Survey Methodology, December 1985
Vol. 11, No. 2, pp. 119–132
Statistics Canada

Estimation of Total for Two Characters in Multiple Frame Surveys

B.C. SAXENA, P. NARAIN, and A.K. SRIVASTAVA[1]

ABSTRACT

In this paper estimation of multiple characters in multiple frame surveys has been investigated. The gain due to two character study in a common survey, over separate surveys for individual characters, has been obtained. Cost comparison is also made between two character multi frame survey and two character single frame survey.

KEY WORDS: Multi-character survey; Post-stratified estimate; Optimization; Cost comparison.

1. INTRODUCTION

The technique of multiple frame surveys was suggested by Hartley (1962) and subsequently discussed by Lund (1968), Hartley (1974), Vogel (1975), Armstrong (1979), etc. Lund suggested an alternate to Hartley's estimator utilizing the actual division in the sample among various domains. Hartley (1974) further considered the problem with more general approach applicable to various sampling designs. He observed that most potential multiple frame situations employed different types of units in their respective frames. Bosecker and Ford (1976) extended Hartley's estimator to take advantage of stratification within the overlap domain. Serrurier and Phillips (1976) and Armstrong (1978) tested multiple frame techniques in agricultural surveys. The utility of multiple frame survey has been demonstrated in a wide variety of situations. In sample surveys, sometimes interest lies not only in the estimation of single character but several characters are required to be studied simultaneously. For a proper utilization of resources this is often achieved through integrated surveys. For instance, for estimating the production of vegetable crops, a single survey is planned to estimate the production of several vegetable crops. Also, besides the frame of all vegetable growers, another incomplete but relatively easily accessible frame of important vegetable growers may be utilized. In this paper, the estimation of total for two characters in multiple frame surveys has been considered. The advantage of studing more than one character in a single survey over the situation when independent surveys are planned for individual characters in a multiple frame situation, is also investigated.

2. ESTIMATOR

Let there be two overlapping frames A and B of sizes N_A and N_B respectively. In multiple frame surveys two samples of sizes n_A and n_B are selected independently by simple random sampling from frames A and B respectively. The overlapping frames generate domains a, b and ab defined as follows:

a: Consisting of units belonging to frame A only,
b: Consisting of units belonging to frame B only,
ab: Units belonging to both A and B frames.

[1] B.C. Saxena, P. Narain, and A.K. Srivastava, Indian Agricultural Statistics Research Institute, New Delhi, India.

The sample sizes n_A and n_B are split into sizes n_a, n_{ab} and n_b, n_{ba} such that n_a and n_{ab} are the number of units out of n_A units belonging to domains a and ab respectively. Similarly n_b and n_{ba} are the split of n_B units belonging to domains b and ab respectively. In the multi-character study, there will be further split of these domains generating sub-domains as follows:

Let there be two characters $y_{(1)}$ and $y_{(2)}$ under study. Then each of the usual domains a, ab and b are further subdivided as $a(1)$, $a(12)$, $a(2)$, $ab(1)$, $ab(12)$, $ab(2)$ and $b(1)$, $b(12)$, $b(2)$ respectively. Here, $a(1)$, $a(12)$ and $a(2)$ are the sub-domains consisting of units having character $y_{(1)}$, both $y_{(1)}$ and $y_{(2)}$, and $y_{(2)}$ only respectively in domain a. Similar explanation holds for other sub-domains $ab(1)$, $ab(12)$ etc. Thus the sample split in two character study will be as follows:

$$n_A = n_a + n_{ab}$$

where

$$n_a = n_{a(1)} + n_{a(2)} + n_{a(12)} \quad \text{and} \quad n_{ab} = n_{ab(1)} + n_{ab(2)} + n_{ab(12)},$$

and

$$n_B = n_b + n_{ba}$$

where

$$n_b = n_{b(1)} + n_{b(2)} + n_{b(12)} \quad \text{and} \quad n_{ba} = n_{ba(1)} + n_{ba(2)} + n_{ba(12)}.$$

