Survey Methodology, December 2006
Vol. 32, No. 2, pp. 133-142
Statistics Canada, Catalogue No. 12-001

Using Calibration Weighting to Adjust for Nonresponse and Coverage Errors

Phillip S. Kott [1]

Abstract

Calibration weighting can be used to adjust for unit nonresponse and/or coverage errors under appropriate quasi-randomization models. Alternative calibration adjustments that are asymptotically identical in a purely sampling context can diverge when used in this manner. Introducing instrumental variables into calibration weighting makes it possible for nonresponse (say) to be a function of a set of characteristics other than those in the calibration vector. When the calibration adjustment has a nonlinear form, a variant of the jackknife can remove the need for iteration in variance estimation.

Key Words: Prediction model; Quasi-randomization model; Quasi-randomization consistent; Instrumental variable; Generalized raking.

1. Introduction

Calibration weighting was originally developed as a method for reducing sampling errors while retaining randomization consistency. Deville and Särndal (1992) demonstrated that many alternative forms of calibration weighting are asymptotically identical in the sampling context. This lead to a breakthrough in our understanding of common weight adjustment methods like raking that do not appear in generalized-regression (GREG) estimator format.

Folsom and Singh (2000) showed that calibration weighting can also be used to adjust for known coverage errors and/or unit nonresponse under appropriate quasi-randomization models. Their work is not in the refereed literature. The heart of this article repeats key results in Folsom and Singh including a necessary modification of the Deville-Särndal approach to model variance/randomization mean-squared-error estimation in this expanded context. An earlier, strictly linear version of calibration weighting for unit-nonresponse adjustment can be found in Fuller, Loughin and Baker (1994). See also Lundström and Särndal (1999).

A distinction is drawn between the prediction model usually underpinning calibration and the quasi-randomization model in Folsom and Singh. Unlike in Folsom and Singh, however, both properties are explored here. Furthermore, the explanatory variables in the quasi-randomization model are allowed to differ from the calibration variables. This is likewise allowed in Lundström and Särndal.

A new jackknife is proposed which is analogous to the Deville-Särndal linearization variance estimator. It employs replicate weights computed in one step even though the calibration weights themselves may be determined iteratively.

After introducing the popular notion of calibration weighting, Section 2 provides a review of the GREG special case in a purely sampling context. Section 3 describes Estevao and Särndal's (2000) extension of calibration weighting in its linear form to include instrumental variables. Section 4 expands Deville and Särndal's treatment of calibration weighting to include the possibility of instrumental variables. Section 5 reviews variance/mean squared error estimation, proposing a new jackknife for certain designs. Section 6 describes how calibration weighting can be used to adjust for nonresponse. In this context, alternative functional forms of calibration weighting need no longer be asymptotically identical. Section 7 discusses quasi-randomization models for coverage errors, that is, frame under- or over-coverage. Section 8 contains a small empirical example supporting the new jackknife. Section 9 provides a discussion of alternative approaches and areas for future research.

2. Calibration Weighting and the GREG Estimator

Suppose we knew the selection probability, π_k, for each sample element k in the sample S. We can 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 $t_{y\_\mathrm{CAL}} = \sum_S w_k y_k$, where $\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k = T_{\mathbf{x}}$ for some

1. Phillip S. Kott, National Agricultural Statistics Service, 3251 Old Lee Highway, Fairfax, VA 22030, U.S.A. E-mail: pkott@nass.usda.gov.

row vector of auxiliary variables, $\mathbf{x}_k = (x_{1k}, \ldots, x_{Pk})$, about which $T_{\mathbf{x}}$ is known. Since there is generally a continuum of sets $\{w_k | k \in S\}$ that satisfy the *calibration equation*:

$$\sum_{k \in S} w_k \mathbf{x}_k = T_{\mathbf{x}}, \tag{1}$$

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

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

$$y_k = \mathbf{x}_k \boldsymbol{\beta} + \varepsilon_k, \tag{2}$$

where $E(\varepsilon_k | \{\mathbf{x}_g, I_g | g \in U\}) = 0$ 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 $E_\varepsilon(t_{y\_\mathrm{CAL}} - T_y) = 0$ (suppressing the conditioning for notational convenience); the subscript ε refers to treating the ε_k as random variables (and the I_k as fixed constants).

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

$$t_{y\_\mathrm{GREG}} = t_{y\_E} + \left(T_{\mathbf{x}} - \sum_{k \in S} a_k \mathbf{x}_k\right)\left(\sum_{k \in S} c_k a_k \mathbf{x}_k' \mathbf{x}_k\right)^{-1} \sum_{k \in S} c_k a_k \mathbf{x}_k' y_k, \tag{3}$$

where c_k is an arbitrary constant which may or may not be a function of $\mathbf{x}_k$, and $\lim_{N \to \infty} \sum_U c_k \mathbf{x}_k' \mathbf{x}_k / N = \boldsymbol{\Lambda}$ 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 $t_{y\_\mathrm{GREG}} = \sum_S w_k y_k$, where

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

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 can be expressed as

$$w_k = a_k (1 + c_k \mathbf{x}_k \mathbf{q}), \tag{4}$$

where $\mathbf{q} = [(\sum_S a_j c_j \mathbf{x}_j' \mathbf{x}_j)^{-1}]'(T_{\mathbf{x}} - \sum_S a_j \mathbf{x}_j)'$ is a column vector, since $\mathbf{x}_k \mathbf{q} = \mathbf{q}' \mathbf{x}_k'$.

