Calibration Estimators in Survey Sampling

JEAN-CLAUDE DEVILLE and CARL-ERIK SÄRNDAL*

This article investigates estimation of finite population totals in the presence of univariate or multivariate auxiliary information. Estimation is equivalent to attaching weights to the survey data. We focus attention on the several weighting systems that can be associated with a given amount of auxiliary information and derive a weighting system with the aid of a distance measure and a set of calibration equations. We briefly mention an application to the case in which the information consists of known marginal counts in a two- or multi-way table, known as *generalized raking*. The general regression estimator (GREG) was conceived with multivariate auxiliary information in mind. Ordinarily, this estimator is justified by a regression relationship between the study variable y and the auxiliary vector $\mathbf{x}$. But we note that the GREG can be derived by a different route by focusing instead on the weights. The ordinary sampling weights of the kth observation is $1/\pi_k$, 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 $1/\pi_k$ 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. That is a consistency check that appeals to many practitioners, because a strong correlation between the auxiliary variables and the study variable means that the weights that perform well for the auxiliary variable also should perform well for the study variable. The GREG uses the auxiliary information efficiently, so the estimates are precise; however, the individual weights are not always without reproach. For example, negative weights can occur, and in some applications this does not make 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 specific weighting system and thereby to a new estimator. These estimators form a family of *calibration estimators*. We show that the GREG is a first approximation to all other members of the family, all are asymptotically equivalent to the GREG, and the variance estimator already known for the GREG is recommended for use in any other member of the family. Numerical features of the weights and ease of computation become more than anything else the bases for choosing between the estimators. The reasoning is applied to calibration on known marginals of a two-way frequency table. Our family of distance measures leads in this case to a family of *generalized raking procedures*, of which classical raking ratio is one.

KEY WORDS Calibration, Multivariate auxiliary information, Raking, Regression estimators

Survey statisticians use auxiliary information in many ways to improve survey estimates. For example, using the general regression estimator for a finite population total or mean requires a vector of auxiliary variables for which the population total is known. The calibration estimators derived in this article are a family of estimators appealing a common base of auxiliary information. A calibration estimator uses calibrated weights, which are as close as possible, according to a given distance measure, to the original sampling design weights π_k^{-1} while also respecting a set of constraints, the calibration equations. For every distance measure there is a corresponding set of calibrated weights and a calibration estimator. In Section 2 we define a family of distance measures and derive the corresponding family of calibration estimators, then establish their properties in a series of results. Variance estimators for calibration estimators are given in Section 3. An important application of these ideas, mentioned in Section 4, is the calibration on known marginal counts in two-way or multiway tables, which leads to generalized raking.

1. DERIVING THE GENERAL REGRESSION ESTIMATOR BY CALIBRATION

Consider a finite population $U = \{1, \ldots, k, \ldots, N\}$, 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 $\pi_{kl} = \Pr(k \,\&\, l \in s)$ are assumed to be strictly positive. Let y_k be the value of the variable of interest, y, for the kth population element, with which also is associated an auxiliary vector value, $\mathbf{x}_k = (x_{k1}, \ldots, x_{kj}, \ldots, x_{kJ})'$. 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 to be accurately known. This knowledge may come from one or more sources, such as census data, administrative data files, and others. If A ($A \subseteq U$) is any set of population elements, $\sum_A$ is our shorthand for $\sum_{k \in A}$ (e.g., $\sum_s y_k$ 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 Horvitz–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 as close as possible, in an average sense for a given metric, to the d_k while respecting the calibration equation

$$\sum_s w_k \mathbf{x}_k = \mathbf{t}_x. \tag{1.1}$$

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 adjusting the sample weights d_k is discussed in the context of the U.S. Consumer Expenditure Survey by Zieschang (1986, 1990), who considered "weighting control procedures" through a generalized least squares weighting algorithm. The work of Bankier (1990) is also related. If $E_p(\cdot)$ denotes expectation with respect to the sampling design

* Jean-Claude Deville is Head of the Statistical Methology and Sampling Division, Institut National de la Statistique et des Études Économiques, 18 Boulevard Adolphe Pinard, 75675 Paris Cedex 14, France. Carl-Erik Sarndal is Professor, Département de Mathématiques et de Statistique, Université de Montréal, C P 6128, Succursale A, Montréal, Québec H3C3J7, Canada. This work was supported in part by the Natural Sciences and Engineering Research Council of Canada

$p(s)$, a measure of average distance reminiscent of the chi-square statistic is $E_p\{\sum_s (w_k - d_k)^2/d_k\}$. For more generality in this expression, we can let the kth term have an individual, known positive weight $1/q_k$, unrelated to d_k, which gives the average distance

$$E_p\left\{\sum_s (w_k - d_k)^2/d_k q_k\right\}. \tag{1.2}$$

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, which follows. Our objective is to derive new weights that modify as little as possible the original sampling weights $d_k = \pi_k^{-1}$, which have the desirable property of yielding unbiased estimates; the survey statistician wants to stay close to these weights. We thus seek the minimum of (1.2) subject to (1.1) holding for every sample s. This is equivalent to minimizing, for any particular s, the quantity $\sum_s (w_k - d_k)^2/d_k q_k = \sum_s d_k(w_k/d_k - 1)^2/q_k$, subject to the single constraint (1.1). In other words, we should minimize the conditional value of the distance, given the realized sample s. Nothing says that the new weights will continue to give unbiased estimates, but a realistic expectation is to remain near unbiasedness. Minimization leads to the calibrated weight