Here $n_{a(1)}$, $n_{a(2)}$, etc. are the split of n_a units belonging to sub-domains $a(1)$, $a(2)$, etc. If we confine to one character then define

$$n_{A(1)} = n_{a(1)} + n_{a(12)} + n_{ab(1)} + n_{ab(12)},$$

$$n_{B(1)} = n_{b(1)} + n_{b(12)} + n_{ba(1)} + n_{ba(12)}.$$

Similarly, for the second character, $n_{A(2)}$ and $n_{B(2)}$ are defined. The estimate of the total for the first character is given by

$$\hat{Y}^{(1)} = \hat{Y}_{a(1)} + \hat{Y}^{(1)}_{a(12)} + p_1\hat{Y}_{ab(1)} + q_1\hat{Y}_{ba(1)} + p_2\hat{Y}^{(1)}_{ab(12)} + $$

$$+ q_2\hat{Y}^{(1)}_{ba(12)} + \hat{Y}_{b(1)} + \hat{Y}^{(1)}_{b(12)} \tag{1}$$

where $\hat{Y}_{a(1)}$, $\hat{Y}^{(1)}_{a(12)}$, etc . are the estimated totals for character $y_{(1)}$ of the respective sub-domains. In the subsequent discussion, for the domains in which both the characters are available, the super script corresponds to the character under consideration. For the domains having only one character the super script is not used since the domain evidently corresponds to the character.

Also, $p_1 + q_1 = 1$ and $p_2 + q_2 = 1$. Define $\bar{y}_{a(1)}$, $\bar{y}_{a(2)}$, etc. as the sample means for respective sub-domains for character $y_{(1)}$ and $y_{(2)}$ respectively.

Thus,

$$\hat{Y}^{(1)} = N_{a(1)}\bar{y}_{a(1)} + N_{a(12)}\bar{y}^{(1)}_{a(12)} + N_{ab(1)}(p_1\bar{y}_{ab(1)} + q_1\bar{y}_{ba(1)})$$
$$+ N_{ab(12)}(p_2\bar{y}^{(1)}_{ab(12)} + q_2\bar{y}^{(1)}_{ba(12)})$$
$$+ N_{b(12)}\bar{y}^{(1)}_{b(12)} + N_{b(1)}\bar{y}_{b(1)}. \tag{2}$$

Similarly for the second character, we have

$$\hat{Y}^{(2)} = N_{a(2)}\bar{y}_{a(2)} + N_{a(12)}\bar{y}^{(2)}_{a(12)} + N_{ab(2)}(p_3\bar{y}_{ab(2)} + q_3\bar{y}_{ba(2)})$$
$$+ N_{ab(12)}(p_4\bar{y}^{(2)}_{ab(12)} + q_4\bar{y}^{(2)}_{ba(12)}) + N_{b(12)}\bar{y}^{(2)}_{b(12)}$$
$$+ N_{b(2)}\bar{y}_{b(2)} \tag{3}$$

where

$$p_3 + q_3 = 1 \text{ and } p_4 + q_4 = 1.$$

2.1 Variance of the Estimator

The conditional variance of the post-stratified estimates $\hat{Y}^{(1)}$, $\hat{Y}^{(2)}$ for given sub-domain sample sizes ignoring the finite population correction may be written as

$$V(\hat{Y}^{(1)} \mid n_{a(1)}, n_{a(12)}, \text{etc.}) = N^2_{a(1)}\frac{\sigma^2_{a(1)}}{n_{a(1)}} + N^2_{a(12)}\frac{\sigma^{(1)^2}_{a(12)}}{n_{a(12)}}$$
$$+ N^{(2)}_{ab(1)}\left(p_1^2\frac{\sigma^2_{ab(1)}}{n_{ab(1)}} + q_1^2\frac{\sigma^2_{ab(1)}}{n_{ba(1)}}\right)$$
$$+ N^2_{ab(12)}\left(p_2^2\frac{\sigma^{(1)^2}_{ab(12)}}{n_{ab(12)}} + q_2^2\frac{\sigma^{(1)^2}_{ab(12)}}{n_{ba(12)}}\right) + N^2_{b(1)}\frac{\sigma^2_{b(1)}}{n_{b(1)}}$$
$$+ N^2_{b(12)}\frac{\sigma^{(1)^2}_{b(12)}}{n_{b(12)}} \tag{4}$$

