Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

An alternative jackknife variance estimator when calibrating weights to adjust for unit nonresponse in a complex survey

by Phillip S. Kott and Dan Liao

Release date: January 6, 2022

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-514-283-9350

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Survey Methodology, December 2021
Vol. 47, No. 2, pp. 349-359
Statistics Canada, Catalogue No. 12-001-X

An alternative jackknife variance estimator when calibrating weights to adjust for unit nonresponse in a complex survey

Phillip S. Kott and Dan Liao[1]

Abstract

Calibration weighting is a statistically efficient way for handling unit nonresponse. Assuming the response (or output) model justifying the calibration-weight adjustment is correct, it is often possible to measure the variance of estimates in an asymptotically unbiased manner. One approach to variance estimation is to create jackknife replicate weights. Sometimes, however, the conventional method for computing jackknife replicate weights for calibrated analysis weights fails. In that case, an alternative method for computing jackknife replicate weights is usually available. That method is described here and then applied to a simple example.

Key Words: Analysis weight; Linearization-based variance estimator; Delete-1 jackknife variance estimator; Replicate weight; Asymptotically unbiased; Bounded logistic response model.

1. Introduction

Calibration weighting is a method for adjusting the weights in probability-sampling theory by forcing the weighted sum of each variable in a set of survey variables to equal a specified target. When that happens, the analysis weights are said to satisfy the calibration equation. There are several reasons to calibrate analysis weights. The reason we focus on here is to remove potential selection bias resulting from unit nonresponse.

It is common in the survey-sampling literature to argue that a survey respondent's calibration-weight adjustment implicitly estimates the inverse of its probability of response (see, for example, Section 5.1 of Fuller, 2009). Kott and Liao (2012) show that using calibration weighting to adjust for unit nonresponse can provide *double protection* against nonresponse bias when estimating a population total. This means that if either a linear outcome model or an implied selection model holds, then the resulting estimator is asymptotically unbiased in some sense. They go on to describe a linearization-based variance estimator for an estimated total based on a stratified multistage (or single-stage) sample with calibration-adjusted analysis weights.

The brief treatment in Sections 2 and 3 of calibration weighting for nonresponse and of linearization-based variance estimation for a calibrated estimator of a population total are developed in more depth in Kott and Liao. Proofs of the various assertions made in these sections can be found there. Here they set up the theory behind variance estimation with a jackknife.

Given a stratified multistage probability sample, a traditional delete-1 jackknife variance estimator creates sets of replicate weights, one set corresponding to each selected primary sampling unit (PSU). One selected PSU is dropped at a time, and the replicate weights of its subsampled elements are set to zero. To

1. Phillip S. Kott, Senior Research Statistician, RTI International, Rockville, Maryland 20852, U.S.A. E-mail: pkott@rti.org; Dan Liao, Research Statistician, RTI International, Rockville, Maryland 20852, U.S.A.

compensate, the replicate probability weights of the remaining elements in the same stratum as the dropped PSU are increased by the factor $n_h / (n_h - 1)$, where n_h is the original number of PSUs selected from the stratum. The replicate probability weights are calibrated in a manner analogous to the original analysis weights. Section 4 describes this jackknife and shows its near equality to the nearly-unbiased linearization-based variance estimator for a population total.

The advantage of a delete-1 jackknife over linearization for variance estimator is that once replicate weights are computed, estimating the variance of smooth function of estimated totals (such as a regression coefficient) is straightforward. Krewski and Rao (1981) provides a rigorous treatment of the delete-1 jackknife and its properties.

Sometimes no solution to a calibration equation exists when starting with a set of replicate probability weights. The main contribution of this paper is contained in the remainder of Section 4, where an alternative method of constructing jackknife replicate weights that can usually overcome this problem is described and justified. This method was introduced in Kott (2006) for another purpose.

Section 6 uses a weight-adjustment function described in Section 5 to illustrate how to implement this method. It then favorably compares the results of the method to those of two popular competitors. Section 7 discusses a variant of the alternative jackknife methodology.