$$w_k = d_k(1 + q_k \mathbf{x}_k'\lambda), \tag{1.3}$$

where the vector of Lagrange multipliers λ is determined from (1.1), that is,

$$\lambda = \mathbf{T}_s^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}), \tag{1.4}$$

assuming that the inverse of

$$\mathbf{T}_s = \sum_s d_k q_k \mathbf{x}_k \mathbf{x}_k'. \tag{1.5}$$

exists. The resulting estimator of t_y is

$$\hat{\mathbf{t}}_{y\text{reg}} = \sum_s w_k y_k = \hat{t}_{y\pi} + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})'\hat{\mathbf{B}}_s, \tag{1.6}$$

where $\hat{\mathbf{t}}_{x\pi} = \sum_s d_k \mathbf{x}_k$ denotes the Horvitz–Thompson estimator for the $\mathbf{x}$-vector and

$$\hat{\mathbf{B}}_s = \mathbf{T}_s^{-1} \sum_s d_k q_k \mathbf{x}_k y_k \tag{1.7}$$

is a weighted estimator of the multiple regression coefficient. Thus, Deville's (1988) calibration technique achieves two things: (1) it provides an alternative derivation of the generalized regression estimator (Cassel, Sarndal, and Wretman 1976; Gourieroux 1981; Särndal 1980; Isaki and Fuller 1982, Wright 1983) and (2) it emphasizes that it is fruitful to view (1.6) as a linear weighting of the observations y_k with weights w_k that are sample-dependent and given by (1.3). This view is also held by Särndal (1982), who used the w_k to create a suitable variance estimator for $\hat{t}_{y\text{reg}}$ (see Section 4), and also by Bethlehem and Keller (1987) and Lemaître and Dufour (1987). The research question addressed in this article is whether useful alternative estimators will result by allowing other distance measures.

Example 1. Derivation of the ratio estimator. Take $\mathbf{x}_k = x_k$, a positive scalar. Then $\mathbf{x}_k'\lambda = x_k\lambda$. Let us take $q_k = 1/x_k$. We obtain $\lambda = (\sum_U x_k)/(\sum_s d_k x_k) - 1 = t_x/\hat{t}_{x\pi} - 1$, 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 (1.6) $\hat{t}_{y\text{reg}} = t_x \hat{t}_{y\pi}/\hat{t}_{x\pi}$, the well-known ratio estimator. Note that the unequal weighting $q_k = 1/x_k$ is essential for obtaining this result.

2. A CLASS OF DISTANCE MEASURES

In (1.2), the distance between the original weight d_k and the new weight w_k was rather arbitrarily taken as $(w_k - d_k)^2/d_k q_k$. It is natural to consider alternative distance measures. These measures should share a few basic features that are easy to accept. For element k, we 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 $G_k(d, d) = 0$; and (2) $g_k(w, d) = \partial G_k(w, d)/\partial w$ is continuous and maps $D_k(d)$ onto an interval $\text{Im}_k(d)$ 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 then measured by $E_p\{\sum_s G_k(w_k, d_k)\}$. To minimize this quantity subject to (1.1) 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 (1.1). If λ denotes a vector of Lagrange multipliers, derivation gives

$$g_k(w_k, d_k) - \mathbf{x}_k'\lambda = 0. \tag{2.1}$$

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

$$w_k = d_k F_k(\mathbf{x}_k'\lambda), \tag{2.2}$$

where $d_k F_k(\cdot)$ is the reciprocal mapping of $g_k(\cdot, d_k)$ that maps $\text{Im}_k(d_k)$ onto $D_k(d_k)$ in an increasing fashion. From our assumptions, $F_k(0) = 1$ and $F_k'(0) > 0$. The important quantity $F_k'(0)$ plays the same role as q_k in (1.2), so we use the notation $F_k'(0) = q_k$.

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 $g(1) = 0$ and $g'(1) = 1$. Examples are found in Table 1. 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 $g(\cdot)$, (2.2) becomes

$$w_k = d_k F(q_k \mathbf{x}_k'\lambda). \tag{2.2a}$$

From (1.1), the calibration equations necessary to determine $\lambda = (\lambda_1, \ldots, \lambda_j, \ldots, \lambda_J)'$ are

$$\mathbf{t}_x = \sum_s w_k \mathbf{x}_k = \sum_s d_k F_k(\mathbf{x}_k'\lambda)\mathbf{x}_k. \tag{2.3}$$

It is convenient to define

$$\phi_s(\lambda) = \sum_s d_k\{F_k(\mathbf{x}_k'\lambda) - 1\}\mathbf{x}_k, \tag{2.4}$$

whereby (2.3) can be written as

$$\phi_s(\lambda) = \mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}. \tag{2.5}$$

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	$q_kG_k(w_k, d_k)$	$g(w_k/d_k) = q_kg_k(w_k, d_k)$	$F_k(u) = F(q_ku)$
1	$(w_k - d_k)^2/2d_k$	$w_k/d_k - 1$	$1 + q_ku$
2	$w_k\log(w_k/d_k) - w_k + d_k$	$\log(w_k/d_k)$	$\exp(q_ku)$
3	$2(\sqrt{w_k} - \sqrt{d_k})^2$	$2\{1 - (w_k/d_k)^{-1/2}\}$	$(1 - q_ku/2)^{-2}$
4	$-d_k\log(w_k/d_k) + w_k - d_k$	$1 - (w_k/d_k)^{-1}$	$(1 - q_ku)^{-1}$
5	$(w_k - d_k)^2/2w_k$	$\{1 - (w_k/d_k)^{-2}\}/2$	$(1 - 2q_ku)^{-1/2}$

