An Overview of Calibration Weighting and the Delete-a-Group Jackknife

Phillip S. Kott

Outline

NASS Applications
Some Calibration Theory
Pre-calibration Examples (of Calibration)
Variance Estimation
The Delete-a-Group Jackknife
Nonresponse and Coverage Adjustment
Where Do We Go From Here?

NASS Applications

The Quarterly Crops/Stocks Survey is based on an mpps sample. The sample before nonresponse adjustment is calibrated to the same frame variables as were used in drawing the sample.

The dag jackknife is used assuming (incorrectly) that NASS reweights for nonresponse treating unit response as an additional phase of sampling.

The Cost and Returns Report (ARMS III) is based on a stratified sample which itself may be a multiphase subsample. The respondent sample is calibrated to targets determined from other NASS surveys. The dag jackknife is used.

The Census of Agriculture uses calibration to adjust for undercoverage. Targets come from various sources. The dag jackknife in not used.

The Production Practices Report (ARMS II) uses a dag jackknife.

Some Calibration Theory

The target:

$T_y = \sum_U y_k$, where U denotes the population of N members.

The expansion estimator:

$t_{y_E} = \sum_S y_k / \pi_k = \sum_S a_k y_k$,

where S is the sample of n members, and $a_k = 1/\pi_k$ is the *sampling weight* of element k. *This same formula applies for any variable y_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 $\sum_S w_k \mathbf{x}_k = \sum_U \mathbf{x}_k$ ← *the calibration equation*

for some row vector of auxiliary variables, $\mathbf{x}_k = (x_{1k}, ..., x_{Pk})$,
about which $\mathbf{T_x} = \sum_U \mathbf{x}_k$ is known.

As with the expansion estimator, the same set of calibration weights can be used no matter what the y_k.

Usually, t_{y_CAL} is nearly randomization unbiased for sufficiently large sample under mild conditions.

The estimator t_{y_CAL} estimates T_y perfectly when $y_k = \mathbf{x}_k\beta$ exactly for all k.

Thus, it is reasonable to expect t_{y_CAL} to be a good estimator when y_k and $\mathbf{x}_k\beta$ are almost always close.

This is formalized by assuming the y_k are random variables satisfying the linear prediction model:

$y_k = \mathbf{x}_k\beta + \epsilon_k$, ← *the (prediction) model*

where $E(\epsilon_k \mid \mathbf{x}_g, I_g) = 0$ for all $k \in U$; $I_g = 1$ when $g \in S$, 0 otherwise.

Under this model, it is easy to see that t_{y_CAL} is an unbiased estimator for T_y in the sense that $E_\epsilon(t_{y_CAL} - T_y)$.

Problem

We are usually interested in estimating totals for a variety of target variables at the same time. It is often unreasonable to assume that different variables satisfy the same linear model.

A simple example demonstrates the solution

Suppose y_{Ck} is the current planted corn acres for farm k, and y_{Wk} the farm's current planted wheat acres.

Suppose there is a list frame of all the farms in the population containing control values for corn and wheat acres.

Denoting the control values for farm k as x_{Ck} and x_{Wk}, the combined linear model inherent in calibration takes the form:

$$y_{Ck} = \begin{pmatrix} 1 & x_{Ck} & x_{Wk} \end{pmatrix} \begin{pmatrix} \beta_{C0} \\ \beta_{CC} \\ \beta_{CW} \end{pmatrix} + \varepsilon_{Ck} \qquad \text{for corn, and}$$

$$y_{Wk} = \begin{pmatrix} 1 & x_{Ck} & x_{Wk} \end{pmatrix} \begin{pmatrix} \beta_{W0} \\ \beta_{WC} \\ \beta_{WW} \end{pmatrix} + \varepsilon_{Wk} \qquad \text{for wheat.}$$

The $\mathbf{x}_k$-vector is common to both the model for corn and wheat. The β-vector is not.

Calibration has its drawbacks.

It may be reasonable to assume that β_{CW} (the effect of the wheat-control acres on the amount of planted corn acres given the corn-control acres) and β_{WC} (the effect of the corn-control acres on the amount of planted wheat acres given the wheat-control acres) are zero and increase the efficiency of both the planted-corn and planted-wheat indications. Calibration does not allow us to do that.

Pre-calibration Examples

Poststratification is a form of calibration where each component of $\mathbf{x}_k$ is a classification variable: $x_{pk} = 1$ when k is in Class p and 0 otherwise.
When each k is in one and only one of the P classes, a poststratified estimator sets

$$w_k = \frac{N_p}{\sum_S a_j x_{pj}} a_k \quad \text{for sampled k in class p,}$$

where N_p is the population size of the class.

The p'th component of the calibration equation, $\sum_S w_k x_{pk} = N_P$, holds for each p.

