife,  $v_j$ , in equation (11) can be computed with these jackknife replicate weights:

$$w_{k(\alpha_j)} = w_k a_{k(\alpha_j)} / a_k + \left( \sum_{m \in U} \mathbf{x}_m - \sum_{m \in S} w_m [a_{m(\alpha_j)} / a_m] \mathbf{x}_m \right) \left( \sum_{m \in S} a_{m(\alpha_j)} f(\mathbf{h}_m \mathbf{q}) \mathbf{h}_m' \mathbf{x}_m \right)^{-1} a_{k(\alpha_j)} f(\mathbf{h}_k \mathbf{q}) \mathbf{h}_k' \quad (16)$$

Again when  $f(\theta) = 1$ ,  $v_j$  can be computed as if there were no nonresponse.

Folsom and Singh (2000) pointed out that the treatment of nonresponse through calibration weighting can also be used to adjust for undercoverage. In the context, the quasi-random phase as sampling occurs conceptually before the actual sample is drawn. The population associated with the sampling frame is assumed to be a Poisson sample from a hypothetical complete population for which the vector  $T_x$  must be known. The frame population becomes  $F$ , while the hypothetical complete population is  $U$ . The probability that element  $k \in U$  is in  $F$  is assumed to be modeled correctly by equation (12). If the first (from  $U$  to  $F$ ) and second (from  $F$  to  $S$ ) phases of sampling are independent, then all the theory developed for using calibration weighting to handle nonresponse carries over to handling undercoverage.

The authors also noted that overcoverage (duplication) or a combination of under and overcoverage can be handled in the same way. The definition of  $p_k$  in equation (12) becomes the expected number of times  $k$  is in the frame, which can now exceed 1 due to potential duplication.

Folsom and Singh further suggested that  $f(\cdot)$  have the flexible form:

$$f(\mathbf{x}_k \phi) = \frac{U(C - L)\exp(\mathbf{x}_k \phi) + L(U - C)}{(U - C) + (C - L)\exp(\mathbf{x}_k \phi)}, \quad (17)$$

where  $L \geq 0$ ,  $1 < U \leq \infty$ , and  $L < C \leq U$  are predetermined constants. They call this the "General Exponential Model" or "GEM." Observe that when  $C = 1$ ,  $U = \infty$ , and  $L = 0$ ,

$p_k = \exp(-\mathbf{x}_k \phi)$ . Similarly, when  $C = 2$ ,  $U = \infty$ , and  $L = 1$ ,  $p_k = [1 + \exp(\mathbf{x}_k \phi)]^{-1}$ ; that is to say, the probability of response (coverage) is logistic. The values  $L$  and  $U$  serve as bounds on the *calibration adjustment*,  $f(\cdot)$ , while  $C = f(0)$  is effectively its center. The authors made the calibration adjustment in GEM even more flexible by postulating three classes of sampling units, each with its own set of  $U$ ,  $C$ , and  $L$  values.

We have seen in this section that calibration weighting can produce estimators with good prediction-model-based properties when the prediction model is correct – in particular, equation (2) holds with  $E(\epsilon_k | \{\mathbf{x}_g, \mathbf{h}_g, I_g | g \in U\}) = 0$ , where  $I_g = 1$  when  $g \in S$  and 0 otherwise, and good quasi-randomization properties when the response or coverage model (in equation (12)) is correct. In some sense, one model provides protection against the failure of the other. See Kott (1994).

As noted, the prediction model is more likely to hold when  $\mathbf{h}_g = \mathbf{x}_g$ . Even then, sometimes the  $\epsilon_k$  in the model in equation (2) satisfy  $E(\epsilon_k | \{\mathbf{x}_g | g \in U\}) = 0$ , but not  $E(\epsilon_k | \{\mathbf{x}_g, I_g | g \in U\}) = 0$ ; that is to say, the sampling mechanism – including response or coverage – is not ignorable with respect to the prediction model.