NOTE The functions are normalized so that $F_k'(0) = q_k$ in all cases

The right side is a known quantity for every sample s. To summarize, the steps of the procedure are:

1. Given the data for the realized sample s and for the chosen $F_k(\cdot)$, solve (2.5) for λ. Iteration may be required.
2. Once λ is determined obtain the resulting *calibration estimator* of t_y as

$$\hat{t}_{yw} = \sum_s w_k y_k = \sum_s d_k F_k(\mathbf{x}_k'\lambda) y_k. \tag{2.6}$$

This estimator will give close estimates of $t = \sum_U y_k$ if there is a strong relationship between y and $\mathbf{x}$. To see this, suppose that $y_k = \mathbf{x}_k'\alpha$ for all k and some constant vector α; that is, y is perfectly explained by $\mathbf{x}$. Then, from (2.3), $\hat{t}_{yw} = t_y$ for every sample, so the variance is nil.

The statistician chooses the distance function $G_k(w_k, d_k)$. Or, equivalently, the uniquely corresponding function $F_k(u) = F_k(\mathbf{x}_k'\lambda)$. Examples of the form $g_k(w, d) = g(w/d)/q_k$ are shown in Table 1, where the functions are normalized to obtain $F_k(0) = 1$ and $F_k'(0) = q_k$. Because $1/q_k$ is a recurring multiplicative factor, the table shows $q_kG_k(w_k, d_k)$ and $q_kg_k(w_k, d_k) = g(w_k/d_k)$.

The cases in Table 1 correspond to well-known distance measures; for example, Hellinger distance in Case 3, and minimum entropy distance in Case 4. In Cases 1, 3, 4, and 5, $F_k(u)$ is of the form $(1 + \alpha q_ku)^{1/\alpha}$, with $\alpha = 1, -1/2, -1, -2$, respectively; Case 2 is obtained when $\alpha \to 0$. What are the relative merits of these cases? The existence of a solution of (2.5) is one aspect that needs to be considered. Cases 1 and 2 always lead to a solution; in Cases 3, 4, and 5, a solution is not guaranteed, but Result 1 on page 379 shows that the probability of a solution tends to 1. More important perhaps is the range of values that the weights $w_k = d_kF(q_k\mathbf{x}_k'\lambda)$ can take. In Case 1, which yields the regression estimator (1.6), the weights can be positive or negative; Cases 2, 3, 4, and 5 guarantee positive weights. In each case, some unrealistic or extreme weights w_k may occur for rare or "unlucky" samples. That Case 1 can yield negative weights w_k may be unacceptable to some users. Case 2 may yield some weights w_k that are extremely large compared to the basic sampling weights $d_k = \pi_k^{-1}$; again, the user may find this unacceptable. One may want to avoid a function $F_k(u)$ that can give overly extreme weights, because applying these weights to make estimates for various subpopulations (domains) may produce unrealistic estimates for some domains. We therefore consider a few additional functions that have the attractive property of yielding weights restricted to an interval that the statistician can specify in advance. Extreme weights can be eliminated, but the resulting estimators retain their favorable properties.

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

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

We then have $F_k(-\infty) = L$; $F_k(\infty) = U$; $F_k(0) = 1$, $F_k'(0) = q_k$. It follows that the weights $w_k = d_kF(q_k\mathbf{x}_k'\lambda)$ are restricted by $Ld_k < w_k < Ud_k$. It is worth noting that the distance function $G_k(w_k, d_k)$ for 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 is large positive, we are close to Case 1. If $L = 0$ and U is large, we are close to Case 2.

Case 7. With the following modification, we can avoid the negative weights and that can arise in Case 1: Specify the constants L and U such that $L < 1 < U$ and define $F_k(u) = 1 + q_ku$ if $(L - 1)/q_k \le u \le (U - 1)/q_k$; $F_k(u) = L$ if $u < (L - 1)/q_k$; and $F_k(u) = U$ if $u > (U - 1)/q_k$. The weights w_k will then be restricted according to $Ld_k \le w_k \le Ud_k$. The corresponding distance function is as in Case 1, if $Ld_k < w_k < Ud_k$, and is defined as infinity otherwise. If we choose that L will be positive, negative weights can never occur

Example 2 Returning to Example 1, we note that the ratio estimator is obtained for *any* $g_k(w, d)$ of the form $g(w/d)/q_k$, if $q_k = 1/x_k$. Then $F_k(\mathbf{x}_k'\lambda) = F(q_kx_k\lambda) = F(\lambda)$, a constant. From (2.3), $F(\lambda) = t_x/\hat{t}_{x\pi}$, so (2.6) gives the ratio estimator $\hat{t}_{yw} = t_x\hat{t}_{y\pi}/\hat{t}_{x\pi}$.

