Jour. Ind. Soc. Ag. Statistics
Vol. XXXVIII, No. 3 (1986), pp. 370-378

ROBUSTNESS OF HARTLEY'S ESTIMATOR FOR MULTIPLE FRAME SURVEYS

B. C. SAXENA, P. NARAIN and A. K. SRIVASTAVA
Indian Agricultural Statistics Research Institute, New Delhi-12

(Received : March, 1983)

SUMMARY

For estimating the population total in multiple frame surveys, Hartley [1], [3] considered the optimum values of 'p' and the sample sizes. The departures in these optimum values are likely to vitiate the optimum nature of Hartley's estimator. In this paper the effect of such departures on the efficiency of the estimator has been investigated. The Hartley's estimator appears to be fairly robust with respect to moderate departures in the values of 'p' but it is not so with departures in optimality of sample sizes.

Keywords : Optimum sample size; Robustness; Departure from optimality; Efficiency of Estimator.

1. Introduction

Hartley [1], [3] formulated the problem of estimating population total of a character through multiple frame surveys approach. In multiple frame survey situations, more than one frames are available, normally with a larger frame in which sampling is costly and with other smaller frames for cheaper sampling methods. These frames are usually overlapping such that all the frames taken together cover the entire population. Instances of multiple frames are quite common in many sample survey situations. For example, for estimating the milk production the frame of all milk producing households may be taken as the larger frame while a list of commercial milk producers as the smaller frame. Evidently sampling from the former frame is costlier than the latter.

In multiple frame situations, the population is considered to be divided into a number of domains depending upon the overlap of various frames. In a two frame situation with frames A and B, three domains emerge as domain (a), (b) and (ab), having respectively set of units belonging to frame A only, B only and to both the frames A and B. Hartley proposed an estimator for the population total based on independent samples from the two frames. His approach consists of estimating the domain totals separately. For the common domain (ab), the domains total is estimated by pooling its independent estimators based on the two samples with weights p and $1 - p$.

In Hartley's approach the sample sizes and the weight p are optimized. These optima consisted of several parametric values, some of which are seldom known in actual practice. It is generally assumed that some knowledge about these parametric values is available through pilot surveys or other alternative sources. When approximate values of these parameters are used, the optimality of the estimator is likely to be vitiated. In this paper, the effect of departures in the optimum values on the efficiency of estimator is investigated. Hartley's estimation procedure is briefly described in the next section.

2. Robustness of the Hartley's Estimator

Consider two independent simple random samples from frames A and B. The populations are assumed to be large enough, so that finite population corrections are ignored. Defining the following notations :

Items	*Frame*		*Domains*		
	A	*B*	*(a)*	*(b)*	*(ab)*
Population size	N_A	N_B	N_a	N_b	N_{ab}
Sample size	n_A	n_B	n_a	n_b	n_{ab}, n_{ba}
Sample mean	$\bar{y}_A$	$\bar{y}_B$	$\bar{y}_a$	$\bar{y}_b$	$\bar{y}_{ab}, \bar{y}_{ba}$
Cost of sampling units	C_A	C_B	—	—	—
Population variances	—	—	σ_a^2	σ_b^2	σ_{ab}^2

where $\bar{y}_{ab}$ and $\bar{y}_{ba}$ are sample means based on n_{ab} and n_{ba} units belonging to domain **(ab)** and coming from frames A and B respectively. Further defining,

$$\alpha = \frac{N_{ab}}{N_A},\ \beta = \frac{N_{ab}}{N_B},\ \rho = \frac{C_A}{C_B} \text{ and } \phi = \frac{\sigma_a^2}{\sigma_{ab}^2}$$

Hartley's [1] estimator is given by

$$\hat{Y}_H = N_a\bar{y}_a + N_{ab}(p\bar{y}_{ab} + q\bar{y}_{ba}) + N_b\bar{y}_b \tag{1}$$

$$V(\hat{Y}_H) = \frac{N_A^2}{n_A}\left[\sigma_a^2(1-\alpha) + p^2\sigma_{ab}^2\alpha\right] + \frac{N_B^2}{n_B}\left[\sigma_b^2(1-\beta) + q^2\sigma_{ab}^2\beta\right] \tag{2}$$

the optimum 'p' was obtained subject to a linear cost function of the form

$$C = c_A n_A + c_B n_B \tag{3}$$

where C is the total cost and $c_A > c_B$.