The unconditional variance of $\hat{Y}^{(1)}$ is approximately given by

$$V(\hat{Y}^{(1)}) = \frac{N_A}{n_A}\left\{N_{a(1)}\sigma^2_{a(1)} + N_{a(12)}\sigma^{(1)^2}_{a(12)} + p_1^2N_{ab(1)}\sigma^2_{ab(1)}\right.$$
$$\left. + p_2^2N_{ab(12)}\sigma^{(1)^2}_{ab(12)}\right\} + \frac{N_B}{n_B}\left\{N_{b(1)}\sigma^2_{b(1)} + N_{b(12)}\sigma^{(1)^2}_{b(12)}\right. \tag{5}$$
$$\left. + q_1^2N_{ab(1)}\sigma^2_{ab(1)} + q_2^2N_{ab(12)}\sigma^{(1)^2}_{ab(12)}\right\}$$

which is equal to the variance for stratified sampling with proportional allocation.

Similarly,

$$V(\hat{Y}^{(2)}) = \frac{N_A}{n_A}\Big\{(N_{a(2)}\sigma^2_{a(2)} + N_{a(12)}\sigma^{(2)2}_{a(12)} + p_3^2 N_{ab(2)}\sigma^2_{ab(2)}$$

$$+ p_4^2 N_{ab(12)}\sigma^{(2)2}_{ab(12)}\Big\} + \frac{N_B}{n_B}\Big\{N_{b(2)}\sigma^2_{b(2)} + N_{b(12)}\sigma^{(2)2}_{b(12)}$$

$$+ q_3^2 N_{ab(2)}\sigma^2_{ab(2)} + q_4^2 N_{ab(12)}\sigma^{(2)2}_{ab(12)}\Big\} \tag{6}$$

where $\sigma^2_{a(1)}$, $\sigma^2_{a(2)}$, etc. are the variances for the two characters in the respective sub-domains.

For optimization of p_i's ($i = 1, 2, 3, 4$) for a common survey a combination of individual variances needs to be minimized subjet to the fixed total cost for the combined survey. Consider the simplest linear combination

$$F = V(\hat{Y}^{(1)}) + V(\hat{Y}^{(2)}).$$

For the common survey, a suitable cost function may be considered as follows:

$$C' = C_1(n_{a(1)} + n_{ab(1)}) + C_2(n_{a(12)} + n_{ab(12)}) + C_3(n_{a(2)} + n_{ab(2)})$$

$$+ C_4(n_{b(1)} + n_{ba(1)}) + C_5(n_{b(12)} + n_{ba(12)}) + C_6(n_{b(2)} + n_{ba(2)}) \tag{7}$$

where C_1 is the cost per unit in sub-domain $a(1)$, $ab(1)$; C_2 in $a(12)$, $ab(12)$; C_3 in $a(2)$, $ab(2)$ of frame A. Similarly C_4, C_5 and C_6 are the cost per unit from frame B. In the above cost function random sample sizes are involved. Consider the expected cost

$$C = E(C') = n_A(C_1\Phi_1 + C_2\Phi_2 + C_3\Phi_3) + n_B(C_4\Phi_4 + C_5\Phi_5 + C_6\Phi_6) \tag{8}$$

where

$$\Phi_1 = \frac{N_{a(1)} + N_{ab(1)}}{N_A}, \quad \Phi_2 = \frac{N_{a(12)} + N_{ab(12)}}{N_A},$$

$$\Phi_3 = \frac{N_{a(2)} + N_{ab(2)}}{N_A}, \quad \Phi_4 = \frac{N_{b(1)} + N_{ba(1)}}{N_B},$$

$$\Phi_5 = \frac{N_{b(12)} + N_{ba(12)}}{N_B}, \quad \Phi_6 = \frac{N_{b(2)} + N_{ba(2)}}{N_B}.$$

