ITERATIVE WEIGHTED RATIO ESTIMATION IN SAMPLE SURVEYS

Alan H. Feiveson, NASA Johnson Space Center

1.0 Introduction

In a sample survey where one is plagued with measurement errors and loss of data, it might be possible to improve the precision of the overall estimate of a population total if observations thought to be of questionable accuracy are down-weighted in the aggregation process.

In this paper, we consider a framework in which the estimate of a stratum total can be obtained in one of two ways: (1) by using direct measurements on sampling units within the stratum, or (2) by ratioing to the estimate of some other stratum (or strata) using a concomitant variable whose values are known at the stratum level.

The first method is unbiased if measurement errors are unbiased; the second procedure is only unbiased if the concomitant variable is proportional to the target variable over strata. In cases where all data within a stratum is lost, one would have no choice but to use the second method; however, even when some data is available, large measurement errors or a small sample size might cause the first estimate to have such a large variance that it is better to use the second estimate if the proportionality condition is even approximately true.

In general, some weighted average of the two types of estimates will produce the smallest mean-squared error. If the resulting stratum estimate is then used in turn to produce a ratio estimate for some other strata a complicated interdependence of the stratum level estimates arises. Section 2 of this paper describes how a weighted ratio estimation technique can be defined and implemented given a set of weights. Section 3 discusses the bias and variance of the overall estimate while Section 4 considers the question of how one might obtain the weights.

2.0 The Weighted Ratio Estimation Process

2.1 Basic Framework

Let $\alpha.$ be a population total of a characteristic which is to be estimated by a sample survey over L strata by $\hat{\alpha}. = \sum_{i=1}^{L} \hat{\alpha}_i$ where $\hat{\alpha}_i$ is the estimate of the population subtotal α_i in the ith stratum. We consider here a situation in which originally n_i sampling units were allocated to the ith stratum (i=1,2,...L) but because of data loss only m_i can be observed where $0 \leq m_i \leq n_i$. For strata with $m_i > 0$, the standard estimator of $\hat{\alpha}_i$ is given by

$$d_i^* = \frac{N_i}{m_i} \sum_{i=1}^{m_i} \hat{y}_{ij} \qquad (2.1)$$

where $\hat{y}_{ij}$ is the estimate of the target characteristic for the jth sampling unit in the ith stratum containing N_i units of which m_i were observed.

For all strata let $d_i = \begin{cases} d_i^* & m_i > 0 \\ 0 & \text{otherwise} \end{cases}$

In addition, let h_i be a positive-valued concomitant variable for the ith stratum. It is assumed that h_i is observable without error for all strata and is approximately proportional to α_i over strata. For example, if the survey is one which is repeated each year, with a census taken every five years, the h_i could be the stratum totals for the last census year.

Also associated with each stratum is a prescribed set of other strata for the purpose of computing a ratio estimate of the total. Specifically, for the ith stratum, there exists a set R_i of indices of other strata such that z_i, the ratio estimate is given by

$$z_i = \left(\frac{\sum_{j \in R_i} d_j}{\sum_{j \in R_i} h_j} \right) h_i \qquad (2.2)$$

Let the Lx1 vector $b_i = (b_{i1}, \ldots b_{iL})^T$ be defined by

$$b_{ij} = \begin{cases} 1, & j \in R_i \\ 0, & \text{otherwise} \end{cases}$$

Then (2.2) can be written

$$z_i = \left(\frac{b_i^T d}{b_i^T h} \right) h_i, \qquad (2.3)$$

where $d = (d_1, \ldots, d_L)^T$ and $h = (h_1, \ldots, h_L)^T$.

2.2 Estimation

We now construct a vector of estimates of the α_i, $\hat{\alpha} = (\hat{\alpha}_1, \ldots \hat{\alpha}_L)^T$ where $\hat{\alpha}$ will be shown to be the limit of a convergent sequence of iterated estimates $\{\hat{\alpha}^{(\nu)}\}_{\nu=0}^{\infty}$. At any stage of the iteration, say the ν+1-st, $\hat{\alpha}^{(\nu+1)}$ will be a weighted average between the direct estimate d_i and a ratio estimate $z_i^{(\nu)}$ which will be defined below.