In addition, $E_I(N_p / \sum_S a_j x_{pj}) \approx 1$ for a sufficiently large sample under mild conditions.

Note that a *stratified sample* is already calibrated:
$\sum_S a_k x_{hk} = N_h$, where $a_k = N_h / n_h$ for $i \in S_h$.

Raking (Deming and Stephan 1940) extends the notion of poststratification to classes that are not mutually exclusive (e.g., size and farm type).
Also called "iterative proportional fitting."
Deming and Stephan called their method "a least-squares adjustment," but it is not.

Ratio Adjustments are like poststratification with $x_{pk} = x_k$ when k is in Class p and 0 otherwise. Here

$$w_k = \frac{\sum_U x_{pj}}{\sum_S a_j x_{pj}} a_k \quad \text{for sampled k in class p,}$$

The estimator is usually written

$$t_{y_r} = \sum_{p=1}^{P} \{ \sum_{k \in U_p} x_k (\sum_{k \in S_p} a_k y_k / \sum_{k \in S_p} a_k x_k) \}.$$

Linear Calibration

Most of the calibration weighting in practice involve a variant of least squares, where the calibration weights have linear the form:

$$w_k = a_k(1 + c_k\mathbf{x}_k\mathbf{g}) \quad \text{or} \quad a_k(1 + c_k\mathbf{g}'\mathbf{x}_k')$$

for some set of constants $\{c_k\}$ constants (c_k is usually 1, but it can be 0 or something else)

and a vector $\mathbf{g} = (\sum_S a_j c_j \mathbf{x}_j'\mathbf{x}_j)^{-1} (\mathbf{T_x} - \sum_S a_j\mathbf{x}_j)'$.

Linear Calibration and Linear Regression

$$t_{y_CAL} = \sum_S w_k y_k = \sum_S a_k (1 + c_k\, \mathbf{g}'\mathbf{x}_k')y_k$$

$$= \sum_S a_k y_k + \mathbf{g}' \sum_S a_k c_k \mathbf{x}_k' y_k$$

$$= t_{y_E} + (\mathbf{T_x} - \sum_S a_j\mathbf{x}_j)(\sum_S a_j c_j \mathbf{x}_j'\mathbf{x}_j)^{-1} \sum_S a_k c_k \mathbf{x}_k' y_k$$

$$= t_{y_E} + (\mathbf{T_x} - \sum_S a_j\mathbf{x}_j) \quad \underset{\substack{\uparrow \\ \text{estimator for } \beta}}{\mathbf{b}}$$

(Notice too that the calibration equation holds when y_k is replaced by $\mathbf{x}_k$.)

Nonlinear Calibration

Raking weights have the form: $w_k = a_k\exp(\mathbf{x}_k\mathbf{g})$,

When $\mathbf{x}_k\mathbf{g}$ is small, raking weights are asymptotically close to linear calibration weights with all the $c_k = 1$.

But there is no obvious link to linear regression and least squares.

Variance Estimation

If the ϵ_k in the model are uncorrelated, each with a variance of σ_k^2, then the model variance of $t_{_CAL}$ is

$$
\begin{aligned}
E_\epsilon[(\textstyle\sum_S w_k y_k - \sum_U y_k)^2] &= E_\epsilon[(\textstyle\sum_S w_k \epsilon_k - \sum_U \epsilon_k)^2] \\
&= \textstyle\sum_S w_k^2 \sigma_k^2 - 2\sum_S w_k \sigma_k^2 + \sum_U \sigma_k^2 \\
&\approx \textstyle\sum_S w_k^2 \sigma_k^2 - \sum_S w_k \sigma_k^2 \\
&\approx \textstyle\sum_S w_k^2 \sigma_k^2 \quad \text{when most } w_k >> 1.
\end{aligned}
$$

This suggests the estimators

$v = \sum_S (w_k^2 - w_k) r_k^2$ and $v_o = \sum_S w_k^2 r_k^2$,

where $r_k = y_k - \mathbf{x_k b}$ (and $E(r_k^2) \approx \sigma_k^2$).

These are also good estimators for the randomization mean squared error under Poisson sampling. In that context,

$(w_k^2 - w_k) \approx (1 - \pi_k)/\pi_k^2$,

and r_k is a proxy for the model-free population residual:

$$
e_k = y_k - \mathbf{x_k B}.
$$

$$
\uparrow
$$

$$
(\textstyle\sum_U c_k \mathbf{x}_j' \mathbf{x}_j)^{-1} \sum_S c_k \mathbf{x}_k' y_k .
$$

By *not* replacing the w_k with the $1/\pi_k$ we retain a stronger model-based property (which is conditioned on the actual sample drawn) without sacrificing the more asymptotic randomization-based property.