Or

$$C = n_A C_A + n_B C_B \tag{9}$$

where

$$C_A = C_1\Phi_1 + C_2\Phi_2 + C_3\Phi_3 \quad \text{and} \quad C_B = C_4\Phi_4 + C_5\Phi_5 + C_6\Phi_6.$$

In order to get the optimum p_i's as also n_A and n_B, the function F is to be minimised subject to the expected cost function as given in (9). The weight variables p_i's and sample sizes are obtained as follow using Lagrange multiplier:

$$\frac{P_1}{q_1} = \frac{P_2}{q_2} = \frac{P_3}{q_3} = \frac{P_4}{q_4} = \frac{N_B n_A}{n_B N_A} = \frac{P}{q} \text{ (say)}, \tag{10}$$

and

$$\frac{n_A^2}{N_A} = \gamma \frac{K_5 + K_1 p_1^2 + K_2 p_2^2 + K_3 p_3^2 + K_4 p_4^2}{C_A},$$

$$\frac{n_B^2}{N_B} = \gamma \frac{K_6 + K_1 q_1^2 + K_2 q_2^2 + K_3 q_3^2 + K_4 q_4^2}{C_B}, \tag{11}$$

with γ determined to meet the expected cost and

$$K_1 = N_{ab(1)} \sigma^2_{ab(1)},\ K_2 = N_{ab(12)} \sigma^{(1)^2}_{ab(12)},$$

$$K_3 = N_{ab(2)} \sigma^2_{ab(2)},\ K_4 = N_{ab(12)} \sigma^{(2)^2}_{ab(12)},$$

$$K_5 = N_{a(1)} \sigma^2_{a(1)} + N_{a(2)} \sigma^2_{a(2)} + N_{a(12)} (\sigma^{(1)^2}_{a(12)} + \sigma^{(2)^2}_{a(12)}),$$

$$K_6 = N_{b(1)} \sigma^2_{b(1)} + N_{b(2)} \sigma^2_{b(2)} + N_{b(12)} (\sigma^{(1)^2}_{b(12)} + \sigma^{(2)^2}_{b(12)}). \tag{12}$$

From (10) and (11), we get

$$\frac{q^2}{p^2} \frac{N_B}{N_A} \frac{C_B}{C_A} = \frac{K_6 + (K_I + K_2 + K_3 + K_4) q^2}{K_5 + (K_1 + K_2 + K_3 + K_4) p^2} \tag{13}$$

This is a bi-quadratic in p and can be solved for p. The optimum sampling fractions can be obtained from (11). A practical case commonly met in multiple frame situations is when one of the frames has got 100% coverage. Consider 100% coverage by the frame A then $N_{b(1)} = N_{b(2)} = N_{b(12)} = 0$.

In this case (13) reduces to

$$p^2 = \frac{\alpha}{\varrho - \alpha} \frac{K_5}{K_1 + K_2 + K_3 + K_4} \tag{14}$$

where

$$\varrho = \frac{C_A}{C_B} \quad \text{and} \quad \alpha = \frac{N_B}{N_A}.$$

Assume that

$$\sigma^2_{a(1)} = \sigma^{(1)2}_{a(12)},\ \sigma^2_{a(2)} = \sigma^{(2)2}_{a(12)},\ \sigma^2_{ab(1)} = \sigma^{(1)2}_{ab(12)},\ \sigma^2_{ab(2)} = \sigma^{(2)2}_{ab(12)}. \tag{15}$$