2. Calibration weighting

Suppose we have a randomly drawn sample S from a finite population U. In the absence of nonresponse (as well as coverage error and measurement error), calibration weighting creates a set of analysis weights, $\{w_k \mid k \in S\}$, not dependent on the survey values of interest that

1. are close to the original inverse-probability weights, $d_k = 1/\pi_k$ where π_k is the selection probability of the k^{th} selected element; and
2. satisfy a set of linear calibration equations, one for each component of $\mathbf{z}_k$, a vector of auxiliary variables with known population totals:

$$\sum_{k \in S} w_k \mathbf{z}_k = \sum_{k \in U} \mathbf{z}_k$$

"Close" means that as the sample grows arbitrarily large, the difference between w_k and d_k vanishes in probability. For a more formal treatment of the assumed asymptotic structure, see Isaki and Fuller (1982).

Most surveys experience unit nonresponse beyond a statistician's control. One is forced to assume, either explicitly or implicitly, some type of model to adjust for the nonresponse. An outcome model (also called a "prediction model") on a survey variable of interest usually assumes the response/nonresponse mechanism, like the sampling design, is ignorable. A response model assumes the response mechanism behaves like a phase of Poisson (i.e., independent) subsampling. Double protection means that if *either* the

prediction or response model is specified correctly, the estimator will be nearly (i.e., asymptotically) unbiased in some sense. Here we will assume a correctly specified response model.

Let R be the subset of S containing respondents to the survey (for simplicity, we ignore the possibility of item nonresponse). The respondent sample can be calibrated to either the full population U:

$$\sum_{k\in R} w_k \mathbf{z}_k = \sum_{k\in U} \mathbf{z}_k \tag{2.1}$$

or to the original sample S:

$$\sum_{k\in R} w_k \mathbf{z}_k = \sum_{k\in S} d_k \mathbf{z}_k. \tag{2.2}$$

We assume a response model in which the probability of response for each $k\in U$, p_k, is an independent function having the form $p_k = p(\boldsymbol{\gamma}^T \mathbf{x}_k)$, where $p(.)$ is a smooth monotonic function, and both the known vector $\mathbf{x}_k$ and unknown parameter vector $\boldsymbol{\gamma}$ have the same number of components as $\mathbf{z}_k$. In much of the literature $\mathbf{x}_k$ is equal to $\mathbf{z}_k$, but most of the theory still follows when it does not.

If there is a vector $\mathbf{g}$ such that inserting $w_k = d_k / p(\mathbf{g}^T \mathbf{x}_k)$ solves either the calibration equation in (2.1) or (2.2), then $\mathbf{g}$ is a consistent estimator for $\boldsymbol{\gamma}$. Kott and Liao (2017) describe what to do when there are fewer components in $\mathbf{x}_k$ than in $\mathbf{z}_k$.

The function $f(\mathbf{g}^T \mathbf{x}_k) = 1/p(\mathbf{g}^T \mathbf{x}_k)$ is called the *weight-adjustment function*. The mean-value theorem tells us that under mild conditions $f(\mathbf{g}^T \mathbf{x}_k) - f(\boldsymbol{\gamma}^T \mathbf{x}_k) \approx f'(\mathbf{g}^T \mathbf{x}_k)(\mathbf{g}-\boldsymbol{\gamma})^T \mathbf{x}_k$. Consequently, as the respondent sample grows arbitrarily large $f(\mathbf{g}^T \mathbf{x}_k)$ converges to $f(\boldsymbol{\gamma}^T \mathbf{x}_k) = 1/p_k$ and $\mathbf{g}$ converges to $\boldsymbol{\gamma}$.

