ESTIMATORS BASED ON SEVERAL STRATIFIED SAMPLES WITH APPLICATIONS TO MULTIPLE FRAME SURVEYS

Michael D. Bankier, Statistics Canada

1. INTRODUCTION

In this paper, a new method for producing estimates from a multiple frame survey is presented. The discussion in this paper is restricted to sample designs where a stratified simple random sample is selected independently from each frame. The estimation technique outlined, however, can be applied to more complex sample designs. It is assumed that units selected in more than one sample can be identified.

Hartley (1962, 1974) developed an estimation technique for multiple frame sample designs. The approach suggested in this paper can provide estimators with significantly smaller variances. Also, calculation of the weights does not require that variances or covariances be estimated from the samples as does the Hartley approach. Finally, computationally and algebraically it is very easy to extend the method in this paper to any number of frames. The method of Hartley, however, rapidly become more complex as the number of frames is increased.

In Sections 2 and 3, a method is given for producing estimates based on two independent stratified samples selected from the same frame. Then, in the following section it is shown how these results can easily be applied to a multiple frame sample design. Next, the estimators for multiple frame sample designs suggested by Hartley are discussed. Finally, a numerical example is given which indicates that the estimators developed in this paper may be better than those suggested by Hartley.

2. ESTIMATOR BASED ON TWO STRATIFIED SAMPLES SELECTED FROM THE SAME FRAME

In this section, it is assumed that two stratified simple random samples are selected independently from the same frame. These will be called Sample A and Sample B. The stratification of the two samples can be completely different. Let N_{Ag} equal the population size of stratum g of Sample A. Let n_{Ag} equal the sample size of stratum g of Sample A. N_{Bh} and n_{Bh} will represent the corresponding quantities for stratum h of Sample B. The N units in the frame can be cross-classified in terms of the strata of Sample A and of Sample B. Let N_{gh} equal the population size of the intersection of stratum g of Sample A and stratum h of Sample B. This intersection will be called cross-stratum gh. Also, let n_{gh} represent the number of distinct units in cross-stratum gh that fall in at least one of the two samples. In addition, let y_{ghi} equal the characteristic of interest for the i-th unit in cross-stratum gh. Then $Y_{gh} = \sum_i^{N_{gh}} y_{ghi}$ and $Y = \sum_g \sum_h Y_{gh}$. It is desired to estimate Y. The probability of a unit from cross-stratum gh being selected in at least one of the two samples is

$W_{gh}^{-1} = 1 - (1 - f_{Ag})(1 - f_{Bh})$ where $f_{Ag} = n_{Ag}/N_{Ag}$ and $f_{Bh} = n_{Bh}/N_{Bh}$. An unbiased estimator of Y is the Horvitz-Thompson estimator.

$$\hat{Y}_{HT} = \sum_g \sum_h \hat{Y}_{gh} \tag{2.1}$$

where

$$\hat{Y}_{gh} = W_{gh} \sum_i^{n_{gh}} y_{ghi}. \tag{2.2}$$

The variance of $\hat{Y}_{HT}$ under this sample design is

$$V(\hat{Y}_{HT}) = \sum_g \sum_h \frac{W_{gh}}{W_{gh}^*} (W_{gh} - 1) \sum_i^{N_{gh}} y_{ghi}^2$$

$$- \sum_g \sum_h \sum_{g'} \sum_{h'} \left(\frac{W_{g'h'}}{W_{ghg'h'}} - 1 \right) (W_{gh} - 1) Y_{gh} Y_{g'h'} \tag{2.3}$$

where

$$W_{gh}^* = 1/\left[1 - (1 - f_{Ag}^*)(1 - f_{Bh}^*)\right]$$

$$f_{Ag}^* = n_{Ag}/(N_{Ag} - 1)$$

$$f_{Bh}^* = n_{Bh}/(N_{Bh} - 1)$$

$$W_{ghg'h'} = 1/\left[1 - (1 - f_{Agg'})(1 - f_{Bhh'})\right]$$

$$f_{Agg'} = \begin{cases} f_{Ag'} & \text{if } g' \neq g \\ f_{Ag}^* & \text{if } g' = g \end{cases}$$

and

$$f_{Bhh'} = \begin{cases} f_{Bh'} & \text{if } h' \neq h \\ f_{Bh}^* & \text{if } h' = h. \end{cases}$$

These results can easily be extended to t independent samples.