To start the iteration process, let $\hat{\alpha}_i^{(0)} = d_i$, i=1,2,...L. Then the ν+1-st iterated estimate is defined by

$$\hat{\alpha}_i^{(\nu+1)} = w_i d_i + (1-w_i) z_i^{(\nu)} \qquad (2.4)$$

where

$$z_i^{(\nu)} = \left(\frac{b_i^T \hat{\alpha}(\nu)}{b_i^T h} \right) h_i \qquad (2.5)$$

and $\{w_i\}_i^L = 1$ $(0 \le w_i \le 1)$ are weights associated with the strata which reflect the degree of accuracy of d_i as an estimate of α_i relative to that of z_i. It assumed in this section that the w_i are known a priori or at least computable; the problem of obtaining them will be discussed in Section 4. When $w_i = 0$, either all data for the stratum is lost (i.e., $m_i = 0$) or for some other reason, measurement errors are regarded as so bad that the direct estimate d_i is considered worthless. Conversely, $w_i = 1$ when d_i is thought to be a much better estimate of α_i than is z_i. (Note that $z_i^{(0)} = z_i$ and hence $\hat{\alpha}_i^{(1)}$ is a simple weighted average between d_i and z_i. At later iterations $z_i^{(\nu)}$ plays the role of z_i but uses $\hat{\alpha}^{(\nu)}$ in place of d in (2.3).)

By letting the LxL matrix $C = (c_{ij}) = b_{ij}(1-w_i)h_i/b_i^T h$, equation (2.4) can be written

$$\hat{\alpha}^{(\nu+1)} = u + C\hat{\alpha}^{(\nu)} \qquad (2.6)$$

where $u = (u_1, \ldots, u_L)^T$ and $u_i = w_i d_i$.

From (2.6) it follows that

$$\hat{\alpha}^{(\nu+1)} = (I+C+C^2+\ldots+C^{\nu})u + C^{\nu+1}\hat{\alpha}^{(0)},$$

hence $\hat{\alpha}^{(\nu)}$ converges to a vector $\hat{\alpha}$ if and only if the series $I+C+C^2+\ldots$ converges.

The following two theorems taken from standard matrix analysis can be used to show that $\hat{\alpha}^{(\nu)}$ converges to $\hat{\alpha}$ for "reasonable" b_{ij}'s where

$$\hat{\alpha} = (I-C)^{-1}u. \qquad (2.7)$$

In the following theorems, the function $\rho(M)$, for any matrix M, is defined by $\rho(M) = \max_i |\lambda_i\ (m)|$ where $\lambda_i\ (M)$ is the ith of eigenvalue of M.

Theorem 1

If M is an arbitrary complex n x n matrix with $\rho(M)<1$, then I - M is nonsingular, and

$$(I-M)^{-1} = I + M + M^2 + \ldots,$$

the series on the right converging. Conversely, if the series on the right converges, then $\rho(M)<1$.

Theorem 2

If $A = (a_{i,j}) \ge 0$ is an n x n matrix, and x is any vector with positive components $x_1, x_2, \ldots, x_n$,

$$\min_{1\le i\le n} \frac{\sum_{j=1}^{n} a_{i,j}x_j}{x_i} \le \rho(A) \le \max_{1\le i\le n} \frac{\sum_{j=1}^{n} a_{i,j}x_j}{x_i}$$

These results may be found on pages 82 and 47 respectively of Varga (1962).

In order to show (2.7), we will first show $\rho(C)<1$ and then use Theorem 1.

