Estimateur par calage et technique de ratissage généralisé dans les enquêtes par sondage

Jean-Claude DEVILLE, Insee
et Carl-Erick SÄRNDAL, Université de Montréal, Canada

Calibration Estimators and Generalized Raking Technique in Survey Sampling

Abstract. This paper is about estimation of the finite population total in the presence of univariate or multivariate auxiliary information. Attention is focused on alternative weighting systems that reflect a given auxiliary information. There are two parts to the paper: (i) derivation of a weighting system with the aid of a distance measure and a set of calibration equations; (ii) an application to the case where the information consists of known marginal counts in a two- or multi-way table; this is called generalized raking.

The general regression estimator (GREG) was conceived with multivariate auxiliary information in mind. It is ordinarily justified by a regression relationship between the study variable y and the auxiliary vector x. But the GREG can be derived by a different route by focusing instead on the weights. The sampling weight of the k:th observation is π_k^{-1}, where π_k is the inclusion probability of k. We show that the weights implied by the GREG are as close as possible, according to a given distance measure, to the π_k^{-1} while respecting side conditions called calibration equations. These state that the sample sum of the weighted auxiliary variable values must equal the known population total for that auxiliary variable. That is, the calibrated weights must give perfect estimates when applied to each auxiliary variable. This consistency check on a weighting system is required by many practitioners. The GREG uses the auxiliary information efficiently, but the weights are not always without reproach. Negative weights can occur; in some applications such weights make no sense. It is natural to seek the root of the dissatisfaction in the

underlying distance measure. Consequently, we allow alternative distance measures that satisfy only a set of minimal requirements. Each distance measure leads, via the calibration equations, to a weighting system and thereby to a new estimator. These form a family of calibration estimators. We show that the GREG is a first approximation to all other members of the family; all of them are asymptotically equivalent to the GREG, and it is the already available variance estimator for the GREG that we recommend to use for any member in the family. Numerical features of the weights and ease of computation may become more than anything else the basis for choosing between the estimators. We apply the theory to calibration on known marginals of a two-way frequency table. Our family of distance measures then leads to a family of generalized raking procedures. Classical raking ratio is one of these.Variance expressions and easily calculated variance estimators are given for generalized raking estimators, and inference conditional on the estimated cell counts is among the topics discussed.

KEY WORDS: Multivariate auxiliary information; Regression estimators; Raking.

1. Introduction

Survey statisticians use auxiliary information in many ways to improve survey estimates. One example is when the general regression estimator is used for the finite population total or mean. It depends on a vector of auxiliary variables for which the population total is known. The calibration estimators derived in this paper are a family of estimators that appeal to a common base of auxiliary information. A calibration estimator uses calibrated weights. These are as close as possible, according to a given distance measure, to the original sampling design weights π_k^{-1}, while respecting a set of constraints, the calibration equations. To every distance measure corresponds a specific calibrated weighting and a calibration estimator.In Section 3, we define a family of distance measures and derive the corresponding family of calibration estimators. Their properties are established in a series of results. Variance estimators for calibration estimators are given in Section 4. An important application is calibration on known marginal counts in multi-way tables.

Two-way tables are examined in Sections 6 to 9. The calculation of the cell weights, given one of our distance measures, can be described as generalized raking. Variance estimation for generalized raking and conditional inference for two-way tables are important topics in Sections 7 to 9.

2. Deriving the general regression estimator by calibration

Consider a finite population $U = \{1, ..., k, ..., N\}$ from which a probability sample s ($s \subseteq U$) is drawn with a given sampling design, $p(\cdot)$. That is, $p(s)$ is the probability that s is selected. The inclusion probabilities $\pi_k = \Pr(k \in s)$ and $\pi_{k\ell} = \Pr(k \& \ell \in s)$ are assumed strictly positive. Let y_k be the value of the variable of interest, y, for the k:th population element, with which is also associated an auxiliary vector value, $\mathbf{x}_k = (x_{k1}, ..., x_{kj}, ..., x_{kJ})'$. For the elements $k \in s$, we observe $(y_k, \mathbf{x}_k)$. The population total of $\mathbf{x}$, $\mathbf{t}_x = \sum_U \mathbf{x}_k$, is assumed accurately known. This knowledge may come from one or more sources: census data, administrative data files, etc. If A ($A \subseteq U$) is any set of population elements, $\sum_A$ is our shorthand for $\sum_{k \in A}$, for example, $\sum_s y_k$ means $\sum_{k \in s} y_k$.

The objective is to estimate the population total $t_y = \sum_U y_k$. Extending an idea of Lemel (1976), Deville (1988) used calibration on known population x-totals to modify the basic sampling design weights, $d_k = 1/\pi_k$, that appear in the Horwitz-Thompson estimator, $\hat{t}_{y\pi} = \sum_s y_k/\pi_k = \sum_s d_k y_k$. A new estimator, $\hat{t}_{yw} = \sum_s w_k y_k$, is sought, with weights w_k that are as close as possible, in an average sense and for a given metric, to the d_k, while respecting the calibration equation

$$\sum_s w_k \mathbf{x}_k = \mathbf{t}_x . \qquad (2.1)$$

Here, w_{ks} would be a more appropriate notation for the sample dependent weights, but for brevity

we write just w_k. The idea of modifying the d_k is found, in a different context, in Zieschang (1986). If $E_p(\cdot)$ denotes expectation with respect to the sampling design p(s), a measure of average distance reminiscient of the chi-square statistic is

$$E_p\{\Sigma_s (w_k - d_k)^2/2d_k\}.$$

For more generality in this expression, we can let the k:th term have an individual, known positive weight $1/q_k$, unrelated to d_k, which gives the average distance

$$E_p\{\Sigma_s (w_k - d_k)^2/2d_kq_k\}. \tag{2.2}$$

The uniform weighting $1/q_k = 1$ is likely to dominate in applications, but unequal weights $1/q_k$ are sometimes motivated; see Example 1 below. To minimize (2.2) subject to (2.1) holding for every possible sample s is equivalent to minimizing, for any particular s, the quantity

$$\Sigma_s (w_k - d_k)^2/2d_kq_k = \Sigma_s d_k(w_k/d_k - 1)^2/2q_k ,$$

subject to the single constraint (2.1). Minimization leads to the calibrated weight

$$w_k = d_k(1 + q_k x_k'\lambda), \tag{2.3}$$

where the Lagrange multiplier λ is determined from (2.1), that is,

$$\lambda = T_s^{-1} (t_x - \hat{t}_{x\pi}) , \tag{2.4}$$

assuming that the inverse of

$$T_s = \Sigma_s d_kq_kx_kx_k' . \tag{2.5}$$

exists. The resulting estimator of t_y is

$$\hat{t}_{yreg} = \Sigma_s w_ky_k = \hat{t}_{y\pi} + (t_x - \hat{t}_{x\pi})'\hat{B}_s , \tag{2.6}$$

where $\hat{t}_{x\pi} = \Sigma_s d_kx_k$ denotes the Horvitz-Thompson estimator for the x-vector, and

$$\hat{B}_s = T_s^{-1} \Sigma_s d_k q_k x_k y_k . \qquad (2.7)$$

is a weighted multiple regression coefficient estimator. Thus, Deville's (1988) calibration technique (a) provides an alternative derivation of the generalized regression estimator, see Cassel, Särndal and Wretman (1976), Gourieroux (1981), Särndal (1980), Isaki and Fuller (1982), Wright (1983) and others, and (b) shows that it is constructive to view (2.6) as a linear weighting method with sample dependent weights w_k given by (2.3). Such a view was taken in Särndal (1982), who used the w_k to create a variance estimator for $\hat{t}_{yreg}$ (see Section 4 below), in Bethlehem and Keller (1987), and in Lemaître and Dufour (1987). The research question addressed in this paper is whether useful alternative estimators will result by using generalized distance measures.