An unbiased estimator of $V(\hat{Y}_{HT})$ is given by

$$v(\hat{Y}_{HT})$$

$$= \sum_g \sum_h \frac{W_{gh}(W_{gh} - 1)}{(W_{gh}^* + 1 - W_{gh})} \sum_i^{n_{gh}} y_{ghi}^2$$

$$- \sum_g \sum_h \sum_{g'} \sum_{h'} \frac{(W_{g'h'} - W_{ghg'h'})(W_{gh} - 1)}{W_{gh} W_{ghg'h'} - W_{g'h'}(W_{gh} - 1)} \hat{Y}_{gh} \hat{Y}_{g'h'} \tag{2.4}$$

For sample designs more complex than stratified simple random sampling, the estimator $\hat{Y}_{HT}$ can still be used. The general form of the estimated variance formulae for $\hat{Y}_{HT}$ (Cochran (1977), pp. 261) may have to be applied, however.

3. USE OF AUXILIARY INFORMATION

Estimators with a smaller variance then Y_{HT} can be calculated using auxiliary information. In this section, the separate, combined and raking ratio estimators will be examined. Let x_{ghi} equal the value of the auxiliary variable for the i-th unit in cross-stratum gh. Also, $X_{gh} = \sum_i^{N_{gh}} x_{ghi}$, $x_{gh} = \sum_i^{n_{gh}} x_{ghi}$ and $y_{gh} = \sum_i^{n_{gh}} y_{ghi}$.

If X_{gh} is known, the separate ratio estimator

$$\hat{Y}_{Rs} = \sum_g \sum_h \frac{X_{gh}}{x_{gh}} y_{gh} \qquad (3.1)$$

can be used. An approximate variance formula can be found by substituting $u_{ghi} = y_{ghi} - R_{gh} x_{ghi}$ for y_{ghi} and $U_{gh} = \sum_i^{N_{gh}} u_{ghi} = 0$ for Y_{gh} in equation (2.3) where $R_{gh} = Y_{gh}/X_{gh}$.

If X_{gh} is not known or some of the cross-strata sample sizes n_{gh} are zero or close to zero, it may not by possible to use the separate ratio estimator. In this case, the combined ratio estimator

$$\hat{Y}_{Rc} = \frac{X}{\hat{X}_{HT}} \hat{Y}_{HT} \qquad (3.2)$$

could be applied with $X = \sum_g \sum_h X_{gh}$. An approximate variance formula can be found by substituting $u_{ghi} = y_{ghi} - Rx_{ghi}$ for y_{ghi} and $U_{gh} = \sum_i^{N_{gh}} u_{ghi}$ for Y_{gh} in equation (2.3) where $R = Y/X$.

In the situation where a separate ratio estimator cannot be used but $X_{g.}$ and $X_{.h}$ are known (summations over g and h are denoted by a dot), it may be possible to use a raking ratio estimator to achieve a lower variance than a combined ratio estimator would give. The raking estimation procedure (RREP) is an iterative process. The p-th interation raking ratio estimator of Y will be defined as

$$\hat{Y}^{(p)} = \sum_g \sum_h W_{gh}^{(p)} y_{gh} \qquad (3.3)$$

where if the raking process starts with the rows

$$W_{gh}^{(p)} = \begin{cases} W_{gh}^{(p-1)} \dfrac{X_{g.}}{\hat{X}_{g.}^{(p-1)}} & \text{if p is odd} \\[2ex] W_{gh}^{(p-1)} \dfrac{X_{.h}}{\hat{X}_{.h}^{(p-1)}} & \text{if p is even} \end{cases} \qquad (3.4)$$

$\hat{X}_{gh}^{(p-1)} = W_{gh}^{(p-1)} x_{gh}$ and $W_{gh}^{(0)} = 1/[1-(1-f_{Ag})(1-f_{Bh})]$.