3. Linearization-based variance estimation

When calibrating the respondent sample to the full sample with (2.2), the calibration estimator for a population total, $t = \sum_R w_k y_k$ can be expressed as

$$\begin{aligned}
t &= \sum_{k\in S} d_k \mathbf{z}_k^T \mathbf{b} + \sum_{k\in R} d_k f(\mathbf{g}^T \mathbf{x}_k)(y_k - \mathbf{z}_k^T \mathbf{b}) \\
&\approx \sum_{k\in S} d_k \mathbf{z}_k^T \mathbf{b} + \sum_{k\in R} d_k f(\boldsymbol{\gamma}^T \mathbf{x}_k)(y_k - \mathbf{z}_k^T \mathbf{b}) + \sum_{k\in R} d_k f'(\mathbf{g}^T \mathbf{x}_k)\left[(\mathbf{g}-\boldsymbol{\gamma})^T \mathbf{x}_k\right](y_k - \mathbf{z}_k^T \mathbf{b}) \\
&= \sum_{k\in S} d_k \mathbf{z}_k^T \mathbf{b} + \sum_{k\in R} d_k f(\boldsymbol{\gamma}^T \mathbf{x}_k)(y_k - \mathbf{z}_k^T \mathbf{b}) + (\mathbf{g}-\boldsymbol{\gamma})^T \sum_{k\in R} d_k f'(\mathbf{g}^T \mathbf{x}_k)\mathbf{x}_k (y_k - \mathbf{z}_k^T \mathbf{b}) \\
&= \sum_{k\in S} d_k \mathbf{z}_k^T \mathbf{b} + \sum_{k\in R} d_k p_k^{-1}(y_k - \mathbf{z}_k^T \mathbf{b}),
\end{aligned} \tag{3.1}$$

where

$$\mathbf{b} = \left[\sum_{k\in R} d_k f'(\mathbf{g}^T \mathbf{x}_k)\mathbf{x}_k \mathbf{z}_k^T\right]^{-1} \sum_{k\in R} d_k f'(\mathbf{g}^T \mathbf{x}_k)\mathbf{x}_k y_k.$$

The key step here is that $\mathbf{b}$ has been defined so that $\sum_R d_k f_k'(\mathbf{g}^T\mathbf{x}_k)\,\mathbf{x}_k(y_k - \mathbf{z}_k^T\mathbf{b}) = 0$. Observe that $f'(.)$ in $\mathbf{b}$ is the *derivative* of the weighting-adjustment function.

Let $\mathbf{b}^*$ be the probability limit of $\mathbf{b}$ as the respondent sample (of PSUs) grows arbitrarily large. The variance of t under the original design and the selection model is nearly equivalent to the variance of $\sum_S d_k q_k^*$, where

$$q_k^* = \mathbf{z}_k^T\mathbf{b}^* + p_k^{-1}(y_k - \mathbf{z}_k^T\mathbf{b}^*)\,I_k,$$

and $I_k = 1$ when k is a unit respondent and 0 otherwise.

For many designs, q_k^* can be approximated by replacing $\mathbf{b}^*$ with $\mathbf{b}$ and p_k^{-1} with $f(\mathbf{g}^T\mathbf{x}_k)$, and the variance of $\sum_S d_k q_k$ estimated under the original design as if the $q_k = \mathbf{z}_k^T\mathbf{b} + f(\mathbf{g}^T\mathbf{x}_k)(y_k - \mathbf{z}_k^T\mathbf{b})\,I_k$ were constants. When calibrating the respondent sample to the population with equation (2.1), the $\sum_S d_k\mathbf{z}_k^T\mathbf{b}$ in equation (3.1) is replaced by $\sum_U \mathbf{z}_k^T\mathbf{b}$, which does not contribute to the variance, so $q_k = f(\mathbf{g}^T\mathbf{x}_k)(y_k - \mathbf{z}_k^T\mathbf{b})\,I_k$. Either way, replacing q_k^* with q_k tends to underestimate variances with *finite* samples (the replacement is asymptotically ignorable) because $e_k = (y_k - \mathbf{z}_k^T\mathbf{b})^2$ tends to be smaller than $e_k^* = (y_k - \mathbf{z}_k^T\mathbf{b}^*)^2$.