Example 1: Derivation of the ratio estimator. Take $x_k = x_k$, a positive scalar. Then $x_k'\lambda = x_k\lambda$. Let us take $q_k = 1/x_k$. We obtain $\lambda = (\Sigma_U x_k)/(\Sigma_s d_k x_k) - 1 = t_x/\hat{t}_{x\pi} - 1$, whereby $w_k = d_k(1+ q_k x_k\lambda) = d_k(1 + \lambda) = d_k t_x/\hat{t}_{x\pi}$, and from (2.6) $\hat{t}_{yreg} = t_x \hat{t}_{y\pi}/\hat{t}_{x\pi}$, the ratio estimator. The unequal weighting $q_k = 1/x_k$ is essential for obtaining this result. □

3. A class of alternative distance measures

In (2.2), the distance between the original weight d_k and the new weight w_k was rather arbitrarily taken as $(w_k - d_k)^2/2d_kq_k$. It is natural to allow alternative distance measures sharing a few basic features. For element k, consider a distance $G_k(w,d)$ such that: (1) for every fixed d > 0, $G_k(w,d)$ is nonnegative, differentiable with respect to w, strictly convex, defined on an interval $D_k(d)$ containing d, and such that $G_k(d,d) = 0$; (2) $g_k(w,d) = \partial G_k(w,d)/\partial w$ is continuous and maps $D_k(d)$ onto an interval $Im_k(d)$ in a one-to-one fashion. It follows that

$g_k(w,d)$ is a strictly increasing function of w and $g_k(d,d) = 0$. Average distance is now measured by $E_p\{\sum_s G_k(w_k,d_k)\}$. To minimize this quantity subject to (2.1) holding for all s is equivalent to seeking the w_k that minimize, for any particular s, the sum $\sum_s G_k(w_k,d_k)$ under the single constraint (2.1). If λ is a Lagrange multiplier, derivation gives

$$g_k(w_k,d_k) - x_k'\lambda = 0 . \tag{3.1}$$

If a solution exists, our assumptions guarantee that it is unique. It can always be written as

$$w_k = d_k F_k(x_k'\lambda) , \tag{3.2}$$

for a certain function $F_k(u)$ such that $F_k(0) = 1$; $F_k'(0) = q_k > 0$. Here, $d_kF_k(\cdot)$ is the reciprocal mapping of $g_k(\cdot,d_k)$. It maps $Im_k(d_k)$ onto $D_k(d_k)$ in an increasing fashion.

In most of our applications, $g_k(w,d) = g(w/d)/q_k$, where $g(\cdot)$ is a function of the single argument w/d, independent of k, continuous, strictly increasing, and such that $g(1) = 0$. Then $g_k(w,d)$ depends on k only through the multiplicative factor $1/q_k$. If $F(u) = g^{-1}(u)$ denotes the inverse function of $g(\cdot)$, (3.2) becomes

$$w_k = d_k F(q_k x_k'\lambda) . \tag{3.2a}$$

From (2.1), the calibration equations necessary to determine $\lambda = (\lambda_1, \ldots, \lambda_j, \ldots, \lambda_J)$ are

$$t_x = \sum_s w_k x_k = \sum_s d_k F_k(x_k'\lambda) x_k . \tag{3.3}$$

It is convenient to define

$$\phi_s(\lambda) = \sum_s d_k\{F_k(x_k'\lambda) - 1\}x_k , \tag{3.4}$$

whereby (3.3) can be written as

$$\phi_s(\lambda) = t_x - \hat{t}_{x\pi} . \qquad (3.5)$$

The right hand side is a known quantity for every sample s. In summary form, the procedure is:

1. Given the data from the realized sample s, and for the chosen $F_k(\cdot)$, solve (3.5) for λ. Iteration may be required, as discussed in Section 10.

2. Once λ is determined, the resulting calibration estimator of t_y is obtained as

$$\hat{t}_{yw} = \Sigma_s w_k y_k = \Sigma_s d_k F_k(x_k'\lambda)\, y_k . \qquad (3.6)$$

Applications of the procedure are given later. The distance function $G_k(w_k,d_k)$ is chosen by the statistician. Alternatively, he chooses the uniquely corresponding function $F_k(u) = F_k(x_k'\lambda)$. Examples of the form $g_k(w,d) = g(w/d)/q_k$ are shown in Table 1. Since $1/q_k$ is a recurring multiplicative factor, the table shows $q_k G_k(w_k,d_k)$ and $q_k g_k(w_k,d_k) = g(w_k/d_k)$.

Table 1. Examples of distance functions $G_k(w_k,d_k)$, with the associated $g_k(w_k,d_k)$ and $F_k(u)$.

Case	$q_k G_k(w_k,d_k)$	$q_k g_k(w_k,d_k) = g(w_k/d_k)$	$F_k(u) = F(q_k u)$
1	$(w_k - d_k)^2/2d_k$	$w_k/d_k - 1$	$1 + q_k u$
2	$w_k \log(w_k/d_k) - w_k + d_k$	$\log(w_k/d_k)$	$\exp(q_k u)$
3	$2(\sqrt{w_k} - \sqrt{d_k})^2$	$2\{1 - (w_k/d_k)^{-1/2}\}$	$(1 - q_k u/2)^{-2}$
4	$-d_k \log(w_k/d_k) + w_k - d_k$	$1 - (w_k/d_k)^{-1}$	$(1 - q_k u)^{-1}$
5	$(w_k - d_k)^2/2w_k$	$\{1 - (w_k/d_k)^{-2}\}/2$	$(1 - 2q_k u)^{-1/2}$

Wellknown distance measures are involved: Hellinger distance in Case 3, minimum entropy distance in Case 4. In Cases 1, 3, 4 and 5, $F_k(u)$ is of the form $(1 + \alpha\, q_k u)^{1/\alpha}$, with $\alpha = 1$, $-1/2$, -1, -2, respectively; Case 2 is obtained when $\alpha \longrightarrow 0$. In Case 1, which yields the regression estimator (2.6), the weights w_k can be positive or negative; Cases 2 to 5 guarantee positive weights w_k. Cases 1 and 2 always lead to a solution of (3.5); in Cases 3 to 5, a solution is not guaranteed, but Result 1 below shows that the probability of a solution tends to one. In a given case, unrealistic weights w_k may occur, although rarely. Negative weights w_k may occur in Case 1; this may be unacceptable. Equally undesirable, Case 2 may yield extremely large positive weights w_k. The cases in Table 1 allow considerable flexibility, but there is reason to consider further alternatives that restrict the values of $F_k(u)$.

Case 6. In Case 2, the values of $F_k(u) = \exp(q_k u)$ range in $(0, \infty)$. To restrict the range, specify constants L and U such that $L < 1 < U$, set $A = (U-L)/\{(1-L)(U-1)\}$ and define

$$F_k(u) = \frac{L(U-1) + U(1-L)\exp(Aq_k u)}{(U-1) + (1-L)\exp(Aq_k u)}$$

We have $F_k(-\infty) = L$; $F_k(\infty) = U$; $F_k(0) = 1$, $F_k'(0) = q_k$. The resulting w_k satisfy $L\, d_k < w_k < U\, d_k$. The distance function $G_k(w_k, d_k)$ in this case is, apart from a multiplicative constant,

$$(x-L)\log\frac{x-L}{1-L} + (U-x)\log\frac{U-x}{U-1}$$

with $x = w_k/d_k$. If L is large negative, and U large positive, we are close to Case 1. If $L = 0$ and U is large, we are close to Case 2. □

Case 7. Case 1 can be similarly restricted by specifying $F(q_k u) = 1 + q_k u$ if $(L - 1)/q_k \leq u \leq (U - 1)/q_k$; $F_k(u) = L$ if $u < (L - 1)/q_k$ and $F_k(u) = U$ if $u > (U - 1)/q_k$, for any suitable constants L and U. The weights w_k will then satisfy $L\, d_k < w_k < U\, d_k$. The corresponding distance function is as in Case 1 if $L\, d_k < w_k < U\, d_k$ and defined as infinity otherwise. A choice $L \geq 0$ eliminates the possibility of negative weights. □

Example 2: The ratio estimator is obtained for any $g_k(w,d)$ of the form $g(w/d)/q_k$, if $x_k = x_k$, a positive scalar, and $q_k = 1/x_k$, as in Example 1. Then $F_k(x_k'\lambda) = F(q_k x_k \lambda) = F(\lambda)$, a constant. From (3.3), $F(\lambda) = t_x/\hat{t}_{x\pi}$, so (3.6) gives the ratio estimator $\hat{t}_{yw} = t_x\, \hat{t}_{y\pi}/\hat{t}_{x\pi}$. □

Example 2 is rather exceptional. Generally, different $F_k(u)$ yield different estimators. However, Result 5 below states that all estimators (3.6), under mild conditions on the underlying $F_k(u)$, are asymptotically equivalent to the regression estimator (2.6), generated by $F_k(u) = 1 + q_k u$. Thus, for medium to large samples, the choice of $F_k(u)$ has only a modest impact on such essential properties as the variance of the estimator. Computational convenience may then more than anything else dictate the choice of $F_k(u)$. We now derive several asymptotic results that are needed later. The setup we use for asymptotics is essentially that of Fuller and Isaki (1981), Isaki and Fuller (1982). Important features of the setup are the following. We consider a sequence of finite populations and sampling designs indexed by n, where n is the sample size (for a fixed size sampling design) or the expected sample size (for a random size sampling design). The finite population size, N, tends to infinity with n, and we assume that

1. $\lim N^{-1} t_x$ exists;

2. $N^{-1} (\hat{t}_{x\pi} - t_x) \longrightarrow 0$ in design probability;

3. $n^{1/2} N^{-1} (\hat{t}_{x\pi} - t_x)$ converges in distribution to the multinormal $N(\mathbf{0}, \Sigma)$.