Let us assume that reasonable regularity conditions hold (see, for example, Kott 2004a for a more thorough treatment) and the sample plan is such that $t_{y\_E} - T_y = O_P(N/\sqrt{n})$, where n is the expected size of S (the actual size can be random), $\sum_S a_k \mathbf{x}_k - T_{\mathbf{x}} = \mathbf{O}_P(N/\sqrt{n})$, and $\sum_S a_k c_k \mathbf{x}_k' \mathbf{f}_k - \sum_U c_k \mathbf{x}_k' \mathbf{f}_k = \mathbf{O}_P(N/\sqrt{n})$, 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 $\sum_S a_k c_k \mathbf{x}_k' e_k = \mathbf{O}_P(N/\sqrt{n})$. We can express the error of t_{y_GREG} as

$$\begin{aligned}
&t_{y_\mathrm{GREG}} - T_y \\
&= \sum_{k \in S} w_k y_k - \sum_{k \in U} y_k \\
&= \sum_{k \in S} w_k e_k - \sum_{k \in U} e_k \quad \left(\text{since } \sum_{k \in S} w_k \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k\right) \\
&= \sum_{k \in S} a_k e_k + \left(T_{\mathbf{x}} - \sum_{k \in S} a_k \mathbf{x}_k\right)\left(\sum_{k \in S} a_k c_k \mathbf{x}_k' \mathbf{x}_k\right)^{-1} \sum_{k \in S} a_k c_k \mathbf{x}_k' e_k \\
&\qquad - \sum_{k \in U} e_k \\
&= \sum_{k \in S} a_k e_k - \sum_{k \in U} e_k + O_P(N/n).
\end{aligned} \tag{5}$$

It is now not hard to see that the GREG estimator is randomization consistent; that is, $p\lim_{n \to \infty} [(t_{y\_\mathrm{GREG}} - T_y)/N] = 0$. Moreover, both the relative randomization bias and relative randomization mean squared error of the GREG estimator are order $1/n$. Since mean squared error $= \mathrm{bias}^2 +$ variance, we can conclude that the randomization bias of the GREG estimator is usually an asymptotically insignificant contributor to its mean squared error.

3. Redefining Calibration Weights

In their original definition of calibration weights, Deville and Särndal (1992) required that the set of calibration weights, $\{w_k | k \in S\}$ 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, $t_{y\_\mathrm{CAL}} = \sum_S w_k y_k$, 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:"

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

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. Equation (6) is a mild generalization of (4) where $\mathbf{h}_k$ effectively replaces $c_k \mathbf{x}_k$.

It is not hard to see that $\mathbf{q} = [(\sum_S a_j \mathbf{h}_j' \mathbf{x}_j)^{-1}]'(T_{\mathbf{x}} - \sum_S a_j \mathbf{x}_j)'$. Moreover, under mild conditions we assume to hold, $t_{y\_\mathrm{CAL}} = \sum_S w_k y_k = \sum_S a_k y_k + (T_{\mathbf{x}} - \sum_S a_j \mathbf{x}_j)(\sum_S a_j \mathbf{h}_j' \mathbf{x}_j)^{-1} \sum_S a_k \mathbf{h}_k' y_k$ is randomization consistent

whenever t_{y_E} is. It is unbiased under the linear prediction model in equation (2) when $E(\varepsilon_k|\{\mathbf{x}_g, \mathbf{h}_g|g\in S\}, \{I_g|g\in U\}) = 0$ 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. $t_{y\_\mathrm{CAL}} = \sum_S w_k y_k$ is randomization consistent whenever t_{y_E} is under mild conditions.

That is the definition we will use. This broadened definition of calibration weighting will prove very helpful when using calibration to adjust for nonresponse or coverage errors.

It follows from our new definition that Estevao and Särdnal'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. Possibly Nonlinear Calibration

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

$$w_{k\_\mathrm{GEN}} = a_k f(\mathbf{h}_k \mathbf{q}^*), \qquad (7)$$

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. Unlike the calibration-weight equation above, the calibration equation itself, $\sum_S w_k \mathbf{x}_k = T_\mathbf{x}$, remains linear. Note that since $f(0) = 1, f'(0) = 1$, $f(\mathbf{h}_k \mathbf{q}^*) \approx 1 + \mathbf{h}_k \mathbf{q}^*$.

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 often be reached 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 $r = 1, 2, \ldots$, one then sets $\mathbf{q}^{(r)} = \mathbf{q}^{(r-1)} + \{[\sum_s f'(\mathbf{h}_k \mathbf{q}^{(r-1)}) a_k \mathbf{x}_k' \mathbf{h}_k]^{-1}\}'(T_\mathbf{x} - \sum_S w_k^{(r-1)} \mathbf{x}_k)'$, and $w_k^{(r)} = a_k f(\mathbf{h}_k \mathbf{q}^{(r)})$. Iteration stops at r^* when $T_\mathbf{x} = \sum_S w_k^{(r^*)} \mathbf{x}_k$ 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 $w_{k\_\mathrm{LIN}} = a_k (1 + \mathbf{h}_k \mathbf{q})$. A Taylor expansion around zero reveals $f(\mathbf{h}_k \mathbf{q}^{(1)}) = 1 + \mathbf{h}_k \mathbf{q}^{(1)} + O_P(1/n)$ under mild conditions, so $\sum_S w_k^{(1)} y_k = \sum_S w_{k\_\mathrm{LIN}} y_k + O_P(N/n) = T_y [1 + O_P(1/n)]$. Furthermore, it is not difficult to see that $w_{k\_\mathrm{GEN}} = w_{k\_\mathrm{LIN}}[1 + O_P(1/n)]$, 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 $x_{1k}, \ldots, x_{Pk}$, 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* model variance/randomization mean-squared-error estimator for t_{y_GREG} under an arbitrary sampling plan:

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

The term derives from r_k being "plugged into" $v_{\mathbf{ssw}}$ in place of the unknown $e_k = y_k - \mathbf{x}_k (\sum_U \mathbf{h}_i' \mathbf{x}_i)^{-1} \sum_U \mathbf{h}_i' y_i$ for randomization-mean-squared-error estimation.