Given a stratified multistage probability sample with n_h sampled PSUs in each of H strata, let S_{hj} denote the subsample of elements within each PSU j in stratum h. A nearly unbiased linearization-based estimator for the variance of t is

$$v(t) = \sum_{h=1}^{H} \frac{n_h}{n_h - 1}\left[\sum_{j=1}^{n_h}\left(\sum_{k\in S_{hj}} d_k q_k\right)^2 - \frac{1}{n_h}\left(\sum_{a=1}^{n_h}\sum_{\kappa\in S_{ha}} d_\kappa q_\kappa\right)^2\right], \tag{3.2}$$

where $q_k = \alpha\mathbf{z}_k^T\mathbf{b} + f(\mathbf{g}^T\mathbf{x}_k)\,e_k I_k$, and $\alpha = 1$ when the respondent sample is calibrated to the original sample and 0 when the respondent sample is calibrated to the population. As is common in practice and continued here, equation (3.2) assumes that the little is lost by treating the PSU selection within strata as if it had been drawn with replacement, obviating the need for finite population correction.

4. Jackknife variance estimation

Let S_{h+} be all the sampled elements in stratum h. The conventional hj^{th} (delete-1) jackknife replicate for an estimated total $t = \sum_R w_k y_k$ is

$$t^{(hj)} = \sum_{k\in R} w_k^{(hj)} y_k = \sum_{k\in S} d_k^{(hj)} f(g^{(hj)T}\mathbf{x}_k)\,I_k y_k, \tag{4.1}$$

where $d_k^{(hj)} = 0$ when $k \in S_{hj}$

$d_k^{(hj)} = [n_h/(n_h - 1)]\,d_k$ when $k \in S_{h+}$ but $k \notin S_{hj}$

$d_k^{(hj)} = d_k$ otherwise,

and the $\mathbf{g}^{(hj)}$ solve the replicate calibration equation:

$\sum_{k \in R} d_k^{(hj)} f(\mathbf{g}^{(hj)T}\mathbf{x}_k)\,\mathbf{z}_k = \sum_{k \in U} \mathbf{z}_k$ when the respondent sample is calibrated to U, or

$\sum_{k \in R} d_k^{(hj)} f(\mathbf{g}^{(hj)T}\mathbf{x}_k)\,\mathbf{z}_k = \sum_{k \in S} d_k^{(hj)}\mathbf{z}_k$ when the respondent sample is calibrated to S

for each hj. Observe that $t^{(hj)}$ is an estimate of the population total $\sum_U y_k$ with the PSU hj removed.

The delete-1 jackknife variance estimator for t is

$$\text{var}_J(t) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h} \sum_{j=1}^{n_h} (t^{(hj)} - t)^2. \tag{4.2}$$

Let $e_k^* = (y_k - \mathbf{z}_k^T\mathbf{b}^*)$. Consequently,

$$\begin{aligned} t^{(hj)} - t &= \sum_{k \in S} \left\{ (d_k^{(hj)} - d_k)\,\alpha\mathbf{z}_k^T\mathbf{b}^* + \left[d_k^{(hj)} f(\mathbf{g}^{(hj)T}\mathbf{x}_k) - d_k f(\mathbf{g}^T\mathbf{x}_k) \right] I_k e_k^* \right\} \\ &\approx \sum_{k \in S} (d_k^{(hj)} - d_k) \left[\alpha\mathbf{z}_k^T\mathbf{b}^* + f(\mathbf{g}^T\mathbf{x}_k)\, I_k e_k^* \right]. \end{aligned}$$