We can factor  $I_k$  into  $I_{k1} I_{k2}$ , where  $I_{k1} = 1$  when and only when  $k$  is in the original sample, and  $I_{k2} = 1$  when and only when  $k$  responds (including when  $k$  is not in the sample, but would have responded if it were) or is covered by the frame. The interested reader can confirm that calibration weighting provides some protection against bias if the prediction model in equation holds when  $E(\epsilon_k | \{\mathbf{x}_g, \mathbf{h}_g, I_{g2} | g \in U\}) = 0$ ; that is when the response (or coverage) mechanism is ignorable with respect to the prediction model but not necessarily the original sampling mechanism.

## 7. A SMALL EMPIRICAL EXAMPLE

Since the jackknife replicate weights expressed in equation (16) are new, it is prudent to investigate whether they actually work with real data. To this end, the author took the MU281 data from Särndal, Swensson, and Wretman (1992) and replicated it 20 times (so  $N = 5,620$ ). Using stratified simple random sampling, 16 units were selected from each of the eight unequally-sized strata. The variable RMT85 served as  $y_k$  and P75 as  $x_k$  in  $\mathbf{x}_k = (1, x_k)$ . Each of the 128 sampled units was given a probability of being in the respondent subsample,  $S$ , which decreased with the size of  $x_k$ ; in particular,  $p_k = \exp(-.35 x_k / M_x)$ , where  $M_x$  was the population mean of the  $x_k$ . In 1,600 simulations, the size of the  $S$  ranged from 78 to 110, with an average of approximately 93.8.

The total  $T_y$  was estimated two ways, with  $t_{y\_LIN} = \sum_S a_k(1 + \mathbf{x}_k \mathbf{q})y_k$  and with  $t_{y\_EXP} = \sum_S a_k \exp(\mathbf{x}_k \mathbf{q}^{(exp)})y_k$ , where  $\mathbf{q}$  and  $\mathbf{q}^{(exp)}$  were respectively selected so that the calibration equation held. The former was a GREG estimator, while the latter was a generalized raking estimator. Both estimators were unbiased under the implied prediction model ( $y_k = \mathbf{x}_k \beta + \epsilon_k$ ), but only  $t_{y\_EXP}$  was randomization consistent under the correct response model. The GREG implicitly assumed  $p_k = 1/(\phi_0^{(LIN)} + \phi_1^{(LIN)}x_k)$  for unknown  $\phi_0^{(LIN)}$  and  $\phi_1^{(LIN)}$ .

The small size of the sample relative to the population in each stratum allowed the ignoring of finite population correction in variance/mean-squared-error estimation (called “variance estimation” from now on). Variances were estimated using,  $i$ , the

linearization estimator,  $v_{ST2}$ , in equation (10) with  $r_k$  defined by equation (15), and, *ii*, the proposed jackknife,  $v_j$ , in equation (11) with replicate weights defined by equation (16). To make the jackknife computations easier, the 16 samples in each stratum were randomly assigned to one of four clusters, so that only 32 jackknife replicates had to be computed.

For comparison purposes, a better version of the linearization variance estimator, labeled  $v_{ST2(e)}$ , was also computed with  $r_k$  replaced by  $e_k = y_k - \mathbf{x}_k (\sum_U f(\mathbf{x}_j \phi) p_j \mathbf{x}_j' \mathbf{x}_j)^{-1} \sum_U f(\mathbf{x}_j \phi) p_j \mathbf{x}_j' y_j$ , where  $\phi$  and  $p_j$  were known. In practice,  $e_k$  is rarely known, but computing  $v_{ST2(e)}$  was useful for comparison purposes.