Paralleling arguments in Deville and Särndal (1992), $v_{\mathbf{ssw}}$ also applies more generally to t_{y_CAL} with calibration weights defined by equation (7) with

$$r_k = y_k - \mathbf{x}_k \left(\sum_{j\in S} a_j \mathbf{h}_j' \mathbf{x}_j \right)^{-1} \sum_{j\in S} a_j \mathbf{h}_j' \mathbf{x}_j. \qquad (9)$$

This is because $w_{k\_\mathrm{GEN}} = w_{k\_\mathrm{LIN}}[1 + O_P(1/n)]$, so $\sum_S w_{k\_\mathrm{GEN}} e_k = \sum_S w_{k\_\mathrm{LIN}} e_k + O_P(N/n) = \sum_S a_k e_k + O_P(N/n)$. The last step uses reasoning exhibited in equation (5) with $\mathbf{h}_j$ serving in place of the $c_j \mathbf{x}_j$.

In their article, Deville and Särndal effectively replace the a_j in equation (9) with $w_j = a_j f(\mathbf{h}_j \mathbf{q}^*)$. A different version is given in Demanti and Rao (2004), where the a_j in the equation are replaced by $a_j f'(\mathbf{h}_j \mathbf{q}^*)$. This author noted in a comment accompanying the latter that all three versions of the r_k are asymptotically identical since $f(0) = f'(0) = 1$ and $\mathbf{q}^*$ is asymptotically $\mathbf{0}$. These asymptotic identities may no longer hold when calibration weighting is used to adjust for nonresponse as we shall see in the following section.

Developing asymptotic properties for $v_{\mathbf{ssw}}$ under stratified simple random sampling is a simple matter. In this context, $v_{\mathbf{ssw}}$ collapses to

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

where S_α denotes the sample of n_α units in stratum $\alpha(\alpha = 1, \ldots, A)$, and U_α the stratum population containing N_α elements.

For a multi-stage sample it makes sense to allow the possibility that ε_k and ε_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 $V_m = \sum_{i \in S'} E_\varepsilon[(\sum_{k \in S(i)} w_k \varepsilon_k)^2]$ 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, not strictly equal to $v_{\mathbf{SSW}}$, often has good randomization and model-based properties (when the first-stage selection probabilities are all small):

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

where α denotes a first-stage stratum of PSU's, $n_{1\alpha}$ the number of sampled PSU's in stratum α, S_α the set of sampled PSU's in α, and $S_{\alpha j}$ the set of subsampled elements from PSU j of stratum α. 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:

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

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

$$w_{k(\alpha j)} = w_k a_{k(\alpha j)} / a_k + \left(\sum_{m \in U} \mathbf{x}_m - \sum_{m \in S} w_m [a_{m(\alpha j)} / a_m] \mathbf{x}_m \right) \times \left(\sum_{m \in S} a_{m(\alpha j)} \mathbf{h}'_m \mathbf{x}_m \right)^{-1} a_{k(\alpha j)} \mathbf{h}'_k, \quad (12)$$

where $a_{k(\alpha j)} = 0$ when k is in PSU j of stratum α, $a_{k(\alpha j)} = a_k$ when k is not in stratum α at all, and $a_{k(\alpha j)} = (n_\alpha / [n_\alpha - 1])\, a_k$ 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 αj.

Let $S(\alpha+)$ be the set of *elements* in stratum α (not PSU's like S_α), and $S(\alpha j)$ the set of elements in PSU j of stratum α. Under mild conditions we assume to hold,

$$\sum_U \mathbf{x}_m - \sum_S w_m [a_{m(\alpha j)} / a_m] \mathbf{x}_m = (n_\alpha / [n_\alpha -]) \left(\sum_{S(\alpha j)} w_k \mathbf{x}_k - \sum_{S(\alpha+)} w_k \mathbf{x}_k / n_\alpha \right) = \mathbf{O}_P(N/n),$$

$$\sum_S a_{m(\alpha j)} \mathbf{h}'_m \mathbf{x}_m = \mathbf{O}_P(N),$$

and $$\sum_S a_{m(\alpha j)} \mathbf{h}'_m e_m = \mathbf{O}_P(N/\sqrt{n}).$$

As a result,

$$t_{y\_\mathrm{CAL}} - t_{y\_\mathrm{CAL}} = \sum_s w_{k(\alpha j)} e_k - \sum_s w_k e_k = (n_\alpha / [n_\alpha - 1]) \left(\sum_{s(\alpha+)} w_k e_k / n_\alpha - \sum_{S(\alpha j)} w_k e_k \right) + O_P(N / n^{3/2}),$$

and $v_J = v_{\mathrm{ST2}}[1 + O_P(1/\sqrt{n})]$ when $p\lim_{n \to \infty} (n v_{\mathrm{ST2}} / N^2) > 0$.

The replicate weights defined in equation (12) do not require iteration even when the calibration weights are themselves produced that way. This is a great computation convenience. It not only saves computer time, it avoids the possibility that at iterative solution for the w_k may exist while one for the replicate weights does not.