Example 2 is exceptional in that the choice of function has no bearing on the estimator. In general, different $F_k(u)$ yield different estimators. However, one can expect these estimators to produce only slightly different estimates in medium to large samples. The theoretical backing for this claim is the important Result 5 (p. 379) stating that all estimators (2.6), under mild conditions on the underlying $F_k(u)$, are asymptotically equivalent to the regression estimator (1.6) generated by the linear function $F_k(u) = 1 + q_ku$. 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. In a small Monte Carlo study with simple random sampling without replacement of $n = 200$ from a population of $N = 2{,}000$, we found practically no difference in variance among the estimators generated by several of the functions that we have described. Computational convenience more than anything else may then dictate the choice of $F_k(u)$. We now derive several asymptotic results that are needed later. Our setup for asymptotics is essentially that of Fuller and Isaki (1981) and Isaki and Fuller (1982). This setup has the following important features: We consider a sequence of finite populations and sampling designs indexed by n, where n is the sample size (for a fixed-sized sampling design) or the expected sample size (for a random-sized sampling design). The finite population size, N, tends to infinity with n, and we assume that for any vector valued variable $\mathbf{x}$ of interest to this article

1. $\lim N^{-1}\mathbf{t}_x$ exists.
2. $N^{-1}(\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x) \to \mathbf{0}$ in design probability.
3. $n^{1/2}N^{-1}(\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x)$ converges in distribution to the multinormal $N(\mathbf{0}, \mathbf{A})$.

Here (3) is to justify the use of the normal approximation in confidence intervals based on $\hat{t}_{yw}$. From a practical standpoint, the assumptions mean that: (a) the components of $\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x$ are considered small and quantities on the order of $\|\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x\|^2$ are considered negligible and (b) $\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x$ follows an approximately normal distribution with covariance matrix $n^{-1}N^2\mathbf{A}$. Let $\sum \sum_U$ be shorthand for $\sum_{k=1}^N \sum_{l=1}^N$; set $\Delta_{kl} = \pi_{kl} - \pi_k\pi_l$. Now (1), (2), and (3) imply that

$$nN^{-2} \sum \sum_U \Delta_{kl}(\mathbf{x}_k/\pi_k)(\mathbf{x}_l/\pi_l)' = nN^{-2}V(\hat{\mathbf{t}}_{x\pi})$$

converges to the fixed matrix $\mathbf{A}$ and that $N^{-1}(\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x) = O_p(n^{-1/2})$. We can view $\mathbf{A}$ as a matrix that describes an asymptotic effect of the sampling design used for the survey.

Before proving any asymptotic properties of $\hat{t}_{yw}$, we discuss the existence of a solution of (2.5). Now, (2.4) defines a function of λ on $C = \cap_{k\in U}\{\lambda : \mathbf{x}_k'\lambda \in \mathrm{Im}_k(d_k)\}$, a convex domain. We assume that C is an open neighborhood of $\mathbf{0}$, for every n. Five results are now stated and proved; the proofs of Results 1, 2, and 3 are given in the Appendix.

Result 1 Equation (2.5) has a unique solution belonging to C, with probability tending to one as $n \to \infty$.

Result 2 Let λ_s be the solution of (2.5), if one exists; otherwise, let λ_s be an arbitrary fixed value. Then, λ_s tends to $\mathbf{0}$ in design probability, and $\lambda_s = O_p(n^{-1/2})$.

To obtain Results 3, 4, and 5, we add the assumptions (a) $\max\|\mathbf{x}_k\| = M < \infty$, where max is over n as well as over k, and (b) $\max F_k''(0) = M' < \infty$. The assumption (b) is verified for Cases 1–7.

Result 3 We have $\lambda_s = \mathbf{T}_s^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}) + O_p(n^{-1})$.

Result 4 The calibration estimator $\hat{t}_{yw}$ given by (2.6) is design-consistent, and $N^{-1}(\hat{t}_{yw} - \hat{t}_{y\pi}) = O_p(n^{-1/2})$.

Proof Because $F_k'(0) = q_k$, we have, using the assumption (b),

$$F_k(u) = 1 + q_k u + \theta_k(u). \tag{2.7}$$

If (2.5) has a solution, λ_s, then $\hat{t}_{yw} - \hat{t}_{y\pi} = \sum_s d_k y_k \times \{q_k\mathbf{x}_k'\lambda_s + \theta_k(\mathbf{x}_k'\lambda_s)\}$. Now, using condition 3 for the variable $y_k\theta_k(\mathbf{x}_k'\lambda_s)$, and given the fact that $\max_{k\in U}\theta_k(u) = O(u^2)$, we obtain $N^{-1}|\hat{t}_{yw} - \hat{t}_{y\pi}| \le N^{-1}\{(\sum_s d_k q_k \times |y_k|\|\mathbf{x}_k\|)\|\lambda_s\|\} + O_p(n^{-1})$, where $N^{-1}\{\sum_s d_k q_k \times |y_k|\|\mathbf{x}_k\|\} = O_p(1)$, and $\lambda_s = O_p(n^{-1/2})$ by Result 2. Result 4 follows, because $\hat{t}_{y\pi}$ is design-consistent and $N^{-1}(\hat{t}_{y\pi} - t_y) = O_p(n^{-1/2})$.

Remark. Because $\hat{t}_{yw}$ is the nearest estimator to $\hat{t}_{y\pi}$ in a given sense, it can be expected to inherit some of the properties of $\hat{t}_{y\pi}$. Design unbiasedness is a property of $\hat{t}_{y\pi}$, so we may expect to find that $\hat{t}_{yw}$ is at least asymptotically design-unbiased (ADU). This property can in fact be obtained, if attention is paid to one detail. It is not certain that (2.5) 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 (2.5) has a solution; if it does not, use $\hat{t}_{y\pi}$ (that is, set $\lambda_s = \mathbf{0}$). This gives an ADU estimator. Note, for example, that the regression estimator (1.6) is undefined if $\mathbf{T}_s$ is singular. The usual poststratification estimator, a special case of (1.6), is undefined if there is at least one zero poststratum count.