These assumptions appear plausible since the variability of one character is not likely to be affected by the presence or absence of the other character. Then p^2 reduces to

$$p^2 = \frac{\alpha}{\varrho - \alpha}\left\{\frac{\sigma^2_{a(1)}(N_{a(1)} + N_{a(12)}) + \sigma^2_{a(2)}(N_{a(2)} + N_{a(12)})}{\sigma^2_{ab(1)}(N_{ab(1)} + N_{ab(12)}) + \sigma^2_{ab(2)}(N_{ab(2)} + N_{ab(12)})}\right\}$$

or

$$p^2 = \frac{(1-\alpha)\Phi_2'}{(\varrho - \alpha)}\left\{\frac{\Phi_3'(\xi_1 + \xi_2) + (1 - \xi_1)}{\Phi_4'(\xi_3 + \xi_4) + (1 - \xi_3)}\right\} \tag{16}$$

where

$$\Phi_1' = \frac{\sigma^2_{a(1)}}{\sigma^2_{ab(1)}},\ \Phi_2' = \frac{\sigma^2_{a(2)}}{\sigma^2_{ab(2)}},\ \Phi_3' = \frac{\sigma^2_{a(1)}}{\sigma^2_{a(2)}},\ \Phi_4' = \frac{\sigma^2_{ab(1)}}{\sigma^2_{ab(2)}}$$

and

$$\xi_1 = \frac{N_{a(1)}}{N_a},\ \xi_2 = \frac{N_{a(12)}}{N_a},\ \xi_3 = \frac{N_{ab(1)}}{N_{ab}},\ \xi_4 = \frac{N_{ab(12)}}{N_{ab}}.$$

Using that $N_{ab} = N_B$, $N_{a(2)} + N_{a(12)} = N_a - N_{a(1)}$ and $N_{ab(2)} + N_{ab(12)} = N_{ab} - N_{ab(1)}$, it may be seen that the above expression of p^2 reduces to the usual form in uni-character case since $\xi_1 = \xi_3 = 1$ and $\xi_2 = \xi_4 = 0$. It may be remarked that the domain variances are generally not known as such these values are based either on prior knowledge or some guessed values. The optimality of p^2 is effected to that extent.

3. COMPARISON OF MULTI-CHARACTER SURVEY WITH INDEPENDENT UNI-CHARACTER SURVEYS IN MULTIPLE FRAME SITUATIONS

Multi-character surveys are planned with a view to economise the available resources and it is expected that a common survey is likely to score over independent uni-character surveys taking into account the cost and efficiency. In this situation the extent of gain due to a common multiple frame survey is investigated.

In a single character study for character $y_{(1)}$ (say), consider simple random samples of sizes n_A and n_B from the frames A and B respectively. Here we assume that the only frames used before are available, not the reduced frame for each character. Define N_{A1}, N_{B1}, n^*_{A1}, and n^*_{B1} as the population sizes and sample sizes respectively with character $y_{(1)}$. Here,

n_{A1}^* and n_{B1}^* are the random sample sizes with $E(n_{A1}^*) = n_A N_{A1}/N_A$ and $E(n_{B1}^*) = n_B N_{B1}/N_B$. In this case, the estimator $\hat{Y}^{(1)*}$ and its variance are as follows:

$$\hat{Y}^{(1)*} = (N_{a(1)} + N_{a(12)})\bar{y}_{(a(1),\, a(12))}$$

$$+ (N_{ab(1)} + N_{ab(12)})(p'\bar{y}_{(ab(1),\, ab(12))} + q'\bar{y}_{(ba(1),\, ba(12))})$$

$$+ (N_{b(1)} + N_{b(12)})\bar{y}_{(b(1),\, b(12))}$$

where p', q' are weight variables such that $p' + q' = 1$ and $\bar{y}_{(a(1),\, a(12))}$, $\bar{y}_{(ab(1),\, ab(12))}$, etc. are sample means for the sample from combined respective domains, e.g. $\bar{y}_{(a(1),\, a(12))}$ is the mean of sample units coming from domain $a(1)$ and $a(12)$.