6. Unit Nonresponse

6.1 Quasi-randomization and Prediction Modeling

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 would respond (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 $T_{\mathbf{x}} = \sum_U \mathbf{x}_j$ is known. Finally, each p_k is assumed to have the form:

$$p_k = 1 / f(\mathbf{h}_k \boldsymbol{\phi}), \quad (13)$$

where $\boldsymbol{\phi}$ is an unknown column vector, $\mathbf{h}_k$ is a row vector with the same dimension as $\mathbf{x}_k$, and $\sum_S a_k \mathbf{h}'_k \mathbf{x}_k / N$, where S now denotes the "subsample" of respondents, is invertible both for the realized population size, N, and in the probability limit.

The function $f(\cdot)$ in equation (13) is assumed to be monotonic and twice differentiable. Its functional form is known, but the value of the governing parameter, $\boldsymbol{\phi}$, is not. When plugged into the calibration-weight equation,

$w_k = a_k f(\mathbf{h}_k \mathbf{q})$, so that the calibration equation itself, $\sum_s w_k \mathbf{x}_k = T_\mathbf{x}$ holds, $f(\mathbf{h}_k \mathbf{q})$ implicitly estimates the inverse of the element response probabilities. Unlike when calibration is used to correct for $\sum_s a_k \mathbf{x}$ differing from $T_\mathbf{x}$ due purely to sampling error, $f(0)$ and $f'(0)$ do not need to be 1 nor does $\mathbf{h}_k \boldsymbol{\phi}$ need to be zero.

The most obvious choice for $\mathbf{h}_k$ when postulating the response model in equation (13) is $\mathbf{x}_k$ itself. In a common example of calibration weighting for nonresponse, the components of $\mathbf{x}_k$ are indicator variables: $x_{gk} = 1$ when k is in group g and zero otherwise. When the groups are mutually exclusive, calibration weighting is the same thing as reweighting within post-stratification classes. See, for example, Särndal, Swensson and Wretman (1992, page 585). The prediction model usually underpinning calibration (the prefix "prediction" is needed to distinguish this model from the quasi-randomization one) assumes that every element k in group g, whether or not it would respond, has a common mean: $E_\varepsilon(y_k) = \beta_g$. The quasi-random response model is analogous: $p_k = 1/\phi_g$. The two models are conceptually very different, however.

When the groups are not mutually exclusive, raking is one method of determining calibration weights. There are others depending on the exact form of the assumed response function $f(\cdot)$. The prediction model remains linear, $E_\varepsilon(y_k) = \mathbf{x}_k \boldsymbol{\beta}$, while the response model that leads to raking, $p_k = \exp\{-\mathbf{x}_k \boldsymbol{\phi}\}$, does not. Berry, Flatt, and Pierce (1996) provides an example of using raking to adjust for nonresponse.

In many applications of calibration weighting the components of $\mathbf{x}_k$ are continuous or semi-continuous rather than dichotomous. In an annual crop survey, for example, let x_{1k} be the quantity of corn harvested in the previous census of agriculture by farm k, x_{2k} be the farm's harvested wheat, x_{3k} its harvested potatoes, and so forth. The annual crop survey has an assumed prediction model for farm k's planted corn acres, y_{1k}, of the form: $y_{1k} = \mathbf{x}_k \beta_{1k} + \varepsilon_{1k}$. The subscript, 1, is corn-specific. There are other survey values of interest, like planted wheat acres, and potentially assumed prediction models for each.

The quasi-random response model for the crop survey depends on assumptions about $f(\cdot)$ and $\mathbf{h}_k$ in equation (13) with $\mathbf{h}_k$ possibly equal to $\mathbf{x}_k$. Unlike the prediction model, the same assumed quasi-randomization model applies for all survey variables.

Promising choices for $f(\cdot)$ are $\exp(\cdot)$ and $1 + \exp(\cdot)$, the latter corresponding to a response probabilities being fit by a logistic function of $\mathbf{h}_k \boldsymbol{\phi}$. It may also be reasonable to assume $h_{gk} = x_{gk}^\lambda$ for $\lambda < 1$. In particular, setting $\lambda = 0$ means that the probability of farm k responding to the annual crop survey depends only on whether the farm had corn, wheat, or potatoes on the previous census of agriculture rather than on how much of those crops it had.

In the crop-survey example, the components of $\mathbf{x}_k$ from the previous census were the best predictors available for the corresponding annual survey values *before* sampling. Whether farm k responds to the survey, however, is more likely a function of the farm's current planted corn acres, if any, than on a predetermined proxy for that value. As a result, placing survey values in $\mathbf{h}_k$ rather than corresponding census values is tempting. There is a theoretical problem with this procedure as we shall see.

Given an $f(\cdot)$, the iterative method described in Section 4 will often be able to uncover a row vector $\mathbf{q}$ such that $T_\mathbf{x} = \sum_s a_k f(\mathbf{h}_k \mathbf{q}) \mathbf{x}_k$. When that happens, estimating T_y with $t_{y\_\mathrm{CAL}} = \sum_s w_k y_k$, where $w_k = a_k f(\mathbf{h}_k \mathbf{q})$, will have good properties under the linear prediction model: $y_k = \mathbf{x}_k \boldsymbol{\beta} + \varepsilon_k$, where $E(\varepsilon_k | \{\mathbf{x}_g, \mathbf{h}_g, I_g | g \in U\}) = 0$ 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. Note, however, that if components of $\mathbf{h}_k$ come from the survey rather than $\mathbf{x}_k$, the prediction-model assumption that $E(\varepsilon_k | \mathbf{h}_k) = 0$ can be problematic. At the extreme, consider the case where one such component is y_k itself. Usually, $E(\varepsilon_k | y_k)$ is not 0. In the crop-survey example described earlier, y_k can be the annual corn acres planted on farm k. Putting this value in $\mathbf{h}_k$ makes the associated calibration estimator for corn prediction-model biased.