Here, (3) is to justify the use of the normal approximation in confidence intervals based on $\hat{t}_{yw}$.

Let $\Sigma\Sigma_U$ be shorthand for $\Sigma_{k=1}^{N}\Sigma_{\ell=1}^{N}$; set $\Delta_{k\ell} = \pi_{k\ell} - \pi_k \pi_\ell$. Now, (1) to (3) imply that

$$n N^{-2} \Sigma\Sigma_U \Delta_{k\ell} (x_k/\pi_k)(x_\ell/\pi_\ell)' = n N^{-2} V(\hat{t}_{x\pi})$$

converges to the fixed matrix Σ and that

$$N^{-1}(\hat{t}_{x\pi} - t_x) = O_p(n^{-1/2}) .$$

In particular, under simple random sampling without replacement (SRS) with sampling fraction $f = n/N$, (1) to (3) imply that

$$N^{-1} \Sigma_U x_k x_k' - N^{-2} t_x t_x'$$

converges to a fixed positive definite matrix, V_0, and that

$$\Sigma = \lim (1 - f) V_0 .$$

We can view Σ as a matrix that describes an asymptotic effect of the sampling design in use.

From a practical point of view the assumptions mean that: (1) the components of $\hat{t}_{x\pi} - t_x$ are considered small, and that quantities of the order of $\| \hat{t}_{x\pi} - t_x \|^2$ are considered negligible; and (2) $\hat{t}_{x\pi} - t_x$ follows approximately a normal distribution with covariance matrix $n^{-1} N^2 \Sigma$.

Before proving asymptotic properties of $\hat{t}_{yw}$, we discuss the existence of a solution of (3.5).

Now, (3.4) defines a function of λ on $C = \cap_{k\in U} \{\lambda : x_k'\lambda \in \mathrm{Im}_k(d_k)\}$, a convex domain. Assuming that C is an open neighbourhood of $\mathbf{0}$, independently of n, we have:

Result 1. Equation (3.5) has a unique solution belonging to C, with probability tending to one as $n \longrightarrow \infty$. □

Result 2. Let λ_s be the solution of (3.5), if one exists; otherwise, let λ_s be an arbitrary fixed value. Then λ_s tends to $\mathbf{0}$ in design probability, and $\lambda_s = O_p(n^{-1/2})$. □

In order to obtain Results 3, 4, and 5 below, we add the assumptions

$$\text{(i)}\ \max \| \mathbf{x}_k \| = M < \infty\ , \quad \text{(ii)}\ \max F_k''(0) = M' < \infty\ ,$$

where max is over n as well as over k.

Result 3. We have

$$\lambda_s = \mathbf{T}_s^{-1} (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}) + O_p(n^{-1})\ . \qquad \square$$

The proofs of Results 1, 2 and 3 are given in the Appendix.

Result 4. The calibration estimator $\hat{t}_{yw}$ given by (3.6) is design consistent, and

$$N^{-1}(\hat{t}_{yw} - \hat{t}_{y\pi}) = O_p(n^{-1/2})\ . \qquad \square$$

Proof. With $F_k'(0) = q_k$, we have

$$F_k(u) = 1 + q_k u + \theta_k(u)\ , \tag{3.7}$$

where $\max \theta_k(u) = O(u^2)$. If (3.5) has a solution, λ_s, then

$$\hat{t}_{yw} - \hat{t}_{y\pi} = \Sigma_s\, d_k\, y_k \{ q_k \mathbf{x}_k' \lambda_s + \theta_k(\mathbf{x}_k' \lambda_s) \}\ ,$$

so

$$N^{-1} \left| \hat{t}_{yw} - \hat{t}_{y\pi} \right| \le N^{-1} \{ (\Sigma_s\, d_k\, q_k \left| y_k \right| \| \mathbf{x}_k \|) \| \lambda_s \| \} + O_p(n^{-1}),$$

where $N^{-1} \{ \Sigma_s\, d_k\, q_k \left| y_k \right| \| \mathbf{x}_k \| \} = O_p(1)$ and $\lambda_s = O_p(n^{-1/2})$ by Result 2. The result follows, since $\hat{t}_{y\pi}$ is design consistent and $\hat{t}_{y\pi} - t_y = O_p(n^{-1/2})$. □

Remark. Since $\hat{t}_{yw}$ is the nearest estimator to $\hat{t}_{y\pi}$ in a given sense, it can be expected to inherit some of the properties of $\hat{t}_{y\pi}$. By definition, $\hat{t}_{y\pi}$ is design unbiased. We expect to find that $\hat{t}_{yw}$ is asymptotically design unbiased (ADU). This property can be obtained, if attention is paid to one detail: it is not certain that (3.5) has a solution. With a small probability, there is none, and $\hat{t}_{yw}$ is undefined. We therefore modify the estimator as follows: Use $\hat{t}_{yw}$ if (3.5) has a solution; if not, use $\hat{t}_{y\pi}$ (that is, set $\lambda_s = 0$). This gives an ADU estimator. Undefined estimators occur in simple cases, too. The usual poststratification estimator is undefined if there is one or more zero poststratum counts. The regression estimator (2.6) is undefined if T_s is singular. □

Result 5. For any $F_k(\cdot)$ obeying our conditions, the estimator $\hat{t}_{yw}$ given by (3.6) is asymptotically equivalent to the regression estimator $\hat{t}_{yreg}$ given by (2.6), in the sense that

$$N^{-1} (\hat{t}_{yw} - \hat{t}_{yreg}) = 0_p(n^{-1}) .$$

As a consequence, the two estimators share the same asymptotic variance. □

Proof: From (3.6) and (3.7),

$$N^{-1}\hat{t}_{yw} = N^{-1} \hat{t}_{y\pi} + N^{-1} (t_x - \hat{t}_{x\pi})' T_s^{-1} \Sigma_s d_k q_k x_k y_k + 0_p(n^{-1}) + N^{-1} \Sigma_s d_k y_k \theta_k (x_k' \lambda_s) .$$

The first two terms of the r.h.s. equal $N^{-1} \hat{t}_{yreg}$, where $\hat{t}_{yreg}$ is the regression estimator (2.6). The last term was found in the proof of Result 4 to be $0_p(n^{-1})$. Therefore, $n^{1/2} N^{-1} (\hat{t}_{yw} - \hat{t}_{yreg}) = 0_p(n^{-1/2})$, with a zero asymptotic variance. □

4. Variance and variance estimation

Result 5 states that $\hat{t}_{yw}$ is asymptotically equivalent to $\hat{t}_{yreg}$, which is the special case of $\hat{t}_{yw}$ generated by $F_k(u) = 1 + q_k u$. For any $F_k(u)$ satisfying our conditions, the asymptotic variance (AV) of $\hat{t}_{yw}$ is thus the same as that of the regression estimator, namely,

$$AV(\hat{t}_{yw}) = \sum\sum_U \Delta_{k\ell} (d_k E_k)(d_\ell E_\ell) , \tag{4.1}$$

where $\Delta_{k\ell} = \pi_{k\ell} - \pi_k \pi_\ell$ and $E_k = y_k - \mathbf{x}_k' \mathbf{B}$, with $\mathbf{B}$ satisfying the normal equation

$$(\sum_U q_k \mathbf{x}_k \mathbf{x}_k') \mathbf{B} = \sum_U q_k \mathbf{x}_k y_k. \tag{4.2}$$

Clearly, $\mathbf{B}$ minimizes the weighted least squares expression

$$SS_U = \sum_U q_k (y_k - \mathbf{x}_k' \mathbf{B})^2 = \sum_U q_k E_k^2 . \tag{4.3}$$

The residuals E_k cannot be used for variance estimation, since $\mathbf{B}$ is unknown. Let $\hat{\mathbf{B}}_s$ be an estimator. Two alternatives are given below. Sample-based residuals can then be calculated as

$$e_k = y_k - \mathbf{x}_k' \hat{\mathbf{B}}_s . \tag{4.4}$$

The variance estimator that we advocate uses these residuals as follows:

$$\hat{V}(\hat{t}_{yw}) = \sum\sum_s (\Delta_{k\ell}/\pi_{k\ell})(w_k e_k)(w_\ell e_\ell). \tag{4.5}$$

In (4.5), calibrated weights w_k are given to the residuals. The advantage that these weights have over the simple design weights d_k is that (4.5) has attractive properties with respect to both the sampling design and the underlying regression model, as Särndal, Swensson and Wretman (1989) show. These properties are design consistency and approximate model unbiasedness.