From which we can conclude that $\text{var}_J(t)$ is approximately equal to the delete-1 jackknife for $\sum_S d_k q_k^*$, where $q_k^* = \alpha\mathbf{z}_k^T\mathbf{b}^* + f(\mathbf{g}^T\mathbf{x}_k)\, I_k e_k^*$, based on a stratified multistage sample with H strata and n_h PSUs in stratum h. A little algebra will show that the delete-1 jackknife for $\sum_S d_k q_k^*$ is equal to $\text{var}(t)$ in equation (3.2) with q_k^* replacing q_k because

$$\sum_{k \in S} (d_k^{(hj)} - d_k)\, q_k^* = \sum_{h=1}^{H} \sum_{k \in S_{h+}} (d_k^{(hj)} - d_k)\, q_k^*,$$

where

$$\begin{aligned} \sum_{k \in S_{h+}} (d_k^{(hj)} - d_k) q_k^* &= \sum_{k \in S_{h+}} \frac{1}{n_h - 1} d_k q^* - \frac{n_h}{n_h - 1} \sum_{k \in S_{hj}} d_k q^* \\ &= -\frac{n_h}{n_h - 1} \left(\sum_{k \in S_{hj}} d_k q^* - \frac{1}{n_h} \sum_{k \in S_{h+}} d_k q^* \right). \end{aligned}$$

Note that the contribution to the jackknife variance estimator for the hj^{th} replicate comes mostly from the hj^{th} PSU.

Observe that the small downward bias in finite samples caused by q_k replacing q_k^* in $\text{var}(t)$ does not apply to $\text{var}_J(t)$ in equation (4.2). The latter may have a slight tendency to be upwardly biased in finite samples because $\mathbf{g}^{(hj)}$ and $\mathbf{g}$, while both consistent estimators for $\boldsymbol{\gamma}$, need not be exactly equal.

There is sometimes a problem with computing the jackknife variance estimator $\text{var}_J(t)$ in practice. That problem occurs when $f(.)$ is such that while there is a $\mathbf{g}$ satisfying the calibration equation in (2.1)

or (2.2), no $\mathbf{g}^{(hj)}$ satisfies its analogue for at least one hj jackknife replicate. When that happens, one can follow a suggestion in Kott (2006) and compute the $w_k^{(hj)}$ in equation (4.1) with this alternative:

$$\tilde{w}_k^{(hj)} = d_k^{(hj)} f(g^T \mathbf{x}_k) + \left[\mathbf{c}^{(hj)} - \sum_{\kappa \in R} d_\kappa^{(hj)} f(g^T \mathbf{x}_\kappa) \mathbf{z}_\kappa^T \right] \left[\sum_{\kappa \in R} d_\kappa^{(hj)} f'(g^T \mathbf{x}_\kappa) \mathbf{x}_\kappa \mathbf{z}_\kappa^T \right]^{-1} d_k^{(hj)} f'(g^T \mathbf{x}_k) \mathbf{x}_k, \quad (4.3)$$

where $\mathbf{c}^{(hj)}$ is the calibration target for the hj^{th} replicate:

$$\mathbf{c}^{(hj)} = \sum_{\kappa \in U} \mathbf{z}_\kappa^T \text{ when the respondent sample is calibrated to } U,$$

and

$$\mathbf{c}^{(hj)} = \sum_{\kappa \in S} d_\kappa^{(hj)} \mathbf{z}_\kappa^T \text{ when the respondent sample is calibrated to } S.$$

By design $\sum_R \tilde{w}_k^{(hj)} \mathbf{z}_k = \mathbf{c}^{(hj)}$.

Letting $e_k = y_k - \mathbf{z}_k^T \mathbf{b}$, one can see that with $\tilde{t}^{(hj)} = \sum_R \tilde{w}_k^{(hj)} y_k$,

$$\begin{aligned} \tilde{t}^{(hj)} - t &= \sum_{k \in S} (d_k^{(hj)} - d_k) \left[\alpha \mathbf{z}_k^T \mathbf{b} + f(\mathbf{g}^T \mathbf{x}_k) I_k e_k \right] \\ &\approx \sum_{k \in S} (d_k^{(hj)} - d_k) \left[\alpha \mathbf{z}_k^T \mathbf{b}^* + f(\mathbf{g}^T \mathbf{x}_k) I_k e_k^* \right] \end{aligned}$$

so the alternative jackknife variance estimator $\text{var}_{AJ}(t)$ computed with $\tilde{t}^{(hj)}$ in place of $t^{(hj)}$ is nearly unbiased. Observe that the only possible restriction on the computation of $\tilde{w}_k^{(hj)}$ is that $\sum_R d_k^{(hj)} f'(g^T \mathbf{x}_k) \mathbf{x}_k \mathbf{z}_k^T$ be non-singular.