When the prediction model is correct (treating $E(\varepsilon_k | \{\mathbf{x}_g, \mathbf{h}_g, I_g | g \in U\}) = 0$ as an integral part of the model), however, calibration weighting based on any choice of $f(\cdot)$ will produce estimators with good prediction-model-based properties. These estimators will also have good quasi-randomization properties when the response model in equation (13) is correct for that choice of $f(\cdot)$. 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 ε_k in the model in equation (2) satisfy $E(\varepsilon_k | \{\mathbf{x}_g | g \in U\}) = 0$, but not $E(\varepsilon_k | \{\mathbf{x}_g I_g | g \in U\}) = 0$; that is to say, the sampling mechanism – including response – is not ignorable with respect to the prediction model.

We can factor I_k into $I_{k1} I_{k2}$, where $I_{k1} = 1$ if and only if k is in the original sample, and $I_{k2} = 1$ if and only if k would respond if sampled. The interested reader can confirm that calibration weighting provides some protection against bias if the prediction model in equation (2) holds when $E(\varepsilon_k | \{\mathbf{x}_g, \mathbf{h}_g, I_{g2} | g \in U\}) = 0$; that is when the response mechanism is ignorable with respect to the

prediction model but not necessarily the original sampling mechanism.

6.2 Quasi-randomization Mean Squared Error Estimation

Whether or not t_{y_CAL} can reasonably be called prediction-model unbiased has no effect on its quasi-randomization-based properties. Note that $\mathbf{h}_k\boldsymbol{\phi}$ are $\mathbf{h}_k\mathbf{q}$ are scalar values not vectors. Since $T_{\mathbf{x}} = \sum_s a_k f(\mathbf{h}_k\mathbf{q})\mathbf{x}_i$, our assumptions and the mean value theorem $(f(\mathbf{h}_k\boldsymbol{\phi}) = f(\mathbf{h}_k\mathbf{q}) + f'(\theta_k)(\mathbf{h}_k\boldsymbol{\phi} - \mathbf{h}_k\mathbf{q}))$ reveal

$$T_{\mathbf{x}} - \sum_{k\in S} a_k f(\mathbf{h}_k\boldsymbol{\phi})\mathbf{x}_k = \sum_{k\in S} a_k[f'(\theta_k)\mathbf{h}_k(\mathbf{q}-\boldsymbol{\phi})]\mathbf{x}_k = \mathbf{O}_P(N/\sqrt{n})$$

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

$$\begin{aligned}\mathbf{q}-\boldsymbol{\phi} &= \left\{\left[\sum_{j\in S} a_j f'(\mathbf{h}_j\mathbf{q})_j\mathbf{h}'_j\mathbf{x}_j\right]^{-1}\right\}'\left[T_{\mathbf{x}} - \sum_{i\in S} a_i f(\mathbf{h}_i\boldsymbol{\phi})\mathbf{x}_i\right]' \\ &= \mathbf{O}_P(1/\sqrt{n}) \\ &= \left\{\left[\sum a_j f'(\mathbf{h}_j\boldsymbol{\phi})_j\mathbf{h}'_j\mathbf{x}_j\right]^{-1}\right\}'\left[T_{\mathbf{x}} - \sum a_i f(\mathbf{h}_i\boldsymbol{\phi})\mathbf{x}_i\right]' \\ &\quad + \mathbf{O}_P(1/n).\end{aligned}$$

The estimator t_{y_CAL} has an error of

$$t_{y\_\mathrm{CAL}} - T_y = \sum_{k\in S} a_k f(\mathbf{h}_k\mathbf{q})y_k - \sum_{k\in U} y_k = \sum a_k f(\mathbf{h}_k\mathbf{q})e_k - \sum e_k,$$

where

$$e_k = y_k - \mathbf{x}_k\left(\sum\nolimits_U f'(\mathbf{h}_j\boldsymbol{\phi})p_j\mathbf{h}'_j\mathbf{x}_j\right)^{-1}\sum\nolimits_U f'(\mathbf{h}_j\boldsymbol{\phi})p_j\mathbf{h}'_j y_j,$$

and $p_j = 1/f(\mathbf{h}_j\boldsymbol{\phi})$. The e_k are again unknown. They have been design so that $\sum_S a_k f'(\mathbf{h}_k\boldsymbol{\phi})\mathbf{h}'_k e_k = \mathbf{O}_P(N/\sqrt{n})$. Continuing:

$$\begin{aligned}&t_{y\_\mathrm{CAL}} - T_y \\ &= \sum_{k\in S} a_k f(\mathbf{h}_k\boldsymbol{\phi})e_k - \sum_{k\in U} e_k + \sum_{k\in S} a_k\{f(\mathbf{h}_k\mathbf{q}) - f(\mathbf{h}_k\boldsymbol{\phi})\}e_k \\ &= \sum a_k f(\mathbf{h}_k\boldsymbol{\phi})e_k - \sum e_k + \sum a_k f'(\mathbf{h}_k\boldsymbol{\phi})\mathbf{h}_k(\mathbf{q}-\boldsymbol{\phi})e_k \\ &\quad + O_P(N/n) \\ &= \sum a_k f(\mathbf{h}_k\boldsymbol{\phi})e_k - \sum e_k + (\mathbf{q}-\boldsymbol{\phi})'\sum a_k f'(\mathbf{h}_k\boldsymbol{\phi})\mathbf{h}'_k e_k \\ &\quad + O_P(N/n) \\ &= \sum a_k f(\mathbf{h}_k\boldsymbol{\phi})e_k - \sum e_k + O_P(N/n). \qquad (14)\end{aligned}$$

Thus, t_{y_CAL} is quasi-randomization consistent under mild conditions whenever $t = \sum_S a_k f(\mathbf{h}_k\boldsymbol{\phi})y_k$ is.