Now consider the calculation of $\hat{\mathbf{B}}_s$ in (4.4). Note that SS_U given by (4.3) is the unknown population total of the fixed quantities $q_k E_k^2$. The calibrated weights estimator of this total is SS_{sw}

$= \sum_s w_k q_k E_k^2$, which is minimized by $\hat{\mathbf{B}}_{sw}$ satisfying the sample-based normal equations

$$(\sum_s w_k q_k \mathbf{x}_k \mathbf{x}_k') \hat{\mathbf{B}}_{sw} = \sum_s w_k q_k \mathbf{x}_k y_k \ . \qquad (4.6)$$

An alternative is to take $SS_{sd} = \sum_s d_k q_k E_k^2$ to estimate SS_U. This leads to $\hat{\mathbf{B}}_{sd}$ satisfying

$$(\sum_s d_k q_k \mathbf{x}_k \mathbf{x}_k') \hat{\mathbf{B}}_{sd} = \sum_s d_k q_k \mathbf{x}_k y_k \ .$$

Either $\hat{\mathbf{B}}_s = \hat{\mathbf{B}}_{sw}$ or the computationally slightly simpler $\hat{\mathbf{B}}_s = \hat{\mathbf{B}}_{sd}$ may be used in (4.4) and (4.5). The difference in the calculated value of $\hat{V}(\hat{t}_{yw})$ is negligible in most cases.

Example 3. Let us return to $\hat{t}_{yw} = t_x \hat{t}_{y\pi}/\hat{t}_{x\pi}$ in Example 2. Under SRS, (4.5) yields

$$\hat{V}(\hat{t}_{yw}) = \left(\frac{\bar{x}_U}{\bar{x}_s}\right)^2 \frac{1-f}{n} \frac{\sum_s e_k^2}{n-1}$$

where $e_k = y_k - \hat{B}_s x_k$ with $\hat{B}_s = (\sum_s y_k)/(\sum_s x_k) = \hat{B}_{sw} = \hat{B}_{sd}$. This is an often recommended variance estimator for the ratio estimator. Note that $\hat{B}_{sw}$ and $\hat{B}_{sd}$ agree in this case. □

5. Implications for poststratification.

Poststratification is an important practice. To apply the preceding, let there be H population groups (the poststrata). Let $\mathbf{x}_k$ be composed of H-1 "zeroes" and a single "one", indicating the group to which k belongs. Then $\mathbf{t}_x = \sum_U \mathbf{x}_k = (N_1, \ldots, N_h, \ldots, N_H)'$, the vector of known population group counts. With $\lambda = (\lambda_1, \ldots, \lambda_h, \ldots, \lambda_H)'$, we get $\mathbf{x}_k'\lambda = \lambda_h$ whenever k is in group h, so $\mathbf{x}_k'\lambda$ depends only on the group, not on the label k within the group. We assume

$g_k(w,d) = g(w/d)$, that is, $q_k = 1$ for all k. The calibration equations (3.3) give $F_k(x_k'\lambda) =$

$F(\lambda_h) = g^{-1}(\lambda_h) = N_h/\hat{N}_h$, where $\hat{N}_h = \Sigma_{s_h} 1/\pi_k$. For any element k in group h, the weight is

$$w_k = d_k F(\lambda_{gh}) = d_k N_h/\hat{N}_h.$$

Let $\tilde{y}_{s_h} = (\Sigma_{s_h} y_k/\pi_k) /\hat{N}_h$. From (3.6) we obtain the usual poststratified estimator,

$$\hat{t}_{yw} = \Sigma_{h=1}^{H} N_h \tilde{y}_{s_h} = \hat{t}_{ypos}. \qquad (5.1)$$

This estimator obtains for any function $g_k(w,d)$ of the form $g(w/d)$.

6. Calibration for a two-way table

The technique of this paper can be used to calibrate on the known marginal counts for a frequency table in any number of dimensions. In the case of a two-way table with c columns and r rows, there are $c + r - 1$ linearly independent components in the λ-vector to be determined. For a three-way table, the λ-vector has $r + c + f - 2$ components to be determined, where f is the number of levels of the additional third factor, and so on.

For simplicity, we limit the discussion to two-way tables. With r rows and c columns, there are $r \times c$ cells. The typical population cell, U_{ij}, contains N_{ij} elements; $i = 1, ..., r$; $j = 1, ..., c$; so $N = \Sigma\Sigma_{i,j} N_{ij}$, where $\Sigma\Sigma_{i,j}$ means $\Sigma_{i=1}^{r}\Sigma_{j=1}^{c}$. We distinguish two levels of calibration: (a) at the higher level, calibration is on the known cell counts N_{ij}; this is complete poststratification; (b) at the lower level, calibration is on known marginal counts, leading to a class of raking procedures; this is incomplete poststratification. In the following, we assume $q_k = 1$ for all k, and distance measures such that $g_k(w,d) = g(w/d)$. This implies $F_k(u) = F(u) = g^{-1}(u)$.

Case a: Calibration on known population cell counts (complete poststratification). The results

of Section 5 apply directly, with $H = r\,c$ groups. From (5.1), the calibrated weights are

$w_k = d_k\, N_{ij}/\hat{N}_{ij}$ for all k in cell ij, with $\hat{N}_{ij} = \Sigma_{s_{ij}}\, 1/\pi_k$, so the calibration estimator is

$$\hat{t}_{ypos} = \Sigma\Sigma_{i,j}\; N_{ij}\, \tilde{y}_{s_{ij}}\ , \qquad (6.1)$$

where $\tilde{y}_{s_{ij}} = (\Sigma_{s_{ij}}\, y_k/\pi_k)/\hat{N}_{ij}$ is the π-weighted y-mean of the sample cell $s_{ij} = U_{ij} \cap s$.

If all the N_{ij} are known, and none of the s_{ij} are empty or extremely small, (6.1) exhausts t available information and is the preferred estimator. But situations often arise where calibration at the lower level is either necessary or preferred:

1. The population marginal counts N_{ij} are known, but the cell counts are not. The marginal counts may come from different data files, for example, age group counts from one file, professional group counts from another, but crossclassification counts are lacking. By necessity, calibration is on the known marginals.

2. There are some zero or extremely small sample cell counts. Then $\hat{t}_{ypos}$ is undefined or may be unstable. Calibration on the cell counts, although perhaps feasible, is abandoned in favor of the more reliable calibration obtained from the known marginals. This is of particular interest when a table has three or more dimensions.

3. The auxiliary information comes from an independent, large survey. Suppose this survey provides precise marginal count estimates, but modest precision for cell count estimates. For example, the annual French survey on employment uses a sample of about 130,000 individuals. With a sample this large, excellent precision is obtained for estimated counts in socio-professional categories, head-of-household age groups, and educational levels, but estimated cell counts for the crossclassification of any two dimensions have modest precision. The marginal count estimates are treated as true values, and are used for calibration, instead of the volatile cell count estimates.

4. Stratified sampling in conjunction with poststratification. One wishes to preserve the

advantages of the stratification (say, geographical strata) and at the same time benefit from poststratification on another dimension (say, known age group counts). Then the cell counts are typically unknown, and part of the interest lies in estimating them.

Now, (1) to (4) are reasons to consider the following

Case b. Calibration on known marginals (incomplete poststratification). The considerable literature on the subject starts with Deming and Stephan (1940). References to this development are given later. We consider any function $F(\cdot)$ obeying the conditions in Section 3, and assume $q_k = 1$ for all k. The $\mathbf{x}_k$-vector must be defined so that $\Sigma_U \mathbf{x}_k$ captures (but does not go beyond) the information used for calibration, which is now the vector of marginal counts. This implies that

$$\mathbf{x}_k = (\delta_{1\cdot k}, \ldots, \delta_{r\cdot k}, \delta_{\cdot 1k}, \ldots, \delta_{\cdot ck})' , \qquad (6.2)$$

where $\delta_{i\cdot k} = 1$ if the element k is in row i, and 0 otherwise; $\delta_{\cdot jk} = 1$ if k is in column j, and 0 otherwise. Then, $\Sigma_U \mathbf{x}_k = (N_{1+}, \ldots, N_{r+}, N_{+1}, \ldots, N_{+c})'$, where $N_{i+} = \Sigma_{j=1}^{c} N_{ij}$, $N_{+j} = \Sigma_{i=1}^{r} N_{ij}$. Letting $\lambda = (a_1, \ldots, a_r, b_1, \ldots, b_c)$, we have $\mathbf{x}_k'\lambda = a_i + b_j$ whenever k belongs to cell ij. That is, $F(\mathbf{x}_k'\lambda) = F(a_i + b_j)$ depends on the cell, but not on the label within the cell.