Result 5. For any $F_k(\cdot)$ obeying our conditions, $\hat{t}_{yw}$ given by (2.6) is asymptotically equivalent to the regression estimator $\hat{t}_{y\mathrm{reg}}$ given by (1.6), in the sense that $N^{-1}(\hat{t}_{yw} - \hat{t}_{y\mathrm{reg}}) = O_p(n^{-1})$. As a consequence, the two estimators share the same asymptotic variance.

Proof From (2.6) and (2.7),

$$N^{-1}\hat{t}_{yw} = N^{-1}\hat{t}_{y\pi} + N^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})'\mathbf{T}_s^{-1}\sum_s d_k q_k \mathbf{x}_k y_k + O_p(n^{-1}) + N^{-1}\sum_s d_k y_k \theta_k(\mathbf{x}_k'\lambda_s).$$

The first two terms of the right side equal $N^{-1}\hat{t}_{y\mathrm{reg}}$, where $\hat{t}_{y\mathrm{reg}}$ is given by (1.6). The last term was found in the proof of Result 4 to be $O_p(n^{-1})$. Therefore, $n^{1/2}N^{-1}(\hat{t}_{yw} - \hat{t}_{y\mathrm{reg}}) = O_p(n^{-1/2})$, with a zero asymptotic variance.

3. VARIANCE AND VARIANCE ESTIMATION

Result 5 states that $\hat{t}_{yw}$ is asymptotically equivalent to $\hat{t}_{y\mathrm{reg}}$, 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

$$\mathrm{AV}(\hat{t}_{yw}) = \sum \sum_U \Delta_{kl}(d_k E_k)(d_l E_l), \tag{3.1}$$

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

$$\left(\sum_U q_k \mathbf{x}_k \mathbf{x}_k'\right)\mathbf{B} = \sum_U q_k \mathbf{x}_k y_k. \tag{3.2}$$

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

$$SS_U = \sum_U q_k(y_k - \mathbf{x}_k'\mathbf{B})^2 = \sum_U q_k E_k^2. \tag{3.3}$$

To estimate (3.1), the residuals E_k cannot be used, because $\mathbf{B}$ is unknown. An estimator, $\hat{\mathbf{B}}_{ws}$, is obtained by noting that SS_U given by (3.3) 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 the vector $\hat{\mathbf{B}}_{ws}$ satisfying the sample-based normal equation $(\sum_s w_k q_k \mathbf{x}_k \mathbf{x}_k')\hat{\mathbf{B}}_{ws} = \sum_s w_k q_k \mathbf{x}_k y_k$. Sample-based residuals now can be calculated as $e_k = y_k - \mathbf{x}_k'\hat{\mathbf{B}}_{ws}$. The variance estimator that we advocate is given by

$$\hat{V}(\hat{t}_{yw}) = \sum_s \sum (\Delta_{kl}/\pi_{kl})(w_k e_k)(w_l e_l). \qquad (3.4)$$

The calibration weights w_k are used in (3.4) to weight the residuals e_k. From a strictly design-based point of view, one can simply weight the e_k by the standard design weights d_k and obtain a design-consistent variance estimator. When model-based as well as design-based properties are considered, however, there is reason to prefer the w_k over the d_k. The model, ξ, that underlies the regression estimator (1.6) states that $E_\xi(y_k) = \boldsymbol{\beta}'\mathbf{x}_k$, $V_\xi(y_k) = \sigma_k^2$. Now (3.4) not only is a design-consistent variance estimator but also is nearly model-unbiased for the model mean squared error, $E_\xi(\hat{t}_{yw} - t)^2$, as Särndal, Swensson, and Wretman (1989) show in the case where σ_k^2 is of the form $\boldsymbol{\alpha}'\mathbf{x}_k$, for a constant vector $\boldsymbol{\alpha}$. Under simple random sampling without replacement, the model bias of (3.4) is negligible if the sampling fraction is small.

Example 3 Let us return to $\hat{t}_{yw} = t_x\hat{t}_{y\pi}/\hat{t}_{x\pi}$ in Example 2. Under SRS, (3.4) 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 $\hat{B}_s = (\sum_s y_k)/(\sum_s x_k)$. This is an often-recommended variance estimator for the ratio estimator.

To calculate the calibration estimator (2.6), we must first solve (2.5) for $\boldsymbol{\lambda}$. A note on the computational aspects is in order. We suggest an algorithm based on Newton's method; in the examples where we tried it, convergence was quick. Let $\boldsymbol{\phi}_s'(\boldsymbol{\lambda}) = \partial\boldsymbol{\phi}_s(\boldsymbol{\lambda})/\partial\boldsymbol{\lambda}$. Start with $\boldsymbol{\lambda}_0 = \mathbf{0}$. Subsequent iterative values, $\boldsymbol{\lambda}_\nu$, $\nu = 1, 2, \cdots$ are obtained by

$$\boldsymbol{\lambda}_{\nu+1} = \boldsymbol{\lambda}_\nu + \{\boldsymbol{\phi}_s'(\boldsymbol{\lambda}_\nu)\}^{-1}\{\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi} - \boldsymbol{\phi}_s(\boldsymbol{\lambda}_\nu)\}. \qquad (3.5)$$