To estimate the quasi-randomization mean squared error of t_{y_CAL} (*i.e.*, the estimator's randomization mean squared error under the response 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 $1/p_k = f(\mathbf{h}_k\boldsymbol{\phi})$. From equation (14), we see that the quasi-randomization mean squared error of t_{y_CAL} is approximately

$$\begin{aligned}&E_1[(t_{y\_\mathrm{CAL}} - T_y)^2] \\ &\approx \sum_{k\in U}\sum_{j\in U}(\pi^*_{kj} - \pi^*_k\pi^*_j)(e_k/\pi^*_k)(e_j/\pi^*_j) \\ &= \sum_{k\in U}(1-\pi^*_k)e_k^2/\pi^*_k \\ &\quad + \sum_{k\in U}\sum_{\substack{j\in U\\ k\neq j}}(\pi_{kj} - \pi_k\pi_j)(e_k/\pi_k)(e_j/\pi_j). \qquad (15)\end{aligned}$$

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

$$r_k = y_k - \mathbf{x}_k\left[\sum_{j\in S} a_j f'(\mathbf{h}_j\mathbf{q})\mathbf{h}'_j\mathbf{x}_j\right]^{-1}\sum_{j\in S} a_j f'(\mathbf{h}_j\mathbf{q})\mathbf{h}'_j y_j, \quad (16)$$

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 (16) can be found in Folsom and Singh (2000). See Kott (2004a) for a further discussion of v_m in a purely sampling context.

For a general design, we can get close to a good variance/mean-squared-error estimator with

$$\begin{aligned}v_{\mathrm{com}} &= \sum_{k\in S}(w_k^2 - w_k)r_k^2 \\ &\quad + \sum_{k\in S}\sum_{\substack{j\in S\\ k\neq j}}[(\pi_{kj} - \pi_k\pi_j)/\pi_{kj}](w_k r_k)(w_j r_j). \qquad (17)\end{aligned}$$

The right hand side of equation (17) estimates the right hand side of equation (15) with r_k replacing e_k. Note that $\sum_U (1-\pi^*_k)e_k^2/\pi^*_k$ in equation (15) is estimated by $\sum_S (w_k^2 - w_k)r_k^2$ rather than $\sum_S w_k^2(1-\pi^*_k)r_k^2$, which would make v_{com} more consistent with v_{SSW} in equation (8). This substitution results in a variance estimator with good prediction-model-based properties when the ε_k are uncorrelated, and $\sigma_k^2 = \mathbf{x}_k\boldsymbol{\zeta}$, for some $\boldsymbol{\zeta}$. 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_{ST2} in equation (10) can be used as the variance/mean-squared-error estimator with r_k defined once more by equation (16).

When f is linear, $f'(\theta) = 1$, and the r_k in equation (16) are computed as if there were no nonresponse. The same holds true for the variance/mean-squared-error estimator v_{ST2}. Unfortunately, this f corresponds to an awkward-looking response-probability function: $p_k = 1/\mathbf{h}_k\boldsymbol{\phi}$. 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:

$$w_{k(\alpha j)} = w_k a_{k(\alpha j)}/a_k + \left(\sum_{m\in U}\mathbf{x}_m - \sum_{m\in S} w_m[a_{m(\alpha j)}/a_m]\mathbf{x}_m\right) \times \left(\sum_{m\in S} a_{m(\alpha j)} f'(\mathbf{h}_m\mathbf{q})\mathbf{h}'_m\mathbf{x}_m\right)^{-1} a_{k(\alpha j)} f'(\mathbf{h}_k\mathbf{q})\mathbf{h}'_k, \qquad (18)$$

an obvious generalization of the jackknife replicate weights in equation (12). Again when $f'(\theta)=1, v_J$ can be computed as if there were no nonresponse.

7. Coverage Modeling

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 $T_{\mathbf{x}}$ 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 (13). 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.

It should be noted that coverage adjustment through calibration is a extension of the well-known practice of coverage adjustment through post-stratification often used with telephone surveys. As with the post-stratification special case, one needs quantities for the calibration targets for *U* that can be assumed to be free of error or to have very little mean squared error compared to the calibration estimators themselves.

Folsom and Singh noted that overcoverage (duplication) or a combination of under and overcoverage can be handled with their methodology. The definition of p_k in equation (13) 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:

$$f(\mathbf{x}_k\boldsymbol{\phi}) = \frac{U(C-L)\exp(\mathbf{x}_k\boldsymbol{\phi}) + L(U-C)}{(U-C)+(C-L)\exp(\mathbf{x}_k\boldsymbol{\phi})}, \qquad (19)$$

where $L\ge 0, 1<U\le\infty$, and $L<C\le U$ are predetermined constants. They call this the "General Exponential Model" or "GEM." Observe that when $C=1, U=\infty$, and $L=0$, $p_k = 1/f(\mathbf{x}_k\boldsymbol{\phi}) = \exp(-\mathbf{x}_k\boldsymbol{\phi})$. Similarly, when $C=2$, $U=\infty$, and $L=1$, $p_k = [1+\exp(\mathbf{x}_k\boldsymbol{\phi})]^{-1}$; that is to say, the probability of coverage (or response) 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. They proposed its use both for coverage-error and unit-non-response adjustment