For p even (using arguments similar to those given by Brackstone and Rao (1979))

$$V(\hat{Y}^{(p)}) \doteq V\left(\sum_g \sum_h W_{gh}^{(p-1)} (y_{gh} - R_{.h} x_{gh})\right) \qquad (3.5)$$

where $R_{.h} = Y_{.h}/X_{.h}$. Thus $V(\hat{Y}^{(p)})$ for even p cab be obtained from $V(\hat{Y}^{(p-1)})$ by substituting $y_{gh} - R_{.h} x_{gh}$ for y_{gh} in the latter. Similarly, for p odd

$$V(\hat{Y}^{(p)}) \doteq V\left(\sum_g \sum_h W_{gh}^{(p-1)} (y_{gh} - R_{g.} x_{gh})\right) \qquad (3.6)$$

where $R_{g.} = Y_{g.}/X_{g.}$. Equations (3.5) and (3.6) can be applied repeatedly until the linearized expression contains the weight $W_{gh}^{(0)}$. Then the variance formula $V(\hat{Y}_{HT})$ (equation (2.3)) can be applied to the linearized data. In Section 6 this is discussed further.

In estimating the variance from a sample, consistent estimators of $R_{g.}$ and $R_{.h}$ are used in the linearization process. Also, the formula for the unbiased estimator of $V(\hat{Y}_{HT})$ (equation (2.4)) is applied rather than the variance formula of $\hat{Y}_{HT}$ (equation (2.3)).

Because the RREP is an iterative procedure, it is useful to know when it converges and what are the characteristics of the estimates to which it converges. Ireland and Kullback (1968) have studied this question. An extension of a result presented in their paper is as follows. Assume all $n_{gh} > 0$. Then the RREP converges to the set of estimates $\hat{X}_{gh}$ which minimizes the function

$$\sum_g \sum_h \hat{X}_{gh} \ln (\hat{X}_{gh}/\hat{X}_{gh}^{(0)}) \qquad (3.7)$$

subject to the constraints $\hat{X}_{g.} = X_{g.}$ for all g and $\hat{X}_{.h} = X_{.h}$ for all h.

4. APPLICATION OF THE RESULTS TO A MULTIPLE FRAME SAMPLE DESIGN

In a two frame sample design (these will be called Frame A and Frame B), usually Frame A and Frame B each contains units the other frame does not have as well as units in common. In Section 2, it was assumed that two independent samples were selected from one frame. The situation of two independent samples selected from two frames can be accommodated by setting $n_{Ag} = 0$ for one stratum in Sample A of Section 2. This stratum will contain the units in Frame B that are not in Frame A. Similarly, for one stratum in Sample B, it is possible to set $n_{Bh} = 0$. This stratum will contain the units in Frame A that are not in Frame B.

It should be noted that if the Horvitz-Thompson estimator is used then $\sum_h \hat{N}_{gh}^{(0)}$ may not equal N_{Ag} and $\sum_g \hat{N}_{gh}^{(0)}$ may not equal N_{Bh}. This is because the cross-strata sample sizes n_{gh} are random variables. For this reason, there are potential reductions in variance possible by using the ratio estimators $\hat{Y}_{Rs}$ and $\hat{Y}^{(p)}$ of

Section 3 with $X_{gh} = N_{gh}$. The reductions in variance will be demonstrated in the numerical example given in Section 6.