From (2.4), $\boldsymbol{\phi}_s(\mathbf{0}) = \mathbf{0}$; $\boldsymbol{\phi}_s'(\mathbf{0}) = \mathbf{T}_s$. The first iteration gives $\boldsymbol{\lambda}_1 = \mathbf{T}_s^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})$; subsequent iterations, $\nu = 2, 3, \ldots$, obey (3.5) until convergence. Now $\boldsymbol{\lambda}_1$ is the vector (1.4) that yields the regression estimator (1.6). Thus, (1.6) is a first approximation to (2.6); Result 5 showed them to be asymptotically equivalent. Put somewhat differently, the first adjustment of $\boldsymbol{\lambda}$ is essential, but the remaining adjustments are relatively unimportant. If $F(u) = 1 + u$ is the chosen function, the iteration stops after the first step. (Note: For cases where F^{-1} maps C onto an interval I of $\mathbb{R}$, one must check that $\mathbf{x}_k'\boldsymbol{\lambda}_{\nu+1}$ really belongs to I. For instance, if $\mathbf{x}_k'\boldsymbol{\lambda}_{\nu+1} \geq \sup I$, it is a good idea to replace $\boldsymbol{\lambda}_{\nu+1}$ by $\boldsymbol{\lambda}_{\nu+1}^0 = \boldsymbol{\lambda}_\nu + \theta_\nu(\boldsymbol{\lambda}_{\nu+1} - \boldsymbol{\lambda}_\nu)$ for some $\theta_\nu < 1$ such that $\boldsymbol{\lambda}_{\nu+1}^0$ is near the border of the set of permissible values.) A SAS computer program by Sautory (1991), based on (3.5), is now routinely used in certain large surveys at the French national statistics bureau (I.N.S.E.E.) to produce estimates calibrated on known marginals.

4. CALIBRATION ON KNOWN COUNTS IN FREQUENCY TABLES

An important application of the technique in this article occurs in connection with calibration on the known counts (cell counts or marginal counts) of a frequency table in any number of dimensions. Here we limit ourselves to a brief discussion of two-way tables. We assume distance measures of the form $g_k(w, d) = g(w/d)$ and $q_k = 1$ for all k. This implies that $F_k(u) = F(u) = g^{-1}(u)$. With r rows and c columns, the population elements are classified into $r \times c$ cells—for example, individuals classified into age group by socioprofessional category. Suppose the typical population cell, U_{ij}, contains N_{ij} elements: $i = 1, \ldots, r$; $j = 1, \ldots, c$; so that $N = \sum\sum_{i,j} N_{ij}$, where $\sum\sum_{i,j}$ means $\sum_{i=1}^r \sum_{j=1}^c$. We can distinguish (a) calibration on known cell counts N_{ij}, which may be called "complete poststratification" and (b) calibration on known marginal counts, which may be called "incomplete poststratification." In case (a), the calibration estimator (2.6) equals the well-known poststratification estimator with the $r \times c$ cells as poststrata; this holds for any distance measure such that $g_k(w, d) = g(w/d)$. Case (b) is more interesting, because here the theory of calibration estimators leads to a new class of estimators corresponding to the class of distance measures discussed in Section 2. These estimators can be described as *generalized raking procedures,* the classical raking ratio is a simple special case. Case (b) has at least two important practical applications. First, the marginal counts are known, but the cell counts N_{ij} are not. The marginal counts may stem from different sources; for example, age group counts from one data file and professional group counts from another, with cross-classification counts lacking. By necessity, calibration is then on the known marginals. Second, there are some zero or extremely small sample cell counts. Calibration on the cell counts, although perhaps feasible, is abandoned in favor of the more reliable calibration obtained from the known marginals. (The need to calibrate on marginal counts rather than on cell counts would be even more strongly felt for a table with three or more dimensions.)

To calibrate on the marginals, first identify the $\mathbf{x}_k$-vector having the property such that $\sum_U \mathbf{x}_k$ summarizes (and does not go beyond) the population totals used in the calibration—in this case, the marginal counts. It is easy to see that $\mathbf{x}_k = (\delta_{1\cdot k}, \ldots, \delta_{r\cdot k}, \delta_{\cdot 1k}, \ldots, \delta_{\cdot ck})'$, where $\delta_{i\cdot k} = 1$ if the element k is in row i and 0 otherwise, and $\delta_{\cdot jk} = 1$ if k is in column j and 0 otherwise. Then, $\sum_U \mathbf{x}_k = (N_{1+}, \ldots, N_{r+}, N_{+1}, \ldots, N_{+c})'$, where $N_{i+} = \sum_{j=1}^c N_{ij}$, $N_{+j} = \sum_{i=1}^r N_{ij}$. Letting $\boldsymbol{\lambda} = (u_1, \ldots, u_r, v_1, \ldots, v_c)'$, we have $\mathbf{x}_k'\boldsymbol{\lambda} = u_i + v_j$ whenever k belongs to cell ij. That is, $F(\mathbf{x}_k'\boldsymbol{\lambda}) = F(u_i + v_j)$ depends on the cell but not on the label within the cell. With $\hat{N}_{ij} = \sum_{s_{ij}} 1/\pi_k$, the calibration equations (2.3) take the form

$$\sum_{j=1}^{c} \hat{N}_{ij}F(u_i + v_j) = N_{i+} \qquad (i = 1, \ldots, r) \qquad (4.1)$$

and

$$\sum_{i=1}^{r} \hat{N}_{ij}F(u_i + v_j) = N_{+j} \qquad (j = 1, \ldots, c). \qquad (4.2)$$