8. A Small Empirical Example

Since the jackknife replicate weights expressed in equation (18) 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, $p_k = \exp(-0.35\, x_k/M_x)$, 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 $t_{y\_\mathrm{LIN}} = \sum_S a_k(1+\mathbf{x}_k\mathbf{q})y_k$ and with $t_{y\_\mathrm{EXP}} = \sum_S a_k \exp(\mathbf{x}_k\mathbf{q}^{(\exp)})y_k$, 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 $(y_k = \mathbf{x}_k\boldsymbol{\beta}+\varepsilon_k)$, but only t_{y_EXP} was randomization consistent under the correct response model. The GREG implicitly assumed $p_k = 1/(\phi_0^{(\mathrm{LIN})} + \phi_1^{(\mathrm{LIN})}\mathbf{x}_k)$ for unknown $\phi_0^{(\mathrm{LIN})}$ and $\phi_1^{(\mathrm{LIN})}$.

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 (16), and, *ii*, the proposed jackknife, v_J, in equation (11) with replicate weights defined by equation (18). 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 $v_{\mathrm{ST2}(e)}$, was also computed with r_k replaced by $e_k = y_k - \mathbf{x}_k(\sum_U f'(\mathbf{x}_j\boldsymbol{\phi})p_j\mathbf{x}'_j\mathbf{x}_j)^{-1}\sum_U f'(\mathbf{x}_j\boldsymbol{\phi})p_j\mathbf{x}'_j y_j$, where $\boldsymbol{\phi}$ and p_j were known. In practice, e_k is rarely known, but computing $v_{\mathrm{ST2}(e)}$ is useful here for comparison.

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}, $f'(\mathbf{x}_j \boldsymbol{\phi}) = f'(\mathbf{x}_j \mathbf{q}) = 1$; for t_{y_EXP}, $f'(\mathbf{x}_j \mathbf{q}^{(\mathrm{exp})}) = \exp(\mathbf{x}_j \mathbf{q}^{(\mathrm{exp})})$, and $f'(\mathbf{x}_j \boldsymbol{\phi}) = 1/p_j$.

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 0.05 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 $v_{\mathrm{ST2}(e)}$'s were 100. Neither $v_{\mathrm{ST2}(e)}$ 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 $v_{\mathrm{ST2}(e)}$ was not significant.

The $v_{\mathrm{ST2}(e)}$ were chosen as benchmarks for the table rather than the empirical mean squared errors because each $v_{\mathrm{ST2}(e)}$ had roughly half the empirical standard error of the corresponding EMSE (which itself was the average of 1,600 squared differences) and correlated more strongly with the variance estimators. The $t-$values for this part of the table were also computed with respect to the $v_{\mathrm{ST2}(e)}$.

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 $r_{k(\mathrm{adjusted})} = r_k / \omega_k$, where $\omega_k^2 = 1 - \mathbf{x}_k (\sum_S a_j f'(\mathbf{x}_j \mathbf{q})\, \mathbf{x}'_j \mathbf{x}_j)^{-1} a_k f'(\mathbf{x}_k \mathbf{q})\, \mathbf{x}'_k = 1 + O_P(1/n)$. Observe that under the prediction model with the ε_k uncorrelated and $E(\varepsilon_k^2) = \sigma_k^2$, $E(r_{k(\mathrm{adjusted})}^2) \approx \sigma_k^2$. The near equality is exact when all the $a_j f'(\mathbf{x}_j \mathbf{q})$ and σ_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 $v_{\mathrm{ST2}(e)}$ 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. See Crouse and Kott (2004) for a different set of results.

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 (0.02)	difference from
t_{y_EXP}	100.04 (0.06)	0.58 (0.56)	T_y
Variance Estimators for $t_{y\_\mathrm{LIN}} (E_{\mathrm{EMP}}(v_{\mathrm{ST2}(e)}) = 100)$			
v_{ST2}	83.59 (1.53)	−19.96 (< 0.0001)	difference from
$v_{\mathrm{ST2(adjusted)}}$	95.53 (1.80)	−6.09 (< 0.0001)	$v_{\mathrm{ST2}(e)}$
v_J	104.69 (2.28)	3.60 (0.0003)	
EMSE	99.35 –	−0.18 (0.85)	
Variance Estimators for $t_{y\_\mathrm{EXP}} (E_{\mathrm{EMP}}(v_{\mathrm{ST2}(e)}) = 100)$			
v_{ST2}	73.12 (1.54)	−18.22 (< 0.0001)	difference from
$v_{\mathrm{ST2(adjusted)}}$	88.79 (1.98)	−8.57 (< 0.0001)	$v_{\mathrm{ST2}(e)}$
v_J	107.00 (2.73)	4.09 (< 0.0001)	
EMSE	101.21 –	0.33 (0.74)	
Other Statistics			
relvar $(v_{\mathrm{ST2}(e)[\mathrm{LIN}]})$	0.051 –	–	
relvar $(v_{\mathrm{ST2}(e)[\mathrm{EXP}]})$	0.059 –	–	
$\dfrac{(v_{\mathrm{ST2}(e)[\mathrm{LIN}]} - v_{\mathrm{ST2}(e)[\mathrm{EXP}]})}{(E_{\mathrm{EMP}}(v_{\mathrm{ST2}(e)[\mathrm{EXP}]})}$	−0.1340 (0.010)	−13.87 (< 0.0001)	

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

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 of 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.*, $\mathrm{var}_{\mathrm{ST2}}$ 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 $\mathrm{var}_{\mathrm{ST2}(e)}$) 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%.