One should note that computations of  $r_k$  and  $e_k$  were slightly different depending on whether the variance estimator for  $t_{y\_LIN}$  or for  $t_{y\_EXP}$  was of interest. For  $t_{y\_LIN}$ ,  $f(\mathbf{x}_j \phi) = f(\mathbf{x}_j \mathbf{q}) = 1$ ; for  $t_{y\_EXP}$ ,  $f(\mathbf{x}_j \mathbf{q}^{(exp)}) = \exp(\mathbf{x}_j \mathbf{q}^{(exp)})$ , and  $f(\mathbf{x}_j \phi) = 1/p_j$ .

Table 1 displays the empirical means (the mean over the 1,600 simulations) of the two estimators for  $T_y$  normalized so that  $T_y = 100$ . Although both are close to unbiased,  $t_{y\_LIN}$  is significantly different from 100 at the .05 level;  $t_{y\_EXP}$  is not. This is not surprising, since only the latter is based on the correct response model.

The variance estimators and empirical mean squared errors of each estimator were normalized so that the empirical means of the respective  $v_{ST2(e)}$ 's were 100. Neither  $v_{ST2(e)}$  had an empirical mean significantly different from the empirical mean squared error (EMSE) of the associated estimator. This was a bit disappointing. It seems that although  $t_{y\_LIN}$  had a significant empirical bias, this bias was such a small component of the estimator's mean squared error, that the difference between its

EMSE and the empirical mean of  $t_{ST2(e)}$  was not significant.

The  $v_{ST2(e)}$  were chosen as benchmarks for the table rather than the empirical mean squared errors because the former had roughly half the empirical standard errors of the EMSE's and correlated more strongly with the variance estimators. The t-values for this part of the table were also computed with respect to the  $v_{ST2(e)}$ .

The two linearization variance estimators had surprisingly large downward biases. Apparently, there was a tendency for unusually large  $w_{k\_LIN}$  and  $w_{k\_EXP}$  to cause associated  $r_k$  to be appreciably smaller than  $e_k$  in absolute terms. The problems associated with unusually large  $w_{k\_LIN}$  and  $w_{k\_EXP}$  seem to be more muted with the jackknives.

To speed up the asymptotics of the linearization variance estimators (i.e., reduce the difference between  $r_k$  and  $e_k$ ), an *ad-hoc* adjustment of  $v_{ST2}$  was computed by replacing each  $r_k$  with  $r_{k(\text{adjusted})} = r_k / \omega_k$ , where  $\omega_k^2 = 1 - \mathbf{x}_k(\sum_S a_j \mathbf{f}'(\mathbf{x}_j, \mathbf{q}) \mathbf{x}_j' \mathbf{x}_j)^{-1} a_k \mathbf{f}'(\mathbf{x}_k, \mathbf{q}) \mathbf{x}_k' = 1 + O_p(1/n)$ . Observe that under the prediction model with the  $e_k$  uncorrelated and  $E(e_k^2) = \sigma_k^2$ ,  $E(r_{k(\text{adjusted})}^2) \approx \sigma_k^2$ . The near equality is exact when all the  $a_j \mathbf{f}'(\mathbf{x}_j, \mathbf{q})$  and  $\sigma_j$ , respectively, are equal.

The adjusted  $v_{ST2}$  for both  $t_{y\_LIN}$  and  $t_{y\_EXP}$  remained biased downward, while the  $v_j$  were biased upward by a slightly smaller amount. Although these biases were significant, they were reasonably small (from 4.5 to 11.2%) and suggest that the variance estimators may have indeed been asymptotically unbiased as theoretically demonstrated in previous sections.

Using  $v_{ST2(e)}$  as an efficient proxy for EMSE, the empirical mean squared error of

$t_{y\_EXP}$ , which incorporated the correct response model, was more than 13% larger than that of the  $t_{y\_LIN}$ , which did not. One should not generalize broadly based on one data set involving only two calibration variables, however.

Whether or not one is better off incorporating the correct response model in the calibration estimator, if one does so, then the variance estimators discussed in the previous section, perhaps with the linearization estimator adjusted as suggested in this section, appear to be serviceable.