We can attempt to show $\rho(C)<1$, by using Theorem 2 with $x_j = h_j$. Then for fixed i,

$$\sum_j c_{ij}x_j/x_i = (1-w_i)h_i\ (\sum_j b_{ij}h_j)/h_i(\sum_k b_{ik}h_k) = 1-w_i.$$

Thus, if $w_i>0\ \forall i$, by Theorem 2 we have $\rho(C)<1$. In practice, however, some w_i will be zero; hence some restrictions must be put on the B matrix to ensure $\rho(C)<1$. A natural restriction to require is that $b_{ij} = 0$ if $w_j = 0$; i.e., if the "direct" estimate in the jth stratum is nonexistent or thought to be completely worthless ($w_j = 0$) then that stratum should not be used in a ratio estimate for some other stratum. With this restriction, the C matrix now takes on the form

$$C = \begin{bmatrix} \underset{r x r}{0} & \underset{r x(L-r)}{C_1} \\ \underset{(L-r) x r}{0} & \underset{(L-r) x(L-r)}{C_2} \end{bmatrix} \qquad (2.8)$$

where the order of the strata has been rearranged so that $w_1 = w_2 = w_r = 0$ and $w_j > 0$ for $j = r+1, \ldots, L$.

Since the non-zero eigenvalues of C_2 are also the non-zero eigenvalues of C, it is clear that $\rho(C) = \rho(C_2)$. By applying Theorem 2 to C_2 as in the previous paragraph, we now have $\rho(C) = \rho(C_2)<1$.

By Theorem 1, the series $I+C+C^2+\ldots$ converges to $(I-C)^{-1}$; hence, $\lim_{\nu \to \infty} \hat{\alpha}^{(\nu)} = (I-C)^{-1}u = \hat{\alpha}$.

3.0 Properties of the Estimate $\hat{\alpha}$

Strictly speaking, $w = (w_1, \ldots w_L)^T$ is random because the weights are usually functions of random events which determine (a) how much data is lost, or (b) the magnitude of measurement errors. If w is considered fixed or the conditional mean and variance of d_i given w is known, we can make straightforward statements about the bias and variance of $\hat{\alpha}$. In the following discussion all expectations and variances are assumed conditional on w.

3.1 Bias

If $E(d_i) = \hat{\alpha}_i$ for $w_i \ge 0$ and h_j is proportional to α_i for $j \varepsilon R_i$ then $\hat{\alpha}_i$ is unbiased.

To show unbiasedness under the above conditions, we will use an induction argument. First, from (2.4) we have $E(\hat{\alpha}_i^{(1)}) = w_i\alpha_i + (1-w_i)E(z_i)$. From the second assumption, $\alpha_i/\sum_j b_{ij}\alpha_j = h_i/\sum_j b_{ij}h_j$. As a consequence, remembering that $w_i = 0 \Rightarrow b_{ij} = 0$, we have

$$E(z_i) = h_i \sum_j E(b_{ij}d_j)/\sum_k b_{ik}h_k.$$

$= h_i \sum_j b_{ij}\alpha_j / \sum_k b_{ik}h_k$

$= \alpha_i$. (By second assumption)

It thus follows that

$E(\hat{\alpha}^{(1)}) = w_i\alpha_i + (1-w_i)\alpha_i = \alpha_i$.

Suppose now that $E(\hat{\alpha}_i^{(\nu)}) = \alpha_i$. Then similarly,

$$E(\hat{\alpha}_i^{(\nu+1)}) = E(w_i d_i) + (1-w_i)E(z_i^{(\nu)})$$
$$= w_i\alpha_i + (1-w_i)h_i \sum_j E(b_{ij}\hat{\alpha}_j^{(\nu)})/\sum_k b_{ik}h_k$$
$$= w_i\alpha_i + (1-w_i)h_i \sum_j b_{ij}\alpha_j/\sum_k b_{ik}h_k$$
$$= w_i\alpha_i + (1-w_i)\alpha_i$$
$$= \alpha_i$$

Hence, $\hat{\alpha}_i^{(\nu)}$ is unbiased for all ν which implies $E(\hat{\alpha}_i) = \alpha_i$.