A special case of 100% coverage by the frame A commonly met in practice leads to $N_b = 0$ and $\beta = 1$. In this case the optimum 'p' (p_0, say) is given by

$$p_0^2 = \frac{\phi(1-\alpha)}{(\rho-\alpha)} \tag{4}$$

Also the optimum sample sizes n_{A_0}, n_{B_0} are as follows :

$$\frac{n_{A0}}{N_A} = K_0\sqrt{\frac{\sigma_a^2(1-\alpha) + \alpha p_0^2\sigma_{ab}^2}{C_A}} \tag{5}$$

$$\frac{n_{B0}}{N_B} = K_0\sqrt{\frac{q_0^2\sigma_{ab}^2}{C_B}} \tag{6}$$

With K_0 determined to meet the budget cost C given by

$$C = C_A n_{A_0} + C_B n_{B_0} \tag{7}$$

When some of the population parameters required for optimum n_A, n_B and 'p' are not known, their close guessed values are commonly used. One situation may be when all the three values are arbitrary. Alternatively formula (4), (5) and (6) may be used with guess values of the unknown parameters. It is seen that p_0, n_{A_0} and n_{B_0} are function of σ_a^2, σ_{ab}^2, α, C_A, C_B and C. Assume that α, C_A, C_B and C are known whereas σ_a^2 and σ_{ab}^2 are not known, with $\sigma_a'^2$ and $\sigma_b'^2$ as their guessed values let p', n_A' and n_B' be the values of p, n_A and n_B as obtained from (4), (5) and (6) respectively. In the situation when p, n_A and n_B are arbitrary guessed values, denote them by p'', n'' and n_B'' respectively. Let $V(\hat{Y}_H)$ as given in (2), be denoted by $V(n_A, n_B, p)$. Denote this variance with $V(n_{A_0}, n_{B_0}, p_0)$, $V(n_A', n_B', p')$ and $V(n_A'', n_B'', p'')$ for the values of (n_A, n_B, p) as (n_{A_0}, n_{B_0}, p_0), (n_A', n_B', p') and (n_A'', n_B', p'') respectively. The percentage relative departure in variance corresponding to the two situations, when (n_A', n_B', p') and

(n_A, n''_B, p'') are used, may be given as

$$D^* = \left[\frac{V(n'_A, n'_B, p')}{V(n_{A_0}, n_{B_0}, p_0)} - 1\right] \times 100 \tag{8}$$

and

$$D^{**} = \left[\frac{V(n''_A, n_B, p'')}{V(n_{A_0}, n_{B_0}, p_0)} - 1\right] \times 100 \tag{9}$$

These expressions are simplified as follows :

*Calculation of D**

When σ'^2_a and σ'^2_{ab} are used as the guessed values for σ^2_a and σ^2_{ab}, n'_A and n'_B are given by

$$\frac{n'_A}{n_A} = K' \sqrt{\frac{\sigma'^2_r (1-\alpha) + p'^2 \sigma'^2_{ab} \alpha}{C_A}} \tag{10}$$

$$\frac{n'_B}{N_B} = K' \sqrt{\frac{q'^2 \sigma'^2_{ab}}{C_B}} \tag{11}$$

where K', is determined to meet the cost

$$C = C_A n'_A + C_B n'_B \tag{12}$$

Therefore,

$$\frac{V(n'_A, n'_B, p')}{V(n_{A_0}, n_{B_0}, p_0)} = \frac{K_0\,\sigma_{ab}}{K'\,\sigma'_{ab}} \left[\frac{(\xi_2^2 + \alpha\lambda'^2)\,\xi_1 + \alpha\lambda'\,(\xi_2 - \lambda'\xi_1)}{\lambda'(\rho\xi_1 + \alpha\,(\xi_2 - \xi_1))}\right]$$

where

$$\lambda' = \frac{p'}{p_0} \text{ and } p_0 = \frac{\xi_1}{\xi_2} \; (< 1)$$

and

$$\xi_1 = \sqrt{\phi(1-\alpha)},\; \xi_2 = \sqrt{\rho - \alpha}$$

From (7) and (12) we get

$$C_A n_{A_0} + C_B n_{B_0} = C_A n'_A + C_B n'_B$$

On substituting the values of n'_A, n'_B, n_{A_0} and n_{B_0} and after a little simplification, we get

$$\frac{K_0\,\sigma_{ab}}{K'\,\sigma'_{ab}} = \frac{\rho\,\lambda'\xi_1 + \alpha\,(\xi_2 - \lambda'\xi_1)}{\rho\xi_1 + \alpha(\xi_2 - \xi_1)} \tag{13}$$

Thus,

$$\frac{V(n_A, n'_B, p')}{V(n_{A_0}, n_{B_0}, p_0)} = \frac{[(\rho\lambda'\xi_1 + \alpha(\xi_2 - \lambda'\xi_1)] [(\xi_2^2 + \alpha\lambda'^2)\xi_1 + \alpha\lambda'(\xi_2 - \lambda'\xi_1)]}{\lambda' [\rho\xi_1 + \alpha(\xi_2 - \xi_1)]^2}$$