With $\hat{N}_{ij} = \Sigma_{s_{ij}} 1/\pi_k$, the calibration equations (3.3) take the form

$$\Sigma_{j=1}^{c} \hat{N}_{ij} F(a_i + b_j) = N_{i+} \quad (i = 1, \ldots, r) \quad ; \qquad (6.3)$$

$$\Sigma_{i=1}^{r} \hat{N}_{ij} F(a_i + b_j) = N_{+j} \quad (j = 1, \ldots, c) \quad . \qquad (6.4)$$

This system must be solved for $a_1, \ldots, a_r, b_1, \ldots, b_c$, for the function $F(\cdot)$ chosen by the statistician. Iterative solution is often required; see Section 10. One equation is redundant, so we fix one component, say, $b_c = 0$, and solve the system for $i = 1, \ldots, r$; $j = 1, \ldots, c-1$. Note that

$a_i + b_j$ remains invariant to the elimination of one equation. Having solved for the a_i and b_j, we calculate the cell effects $F(a_i + b_j)$, the calibrated cell count estimates

$$\hat{N}_{ij}^{w} = \hat{N}_{ij}\, F(a_i + b_j)\,, \tag{6.5}$$

the calibrated weights

$$w_k = d_k\, \hat{N}_{ij}^{w}/\hat{N}_{ij}\,, \tag{6.6}$$

and finally the estimator obtained from (3.6), denoted $\hat{t}_{ymarg}$, is

$$\hat{t}_{yw} = \Sigma_s\, w_k y_k = \Sigma\Sigma_{i,j}\ \hat{N}_{ij}^{w}\ \tilde{y}s_{ij} = \hat{t}_{ymarg}\ . \tag{6.7}$$

If we compare with complete poststratification, the difference is that the known cell counts N_{ij} in (6.1) are replaced in (6.7) by estimates, $\hat{N}_{ij}^{w}$. If the information content is high in the marginals, the $\hat{N}_{ij}^{w}$ are excellent estimates that improve substantially on the naive estimates $\hat{N}_{ij}$.

In (6.5), $F(a_i + b_j)$ measures the effect of cell ij (factor one at level i, factor two at level j). This effect can be given a group theoretical representation. Let $F(a_i) = \alpha_i$; $F(b_j) = \beta_j$. Then if $b_j = 0$, the cell effect reduces to $F(a_i + b_j) = F(a_i) = \alpha_i$. Similarly, if $a_i = 0$, then $F(a_i + b_j) = F(b_j) = \beta_j$. Introduce $\alpha_i{*}\beta_j$ as a symbolic notation for $F(a_i + b_j)$. Then

$$\alpha_i{*}\beta_j = F\{F^{-1}(\alpha_i) + F^{-1}(\beta_j)\} = F(a_i + b_j)\ , \tag{6.8}$$

and the calibrated weight can be written as $w_k = d_k\, \alpha_i{*}\beta_j$. It can be shown that if F^{-1} maps D onto $\mathbb{R}$, then $\alpha_i{*}\beta_j$ defines a group on D, with unity as neutral element. Conversely, one can show that a group on D with 1 as neutral element can be expressed on the form (6.8). This is

true of Cases 1, 2, and 6 in Section 3. When F^{-1} maps D onto an interval I of $\mathbb{R}$ containing 0, as in Cases 3, 4 and 5, then $\alpha_i * \beta_j$ defines a "pseudo-group" in the following sense: (i) $\alpha*\beta$ is defined on the domain $F^{-1}(\alpha) + F^{-1}(\beta) \in I$, and when defined, then $\alpha*\beta = \beta*\alpha$, (ii) unity is neutral element, that is, $\alpha*1 = \alpha$; (iii) α has an inverse, $F\{-F^{-1}(\alpha)\}$, if $-F^{-1}(\alpha) \in I$; (iv) if both $(\alpha*\beta)*\gamma$ and $\alpha*(\beta*\gamma)$ are defined, then they are equal.

A compact restatement of the weighting problem is then: If D is an interval containing 1, we seek cell effects of the form $\alpha_i * \beta_j$ (with $\beta_c = 1$) where * represents a group or a pseudo-group on D, with 1 as neutral element (so that $\alpha_i * 1 = \alpha_i$), such that

$$\sum_{j=1}^{c} \hat{N}_{ij}\ \alpha_i*\beta_j = N_{i+}\ \ (i = 1, \ldots, r)\ \ ;\ \ \sum_{i=1}^{r} \hat{N}_{ij}\ \alpha_i*\beta_j = N_{+j}\ \ (j = 1, \ldots, c)$$

Cases of particular interest are:

1. The linear case $F(u) = 1 + u$ yields additive cell effects, $\alpha_i*\beta_j = \alpha_i + \beta_j - 1 = 1 + a_i + b_j$, which are not necessarily positive. The calibration equations (6.3) and (6.4) that result from this case were presented in Deming and Stephan (1940).

2. The exponential case $F(u) = \exp(u)$ gives positive, multiplicative cell effects, $\alpha_i*\beta_j = \alpha_i\beta_j = \exp(a_i+b_j)$. The solution to (6.3) and (6.4) in this case can be obtained by carrying out until convergence the classical raking ratio algorithm of Deming and Stephan (1940). (Practitioners often stop it after two iterations.) However, as pointed out in Huang (1976), they suggested the algorithm apparently thinking it converges to the solution for the linear case, for which they had presented the equations. This was later noted by Deming (1943).

3. Other solutions have been suggested. Smith (1947) gave a method corresponding to our Case

4, for which $\alpha_i*\beta_j = \alpha_i\beta_j/(\alpha_i + \beta_j - \alpha_i\beta_j) = (1 - a_i - b_j)^{-1}$. To each function $F(\cdot)$ in Table 1 corresponds a cell effect representation $\alpha_i*\beta_j$. For Cases 3 and 5, they are, respectively,

$$\alpha_i*\beta_j = \alpha_i\beta_j/(\sqrt{\alpha_i} + \sqrt{\beta_j} - \sqrt{\alpha_i\beta_j})^2 \; ; \quad \alpha_i*\beta_j = \alpha_i\beta_j/(\alpha_i^2 + \beta_j^2 - \alpha_i^2\beta_j^2)^{-1/2} .$$

7. A parametrization in terms of additive finite population effects

Calibration on known marginals is often almost as efficient as calibration on known population cell counts. That is, the variance of $\hat{t}_{ymarg}$ is often just slightly greater than that of $\hat{t}_{ypos}$. It is illustrative to go a conditional route to show this. Let us condition on the vector of cell count estimates, $\hat{\mathbf{N}} = (\hat{N}_{11}, \hat{N}_{12}, ..., \hat{N}_{ij}, ..., \hat{N}_{rc})'$, where $\hat{N}_{ij} = \Sigma_{s\,ij}\, d_k$. We need expressions for the two estimators that will facilitate an analysis of their conditional bias and conditional variance.

We associate with $\hat{t}_{ymarg}$ a parameterization of the finite population obtained by a two-way additive effects ANOVA model saying that, for elements k in the population cell U_{ij}, $y_k = A_i + B_j + E_k$, where the A_i and B_j are fixed unknown finite population parameters and E_k is a residual. With $\mathbf{x}_k$ defined by (6.2) and $\mathbf{B} = (A_1, ..., A_r, B_1, ..., B_c)$, we have $\mathbf{x}_k'\mathbf{B} = A_i + B_j$. Since all $q_k = 1$, (4.3) takes the form $SS_U = \Sigma_U E_k^2$, so the normal equations (4.2) are

$$\Sigma_{j=1}^{c} N_{ij} (A_i + B_j) = \Sigma_{U_{i+}} y_k = N_{i+}\overline{y}_{i+} \quad (i = 1, ..., r), \tag{7.1}$$

$$\Sigma_{i=1}^{r} N_{ij} (A_i + B_j) = \Sigma_{U_{+j}} y_k = N_{+j}\overline{y}_{+j} \quad (j = 1, ..., c), \tag{7.2}$$

where $U_{i+} = \cup_{j=1}^{c} U_{ij}$; $U_{+j} = \cup_{i=1}^{r} U_{ij}$. In the following $A_1, ..., A_r, B_1, ..., B_c$ denote the unique solution of (7.1) and (7.2) obtained after fixing arbitrarily a value for one of the $r + c$

components of $\mathbf{B}$; we take $B_c = 0$. Note that $\mathbf{x}_k'\mathbf{B} = A_i + B_j$ is invariant to the fixing of one component. Here, $A_1, ..., A_r, B_1, ..., B_{c-1}$ and $B_c = 0$ define fixed additive effects particular to the finite population at hand. The approximate unconditional variance of $\hat{t}_{ymarg}$ is now given by (4.1) with $E_k = y_k - (A_i + B_j)$ when k is in cell ij. (To estimate the variance, the unknown A_i and B_j must first be estimated; see Section 9.) We now derive an expression for the error $\hat{t}_{ymarg} - t_y$ needed to establish the conditional properties: Write $E_k = L_{ij} + R_k$, where $L_{ij} = \overline{y}_{ij} - (A_i + B_j)$ represents lack of additive model fit (or interaction), and $R_k = y_k - \overline{y}_{ij}$, where $\overline{y}_{ij} = \Sigma_{U_{ij}} y_k/N_{ij}$. The typical observation is then