Observe that equation (4.3) can be rewritten as

$$\tilde{w}_k^{(hj)} = \tilde{d}_k^{(hj)} \tilde{f}(\tilde{\mathbf{g}}^{(hj)T} \tilde{\mathbf{x}}_k) \quad (4.4)$$

where $\tilde{d}_k^{(hj)} = d_k^{(hj)} f(g^T \mathbf{x}_k)$,

$$\tilde{f}(\tilde{\mathbf{g}}^{(hj)T} \tilde{\mathbf{x}}_k) = 1 + \tilde{\mathbf{g}}^{(hj)T} \tilde{\mathbf{x}}_k,$$

$$\tilde{\mathbf{g}}^{(hj)T} = \left[\mathbf{c} - \sum_{\kappa \in R} \tilde{d}_\kappa^{(hj)} \mathbf{z}_\kappa^T \right] \left[\sum_{\kappa \in R} \tilde{d}_\kappa^{(hj)} \tilde{\mathbf{x}}_\kappa \mathbf{z}_\kappa^T \right]^{-1},$$

and

$$\tilde{\mathbf{x}}_k = \frac{f'(\mathbf{g}^T \mathbf{x}_k)}{f(\mathbf{g}^T \mathbf{x}_k)} \mathbf{x}_k.$$

This equation treats the hj^{th} replicate as the full sample. The weight-adjustment function $\tilde{f}(.)$ is linear, and the $\tilde{f}_k$ are not restricted to positive values even when the f_k are. In addition, observe that even when $\mathbf{x}_k = \mathbf{z}_k$, $\tilde{\mathbf{x}}_k$ will not equal $\mathbf{z}_k$ unless $f(.) = \exp(.)$.

5. The (bounded) logistic response model

Up until this point, we have not specified a response function, $p(.)=1/f(.)$. Consider now a (bounded) logistic (or logit) response model having the form:

$$p(\boldsymbol{\gamma}^T\mathbf{x}_k) = \frac{1 + \exp(\boldsymbol{\gamma}^T\mathbf{x}_k)/U}{L + \exp(\boldsymbol{\gamma}^T\mathbf{x}_k)}. \tag{5.1}$$

where $1 \le L < U$. When $L=1$ and U is infinite, this is a standard logistic response model, where the response probability can range from 0 to 1, not including the endpoints. For finite values of L and U, the bounded probability of response falls between $1/U$ and $1/L$. Consequently, the value of adjustment function ranges from L to U. In practice, L is usually set to 1, while U is frequently set as low as the sample allows for the calibration equation to hold.

The calibration procedures in the SUDAAN® language (Research Triangle Institute, 2012), WTADJUST for when $\mathbf{x}_k = \mathbf{z}_k$ and WTADJX otherwise, fit an equivalent weight-adjustment function:

$$f(\mathbf{g}^T\mathbf{x}_k) = \frac{L + B\exp(A\mathbf{g}^T\mathbf{x}_k)}{1 + B\exp(A\mathbf{g}^T\mathbf{x}_k)/U}, \tag{5.2}$$

where $A = \frac{U-L}{(C-L)(U-C)}$, $B = U\frac{C-L}{U-C}$, and $L < C < U$.

The choice of C helps determine what $\mathbf{g}$ satisfies the calibration equation but will not affect the value of the weight adjustment itself, $f_k = f(\mathbf{g}^T\mathbf{x}_k)$. Consequently, C can be any value between L and U. When $L=1$, $C=2$, and U is infinite, $\mathrm{A}=\mathrm{B}=1$.

A little calculus reveals with the weight-adjustment function in equation (5.2):

$$f_k' = f'(g^T\mathbf{x}_k) = \frac{(U - f_k)(f_k - L)}{(U - C)(C - L)},$$

which is needed to compute equation (4.3) or (4.4). The general exponential model in the SUDAAN calibration procedures allow the L, C, and U to vary from element to element, a flexibility hard to interpret in response modeling and not considered here.