3.2 Variance

Since $\hat{\alpha} = (I-C)^{-1}u$, the conditional covariance matrix of $\hat{\alpha}$ given w is given by

$$\Sigma = (I-C)^{-1}\, \Psi_u\, (I-C^T)^{-1} \qquad (3.1)$$

where Ψ_u is the conditional covariance matrix of u given w. In particular, since the estimate of the population total is $\hat{\alpha}. = \sum_i \hat{\alpha}_i$, the conditional variance of $\hat{\alpha}.$ is given by

$$V(\hat{\alpha}.) = e^T\, \Sigma\, e \qquad (3.2)$$

where $e = (1,1,\ldots,1)^T$.

3.3 Estimation of $V(\hat{\alpha}.)$

If an estimate of Ψ_u is available, it can be used in (3.1) to obtain an estimate of Σ and hence $V(\hat{\alpha}.)$. In most surveys, the d_i's can be assumed independent since they are based on independent data sets, hence Ψ_u is diagonal.

For strata with large enough sample sizes, Var (d_i) can be estimated directly, otherwise pooling or regression methods may have to be used. In any case, if s^2_i represents the estimated variance of d_i and the w's are regarded as fixed, Ψ_u can be estimated by $\hat{\Psi}_u = \text{diag}\,(w_i{}^2 s_i{}^2)$. Thus,

$$\hat{V}(\hat{\alpha}.) = e^T\, [(I-C)^{-1}\, \hat{\Psi}_u (I-C^T)^{-1}]e \qquad (3.3)$$

4.0 <u>Determination of Weights</u>

4.1 The General Case

Other than using the obvious choice of $w_i = 0$ when $m_i = 0$, there is no straightforward procedure for determining w. Ideally, one would like to choose the non-zero w's such that the mean-squared error of $\hat{\alpha}.$ is minimized, but this is impossible without knowledge of the bias. If an estimate of Ψ_u is available, one could minimize (3.3) with respect to the w's.

The unconstrained minimization of $\hat{V}(\hat{\alpha}.)$ leads to the set of equations

$$\frac{\partial\hat{V}(\hat{\alpha}.)}{\partial w} = \begin{bmatrix} \frac{\partial\hat{V}(\hat{\alpha})}{\partial w_1} \\ \vdots \\ \frac{\partial\hat{V}(\hat{\alpha})}{\partial w_L} \end{bmatrix} = 2D(g_k{}^2 s_k{}^2)w - 2D(g_k)P(I-C)^{-1}\hat{\Psi}_u g = 0 \qquad (4.1)$$

where $D(x_k) = \text{diag}\,(x_1,\ldots x_L)$

$$P = (P_{ij}) = \frac{b_{ij}h_i}{\sum_{j'} b_{ij'}h_{j'}}$$

and $g = (g_1 \ldots g_L)^T = (I-C^T)^{-1}e$.

By only considering the equations above where $s_k{}^2 > 0$ (i.e., those corresponding to the non-zero w's), one can obtain the iterative equation $w = f(w)$ where

$$f(w) = D^+\left(\frac{1}{g_k s_k{}^2}\right) P(I-C)^{-1}\, D^+\,(w_k{}^2 s_k{}^2)g \qquad (4.2)$$

and D^+ indicates the collapsed matrix D which includes only the positive elements of the original D. One approach to finding the optimal w is to use successive substitution with a boundary of 1, i.e., the ith element of $w^{(n+1)}$, the n+1-st iterate of w, is equal to the ith element of $f(w^{(n)})$ if the latter is less than or equal to 1, otherwise it is set equal to 1. In practice, this method has always led to a useful positive convergent sequence,; with the value of $\hat{V}(\hat{\alpha}.)$ monotonically decreasing at each step; however, nothing has been formally proven about convergence as yet.

4.2 Special Case

For the case where each column of B is either all zeros or all ones, the w_i which minimize (3.3) can be explicitly derived.