y_k = additive model prediction + lack of additive model fit + deviation from cell mean,

or, equivalently,

$$y_k = (A_i + B_j) + L_{ij} + R_k ,$$

for $k \in U_{ij}$; $i = 1, ..., r$; $j = 1, ..., c$. From (6.7) we now have

$$\hat{t}_{ymarg} = \Sigma\Sigma_{i,j}\ \hat{N}_{ij}^{w}(A_i + B_j + L_{ij} + \tilde{R}_{sij}) , \qquad (7.3)$$

where $\hat{N}_{ij}^{w}$ is defined by (6.5), and

$$\tilde{R}_{sij} = (\Sigma_{s\,ij}\, R_k/\pi_k)/\hat{N}_{ij} , \qquad (7.4)$$

We need a matching expression for t_y. Multiply (6.3) by A_i and sum over i; multiply (6.4) by B_j and sum over j, to obtain

$$\Sigma\Sigma_{i,j}\, \hat{N}_{ij}^{w} A_i = \Sigma_{i=1}^{r}\, N_{i+}\, A_i ; \quad \Sigma\Sigma_{i,j}\, \hat{N}_{ij}^{w} B_j = \Sigma_{j=1}^{c}\, N_{+j}\, B_j .$$

These two equations, together with (7.1), yield

$$t_y = \Sigma_U y_k = \Sigma\Sigma_{i,j} N_{ij} (A_i + B_j) = \Sigma\Sigma_{i,j} \hat{N}^w_{ij} (A_i + B_j) . \qquad (7.5)$$

From (7.3) and (7.5), the error of $\hat{t}_{ymarg}$ is then

$$\hat{t}_{ymarg} - t_y = \Sigma\Sigma_{i,j} \hat{N}^w_{ij} L_{ij} + \Sigma\Sigma_{i,j} \hat{N}^w_{ij} \tilde{R}s_{ij} . \qquad (7.6)$$

For the complete poststratification estimator (6.1), the easily derived counterpart to (7.6) is

$$\hat{t}_{ypos} - t_y = \Sigma\Sigma_{i,j} N_{ij} \tilde{R}s_{ij} . \qquad (7.7)$$

8. Conditional properties of t_{ymarg} and $\hat{t}_{ypos}$

Conditional properties for raking ratio estimators, especially for SRS, are examined, for example, in Oh and Scheuren (1987) and other papers by the same authors. We start from (7.6) and (7.7) and analyze $\hat{t}_{ymarg}$ and t_{ypos}, conditionally on $\hat{N} = (\hat{N}_{11}, ..., \hat{N}_{ij}, ..., \hat{N}_{rc})'$, where $\hat{N}_{ij}$ $= \Sigma_{s_{ij}} d_k$. The objective is to find the conditional bias (c-bias) and the conditional variance (c-variance). They determine the unconditional variance through the relation expected c-variance plus variance of the c-bias. The index C is used to indicate conditional mean and variance operators, that is, $E_C(\cdot) = E(\cdot \mid \hat{N})$ and $V_C(\cdot) = V(\cdot \mid \hat{N})$. If $\hat{t}_y$ is an estimator of t_y, its c-bias is $B_C(\hat{t}_y) = E_C(\hat{t}_y) - t_y$, and its c-variance is $V_C(\hat{t}_y)$. Our conditional analysis is simple, since the $\hat{N}^w_{ij} = \hat{N}_{ij} F(a_i + b_j)$ in (7.6) are fixed, given $\hat{N}$. This is because the a_i and b_j are solutions to (6.4) and (6.5), in which all quantities are fixed once the $\hat{N}_{ij}$ have been fixed. From (7.6),

$$B_C(\hat{t}_{ymarg}) = \Sigma\Sigma_{i,j} \hat{N}^w_{ij} L_{ij} + \Sigma\Sigma_{i,j} \hat{N}^w_{ij} E_C(\tilde{R}s_{ij}); \quad V_C(\hat{t}_{ymarg}) = V_C(\Sigma\Sigma_{i,j} \hat{N}^w_{ij} \tilde{R}s_{ij}) . \quad (8.1)$$

The corresponding expressions for the complete poststratification estimator follow from (7.7):

$$B_C(\hat{t}_{ypos}) = \Sigma\Sigma_{i,j} N_{ij} E_C(\tilde{R}s_{ij}) ; \qquad V_C(\hat{t}_{ypos}) = V_C(\Sigma\Sigma_{i,j} N_{ij} \tilde{R}s_{ij}) . \qquad (8.2)$$

The approximation $\hat{N}_{ij}^{w} \simeq N_{ij}$ is justified in large samples, so then

$$B_C(\hat{t}_{ymarg}) - B_C(\hat{t}_{ypos}) \simeq \Sigma\Sigma_{i,j}\, \hat{N}_{ij}^{w}\, L_{ij} \quad ; \qquad (8.3)$$

$$V_C(\hat{t}_{ymarg}) - V_C(\hat{t}_{ypos}) \simeq 0 \ .$$

The two estimators differ mainly in regard to their c-bias. Now, $\hat{t}_{ypos}$ is c-unbiased if $E_C(\tilde{R}s_{ij}) = 0$ for all i,j, as for SRS. Under certain other designs, $\hat{t}_{ypos}$ is approximately c-unbiased. In any case, the c-bias of $\hat{t}_{ymarg}$ contains $\Sigma\Sigma_{i,j}\, \hat{N}_{ij}^{w}\, L_{ij}$, which is nonzero when there is interaction; compare Oh and Scheuren (1987) in the special case SRS.

The unconditional variance is $V(\cdot) = EB_C(\cdot) + VE_C(\cdot)$, where E and V are with respect to the distribution of $\hat{N}$. Now, $V(\hat{t}_{ymarg})$ will ordinarily exceed $V(\hat{t}_{ypos})$, since the c-bias of the former contains $\Sigma\Sigma_{i,j}\, \hat{N}_{ij}^{w}\, L_{ij}$. But if this interaction term is near zero, then $\hat{t}_{ymarg}$ and $\hat{t}_{ypos}$ have essentially the same unconditional variance. In practice, this is often the case.

9. Variance estimation for generalized raking

Although raking ratio has a long history, the variance of the resulting estimators have been difficult to work out even approximately. For specific designs, Brackstone and Rao (1979), Konijn (1981), Choudhry and Lee (1987) derive formulas for the variance arising for the classical raking ratio algorithm (our exponential case) stopped after a few iteration steps. Bankier (1986) suggests an approach involving repeated linearization techniques. A fairly complex variance estimator is proposed by Binder and Théberge (1988). In the work of Bethlehem and Keller (1987), a variance estimator is implicit; although the formula is not given, it probably resembles the one we get for the linear case.

As pointed out, the variance of $\hat{t}_{ymarg}$ is given by (4.1), which applies with $E_k = y_k - (A_i + B_j)$ for any design and any function $F(\cdot)$. The A_i and B_j verify the normal equations (7.2) and (7.3). The variance estimator is obtained from (4.5), replacing first the unknown A_i and B_j by estimates. These are derived from the sample-based normal equations (4.6). They take the form

$$\Sigma_{j=1}^{c} \hat{N}_{ij}^{w} (\hat{A}_i + \hat{B}_j) = \Sigma_{s\,i+} w_k y_k = \Sigma_{j=1}^{c} \hat{N}_{ij}^{w} \tilde{y}_{s_{ij}} \quad (i = 1, \ldots, r),$$

$$\Sigma_{i=1}^{r} \hat{N}_{ij}^{w} (\hat{A}_i + \hat{B}_j) = \Sigma_{s\,+j} w_k y_k = \Sigma_{i=1}^{r} \hat{N}_{ij}^{w} \tilde{y}_{s_{ij}} \quad (j = 1, \ldots, c).$$

These resemble the calibration equations (6.3) and (6.4), so the same computer routine can be used to solve for $\hat{A}_i$, $\hat{B}_j$ (after fixing $\hat{B}_c = 0$). The resulting sample-based residuals are

$$e_k = y_k - (\hat{A}_i + \hat{B}_j) \tag{9.1}$$

With these e_k, the variance estimator $\hat{V}(\hat{t}_{ymarg})$ is easily calculated from (4.5). We can also write $e_k = \hat{L}_{ij} + \hat{R}_k$, with $\hat{L}_{ij} = \tilde{y}_{s_{ij}} - (\hat{A}_i + \hat{B}_j)$ and $\hat{R}_k = y_k - \tilde{y}_{s_{ij}}$ for k in cell ij.