What will be useful here, although not for modeling, is the possibility that L in equation (5.2) is 0 and U is infinite. When iteratively solving a calibration equation for $\mathbf{g}$ with $f(\mathbf{g}^T\mathbf{x}_k) = \exp(\mathbf{g}^T\mathbf{x})$ using Newton's method, the SUDAAN calibration procedures first solve for $\mathbf{g}_1$ in the calibration equation with $f(\mathbf{g}_1^T\mathbf{x}_k) = 1 + \mathbf{g}_1^T\mathbf{x}_k$, which is a useful result when computing alternative jackknife weights. (The programs set the first iteration of the weight adjustment at $f_{k1} = \exp(\mathbf{g}_1^T\mathbf{x}_k)$ from which $1 + \mathbf{g}_1^T\mathbf{x}_k$ is easily derived.)

6. A simulation example

The MU281 population of municipalities in Särndal, Swensson and Wretman (1992; data from the slightly revised version is contained in http://lib.stat.cmu.edu/datasets/mu284; one of the municipalities

was accidentally dropped in this analysis) has been augmented with an indicator (RESP) for whether an element (municipality) would respond if sampled. Probabilities of element response were generated using a logistic function of one of the data set's covariates (the log of the element's 1975 population in thousands). The average probability of response was roughly 70%.

A stratified simple random sample of 10 elements per each of 8 strata was simulated 1,716 times. In each simulated sample, the elements with RESP = 1 were treated as respondents, and the respondent sample was calibrated to the full sample using the weight-adjustment function in equation (5.2) with a lower bound of 1 and an upper bound of 5. In the calibration model, the two components of $\mathbf{x}_k$ were 1 and the log of the element's 1975 population in thousands; $\mathbf{z}_k$ was set equal to $\mathbf{x}_k$. 1,225 out of the 1,716 simulations had their respondent samples successfully calibrated on both components (i.e., satisfied the calibration equation in 2.2) and produced linearization-based standard-errors.

Estimated means (ratios of two estimated totals) and standard errors (square roots of estimated variances) were computed for four variables:

P85	1985 population (in thousands).
RMT85	Revenues from 1985 municipal taxation (in millions of kronor).
ME84	Number of municipal employees in 1984.
REV84	Real estate values according to 1984 assessment (in millions of kronor).

Although the SUDAAN procedure WTADJUST can compute standard errors when using a delete-1 jackknife, it will fail when one or more replicates fail to calibrate. Therefore, two versions of the conventional delete-1 jackknife standard errors were computed using a macro the authors created. In one, the set of the imperfect "calibrated" weights from the last iteration for the failed replicates were used. In the other, the replicates that failed to calibrate were dropped and this following modified jackknife variance estimator was computed:

$$\mathrm{var}_J^*(t) = \sum_{h=1}^{H} \frac{n_h - 1}{n_h^*} \sum_{j=1}^{n_h^*} (t^{(hj)} - t)^2, \tag{6.1}$$

where n_h^* is the number of replicates in stratum h that successfully calibrated. This revised jackknife variance estimator is suggested by Rust (1985) when replicates are dropped at random, which is not what happens here. The SAS-callable (SAS Institute Inc., 2015) SUDAAN code used in the analysis for a single simulation is available from the authors upon request.

Among the 1,225 analyzable samples, 867 simulations had all the replicates using conventional delete-1 jackknife calibrate, while the remaining 358 simulations had at least one replicate that failed to calibrate after 50 iterations (the default is 10). Table 6.1 averages the results for both situations. When no conventional replicate failed, the alternative and conventional jackknife standard errors are close (on

average) and slightly higher than those produced by linearization as theory predicts (note that the two versions of the conventional delete-1 jackknife are identical).