The Delete-a-Group Jackknife

Suppose the (respondent) sample was randomly divided into 15 mutually exclusive and nearly equal groups denoted Sg where g = 1, ..., 15. We call the complement of each group, the *jackknife replicate group* and denote it by S(g).

Observe that

$$\sum_S w_k \epsilon_k - \sum_{S(g)} (15/14) w_k \epsilon_k = \sum_{Sg} w_k \epsilon_k - \sum_{S(g)} (1/14) w_k \epsilon_k.$$

Consequently,

$$E_M[\sum^{15} (14/15)(\sum_S w_k \epsilon_k - \sum_{S(g)} (15/14) w_k \epsilon_k)^2] = \sum_S w_k^2 \sigma_k^2.$$

Unfortunately, the ϵ_k are unknown. We could replace them by the y_k if the calibration equation

$$\sum_{S(g)} (15/14) w_k \mathbf{x}_k = \sum_U \mathbf{x}_k \quad \text{held for all g.}$$

It does not.

We can, however, replace the $(15/14)w_k$ within S(g) by the (nonstandard) g'th replicate weight:

$$w_{k(g)} = (15/14)w_k + (\sum_U \mathbf{x}_i - \sum_{S(g)}[15/14]w_i\mathbf{x}_i)[\sum_{S(g)} c_i a_i \mathbf{x}_i' \mathbf{x}_i]^{-1} c_k a_k \mathbf{x}_k' .$$

(In the standard formulation, w_k and w_i above are replaced by a_k and a_i.)

The replicate weights within S(g) are not only very close to the $(15/14)w_k$, they also satisfy the calibration equation since:

$$w_{k(g)} = (15/14)w_k + (\sum_U \mathbf{x}_i - \sum_{S(g)}[15/14]w_i\mathbf{x}_i)[\sum_{S(g)} c_i a_i \mathbf{x}_i' \mathbf{x}_i]^{-1} c_k a_k \mathbf{x}_k'$$

implies

$$\sum_{S(g)} w_{k(g)} \mathbf{x}_k = \sum_{S(g)} (15/14) w_k \mathbf{x}_k + (\sum_U \mathbf{x}_i - \sum_{S(g)}[15/14]w_i\mathbf{x}_i) \sum_{S(g)} c_i a_i \mathbf{x}_i' \mathbf{x}_i]^{-1} \sum_{S(g)} c_k a_k \mathbf{x}_k' \mathbf{x}_k$$

Thus, the delete-a-group jackknife variance estimator,

$$v_{dag} = \sum_{g=1}^{15} (14/15)(\textstyle\sum_S w_k y_k - \sum_{S(g)} w_{k(g)} y_k)^2$$

$$= \sum (14/15)(\textstyle\sum_S w_k \epsilon_k - \sum_{S(g)} w_{k(g)} \epsilon_k),^2$$

is a nearly unbiased estimator for the model variance of t_{y_CAL} no matter what the sample design. The model bias caused by the replication weights not exactly equaling $(15/14)w_k$ is usually upward.

Poisson Sampling

Under Poisson sampling v_{dag} is almost unbiased for the randomization mean squared error of t_{y_CAL} when $\sum_U e_k = 0$ (recall e_k is the model-free population residual), which happens when all the $c_k = 1$ and $\mathbf{x}_k$ contains a constant or the equivalent. It also happens with standard ratio adjustments.

The key is that both sides of $\sum_S w_k e_k - \sum_{S(g)} (15/14) w_k e_k$ estimate $\sum_U e_k = 0$.

The Expansion Estimator Under Stratified SRS (before calibration)

Ignoring finite population correction

$$V_{rand} = \sum_{h=1}^{H} (N_h^2/n_h) \sum_{k \in U_h} e_k^2/(N_h - 1) \approx \sum_{k \in U} w_k e_k^2,$$

where

$$e_k = y_k - \sum_{j \in U_h} y_j/N_h, \quad w_k = N_h/n_k \quad \text{for } k \in U_h.$$

A conceptual unbiased estimator,

$$v = \textstyle\sum_S w^2 e_k^2,$$

is approximately by the dag jackknife with $w_{k(g)} = N_h/n_{h(g)}$ when $k \in S_h$. This is approximately $(15/14)N_h/n_h$ as long as n_h is not too small (say n_h is greater than five).

Multi-phase Samples

When the calibration occurs *after* the latest phase of a multi-phase sample, one can use these replicate weights:

$$w_{k(g)} = (a_{k(g)}/a_k)w_k + \left(\sum_U \mathbf{x}_i - \sum_{S(g)}[a_{i(g)}/a_i]w_i\mathbf{x}_i\right)\left[\sum_{S(g)} c_i a_i \mathbf{x}_i'\mathbf{x}_i\right]^{-1} c_k a_k \mathbf{x}_k',$$

where

a_k is the sampling weight before calibration, and

$a_{k(g)}$ is the replicate weight before calibration.