Therefore, D^* as defined in (8), simplified to

$$D^* = \frac{\xi_2^2 (\xi_1\xi_2 \lambda' + \alpha)(\xi_1\xi_2 + \alpha\lambda') - \lambda' [(\xi_2^2 + \alpha)\xi_1 + \alpha(\xi_2 - \xi_1)]^2}{\lambda' [(\rho - \alpha)\xi_1 + \alpha\xi_2]^2} \times 100$$

$$= \frac{\xi_2^2 (\alpha\xi_1 \xi_2 + \alpha\lambda'^2 \xi_1 \xi_2 - 2\lambda' \alpha\xi_1 \xi_2)}{\lambda' (\xi_1 \xi_2^2 + \alpha\xi_2)^2} \times 100$$

or

$$D^* = \frac{\alpha\xi_1 \xi_2 (1 - \lambda')^2}{\lambda' (\xi_1 \xi_2 + \alpha)^2} \times 100 \qquad (14)$$

Calculation of D^{**}

It is clear that,

$$\frac{V(n''_A, n''_B, p'')}{V(n_{A_0}, n_{B_0}, p_0)} = \frac{r_0/r_A [\phi(1 - \alpha) + \alpha\lambda''^2 \xi_1^2/\xi_2^2] + \alpha^2/r_B (1 - \lambda'' \xi_1/\xi_2)^2}{r_0 [\phi(1 - \alpha) + \alpha \xi_1^2/\xi_2^2] + \alpha^2 (1 - \xi_1/\xi_2)^2}$$

where

$$\lambda'' = \frac{p''}{p_0}, \; r_A = \frac{n''_A}{n_{A_0}}, \; r_B = \frac{n''_B}{n_{B_0}},$$

and

$$r_0 = \frac{n_{B_0}}{n_{A_0}} = \frac{N_B \sqrt{c_A q_0^2 \sigma_{ab}^2}}{N_A \sqrt{c_B [\sigma_a^2 (1 - \alpha) + \alpha p_0^2 \sigma_{ab}^2]}}$$

After little simplification r_0 reduces to

$$r_0 = \frac{\alpha(\xi_2 - \xi_1)}{\xi_1}$$

r_A, r_B and r_0 are evidently related as follows :

$$r_B = 1 + \frac{(1 - r_A)\rho}{r_0}$$

Thus if r_A is assigned values less than unity the corresponding r_B can be obtained, Now

$$D^{**} = \left[\frac{r_0/r_A\,(\xi_1^2\,(\xi_2^2 + \alpha\lambda''^2) + \alpha^2/r_B\,(\xi_2 - \lambda''\,\xi_1)^2}{r_0\,\xi_1^2\,\rho + \alpha^2\,(\xi_2 - \xi_1)^2} - 1 \right] \times 100$$

$$= \left[\frac{\xi_1(\xi_2 - \xi_1)\,(\xi_2^2 + \alpha\lambda''^2)/r_A + \alpha(\xi_2 - \lambda''\xi_1)^2/r_B}{\xi_2(\xi_2 - \xi_1)\,(\xi_1\xi_2 + \alpha)} - 1 \right] \times 100 \quad (15)$$

It is seen that the expression of D^* and D^{**} are considerably simplified for different values of ϕ, ρ, λ', λ'', and α. The values of D^* and D^{**} have been presented in Tables 1 and 2 respectively. Table 1 indicates the relative departure in the variance when close guessed values of σ_a^2 and σ_{ab}^2 are used in formulae (4), (5) and (6). Table 2 provides the corresponding departure when arbitrary guessed values of p, n_A and n_B are used. In these tables reasonable values of ϕ, ρ, α, λ' and λ'' are considered. The values of $\phi^{-1} = \sigma_{ab}^2/\sigma_a^2$ are chosen to lie between 1/4 and 4 in order to cover differential variabilities in the domains (ab) and (a). The values of α and $\rho^{-1} = c_B/c_A$ are taken to be less than unity as α is a proportion and the frame A is assumed to be costlier than the frame B. The values of λ' and λ'' are also taken less than unity just for convenience. The corresponding values of D^* for λ' more than unity may be obtained from Table 1, using (14). It is evident that when $n_A'' = n_A'$, D^{**} reduces to D^*. Therefore, for a reasonable value of r_A, it is of interest to examine the ratio n_A'/n_{A0} for