$$V(\hat{Y}^{(1)*}) = \frac{N_A}{n_A}(N_{a(1)}\sigma^2_{a(1)} + N_{a(12)}\sigma^{(1)2}_{a(12)}$$

$$+ (p'^2\frac{N_A}{n_A} + q'^2\frac{N_B}{n_B})(N_{ab(1)}\sigma^2_{ab(1)} + N_{ab(12)}\sigma^{(1)2}_{ab(12)})$$

$$+ \frac{N_B}{n_B}(N_{b(1)}\sigma^2_{b(1)} + N_{b(12)}\sigma^{(1)2}_{b(12)}). \quad (17)$$

In this case, the cost function is of the form

$$C = C_1 n_{A1}^* + C_4 n_{B1}^*$$

and expected cost is given by C^* as

$$C^* = C_1\frac{n_A}{N_A}N_{A1} + C_4\frac{n_B}{N_B}N_{B1} = C'_A n_A + C'_B n_B \quad (18)$$

where $C'_A = C_1 N_{A1}/N_A$ and $C'_B = C_4 N_{B1}/N_B$.

For simplicity, we assume 100% coverage by frame A, equality of variances as in (15), and $C_4/C_1 = C_5/C_2 = C_6/C_3 = K$. Based on these assumptions, the cost C^* with n_A and n_B which minimize the variance (17) is given by (see Appendix for derivation).

$$C^* = \frac{(\xi_1 + \xi_2)[\{C_1(1 + \alpha_1^*)(\Phi'_1 + \alpha_1^* p'^2)\}^{1/2} + \alpha_1^*(C_4 q'^2)^{1/2}]^2}{\frac{1}{1-\alpha}\left\{\frac{(\Phi'_1 + \alpha_1^* p^2)}{n_A} + \frac{\alpha\alpha_1^* q^2}{n_B}\right\}}$$

where

$$\alpha_1^* = \frac{\alpha}{1-\alpha}\,\frac{\xi_3 + \xi_4}{\xi_1 + \xi_2}.$$

Similarly, for the separate survey for the 2nd character, the cost is obtained as

$$C^{**} = \frac{(1-\xi_1)[\{C_3(1+\alpha_2^*)(\Phi_2' + \alpha_2^* p''^2)\}^{1/2} + \alpha_2^*(C_6 q''^2)^{1/2}]^2}{\dfrac{1}{1-\alpha}\left\{\dfrac{(\Phi_2' + \alpha_2^* p^2)}{n_A} + \dfrac{\alpha\alpha_2^* q^2}{n_B}\right\}} \tag{19}$$

where

$$p''^2 = \frac{K\Phi_2'}{1 + \alpha_2^*(1-K)},\quad \alpha_2^* = \frac{\alpha}{1-\alpha}\,\frac{1-\xi_3}{1-\xi_1}.$$

For the combined character study, the total cost C for 100% coverage by the frame A is given by (8).
Thus

$$C = \frac{n_A}{N_A}[C_1(N_{a(1)} + N_{ab(1)}) + C_2(N_{a(12)} + N_{ab(12)}) + C_3(N_{a(2)} + N_{ab(2)})]$$

$$+ \frac{n_B}{N_B}[C_4 N_{ab(1)} + C_5 N_{ab(12)} + C_6 N_{ab(2)}].$$

Using assumptions in costs (i.e. $C_4/C_1 = C_5/C_2 = C_6/C_3 = K$) we get

$$C = C_2 n_A[(1-\alpha)\{\varrho_1\xi_1 + \xi_2 + \varrho_3(1-\xi_1-\xi_2)\} +$$

$$\alpha\{\varrho_1\xi_3 + \xi_4 + \varrho_3(1-\xi_3-\xi_4)\} + \frac{K}{r}\{\varrho_1\xi_3 + \xi_4 + \varrho_3(1-\xi_3-\xi_4)\}] \tag{20}$$

where $r = n_A/n_B$, $\varrho_1 = C_1/C_2$ *and* $\varrho_3 = C_3/C_2$.
But in combined character study (n_A/n_B) Opt. $= p/\alpha q$ where p is given by (16). Thus the gain may be obtained from the ratio.