5. STANDARD EXTIMATORS GIVEN IN THE LITERATURE

In this section, a brief review will be given of the approach suggested in the literature for producing estimates based on two independent stratified samples selected from separate frames. Hartley (1962) suggested using the estimator

$$\dot{Y} = \hat{Y}_a + \hat{Y}_b + p\,\hat{Y}'_{ab} + (1-p)\,\hat{Y}''_{ab} \qquad (5.1)$$

where $\hat{Y}_a$ is the estimator based on Sample A for the domain of units in Frame A not in Frame B. Similarly, $\hat{Y}'_{ab}$ is the estimator based on Sample A for the domain of units in both Frame A and Frame B. For $\hat{Y}_a$ and $\hat{Y}'_{ab}$, sampled units in stratum g are weighted by N_{Ag}/n_{Ag}. $\hat{Y}_b$ is an estimator based on Sample B for the domain of units in Frame B not in Frame A. $\hat{Y}''_{ab}$ estimates the same quantity as $\hat{Y}'_{ab}$ but is based on Sample B. For $\hat{Y}_b$ and $\hat{Y}''_{ab}$, sampled units in stratum h are weighted by N_{Bh}/n_{Bh}. With estimator $\dot{Y}$, it is assumed that sampled units which belong to both frames can be identified. It is also assumed that the number of units in the population which belong to both frames is not known. Duplicate units which fall in both samples are not eliminated. Also p is chosen such that $V(\dot{Y})$ is minimized. The value of p which achieves this is

$$p = \frac{\mathrm{Cov}\,(\hat{Y}_B, \hat{Y}''_{ab}) - \mathrm{Cov}\,(\hat{Y}_a, \hat{Y}'_{ab})}{V\,(\hat{Y}'_{ab}) + V\,(\hat{Y}''_{ab})} \qquad (5.2)$$

Using this value of p, the minimum value of $V(\dot{Y})$ is

$$V\,(\dot{Y}) = V\,(\hat{Y}_a) + V\,(\hat{Y}_B) - p^2\,(V\,(\hat{Y}'_{ab}) + V(\hat{Y}''_{ab})) \qquad (5.3)$$

where $\hat{Y}_B = \hat{Y}_b + \hat{Y}''_{ab}$.

If N_{gh} is known, then another estimator which can be considered is

$$\dot{Y}_s = \sum_g \sum_h \left(p_{Agh} \frac{N_{gh}}{n_{Agh}} y_{Agh} + p_{Bgh} \frac{N_{gh}}{n_{Bgh}} y_{Bgh}\right) \qquad (5.4)$$

where n_{Agh} equals the number of units selected in Sample A that fall in cross-stratum gh and y_{Agh} equals the sum for the characteristic of interest for the units in Sample A that fall in cross-stratum gh. n_{Bgh} and y_{Bgh} have definitions similar to n_{Agh} and y_{Agh}. The sum of p_{Agh} and p_{Bgh} equals one. $V\,(\dot{Y}_s)$ is minimized if

$$p_{Agh} = D_{Bh} \;/\; (D_{Ag} + D_{Bh}) \qquad (5.5)$$

where $D_{Ag} = N_{Ag}\,(N_{Ag} - n_{Ag}) / [n_{Ag}\,(N_{Ag} - 1)]$ and $D_{Bh} = N_{Bh}\,(N_{Bh} - n_{Bh}) / [n_{Bh}\,(N_{Bh} - 1)]$. In the case where $n_{Ag} = 0$ then $p_{Agh} = 0$. Also, if $n_{Bh} = 0$ then $p_{Bgh} = 0$. The minimum value of $V\,(\dot{Y}_s)$ is

$$V\,(\dot{Y}_s) = \sum_g \sum_h D_{Ag} D_{Bh}\,(N_{gh} - 1)\,S^2_{gh}/(D_{Ag} + D_{Bh}) \qquad (5.6)$$

where $S^2_{gh} = \sum_i^{N_{gh}} (y_{ghi} - \overline{Y}_{gh})^2 / (N_{gh} - 1)$ and $\overline{Y}_{gh} = Y_{gh}/N_{gh}$. Using the numerical example in Section 6, the variance of $\dot{Y}_s$ will be compared to the variance of $\hat{Y}_{Rs}$.