This system is to be solved for $u_1, \ldots, u_r, v_1, \ldots, v_c$, using the function $F(\cdot)$ chosen by the statistician. Iterative solution is often required. One of the $r + c$ equations is redundant, so it is possible to fix one component—say, $v_c = 0$—and solve the system for $i = 1, \ldots, r$; $j = 1, \ldots, c - 1$. Note that $u_i + v_j$ remains invariant to the elimination of one equation. Having obtained the u_i and v_j, we calculate the cell factors $F(u_i + v_j)$, the calibrated cell count estimates $\hat{N}_{ij}^w = \hat{N}_{ij} F(u_i + v_j)$, and the calibrated weights $w_k = d_k \hat{N}_{ij}^w / \hat{N}_{ij}$. Finally, the calibration estimator, obtained from (2.6), is

$$\hat{t}_{yw} = \sum_s w_k y_k = \sum \sum_{i,j} \hat{N}_{ij}^w \bar{y}_{s_{ij}}, \tag{4.3}$$

where $\bar{y}_{s_{ij}} = (\sum_{s_{ij}} d_k y_k)/\hat{N}_{ij}$. The cell count estimates $\hat{N}_{ij}^w$ are often substantial improvements on the naive estimates $\hat{N}_{ij}$. In fact, the estimator (4.3) can be nearly as efficient as $\sum \sum_{i,j} N_{ij} \bar{y}_{s_{ij}}$, the poststratified estimator formed when the N_{ij} are known. If the effects on y of the rows and the columns are additive (i.e., the interaction effects are negligible), then (4.3) and $\sum \sum_{i,j} N_{ij} \bar{y}_{s_{ij}}$ have virtually identical variances. Efficiency, variance estimation, computational aspects, the occurrence of extreme weights, and the use of special functions $F(\cdot)$ to restrict the range of the weights (as in Cases 6 and 7) are aspects of generalized raking that we discuss in a forthcoming paper. The theory discussed in Section 2 permits a wide choice of functions $F(\cdot)$. Some simple specifications of $F(\cdot)$ correspond to well-known procedures: First, the linear function $F(u) = 1 + u$ yields additive cell factors, $F(u_i + v_j) = 1 + u_i + v_j$, and the weights are $w_k = d_k(1 + u_i + v_j)$ for the elements k in cell ij. These weights are not necessarily positive. The calibration equations (4.1) and (4.2) that result from this case were presented in Deming and Stephan (1940). Second, the exponential case $F(u) = \exp(u)$ gives multiplicative cell factors $F(u_i + v_j) = \exp(u_i)\exp(v_j)$, and the always-positive weights are $w_k = d_k \exp(u_i + v_j)$. The solution to (4.1) and (4.2) in this case can be obtained by carrying out (until convergence) the classical raking ratio algorithm of Deming and Stephan (1940), sometimes called *iterative proportional fitting*. In practice, the procedure is sometimes stopped after two iterations. As pointed out in Huang (1976), Deming and Stephan suggested the algorithm apparently thinking that it converges to the solution for the linear case, for which they had presented the equations. This was later noted by Deming (1943).

APPENDIX: PROOFS OF RESULTS 1, 2, AND 3

1. Mathematical Preliminaries

1.1 The Function ϕ and Its Properties

Let $C_n = \cap \{\lambda : \mathbf{x}_k'\lambda \in \mathrm{Im}_k(d_k)\}$, 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 $\mathbf{0}$ for every n. Moreover, $C^* = \cap_{n=1}^{\infty} C_n^0$ is convex; we assume it is also open. Let E_n and P_n denote expectation and probability, 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 assumption 3 applied to the variable $F_k(\mathbf{x}_k'\lambda)$, it converges to a fixed function denoted ϕ. Convergence is uniform on every compact set in C^*. Note the properties $N^{-1}\phi_s(\mathbf{0}) = \mathbf{0}$, $\phi(\mathbf{0}) = \mathbf{0}$, and $N^{-1}\phi_s'(\mathbf{0}) = N^{-1}\mathbf{T}_s$; $\phi'(\mathbf{0}) = \mathbf{T} = \lim N^{-1} \sum_U \mathbf{x}_k\mathbf{x}_k'$. Now for every λ, ϕ' is a positive definite matrix, because all F_k are increasing functions. Consequently, ϕ is injective and maps C^* onto an open neighborhood of $\mathbf{0}$ in $\mathbb{R}^J$. Let B be a closed sphere with radius r contained in that neighborhood, and let A be the compact set $\phi^{-1}(B)$. The inverse function ϕ^{-1} is defined on B, continuous, and continuously differentiable. Then $\|\phi^{-1}(\mathbf{x})\|$ is continuously differentiable and bounded on B. Let $K = \max_{\mathbf{x}\in B}\|(\phi^{-1})'(\mathbf{x})\|$.

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

We need a result that justifies the use of an inverse mapping of ϕ_s. Such a result is obtained in this section. 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 $P_n(\|N^{-1}\phi_s - \phi\|_A < \varepsilon) \to 1$ when n increases. Now let $\phi_1 = N^{-1}\phi_s$ for some function verifying $\|\phi_1 - \phi\|_A \le \beta r$; $\|\phi_1' - \phi'\|_A \le \beta K$, with $0 < \beta < \frac{1}{2}$. The probability of this event tends to 1 as n increases. Let $r_1 = (1 - \beta)r$, and let B_1 be the sphere $\|\mathbf{x}\| < r_1$ in $\mathbb{R}^J$. Now ϕ_1 maps the frontier of A onto the crown $r_1 \le \|\mathbf{x}\| \le r(1 - \beta)$, and $\phi_1(A)$ is a bordered manifold homotopic to B. These notions are discussed in Trenoguine (1987). Consequently, $\phi_1(A)$ covers the sphere B_1—in other words, for every $\mathbf{x} \in B_1$, the equation $\phi_1(\lambda) = \mathbf{x}$ has a (unique) solution. Moreover, ϕ_1^{-1}, defined on B_1, is a continuously differentiable function. Because $\|\phi_1' - \phi'\| \le \beta K$ for every λ in C, $(\phi_1^{-1})'(\mathbf{x})$ exists for every $\mathbf{x} \in B_1$, and $\|\phi_1^{-1}(\mathbf{x})\| \le \|\mathbf{x}\|K(1 - \beta)^{-1}$.