$$\frac{C^* + C^{**}}{C} =$$

$$\frac{\dfrac{(\xi_1+\xi_2)\varrho_1 T_1^2}{(\Phi_1' + \alpha_1^* p)} + \dfrac{(1-\xi_1)\varrho_3 T_2^2}{(\Phi_3' + \alpha_2^* p)}}{\{\varrho_1\xi_1 + \xi_2 + \varrho_3(1-\xi_1-\xi_2)\} + \{\varrho_1\xi_3 + \xi_4 + \varrho_3(1-\xi_3-\xi_4)\}\left\{\dfrac{r\alpha + K}{r(1-\alpha)}\right\}} \tag{21}$$

where

$$T_1 = \left\{(\Phi_1' + \alpha_1^* p'^2)(1 + \alpha_1^*)\right\}^{1/2} + \alpha_1^* q' \sqrt{K}$$

$$T_2 = \left\{(\Phi_2' + \alpha_2^* p''^2)(1 + \alpha_2^*)\right\}^{1/2} + \alpha_2^* q'' \sqrt{K}.$$

K can be determined as follows: Using the definitions of C_A, C_B, Φ_i's ($i = 1, \ldots, 6$) and equation (A.1), we obtain

$$\frac{C_A}{C_B} = \frac{1}{K} \frac{\varrho_1 \Phi_1 + \Phi_2 + \varrho_3 \Phi_3}{\varrho_1 \Phi_4 + \Phi_5 + \varrho_3 \Phi_6} = \varrho,$$

and thus

$$K = \varrho^{-1}\left\{\alpha + (1 - \alpha)\frac{\varrho_1 \xi_1 + \xi_2 + \varrho_3(1 - \xi_1 - \xi_2)}{\varrho_1 \xi_3 + \xi_4 + \varrho_3(1 - \xi_3 - \xi_4)}\right\}. \quad (22)$$

The expression in (21) may be used to obtain the gain in cost due to studying both the character simultaneously in comparison to independent individual surveys. The percent gain G is thus given by

$$G = \left(\frac{C^* + C^{**}}{C} - 1\right) \times 100$$

In the above cost comparison, the expected costs, C, C^* and C^{**} do not include the overhead costs for the combined or individual surveys, however, it is expected that the sum of overhead costs pertaining to individual surveys would be much larger than the corresponding overhead cost for the combined survey. Therefore, the actual gain in costs due to common multiple frame surveys compared to independent surveys will be larger than the percent gain G defined above.

The expression (21) reduced substantially under the assumptions $\Phi_1' = \Phi_2' = \Phi$ (say) and $\xi_1 = \xi_2 = \xi_3 = \xi_4 = \xi$ (say).

From (22) $\varrho = 1/K$ and from (16) since $\Phi_1'/\Phi_2' = \Phi_3'/\Phi_4'$, the p^2 reduces as follows:

$$p^2 = \frac{K(1 - \alpha)}{1 - k\alpha} \Phi.$$

Also $\alpha_1^* = \alpha_2^* = \alpha/(1 - \alpha)$.

Therefore, from (A.1)

$$p' = p'' = \left\{\frac{K(1 - \alpha)\Phi}{1 - K\alpha}\right\}^{1/2}.$$

Thus

$$T_1 = T_2 = \left\{\Phi \frac{1 - K\alpha}{1 - \alpha}\right\}^{1/2} + \frac{\alpha\sqrt{K}}{1 - \alpha}.$$

With all these substitutions in (21) $(C^* + C^{**})/C$ simplifies as follows:

$$\frac{C^* + C^{**}}{C} = \frac{T_1^2}{\Phi + \alpha_1^* p} \times \frac{2\xi\varrho_1 + (1-\xi)\varrho_3}{\left\{\frac{r\alpha + K}{K(1-\alpha)} + 1\right\}\left\{\varrho_1\xi + \xi + \varrho_3(1-2\xi)\right\}}$$

$$= \frac{T_1^2}{(\Phi + \alpha_1^* p)\left\{\frac{r\alpha + K}{K(1-\alpha)} + 1\right\}} \times \frac{\varrho_3 + \xi(2\varrho_1 - \varrho_3)}{\varrho_3 + \xi(\varrho_1 + 1 - 2\varrho_3)}$$

$$= \frac{\varrho_3 + \xi(2\varrho_1 - \varrho_3)}{\varrho_3 + \xi(1 + \varrho_1 - 2\varrho_3)}$$

where $r = (n_A/n_B)$ opt. $= p/\alpha q$ from (10).