If N_{gh} is not known or if some of the cross-strata sample sizes n_{Agh} or n_{Bgh} are zero or close to zero, it may not be possible to use $\dot{Y}_s$. In that case, an estimator $\dot{Y}_d$ can be considered. It will have the same form as $\dot{Y}$ given in equation (5.1). The weight N_{aAg}/n_{aAg}, however, will replace N_{Ag}/n_{Ag} in the estimator $\hat{Y}_a$. N_{aAg} is the number of units in stratum Ag that do not belong to Frame B. n_{aAg} is the corresponding sample count. Similarly, the weight N_{abAg}/n_{abAg} will replace N_{Ag}/n_{Ag} in the estimator $\hat{Y}'_{ab}$ where $N_{abAg} = N_{Ag} - N_{aAg}$ and $n_{abAg} = n_{Ag} - n_{aAg}$. Similar modifications will be made to the estimators $\hat{Y}_b$ and $\hat{Y}''_{ab}$. Equation (5.2) gives the value of p which minimizes $V(\dot{Y}_d)$. Using the numerical example in Section 6, the variance of $\dot{Y}_d$ will be compared to the variance of the raking ratio estimator $\hat{Y}^{(p)}$.

6. AN EXAMPLE

A simple numerical example constructed from artificial data is presented in Table 1. In this example, both Sample A and Sample B have two strata. Units that belong to Frame B but not to Frame A are placed in stratum A3 with $n_{A3} = 0$. Similarly, units that belong to Frame A but not to Frame B are placed in stratum B3 with $n_{B3} = 0$.

Estimators which incorporate the same amount of information regarding population counts from Table 1 will be compared. The first group of estimators to be considered are those which require knowledge of only N_{A1}, N_{A2}, N_{B1} and N_{B2}. The estimators $\dot{Y}$, $\hat{Y}_{HT}$ and $\hat{Y}_2^{(20)}$ belong to this group. $\hat{Y}_2^{(20)}$ represents the twentieth iteration raking ratio estimator where $X_{g.} = N_{Ag}$ and $X_{.h} = N_{Bh}$ and only the first two rows and columns of Table 1 had the RREP applied. The weights for $\hat{Y}_2^{(20)}$ were calculated iteratively as follows.

TABLE 1: ARTIFICIAL DATA FOR THE EXAMPLE

	Stratum 1 of Frame B	Stratum 2 of Frame B	Units that do not Belong to Frame B	Total
	$N_{11} = 87.0$	$N_{12} = 63.0$	$N_{13} = 30.0$	$N_{A1} = 180.0$
Stratum 1 of Frame A	$Y_{11} = 430.0$	$Y_{12} = 180.0$	$Y_{13} = 110.0$	$\sum_h Y_{1h} = 720.0$
	$\sum_i y^2_{11i} = 2{,}566.0$	$\sum_i y^2_{12i} = 635.0$	$\sum_i y^2_{13i} = 742.0$	$\sum_{h,i} y^2_{1hi} = 3{,}943.0$
	$s^2_{11} = 5.1$	$s^2_{12} = 1.9$	$s^2_{13} = 11.7$	$n_{A1} = 18.0$
	$N_{21} = 45.0$	$N_{22} = 30.0$	$N_{23} = 15.0$	$N_{A2} = 90.0$
Stratum 2 of Frame A	$Y_{21} = 450.0$	$Y_{22} = 170.0$	$Y_{23} = 100.0$	$\sum_h Y_{2h} = 720.0$
	$\sum_i y^2_{21i} = 5{,}290.0$	$\sum_i y^2_{22i} = 1{,}115.0$	$\sum_i y^2_{23i} = 1{,}134.0$	$\sum_{h,i} y^2_{2hi} = 7{,}539.0$
	$s^2_{21} = 18.0$	$s^2_{22} = 5.2$	$s^2_{23} = 33.4$	$n_{A2} = 18.0$
	$N_{31} = 18.0$	$N_{32} = 12.0$	$N_{33} = 0.0$	$N_{A3} = 30.0$
Units that do not Belong to Frame A	$Y_{31} = 170.0$	$Y_{32} = 70.0$	$Y_{33} = 0.0$	$\sum_h Y_{3h} = 240.0$
	$\sum_i y^2_{31i} = 1{,}728.0$	$\sum_i y^2_{32i} = 449.0$	$\sum_i y^2_{33i} = 0.0$	$\sum_{h,i} y^2_{3hi} = 2{,}177.0$
	$s^2_{31} = 7.2$	$s^2_{32} = 3.7$	$s^2_{33} = 0.0$	$n_{A3} = 0.0$
	$N_{B1} = 150.0$	$N_{B2} = 105.0$	$N_{B3} = 45.0$	$N = 300.0$
Total	$\sum_g Y_{g1} = 1{,}050.0$	$\sum_g Y_{g2} = 420.0$	$\sum_g Y_{g3} = 210.0$	$\sum_{g,h} Y_{gh} = 1{,}680.0$
	$\sum_{g,i} y^2_{g1i} = 9{,}584.0$	$\sum_{g,i} y^2_{g2i} = 2{,}199.0$	$\sum_{g,i} y^2_{g3i} = 1{,}876.0$	$\sum_{g,h,i} y^2_{ghi} = 13{,}659.0$
	$n_{B1} = 53.0$	$n_{B2} = 21.0$	$n_{B3} = 0.0$	