A second set 1,600 simulations (not displayed) were done using the same population and stratified sampling design but with each sampled element given a 70% chance of being in the respondent sample (the average respondent sample size was roughly 89.8). In this set of simulations, both estimators for  $T_y$  are randomization consistent under the response model. Consequently, it is not surprising, that the empirical means of  $t_{y\_LIN}$  and  $t_{y\_EXP}$  were virtually identical (within 0.01% of each other) as were their empirical mean squared errors (within 1% of each other). The empirical means of each pair of variance estimators (e.g.  $var_{ST2}$  for  $t_{y\_LIN}$  and  $t_{y\_EXP}$ ) were likewise very close (within 1% of each other). The relative bias of the adjusted  $v_{ST2}$  (compared to  $var_{ST2(e)}$ ) was -1.3% when estimating the variance of  $t_{y\_LIN}$  and -2.2% when estimating the variance of  $t_{y\_EXP}$ . The relative biases of the unadjusted linearization variances were -9.0% and -10.3%, respectively. The relative bias of both jackknives was 3.6%.

Suppose there had been a true second phase of sampling rather than nonresponse. In this phase, each element had a .7 chance of being subsampled. It is not hard to see that were each  $a_k$  replaced by  $a_k/.7$ , the inverse overall selection probability of element  $k$ , neither estimator for  $T_y$  would change, nor would any of the

variance estimators for  $t_{y\_LIN}$ . The putative variance estimator  $v_{ST2(e)}$  for  $t_{y\_EXP}$  would likewise be unaffected. Thus, the empirical results from the second set of simulations support the contention from Section 6 made for calibration in the absence of nonresponse (or coverage errors): the variance estimators developed for  $t_{y\_LIN}$  can be used to estimate of variance of  $t_{y\_GEN}$  when both employ the same set of  $\mathbf{h}_k$  (here  $\mathbf{h}_k = \mathbf{x}_k$ ).

## 8. DISCUSSION

### 8.1 *Estimating a Response Model Explicitly*

When faced with unit nonresponse, many have attempted to estimate the element probabilities of response,  $p_k = 1/f(\mathbf{h}_k \phi)$ , directly. This method requires one to have information on  $\mathbf{h}_k$  for every element in the sample whether it responded to the survey or not, but  $\mathbf{h}_k$  need not have the same dimension as  $\mathbf{x}_k$ . The direct-adjustment method is generally not available for handling coverage errors.

Fuller (2002) noted that there can be an extra term in the quasi-randomization mean squared error of  $t_{y\_GREG} = \sum_S a_k^* y_k + (T_x - \sum_S a_j^* x_j) (\sum_S c_j a_j^* x_j' x_j)^{-1} \sum_S c_k a_k^* x_k' y_k$ , where  $S$  is the respondent subsample,  $a_k^* = a_k [1 + f(\mathbf{h}_k \mathbf{q})]$ , and  $\mathbf{q}$  is a consistent direct estimator for the quasi-randomization model parameter,  $\phi$ .

## 8.2 Response Homogeneity Groups

To control the magnitude of the weight adjustment due to nonresponse, Little (1986) recommended that one estimate  $\mathbf{q}$  explicitly and then divide the sample into  $C$  mutually exclusive groups (often called “cells” or “poststrata”) based on the sizes of the fitted  $f(\mathbf{h}_k, \mathbf{q})$  values. One then computes the adjusted weight for each element  $k$  in group  $c$  as  $w_{k\_ADJ} = (\sum_{F(c)} w_g / \sum_{S(c)} w_g) w_k$ , where  $F(c)$  is that part of the original sample in group  $c$ ,  $S(c)$  is the subsample of  $F(c)$  that respond, and  $w_k$  is the sampling weight assigned to element  $k$  after sampling but before quasi-random subsampling. This approach assumes that each element in a group has (roughly) the same probability of response, hence the term “response homogeneity group.”