2 Proofs of the Three Results

Result 1 First, $N^{-1}(\hat{\mathbf{t}}_{x\pi} - \mathbf{t}_x) = \mathbf{z}$ belongs to B_1 with a probability tending to 1. Second, $N^{-1}\phi_s$ has an inverse function on B_1 with a probability tending to 1. As (2.5) can be written $N^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}) = N^{-1}\phi_s(\lambda_s)$, the equation has a unique solution with probability tending to 1.

Result 2 Let $\lambda_s = (N^{-1}\phi_s')^{-1}(\mathbf{z})$ if $\mathbf{z}$ belongs to B_1; otherwise, λ_s is arbitrarily defined. Since $\phi_s(\mathbf{0}) = \mathbf{0}$, we have $\lambda_s - \mathbf{0} = (N^{-1}\phi_s')^{-1}(\mathbf{z}) - (N^{-1}\phi_s')^{-1}(\mathbf{0})$ and $\|\lambda_s\| \le \|\mathbf{z}\|K(1 - \beta)^{-1}$. This inequality holds with probability tending to 1 when n increases. But $\mathbf{z} = O_p(n^{-1/2})$, so there exists a constant K' such that $P_n(\|\mathbf{z}\| \le K'n^{-1/2}) \to 1$. Combining the two inequalities, $P_n(\|\lambda_s\| \le KK'(1 - \beta)^{-1}n^{-1/2}) \to 1$, 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 $\theta_k(u) \le K''u^2$. We can write (2.5) as $\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi} = \sum_s d_k\mathbf{x}_k\{q_k\mathbf{x}_k'\lambda_s + \theta_k(\mathbf{x}_k'\lambda_s)\}$, and therefore $\lambda_s - \mathbf{T}_s^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi}) = -\mathbf{T}_s^{-1}\sum_s d_k\mathbf{x}_k\theta_k(\mathbf{x}_k'\lambda_s)$. For λ_s sufficiently small,

$$\|\lambda_s - \mathbf{T}_s^{-1}(\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})\| \le \|(N^{-1}\mathbf{T}_s)^{-1}\|K''\left\{N^{-1}\sum_s d_k\|\mathbf{x}_k\|^3\right\}\|\lambda_s\|^2$$

Here, $\|(N^{-1}\mathbf{T}_s)^{-1}\| = O_p(1)$, and, using assumption (2) in Section 2, $N^{-1}\sum_s d_k\|\mathbf{x}_k\|^3 = O_p(1)$. Moreover, by Result 2, $\|\lambda_s\|^2 = O_p(n^{-1})$. Result 3 follows.

[*Received November 1989. Revised July 1991.*]

REFERENCES

Bankier, M. D. (1990), "Generalized Least Squares Estimation Under Poststratification," in *Proceedings of the Section on Survey Research Methods, American Statistical Association*, pp. 730–755.

Bethlehem, J. G., and Keller, W. J. (1987), "Linear Weighting of Sample Survey Data," *Journal of Official Statistics*, 3, 141–153.

Cassel, C M , Sarndal, C E., and Wretman, J. H (1976), "Some Results on Generalized Difference Estimation and Generalized Regression Estimation for Finite Populations," *Biometrika, 63,* 615–620

Deming, W E. (1943), *Statistical Adjustment of Data,* New York· John 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," *The 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,* eds. A Monfort and J J. Laffond, Paris· Economica, pp. 915–927.

Fuller, W A , and Isaki, C T (1981), "Survey Design Under Superpopulation Models," in *Current Topics in Survey Sampling,* eds D. Krewski, J N K Rao, and R Platek, New York Academic Press, pp 199–226

Gourieroux, C (1981), *Théorie des Sondages,* Paris· Economica

Huang, E (1976), "Regression Estimation of Means and Totals for Sample Survey Data," unpublished Ph D thesis, Iowa State University

Isaki, C T , and Fuller, W A. (1982), "Survey Design Under the Regression Superpopulation Model," *Journal of the American Statistical Association,* 77, 89–96.

Lemaitre, G., and Dufour, J (1987), "An Integrated Method for Weighting Persons and Families," *Survey Methodology,* 13, 199–207.

Lemel, Y (1976), "Une Généralisation de la Méthode du Quotient pour le Redressement des Enquêtes par Sondage, *Annales de l'INSEE,* 22–23, 272–282

Sarndal, C. E (1980), "On π-Inverse Weighting Versus Best Linear Weighting in Probability Sampling," *Biometrika,* 67, 639–650

——— (1982), "Implications of Survey Design for Generalized Regression Estimation of Linear Functions," *Journal of Statistical Planning and Inference,* 7, 155–170.

Sarndal, 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

Sautory, O. (1991), "Redressements d'Échantillons d'Enquêtes auprès des Ménages par Calage sur les Marges," Internal report, Division Méthodes Statistiques et Sondages, INSEE, Paris.

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

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," in *Proceedings of the Section on Survey Research Methods, American Statistical Association,* pp. 64–71.

——— (1990), "Sample Weighting Methods and Estimation of Totals in the Consumer Expenditure Survey," *Journal of the American Statistical Association,* 85, 986–1001.