For p odd, $W^{(p)}_{gh} = W^{(p-1)}_{gh} \; X_{g.} / \hat{X}^{(p-1)}_{g.}$ for g = 1,2 and $W^{(p)}_{gh} = W^{(p-1)}_{gh}$ for g = 3. For p even, $W^{(p)}_{gh} = W^{(p-1)}_{gh} \; X_{.h} / \hat{X}^{(p-1)}_{.h}$ for h = 1,2 and $W^{(p)}_{gh} = W^{(p-1)}_{gh}$ for h = 3. The iterative process is carried out to the twentieth iteration because there are significant reductions in the variance of $\hat{Y}^{(p)}_2$ as a result. Table 2 gives the variances (rather than estimated variances) of the estimators $\dot{Y}$, $\hat{Y}_{HT}$ and $\hat{Y}^{(20)}_2$. The variance of $\hat{Y}^{(20)}_2$ is smaller than the variance of the other estimators. It is significantly smaller when the estimators are restricted to the domain common to both frames.

TABLE 2: VARIANCE OF ESTIMATORS USING DATA FROM TABLE 1

	Domain	
	All Units in Population	Those Units in Both Frame A and Frame B
$V(\dot{Y})$	12,089.6	5,274.9
$V(\hat{Y}_{HT})$	12,966.3	4,832.6
$V(\hat{Y}^{(20)}_2)$	11,242.4	2,173.2
$V(\dot{Y}_d)$	9,045.6	3,700.4
$V(\hat{Y}^{(4)}_3)$	7,551.7	2,152.5
$V(\dot{Y}_s)$	7,663.5	2,318.3
$V(\hat{Y}_{Rs})$	7,400.0	2,054.8

The second group of estimators to be considered are those which require knowledge of N_{abA1}, N_{abA2}, N_{aA1} and N_{aA2} as well as the corresponding counts for Frame B. Since $N_{A3} = N_{bB1} + N_{bB2}$ and $N_{B3} = N_{aA1} + N_{aA2}$, these are known as well. The estimators $\dot{Y}_d$ and $\hat{Y}^{(4)}_3$ belong to this second group. $\hat{Y}^{(4)}_3$ represents a fourth iteration raking ratio estimator where $X_{g.} = N_{Ag}$ and $X_{.h} = N_{Bh}$ and all three rows and columns had the RREP applied. The variance of $\hat{Y}^{(p)}_3$ did not become much smaller after the fourth iteration. Table 2 shows that the variance of $\hat{Y}^{(4)}_3$ is significantly smaller than $\dot{Y}_d$.

The final group of estimators to be considered are those which require knowledge of the cross-strata population counts N_{gh}. The estimators $\hat{Y}_{Rs}$ and $\dot{Y}_s$ belong to this group. Table 2 shows that $\hat{Y}_{Rs}$ has a smaller variance than $\dot{Y}_s$. It can be seen as well that $\hat{Y}^{(4)}_3$ has a smaller variance than $\dot{Y}_s$.