An alternative way of incorporating fitted  $f(\mathbf{h}_k, \mathbf{q})$  values into the estimation based on methodology developed in the text follows. Divide the fitted values into  $P$  groups based in their sizes, where  $P$  is again the dimension of  $\mathbf{x}_k$ , and let  $\mathbf{d}_k$  be a row vector of indicator variables for the  $P$  cells. By setting each  $w_k = a_k [1 + (\mathbf{T}_x - \sum_S a_j \mathbf{x}_j) (\sum_S a_j \mathbf{d}_j' \mathbf{x}_j)^{-1} \mathbf{d}_k']$ , one computes a set of weights for the respondent subsample that, unlike  $\{w_{k\_ADJ}\}$  above, satisfies the calibration equation for the respondent sample. Because of the nature of  $\mathbf{d}_k$ , this linear method returns the same set of calibration weights as fitting  $w_k = a_k \exp(\mathbf{d}_k \mathbf{f})$  would – if both produce a set of weights. Note that since calibration weights can be negative with the linear method, it may be able to find a set that the generalized raking method cannot. The linear method effectively scales the  $a_k$ -value for every element in the same group by a fixed amount. Thus, it is unlikely to produce surprisingly small or surprisingly large weights when the dimension of  $\mathbf{x}_k$  is small

compared to the sample size.

### 8.3 *Breaking Up Sample and Nonresponse Calibration*

In the previous section we noted that it is possible for components of  $\mathbf{h}_k$  in equation (12), the quasi-random response model, to be unknown before enumeration. When such an  $\mathbf{h}_k$  is used in calibration, it might no longer be reasonable to assert that the resulting  $t_{y\_CAL}$  is prediction-model unbiased. This is particularly troublesome when nonresponse is modest compared to the sample size. An intriguing idea is to calibrate in two phases. The first phase, sample calibration, adjusts for the difference between  $T_x$  and  $\sum_F a_k \mathbf{x}_k$ , and would not include any components in  $\mathbf{h}_k$  unavailable at the time of sampling. The second phase, nonresponse calibration, adjusts for the difference between  $\sum_F a_k \mathbf{x}_k$  and  $\sum_S a_k \mathbf{x}_k$  and would include component variables only available after the respondent subsample is enumerated.

A more thorough analysis of this idea must wait for another time.

### 8.4 *Work at NASS*

The National Agricultural Statistics Service (NASS) used variants of the Fuller, Loughin, and Baker (1994) approach for handling undercoverage in the Census of Agriculture and for adjusting an agricultural economics survey with large nonresponse to match totals from more reliable surveys. In this approach,  $f(\cdot)$  has the form:

$$f(\mathbf{x}_k\phi) = \begin{cases} L & \text{when } 1 + \mathbf{x}_k\phi < L \\ 1 + \mathbf{x}_k\phi & \text{when } L \leq 1 + \mathbf{x}_k\phi \leq U \\ U & \text{when } 1 + \mathbf{x}_k\phi > U, \end{cases} \quad (18)$$

which truncates linear calibration at pre-specified values, L and U, to control the size of the weight adjustment. Note that when  $f(.) = U$  or  $L$ ,  $f'(.) = 0$ . Unlike the calibration adjustment in equation (17),  $f(.)$  in equation (18) is not twice differentiable at  $L - 1$  or  $U - 1$ . This does not cause a problem in practice.

The agency's original justification for calibration in these contexts was based on prediction-modeling. Equation (18) is simple to implement and appears to produce weights within an acceptable range more often than readily available alternatives.

NASS is investigating the following questions: How sensitive is  $t_{y\_CAL}$  to the choice of  $f(.)$  in practice? Would a different choice for  $f(.)$  result in less bias, and if so, would the reduction in absolute bias translate into a lower mean squared error? What would be the effect of replacing some component of the vector of calibration variables with a better predictor of nonresponse/undercoverage?