In the case of SRS, the expressions are easy to interpret. To condition on $\hat{N}$ is then equivalent to conditioning on the sample cell counts, $\mathbf{n} = (n_{11}, n_{12}, \ldots, n_{ij}, \ldots, n_{rc})'$, since $\hat{N}_{ij} = f^{-1} n_{ij}$ under SRS with $f = n/N$. The weights are $w_k = \hat{N}_{ij}^{w}/n_{ij}$ for all k in s_{ij}. From (4.5) and (9.1),

$$\hat{V}(\hat{t}_{ymarg}) = \{n/(n-1)\} (1 - f) \Sigma_s w_k^2 e_k^2 = \hat{V}_1 + \hat{V}_2$$

where

$$\hat{V}_1 = \frac{n}{n-1} (1 - f) \Sigma\Sigma_{i,j} (\hat{N}_{ij}^{w})^2 \frac{\hat{L}_{ij}^2}{n_{ij}}; \quad \hat{V}_2 = \frac{n}{n-1} (1 - f) \Sigma\Sigma_{i,j} (\hat{N}_{ij}^{w})^2 \frac{S_{s_{ij}}^2}{n_{ij}},$$

with $S_{s_{ij}}^2 = \Sigma_{s\,ij} (y_k - \overline{y}_{s_{ij}})^2/n_{ij}$. Here, $\hat{V}_2$ estimates the c-variance in (8.1), namely,

$$V_C(\hat{t}_{ymarg}) = \Sigma\Sigma_{i,j} (\hat{N}^w_{ij})^2 (\frac{1}{n_{ij}} - \frac{1}{N_{ij}}) S^2_{ij} ,$$

with $S^2_{ij} = \Sigma_{U_{ij}} (y_k - \overline{y_{ij}})^2/(N_{ij}-1)$. The component $\hat{V}_1$ estimates the variance of the c-bias in (8.1), that is, $VB_C(\hat{t}_{ymarg}) = V(\Sigma\Sigma_{i,j} \hat{N}^w_{ij} L_{ij})$, where V is with respect to the distribution of **n**.

10. Computational aspects

To use the calibration estimator (3.6) we must first solve (3.5) for λ. Solution by Newton's method is first discussed; then techniques for calibration on known marginals are examined.

1. Solution of (3.5). Let $\phi'_s(\lambda) = \partial \phi_s(\lambda)/\partial\lambda$. Start with $\lambda_0 = \mathbf{0}$. Subsequent iterative values, λ_v, $v = 1, 2, ...$ are obtained by

$$\lambda_{v+1} = \lambda_v + \{\phi'_s(\lambda_v)\}^{-1}\{t_x - \hat{t}_{x\pi} - \phi_s(\lambda_v)\}. \qquad (10.1)$$

(Note: For cases where F^{-1} maps D onto an interval I of $\mathbb{R}$, one must check that $\mathbf{x}'_k\lambda_{v+1}$ really belongs to I. For instance, if $\mathbf{x}'_k\lambda_{v+1} \geq \sup I$, it is a good idea to replace λ_{v+1} by $\lambda'_{v+1} = \lambda_v + \theta_v(\lambda_{v+1} - \lambda_v)$ for some $\theta_v < 1$ such that λ'_{v+1} is near the border of the set of permissible values.)

From (3.4), $\phi_s(\mathbf{0}) = \mathbf{0}$; $\phi'_s(0) = \mathbf{T}_s$. The first iteration gives $\lambda_1 = \mathbf{T}_s^{-1}(t_x - \hat{t}_{x\pi})$; subsequent iterations, $v = 2, 3, ...$ obey (10.1) until convergence. Now, λ_1 is the vector (2.4) that yields the regression estimator (2.6). Thus (2.6) is a first approximation to (3.6); Result 5 shows them to be asymptotically equivalent. If $F(u) = 1 + u$, the iteration stops after the first step.

2. Solution of (6.3) and (6.4). We apply (10.1) to calibration on two known marginals. The

equations to solve are (6.3) and (6.4), and $\lambda = (a_1, ..., a_r, b_1, ..., b_c)$. Fix $b_c = 0$ throughout the iteration. This eliminates the last row and the last column of $\phi'_s(\lambda)$. A square r+c-1 matrix remains. Its elements, m_{ij}, are as follows, with $F'(u) = dF(u)/du$: $m_{i,i} = \sum_{i=1}^{r} \hat{N}_{ij} F'(a_i+b_j)$; $m_{r+j,r+j} = \sum_{j=1}^{c} \hat{N}_{ij} F'(a_i+b_j)$; $m_{i,r+j} = m_{r+j,i} = \hat{N}_{ij} F'(a_i+b_j)$; $i = 1, ..., r$; $j = 1, ..., c - 1$; all other off-diagonal elements are zero. Start with $\lambda_0 = \mathbf{0}$. The elements a_i, b_j of the next value, λ_1, are obtained by solving

$$\sum_{j=1}^{c} \hat{N}_{ij} (a_i + b_j) = N_{i+} - \hat{N}_{i+} \quad (i = 1, ..., r); \quad \sum_{i=1}^{r} \hat{N}_{ij} (a_i + b_j) = N_{+j} - \hat{N}_{+j} \quad (j = 1, ..., c-1).$$

These are final equations if $F(u) = 1 + u$; for other $F(u)$, iteration continues until convergence.

3. Alternative solution of (6.3) and (6.4). The system may be solved by noting that each equation (6.3) can be solved for a_i assuming the b_j are known, and conversely for (6.4). With b_C fixed at 0 throughout, follow the algorithm:

1. Set $b_j = 0$; $j = 1, ..., c-1$;

2. Obtain a set of a_i by solving one by one the r equations (6.3), with b_j from the preceding step;

3. Obtain a set of b_j by solving one by one the c-1 equations (6.4), with a_i from the preceding step;

4. Repeat steps 2 and 3 until convergence.

The procedure requires no matrix inversion and can be shown to converge to the proper solution, but slowly compared with Newton's method. Approximately 15 to 20 iterations may be required in cases for which 3 to 5 would typically suffice for Newton's method, with time of execution typically increased by a factor of 3 to 4.

In the linear case, $F(a_i + b_j) = 1 + a_i + b_j$, and (6.3) gives

$$a_i = (N_{i+} - \hat{N}_{i+} - \sum_{j=1}^{c} \hat{N}_{ij}\, b_j)/\hat{N}_{i+} \; ; \qquad (i = 1, \ldots, r) \, .$$

Equations of this form are wellknown from unbalanced two-way ANOVA. In the exponential case, $F(a_i+b_j) = \exp(a_i+b_j)$, and we get the raking ratio algorithm of Deming and Stephan (1940).

11. Concluding comments

Result 5 states that the calibration estimators obtained by different specifications of $F(\cdot)$ all have the same asymptotic variance. If asymptotic variance is the criterion, the theory in this paper does not designate one estimator as superior to the others. Confirming theory, we found in simulations with modest to fairly large samples that from the standpoint of variance alone, there is little to choose between the estimators $\hat{t}_{ymarg}$ corresponding to different $F(\cdot)$. Individual cells weights $\hat{N}_{ij}^{w}$ may change considerably from one specification $F(\cdot)$ to another, but there is little effect on $V(\hat{t}_{ymarg})$, which combines all cells. Readily observed features of the weights, such as their range, may become the overriding factor in the choice of $F(\cdot)$. The weights must make good sense to the user. For example, if certain cells are also domains of study, the negative weights that can occur in the linear case $F(u) = 1 + u$ is not an appealing prospect. Equally undesirable are the excessively large cell weights that can result in the exponential case $F(u) = \exp(u)$ corresponding to classical raking ratio. Therefore, functions $F(\cdot)$ that give weights bounded from above and below are attractive alternatives. Cases 6 and 7 are of this kind. They allow the practitioner to try several specifications of (L, U) and settle on one that gives suitable weights. On the other hand, there is a slight chance that Cases 6 and 7 yield no solution. An extension of our calibration weighting is to also include the reweighting for estimation done in the presence of nonresponse. The classical raking ratio seems promising in this regard; see Binder and Théberge (1987).

REFERENCES

Bankier, M.D. (1986). Estimators based on several stratified samples with applications to multiple frame surveys. Journal of the American Statistical Association **81**, 1074-1079.

Bethlehem, J.G. and Keller, W.J. (1987). Linear weighting of sample survey data. Journal of Official Statistics **3**, 141-153.

Brackstone, G.J. and Rao, J.N.K. (1979). An investigation of raking ratio estimators. Sankhya Ser. C **41**, 97-114.

Binder, D.A. and Théberge, A. (1988). Estimating the variance of raking-ratio estimators. Canadian Journal of Statistics **16** supplement, 47-55.

Cassel, C.M., Särndal, C.E. and Wretman, J.H. (1976). Some results on generalized difference estimation and generalized regression estimation for finite populations. Biometrika **63**, 615-620.

Choudhry, G.H. and Lee, H. (1987). Variance estimation for the Canadaian Labour Force Survey. Survey Methodology **13**, 147-161.