TABLE 1—PERCENTAGE RELATIVE DEPARTURE OF $V(n_A', n_B', p')$ FROM $V(n_{A0}, n_{B0}, p_0)$

For $\rho^{-1} = 0.1$

ϕ^{-1}	α	*0.25*	*0.75*	*0.9*
1/4	.5	20.7	0.8	0.1
	.7	32.2	1.2	0.1
	.95	54.6	2.0	0.3
1	.5	34.2	1.3	0.2
	.7	46.8	1.7	0.2
	.95	54.6	2.0	0.3
4	.5	48.5	1.8	0.2
	.7	55.8	2.0	0.3
	.95	43.5	1.6	0.2

TABLE 2—PERCENTAGE RELATIVE DEPARTURE OF $V(n''_A, n''_B, p'')$ FROM $V(n_{A0}, n_{B0}, P_0)$

For $\rho^{-1} = 0.1$

ϕ^{-1}	r_A / α/λ''	0.5			0.7			0.9		
		.25	.75	.9	.25	.75	.9	.25	.75	.9
1/4	.5	81.6	85.6	87.8	31.0	33.3	34.7	5.5	5.7	6.4
	.7	70.6	75.2	77.9	23.3	27.6	28.9	4.7	3.8	4.3
	.95	41.5	44.2	46.2	13.3	13.4	14.4	3.6	1.4	1.5
	.5	68.6	71.7	73.4	24.3	25.6	26.6	4.2	3.4	3.8
1	.7	54.4	57.4	59.4	18.1	18,8	19.9	3.7	2.1	2.3
	.95	26.3	27.4	28.7	8.2	7.7	8.2	2.5	0.8	0.8
	.5	52.2	54.1	55.4	17.1	17.5	18.2	3.0	2.0	2.1
	.7	37.4	39.0	40.3	11.7	11.6	12.2	2.7	1.2	1.2
4	.95	15.2	15.6	16.3	4.7	4.2	4.4	1.6	0.4	0.4

different values of ϕ, ρ, α and λ'. From formulae (5), (10) and (13) it is seen that n'_A/n_{A_0} simplifies to

$$\frac{n'_A}{n_{A_0}} = \frac{\alpha + \xi_1 \xi_3}{\alpha/\lambda' + \xi_1 \xi_3}$$

These values were calculated and were found to be less than unity for different ϕ, ρ, α and λ' considered here. Therefore, the values of r_A in Table 2 have been chosen less than unity.

From Table 1, it is seen that when sample sizes n_A and n_B are obtained from the formulae (5) and (6) the estimator $\hat{Y}_H$ is fairly robust in respect to moderate departures in p_0. However, from Table 2, it is apparent that when the sample sizes n''_A, n''_B are chosen arbitrarily, the estimator is not always robust with respect to departure in p_0, particularly when r_A is small such as 0 5. It is also seen from Table 1 that D^* which is zero for $\alpha = 1$, by and large goes on increasing as α increases from 0.5 to 0.95. Thus a sharp decline in the value of D^* is expected for α larger than .95.

A graphical representation (Figure 1) of D^* and D^{**} is also presented for $r_A = 0.7$, $\phi^{-1} = 1/4, 1, 4$; $\rho^{-1} = .1, .4$, $\alpha = 0.5, 0.6, 0.7, 0.8, 0.95$ and λ', λ'' = (say, λ) = .25, .5, .75, .9. It is seen that with increase in λ D^* decreases while D^{**} increases. For larger values of λ, D^* is smaller than D^{**} while for smaller λ such as 0.5 and below, D^{**} is usually smaller than D^*. This indicates that if σ'^2_a and σ'^2_{ab} are expected to be close enough to σ^2_a and σ^2_{ab} respectively resulting in values of λ closer to unity. For $r_A = .9$, D^{**} is fairly small for all values of λ. For $r_A = .7$ and .5, D^{**} is quite substantial and it increases with λ. Thus, departures in optimality of sample sizes effects the variances more effectively as compared to departures in p_0. It is, therefore, advisable to get the sample sizes using the formulae for n'_A and n'_B when close guess values of optimum sample sizes are not available. Increase in D^{**} with increase in λ is expected since the values of n'_A/n_{A0} increase with λ.

REFERENCES

[1] Hartley, H. O. (1962) : Multiple frame surveys. *Proceedings of the Social Statistics Sections*, A.S.A. 203-208.

[2] Sukhatme, P. V. and Sukhatme B. V. (1970) : Sampling theory of surveys with applications. Iowa State University Press. AMES Iowa U.S.A.

[3] Hartley, H. O. (1974) : Multiple frame methodology and selected applications. *Sankhya*, 36, **Series C**, (3) : 99-113.

[4] Cochran, W. G. (1977) : Sampling technique (3rd Ed.) John Wiley and Sons, New York.