Suppose for each stratum, the ratio estimation set is the totality of all strata which have at least one observation available. Then the matrix B is of the Form $B = (b_{ij}) = \delta_j$ where $\delta_j = \begin{cases} 0 & m_j = 0 \\ 1 & m_j > 0 \end{cases}$. Given $w_i = 0$ if $m_i = 0$, the problem is to find the other elements of w such that $\hat{V}(\hat{\alpha}.)$ is minimized.

In this case, the Matrix C can be written $C = v\delta^T$ where $v = (v_1,\ldots v_2)^T$, $v_i = (1-w_i)h_i/\delta^T h$ and $\delta = (\delta_1,\ldots\delta_L)^T$. It then follows that $(I-C)^{-1} = I + \left(\frac{1}{1-\delta^T v}\right)v\delta^T$, and thus

$$\hat{\alpha}. = e^T [I+(\frac{1}{1-\delta^T v}) v \delta^T]_u. \qquad (4.3)$$

Note that $\delta^T v = \sum_{i=1}^{L} \delta_i (1-w_i)h_i / \sum_{i=1}^{L} \delta_i h_i$ and hence

$$1-\delta^T v = \frac{\sum_i w_i \delta_i h_i}{\sum_i \delta_i h_i}$$

Substitution into (4.3) yields,

$$\hat{\alpha}. = \sum_i w_i d_i + \frac{(\Sigma v_i)(\Sigma \delta_i u_i)(\Sigma \delta_i h_i)}{\Sigma \delta_i h_i w_i}$$

From the definitions of δ_i and of w_i when $m_i = 0$, it is clear that $\delta_i w_i = w_i$, hence

$$\hat{\alpha}. = \Sigma w_i d_i [1+ \frac{(\Sigma(1-w_i)h_i)(\Sigma \delta_i h_i)}{(\Sigma \delta_i h_i)(\Sigma h_i w_i)}]$$

$$= (\sum_{i=1}^{L} h_i)(\sum_{i=1}^{L} w_i d_i) / \sum_{i=1}^{L} w_i h_i \qquad (4.4)$$

In terms of the (unknown) w_i corresponding to $m_i>0$, (4.4) can be written

$$\hat{\alpha}. = \frac{(\sum_{i=1}^{L} h_i)(\sum_{+} w_i d_i)}{\sum_{+} w_i h_i} \qquad (4.5)$$

where $\sum_{+}$ means summation only over those i such that $m_i > 0$.

The estimated variance of $\hat{\alpha}.$ is then given by

$$\hat{V}(\hat{\alpha}.) = \frac{(\sum_{i=1}^{L} h_i)^2 (\sum_{+} w_i^2 s_i^2)}{(\sum_{+} w_i h_i)^2} \qquad (4.6)$$

which is minimum when $w_i \propto (h_i/s_i^2)$. The corresponding optimal variance is thus

$$\hat{V}_{opt} = \frac{(\sum_{i=1}^{L} h_i)^2}{\sum_{+} h_i^2/s_i^2} \qquad (4.7)$$

5.0 Conclusions

We have presented a procedure for combining ratio-based and direct estimates of population totals in certain types of surveys. By judicious choosing of the weights, the variance of the estimate can be reduced; however, the ratio estimation sets must be such that the hypothesis of proportionality of h_i to α_i over strata approximately holds, otherwise a bias can result.

In the case of general B matrices and weights, the procedure, which is the limit of an iterative process provides an approach to the problem of accounting for missing data or large measurement errors; particularly when there are large differences between the coefficients of variation of the individual stratum estimates. Also provided is an approximate technique for determining the weights.

Further topics for investigation should include the problems of how to select the B matrix and how to more accurately determine or estimate weights.

REFERENCES

Cochran, William G. (1963), Sampling Techniques, New York, John Wiley & Sons, Inc.

Varga, Richard S. (1962), Matrix Iterative Analysis, Englewood Cliffs, N.J.: Prentice Hall.