Deming, W.E. (1943). Statistical Adjustment of Data. New York: Wiley.

Deming, W. E. and Stephan, F.F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Annals of Mathematical Statistics **11**, 427-444.

Deville, J.-C. (1988). Estimation linéaire et redressement sur informations auxiliaires d'enquêtes par sondage. In Essais en l'honneur d'Edmond Malinvaud, edited by A. Monfort and J.J. Laffond. Paris: Economica, 915-927.

Fuller, W.A. and Isaki, C.T. (1981). Survey design under superpopulation models. In Current Topics in Survey Sampling, edited by D. Krewski, J.N.K. Rao and R. Platek. New York: Academic Press.

Huang, E. (1976). Regression estimation of means and totals for sample survey data. Ph.D. Thesis, Iowa State University.

Konijn, H.S. (1981). Biases, variances, and covariances of raking ratio estimators for marginal and cell totals and averages of observed characteristics. Metrika 28, 109-121.

Lemel, Y. (1976). Une généralisation de la méthode du quotient pour le redressement des enquêtes par sondage. Annales de l'INSEE , n° 22-23, 272-282.

Gourieroux, C. (1981). Théorie des sondages. Paris: Economica.

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

Lemaître, G. and Dufour, J. (1987). An integrated method for weighting persons and families. Survey Methodology 13, 199-207.

Oh, H.L. and Scheuren, F. (1987). Modified raking ratio estimation. Survey Methodology 13, 209-219.

Särndal, C. E. (1980). On π-inverse weighting versus best linear weighting in probability sampling. Biometrika 67, 639-650.

Särndal, C. E. (1982). Implications of survey design for generalized regression estimation of linear functions. Journal of Statistical Planning and Inference 7, 155-170.

Särndal, C.E., Swensson, B. and Wretman, J.H. (1989). The weighted residual technique for estimating the variance of the general regression estimator of the finite population total. Biometrika 76, 527-537.

Smith, J. H. (1947). Estimation of linear function of cell proportions. Annals of Mathematical Statistics 18, 231-254.

Wright, R.L. (1983). Finite population sampling with multivariate auxiliary information. Journal of the American Statistical Association 78, 879-884.

Zieschang, K.D. (1986). A generalized least squares weighting system for the consumer expenditure survey. Section on Survey Research Methods, American Statistical Association, 64-71.

APPENDIX

Proofs of Results 1, 2 and 3.

1. Mathematical preliminaries

1.1. The function ϕ and its properties

Let $C_n = \cap \{\lambda : x_k'\lambda \in \mathrm{Im}_k(d_k)\}$, where $\cap$ is over $k \in U_n$, the finite population associated with the (expected) sample size n. The interior C_n^0 of C_n is an open convex set containing 0 for every n. Moreover, $C^* = \cap_{n=1}^{\infty} C_n^0$ is convex; we assume that it is also open. Let E_n and P_n denote expectation and probability, respectively, with respect to the sampling design indexed by n. For $\lambda \in C^*$, $N^{-1} E_n\{\phi_s(\lambda)\}$ is a well defined continuously differentiable function. By our assumptions, it converges to a fixed function denoted ϕ. Convergence is uniform on every compact set in C^*. Note the properties

$$N^{-1}\phi_s(0) = 0\,; \quad \phi(0) = 0\ ,$$

$$N^{-1}\phi_s'(0) = N^{-1}T_s\,; \quad \phi'(0) = T = \lim\ N^{-1}\Sigma_U\, x_k x_k'\ .$$

Now, for every λ, ϕ' is a positive definite matrix because all F_k are increasing functions. As a consequence, ϕ is injective and maps C^* onto an open neighbourhood of 0 in $\mathbb{R}^J$. Let B be a closed sphere with radius r contained in that neighbourhood, and let A be the compact set $\phi^{-1}(B)$. The inverse function ϕ^{-1} is defined on B, continuous, and continuously differentiable. Then $\| \phi^{-1}(x) \|$ is continuously differentiable and bounded on B. Let $K = \max_{x \in B} \| (\phi^{-1})'(x) \|$.

1.2. Properties of $N^{-1}\,\phi_s'(\lambda)$

All functions $N^{-1}\,\phi_s'(\lambda)$ are defined on C^* and therefore on A. For a continuous ψ defined on C^*, let $\|\psi\|_M = \sup_{\lambda\in M}\|\psi(\lambda)\|$ for M compact in C^*. By our general properties of convergence we have for every $\varepsilon > 0$ that $P_n(\|N^{-1}\phi_s - \phi\|_A < \varepsilon) \longrightarrow 1$ when n increases.

Now let $\phi_i = N^{-1}\phi_s$ for some function verifying $\|\phi_1 - \phi\|_A \leq \beta r$; $\|\phi_1' - \phi'\|_A \leq \beta K$, with $0 < \beta < 1/2$. The probability of this event tends to 1 as n increases. Let $r_1 = (1-\beta)\,r$, let B_1 be the sphere $\|x\| < r_1$ in $\mathbb{R}^J$. Now, ϕ_1 maps the frontier of A onto the crown $r_1 \leq \|x\| \leq r(1+\beta)$ and $\phi_1(A)$ is a bordered manifold homotopic to B. These notions are discussed in Trenoguine (1987). A consequence is that $\phi_1(A)$ covers the sphere B_1, and, in other words, that for every $x \in B_1$, the equation $\phi_1(\lambda) = x$ has a (unique) solution. Moreover, ϕ_1^{-1}, defined on B_1, is a continuously differentiable function. Since $\|\phi_1' - \phi'\| \leq \beta K$ for every λ in C, $(\phi_1^{-1})'(x)$ exists for every $x \in B_1$, and $\|\phi_1^{-1}(x)\| \leq \|x\|\,K\,(1-\beta)^{-1}$.

2. Proofs of the three results

Result 1. First, $N^{-1}(\hat{t}_{x\pi} - t_x) = z$ belongs to B_1 with a probability tending to 1. Secondly, $N^{-1}\phi_s$ has an inverse function on B_1 with probability tending to 1. As (3.5) can be written

$$N^{-1}(\hat{t}_{x\pi} - t_x) = N^{-1}\,\phi_s(\lambda),$$

the equation thus has a unique solution with probability tending to one. □

Result 2. Let $\lambda_s = (N^{-1}\,\phi_s')^{-1}(z)$ if z belongs to B_1; otherwise, λ_s is arbitrarily defined.

Since $\phi_s(0) = 0$, we have

$$\lambda_s - 0 = (N^{-1}\,\phi_s')^{-1}(z) - (N^{-1}\,\phi_s')^{-1}(0)$$

and $\| \lambda_s \| \leq \| z \| K (1 - \beta)^{-1}$. This inequality holds with probability tending to one when n increases. But $z = O_p(n^{-1/2})$, so there exists a constant K' such that $P_n(\| z \| \leq K' n^{-1/2}) \longrightarrow 1$. Combining the two inequalities, $P_n(\| \lambda_s \| \leq K K' (1 - \beta)^{-1} n^{-1/2}) \longrightarrow 1$, which implies, by definition, that $\lambda_s = O_p(n^{-1/2})$. □

Result 3. Let $\theta_k(u) = F_k(u) - 1 - q_k u$. We assume that $\theta_k(u) = O(u^2)$ holds uniformly, which is equivalent to our assumption that $F_k''(0)$ is uniformly bounded. Thus, $\theta(u) = \max \theta_k(u) = O(u^2)$. Otherwise, for any $\varepsilon > 0$, there exists K'' such that, for all k, $|u| < \varepsilon$ will imply that $\theta_k(u) \leq K'' u^2$. We can write (3.5) as $\hat{t}_{x\pi} - t_x = \Sigma_s d_k x_k \{q_k x_k' \lambda_s + \theta_k(x_k' \lambda_s)\}$, and therefore

$$\lambda_s - T_s^{-1}(\hat{t}_{x\pi} - t_x) = T_s^{-1} \Sigma_s d_k x_k \theta_k(x_k' \lambda_s).$$

For λ_s sufficiently small,

$$\| \lambda_s - T_s^{-1}(\hat{t}_{x\pi} - t_x) \| \leq \| (N^{-1} T_s)^{-1} \| K'' \{N^{-1} \Sigma_s d_k \| x_k \|^3\} \| \lambda_s \|^2 .$$

Here, $\| (N^{-1} \mathbf{T}_s)^{-1} \| = 0_p(1)$, $N^{-1}\Sigma_s d_k \| \mathbf{x}_k \|^3 = 0_p(1)$, and, by Result 2, $\| \lambda_s \|^2 = 0_p(n^{-1})$.

Result 3 follows. □

REFERENCE FOR THE APPENDIX

Trenoguine, V. (1987) Analyse Fonctionnelle. Moscow: Editions M.I.R.