Table 6.1
Standard errors based on jackknife methods when calibrating for nonresponse with a bounded logistic model

	Variable	Estimated Mean	Linearization-based Standard Error	Alternative Jackknife Standard Error	Conventional Jackknife Standard Error Including Failed Replicates	Conventional Jackknife Standard Errors Dropping Failed Replicates
867 Simulations Where No Conventional Replicate Failed to Calibrate	P85	22.41	2.06	2.09	2.11	2.11
	RMT85	167.70	17.31	17.72	17.86	17.86
	ME84	1,215.87	124.67	127.27	128.15	128.15
	REV84	2,425.52	212.83	217.00	219.67	219.67
358 Simulations Where at Least One Conventional Replicate Failed to Calibrate	P85	22.78	2.24	2.31	2.95	2.04
	RMT85	170.48	18.93	19.47	24.39	16.60
	ME84	1,236.75	135.94	139.65	175.99	121.31
	REV84	2,451.95	239.27	239.18	296.77	208.45

When at least one replicate failed to calibrate for the conventional delete-1 jackknife, the alternative jackknife's standard errors are again close to linearization-based ones, even though it failed to calibrate in 114 out of these 358 simulations due to a (near) singularity in at least one of the replicates. However, including the failed replicates clearly overestimates standard error and dropping them clearly underestimates relative to linearization. It appears that the alternative jackknife variance estimator produces the more useful set of replicate weights in this situation.

Table 6.1 compares standard-errors from competing jackknifes to linearization-based standard errors rather than empirical standard errors because finite-population correction has been ignored. Moreover, the bounded logistic response model fit in the simulations was not the unbounded response model used to generate responses.

7. Discussion

There is a small chance (about 1.5% in our simulations) for equation (4.4) to return negative replicate weights. The canned procedures of many statistical packages (like SAS) cannot handle negative weights. Consequently, estimated totals computed from replicate weights may need to be calculated without the help of a canned procedure.

One does not need access to SUDAAN to compute alternative jackknife weights for calibration estimators. The *gencalib* routines in the 'Sampling' package in R (Tillé and Matei, 2016) can perform

calibration not only under a bounded logistic response model but under linear calibration as well. Although there are SAS macros equivalent to WTADJUST, to our knowledge, there is currently no publicly-available SAS calibration-weighting macro that can be used when $\tilde{\mathbf{x}}_k$ in the weight-adjustment (equation 4.4) does not equal $\mathbf{z}_k$. Let us hope this is reversed soon.

Acknowledgements

The authors would like to thank the editors whose helpful suggestions improved the quality of this note.

References

Fuller, W. (2009). *Sampling Statistics*, New York: John Wiley & Sons, Inc., Hoboken.

Isaki, C., and Fuller, W. (1982). Survey design under the regression super-population model. *Journal of the American Statistical Association*, 77, 89-96.

Kott, P.S. (2006). Using calibration weighting to adjust for nonresponse and coverage errors. *Survey Methodology*, 32, 2, 133-142. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2006002/article/9547-eng.pdf.

Kott, P., and Liao, D. (2012). Providing double protection for unit nonresponse with a nonlinear calibration weighting routine. *Survey Research Methods*, 6, 105-111.

Kott, P., and Liao, D. (2017). Calibration weighting for nonresponse that is not missing at random: Allowing more calibration than response-model variables. *Journal of Survey Statistics and Methodology*, 5, 159-174.

Krewski, D., and Rao, J.N.K. (1981). Inference from stratified samples: Properties of the linearization, jackknife, and balanced repeated replication methods. *Annals of Statistics*, 9, 1010-1019.

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

Rust, K. (1985). Variance estimation for complex estimators in sample surveys. *Journal of Official Statistics*, 1, 381-397.

Särndal, C.-E., Swensson, B. and Wretman, J. (1992). *Model Assisted Survey Sampling*. New York: Springer.

SAS Institute Inc. (2015). SAS/STAT® 14.1 User's Guide. SAS Institute Inc., Cary, NC.

Tillé, Y., and Matei, A. (2016). Package 'sampling' is available online at http://cran.r-project.org/web/packages/sampling/sampling.pdf.