When $a_{k(g)}/a_k \approx 15/14$, the approximate model unbiasedness of the dag jackknife holds, but there is an additional tendency for the bias to be upward.

If the dag jackknife was almost randomization unbiased before calibration, it remains so after. (Whether the dag jackknife applies under two-phase sampling for an estimator *before* calibration is another question.)

Smooth function of linear Estimators

The dag retains its properties for a for smooth function of linear estimators, like

the ratio estimator: $r = \sum_S w_k y_k / \sum_S w_k z_k$

a composite estimator: $t_{Composite} = \lambda\, t_1 + (1 - \lambda)\, t_2$ or

an estimated regression coefficient: $\mathbf{b} = \left(\sum_S a_j \mathbf{z}_j'\mathbf{z}_j\right)^{-1} \sum_S a_k \mathbf{z}_k' y_k.$

Nonresponse and Coverage Adjustment

We can use the same prediction model to justify both calibration weight and the dag jackknife if the sample is subject to nonresponse or the frame to coverage errors (mostly undercoverage). For the latter, $\sum_U \mathbf{x}_k$ is everywhere replaced by a more generic population total, $\mathbf{T_x}$.

The group-mean model (poststratification) is the most common form of calibration for both nonresponse adjustment and undercoverage (e.g., ratioing to state crop totals in the ARMS II).

Quasi-random models

An alternative approach is to treat unit response or frame coverage as another phase of Poisson sampling. This is called quasi-randomization or quasi-random response (coverage) modeling.

Consider $w_k = a_k(1 + \mathbf{x}_k\mathbf{g})$ (note the c_k are set to 1)

or more generally, $w_k = a_k f(\mathbf{x}_k\mathbf{g})$ (with f(.) = exp(.) for raking).

Just as $a_k = 1/\text{Prob}(k \in S)$,

assume $f(\mathbf{x}_k\gamma) = 1/\text{Prob}(k \in R \mid k \in S)$ or
$1/\text{Prob}$ (k is in the frame)

The components of γ are unknown and estimated implicitly by finding a **g** (perhaps through iteration) such that the calibration equations hold.

For calibration weighting to lead to systematic bias, *both* the prediction and quasi-random models must fail.

That is to say calibration is effective at handling nonresponse (or undercoverage) unless the probability of response is systematically higher than (or lower than) $1/ f(\mathbf{x}_k\mathbf{g})$ for farms where $y_k - \mathbf{x}_k\mathbf{b}$ is greater than zero.

When f(z) is nonlinear, the replicate weights become

$w_{k(g)} = (a_{k(g)}/a_k)w_k + (\mathbf{T_x} - \sum_{S(g)} [a_{i(g)}/a_i]w_i\mathbf{x}_i) [\sum_{S(g)} f'(\mathbf{x}_i\mathbf{g})a_i\mathbf{x}_i'\mathbf{x}_i]^{-1} f'(\mathbf{x}_k\mathbf{g})a_k\mathbf{x}_k'$,

where the $f'(\mathbf{x}_i\mathbf{g})$ is there for quasi-randomization reasons.

$\mathbf{T_x}$ can be replaced by $\mathbf{T}_{\mathbf{x}(g)}$ when $\mathbf{T_x}$ itself is based on a sample.

Where Do We Go From Here?

Should we increase the number of replicates from 15?

Can we find a practical way to assess and perhaps remove the upward bias in the dag jackknife?

Can we develop a rigorous way of dealing with calibration- weight outliers, such as calibration weights less than 1 (or 0) or greater than six times the pre-calibration weight. Presently, we set a_k at the allowable boundary for a calibration weight, change c_k to 0, and recompute the calibration weights.

What, if anything, can we do about replication-weight outliers, such as replicate weights with negative values or with values more than twice 15/14'th of the associated calibration weight?

Is it practical to compute an extended dag jackknife for handling area-frame strata with five or fewer sampled segments?

Should we be using nonlinear f(z) when adjusting for nonresponse or coverage errors; in particular,

$f(z) = \exp(z)$ (generalized raking; note f must be positive),

$f(z) = 1 + \exp(z)$ (the logistic model; note $f \geq 1$),

or, more generally, the c-centered, bounded from below by $\ell \geq 0$ and above by u:

$$f(z) = \frac{(c - \ell)\exp(z) + \ell(u - c)}{(u - c) + (c - \ell)\exp(z\gamma)}$$

(Singh and Folsom, 2000)?

Should be consider "back-link" functions of the form

$1 / \text{Prob(k responds if sampled)} = f(\mathbf{h}_k\gamma)$,

where some components of $\mathbf{h}_k$ are not exactly the same as components of the benchmark vector $\mathbf{x}_k$?

For example, we may want to benchmark a respondent sample to control-variable totals but allow response to be a function of survey variables.