## REFERENCES

- DEMING, W.E. and STEPHAN, F.F. (1940). On a least squares adjustment of a sample frequency table when the expected marginal total are known. *Annals of Mathematical Statistics*, 11, 427-444.
- DEVILLE, J-C. and SÄRNDAL, C-E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, 87, 376-382.
- ESTEVAO, V.M., and SÄRNDAL, C-E. (2000). A functional form approach to calibration. *Journal of Official Statistics*, 16, 379-399.
- FOLSOM., R.E. and SINGH, A.C. (2000). The generalized exponential model for sampling weight calibration for extreme values, nonresponse, and poststratification. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Washington DC, 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.
- HUANG, E.T. and FULLER, W.A. (1978). Nonnegative regression estimation for sample survey data. In *Proceedings of the Section on Social Statistics*, American Statistical Association, Washington DC, 300-305.
- ISAKI, C.T. and FULLER, W.A. (1982). Survey design under the regression super-population model. *Journal of the American Statistical Association*, 77, 89-96.
- KOTT, P.S. (1990). The design consistent regression estimator and its conditional

- variance. *Journal of Statistical Planning and Inference*, 24, 287-296
- KOTT, P.S. (1994). A note on handling nonresponse in surveys. *Journal of the American Statistical Association*, 89, 693-696.
- KOTT, P.S. (2004). Randomization-assisted model-based survey sampling. *Journal of Statistical Planning and Inference*. To appear.
- LITTLE, R.J.A. (1986). Survey nonresponse adjustments for estimates of means, *International Statistical Review*, 54, 139-157.
- RAO, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimations stage. *Journal of Official Statistics*, 25, 1-21.
- SÄRNDAL, C-E, SWENSSON, B., and WRETMAN, J. (1989). The weighted residual technique for estimating the variance of the general regression estimator of a finite population total. *Biometrika*, 76, 527-537.
- SÄRNDAL, C-E, SWENSSON, B., and WRETMAN, J. (1992). *Model Assisted Survey Sampling*. Springer, New York.

**Table 1**  
Empirical Means of Estimators Based on 1,600 Simulations\*

	Empirical mean (standard error)	t-value (two-sided significance)	
The Estimators for $T_y$ ( $T_y = 100$ )			
$t_{y\_LIN}$	99.84 (0.06)	-2.79 (.02)	difference from $T_y$
$t_{y\_EXP}$	100.04 (0.06)	0.58 (.56)	
Variance Estimators for $t_{y\_LIN}$ ( $E_{EMP}(v_{ST2(e)}) = 100$ )			
$V_{ST2}$	83.59 (1.53)	-19.96 (<.0001)	difference from $V_{ST2(e)}$
$V_{ST2(adusted)}$	95.53 (1.80)	-6.09 (<.0001)	
$V_J$	104.69 (2.28)	3.60 (.0003)	
EMSE	99.35 -	-0.18 (.85)	
Variance Estimators for $t_{y\_EXP}$ ( $E_{EMP}(v_{ST2(e)}) = 100$ )			
$V_{ST2}$	73.12 (1.54)	-18.22 (<.0001)	difference from $V_{ST2(e)}$
$V_{ST2(adusted)}$	88.79 (1.98)	-8.57 (<.0001)	
$V_J$	107.00 (2.73)	4.09 (<.0001)	
EMSE	101.21 -	0.33 (.74)	
Other Statistics			
$relvar(v_{ST2(e)[LIN]})$	.051 -	-	
$relvar(v_{ST2(e)[EXP]})$	.059 -	-	
$\frac{V_{ST2(e)[LIN]} - V_{ST2(e)[EXP]}}{E_{EMP}(v_{ST2(e)[EXP]})}$	-.1340 (.010)	-13.87 (<.0001)	

\* In four additional simulations, convergence was not reached in 10 iterations for  $t_{y\_EXP}$ . They were excluded from the analysis.