9. Discussion

9.1 Estimating a Response Model Explicitly

When faced with unit nonresponse, many have attempted to estimate the element probabilities of response, $p_k = 1/f(\mathbf{h}_k\boldsymbol{\phi})$, 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 $t_{y\_\mathrm{GREG}} = \sum_S a_k^* y_k + (T_{\mathbf{x}} - \sum_S a_j^* \mathbf{x}_j)(\sum_S c_j a_j^* \mathbf{x}_j' \mathbf{x}_j)^{-1} \sum_S c_k a_k^* \mathbf{x}_k' \mathbf{x}_k$, 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, $\boldsymbol{\phi}$. This does not imply that direct estimation of the response model based on a given $f(\cdot)$ and $\mathbf{h}_k$ is less efficient than analogous calibration when $\mathbf{h}_k$ has the same dimension of $\mathbf{x}_k$. See Kim (2004) for a suggestion otherwise. Nevertheless, the convenience of incorporating nonresponse adjustment into calibration is appealing when variance estimates need to be produced.

A reasonable compromise is to choose the form of $f(\cdot)$ and $\mathbf{h}_k$ by modeling the response behavior of the entire sample and then estimating the parameter of $f(\cdot)$ implicitly through calibration. This compromise also overcomes a striking weakness of using calibration weighting to adjust for nonresponse (as well as for coverage errors). The choices for $f(\cdot)$ and $\mathbf{h}_k$ are motivated primarily by plausibility and convenience and not by a statistical analysis of the data.

9.2 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 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 with post-stratification: $w_{k\_\mathrm{ADJ}} = (\sum_{F(c)} w_g / \sum_{S(c)} w_g) w_k$, 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 $w_k = a_k[1 + (T_{\mathbf{x}} - \sum_S a_j \mathbf{x}_j) \times (\sum_S a_j \mathbf{d}_j' \mathbf{x}_j)^{-1} \mathbf{d}_k']$, one computes a set of weights for the respondent subsample that, unlike $\{w_{k\_\mathrm{ADJ}}\}$ 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 may not produce surprisingly small or surprisingly large weights when the dimension of $\mathbf{x}_k$ is small compared to the sample size.

9.3 Breaking Up Sample and Nonresponse Calibration

In the previous section we noted that it is possible for components of $\mathbf{h}_k$ in equation (13), the quasi-random response model, to be unknown before enumeration. When such an $\mathbf{h}_k$ is used in calibration, it might no longer to 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 $T_{\mathbf{x}}$ 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.

9.4 Work at NASS

The National Agricultural Statistics Service (NASS) used variants of the Fuller *et al*. (1994) approach for handling undercoverage in the 2002 Census of Agriculture (see Fetter and Kott 2003) and for adjusting an agricultural economics survey with large nonresponse to match totals from more reliable surveys (see Crouse and Kott 2004). In this approach, $f(\cdot)$ has the form:

$$f(\mathbf{x}_k\boldsymbol{\phi}) = \begin{cases} L & \text{when} \quad \mathbf{x}_k\boldsymbol{\phi} < L \\ \mathbf{x}_k\boldsymbol{\phi} & \text{when} \quad \mathrm{L} \le \mathbf{x}_k\boldsymbol{\phi} \le U \\ U & \text{when} \quad \mathbf{x}_k\boldsymbol{\phi} > U, \end{cases} \qquad (20)$$

which truncates linear calibration at pre-specified values, L and U, to control the size of the weight adjustment. Note that when $f(\cdot) = U$ or L, $f'(\cdot) = 0$. Unlike the calibration adjustment in equation (19), $f(\cdot)$ in equation (20) is not twice differentiable at L or U. This does not cause a problem in practice.

The agency's original justification for calibration in these contexts was based on prediction-modeling. Equation (20) 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 $f(\cdot)$ in practice? Would a different choice for $f(\cdot)$ 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

Berry, C.C., Flatt, S.W. and Pierce, J.P. (1996). Correcting unit nonresponse via nonresponse modeling and raking in the California Tobacco Survey. *Journal of Official Statistics*, 12, 349-363.

Crouse, C., and Kott, P.S. (2004). Evaluation alternative calibration schemes for an economic survey with large nonresponse. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

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.

Demnati, A., and Rao, J.N.K. (2004). Linearization variance estimation for survey data. *Survey Methodology*, 30, 17-26.

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.

Fetter, M.J., and Kott., P.S. (2003). Developing a coverage adjustment strategy for the 2002 Census of Agriculture. Presented at 2003 Federal Committee on Statistical Methodology Research Conference, http://www.fcsm.gov/03papers/fetter_kott.pdf.

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 Section on Survey Research Methods*, American Statistical Association, 598-603.

Fuller, W.A. (2002). Regression estimation for survey samples. *Survey Methodology*, 28, 5-23.

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. (2004). Efficient nonresponse weighting adjustment using estimated response probability. *Proceedings of the Survey Research Methods Section*, American Statistical Association.

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. (2004a). Randomization-assisted model-based survey sampling. *Journal of Statistical Planning and Inference*, 48, 263-277.

Kott, P.S. (2004b). Comment. *Survey Methodology*, 30, 27-28.

Little, R.J.A. (1986). Survey nonresponse adjustments for estimates of means. *International Statistical Review*, 54, 139-157.

Lundström, S., and Särndal, C.-E. (1999). Calibration as a standard method for the treatment of nonresponse. *Journal of Official Statistics*, 15, 305-327.

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.