In the calculation of $V(\hat{Y}^{(4)}_3)$ and $V(\hat{Y}^{(20)}_2)$, the values of the observations were linearized repeatedly. Then the linearized results were substituted into the formula for $V(\hat{Y}_{HT})$. This was done rather than explicitly deriving the formula for $V(\hat{Y}^{(4)}_3)$ and $V(\hat{Y}^{(20)}_2)$. In general, to calculate $V(\hat{Y}^{(p)}_k)$, the following operations are performed. Define

$$\sum_i^{N_{gh}} (u^{(t-1)}_{ghi})^2 = \sum_i^{N_{gh}} (u^{(t)}_{ghi})^2 - 2\, R^{(t)}_{.h}\, U^{(t)}_{gh} + N_{gh}\, (R^{(t)}_{.h})^2 \qquad (6.1)$$

and

$$U^{(t-1)}_{gh} = U^{(t)}_{gh} - N_{gh}\, R^{(t)}_{.h} \quad \text{for } h = 1 \text{ to } k \qquad (6.2)$$

if t is even and

$$\sum_i^{N_{gh}} (u^{(t-1)}_{ghi})^2 = \sum_i^{N_{gh}} (u^{(t)}_{ghi})^2 - 2\, R^{(t)}_{g.}\, U^{(t)}_{gh} + N_{gh}\, (R^{(t)}_{g.})^2 \qquad (6.3)$$

and

$$U^{(t-1)}_{gh} = U^{(t)}_{gh} - N_{gh}\, R^{(t)}_{g.} \quad \text{for } g = 1 \text{ to } k \qquad (6.4)$$

if t is odd where $u^{(t)}_{ghi} = y_{ghi}$ if t = p, $R_{.h} = U^{(t)}_{.h} / N_{.h}$ and $R^{(t)}_{g.} = U^{(t)}_{g.} / N_{g.}$. Equations (6.1) to (6.4) are applied iteratively starting with t = p until $\sum_i^{N_{gh}} (u^{(0)}_{ghi})^2$ and $U^{(0)}_{gh}$ are found. Then $\sum_i^{N_{gh}} y^2_{ghi}$ and Y_{gh} are replaced with $\sum_i^{N_{gh}} (u^{(0)}_{ghi})^2$ and $U^{(0)}_{gh}$ in the variance formula for $V(\hat{Y}_{HT})$. The result will be the approximate variance of $\hat{Y}^{(p)}_k$. Linearizing the data rather than explicitly deriving the formula for $V(\hat{Y}^{(p)}_k)$ has the advantage that it is very simple to program. Also, $V(\hat{Y}^{(p)}_k)$ for any p can be calculated.

7. CONCLUSIONS

When stratified samples are selected from different frames, the raking ratio estimator

$\hat{Y}_k^{(p)}$ offers the potential of variances considerably lower than those of the estimators suggested by Hartley (1962). Also, it is computationally easy to extend $\hat{Y}_k^{(p)}$ to a sample design based on any number of frames. The approach of Hartley, however, becomes complex as the number of frames is increased.

REFERENCES

Brackstone, G.J. and Rao, J.N.K. (1979), 'An Investigation of Raking Ratio Estimators', *Sankhyā*, Volume 41, Series C. pt. 2, 97-114.

Cochran, William G. (1977), *Sampling Techniques*, New York: John Wiley and Sons.

Ireland, C.T. and Kullback, S. (1968), 'Contingency Tables with Given Marginals', *Biometrika*, 55, 179-188.

Hartley, H.O. (1962), 'Multiple Frame Surveys', *Proceedings of the Social Science Section of the American Statistical Association Meeting*, Minneapolis, Minnesota.

Hartley, H.O. (1974), 'Multiple Frame Methodology and Selected Applications', *Sankhyā*, Volume 36, Series C. Pt. 3, 99-118.