Hence,

$$G = \frac{\xi(\varrho_1 + \varrho_3 - 1)}{\varrho_3 + \xi(1 + \varrho_1 - 2\varrho_3)} \times 100.$$

The equality of ξ_i's does not seem to be realistic assumption. The value of G, has therefore been calculated using (19) for realistic and representative combinations of parameters and are presented in Table. 1.

This table indicates that there is a definite gain due to integration of multiple frame surveys for both the characters in comparison to separate individual surveys. The gain increases with increasing values of ϱ_1 and ϱ_3.

4. COMPARISON OF TWO CHARACTER MULTIPLE FRAME SURVEYS WITH SINGLE FRAME SURVEY

Comparison of two frame survey with single frame surveys for study of two characters is of practical interest. For single character a similar study was carried out by Hartley (1962). On similar lines the relative reduction in cost was obtained as

$$R = \left(1 + \frac{\alpha q}{p\varrho}\right)^2 \Big/ \left\{1 + \frac{\alpha q(1 + p)}{p^2\varrho}\right\}$$

where p^2 is given by (16), $\varrho = C_A/C_B$ and $\alpha = N_{ab}/N_A$

The reduction in cost due to multiple frame over a single frame survey is tabulated in Table 2 for some set of parametric values. The table indicates considerable cost reduction.

Table 1

Percent Gain in Cost for Common multiple Frame Survey for Both Characters over Individual Surveys,

When $\varrho = 10$, $\Phi_1' = 0.25$, $\Phi_2' = 0.5$, $\Phi_3' = 1$, $\alpha = 0.5$.

	ϱ_3						
ϱ_1	0.3	0.4	0.5	0.6	0.7	0.8	0.9
			$\xi_1 = 0.2$, $\xi_2 = 0.2$, $\xi_3 = 0.4$, $\xi_4 = 0.2$				
0.3						1.5	3.9
0.4					1.7	4.2	6.4
0.5				1.8	4.4	6.7	8.7
0.6			1.8	4.5	6.9	8.9	10.7
0.7		1.7	4.6	7.0	9.1	10.9	12.6
0.8	1.7	4.6	7.1	9.3	11.2	12.8	14.3
0.9	4.5	7.1	9.4	11.3	13.0	14.5	15.9
			$\xi_1 = 0.2$, $\xi_2 = 0.4$, $\xi_3 = 0.2$, $\xi_4 = 0.4$				
0.3						4.5	9.1
0.4					4.6	9.3	13.6
0.5				4.8	9.6	14.0	17.9
0.6			4.9	9.9	14.3	18.3	22.0
0.7		5.1	10.1	14.7	18.8	22.5	25.9
0.8	5.2	10.4	15.1	19.3	23.1	26.5	29.7
0.9	10.8	15.5	19.8	23.6	27.1	30.3	33.2

Table 2

Reduction in Cost for Constant Variances

When $\Phi_1' = 0.25$, $\Phi_2' = 0.5$, $\Phi_3' = 1$, and $\xi_1 = 0.2$, $\xi_2 = 0.3$, $\xi_4 = 0.4$.

	α					
ϱ	0.5	0.6	0.7	0.8	0.9	0.95
100	.227	.175	.132	.094	.059	.040
20	.304	.254	.200	.169	.127	.101
10	.367	.321	.279	.238	.193	.164
5	.462	.423	.387	.351	.308	.277
2	.661	.646	.634	.621	.599	.578
1	.876	.895	.918	.943	.971	.985

APPENDIX

Minimizing the variance (17) with respect to C^* with the assumption of 100% coverage by frame A and the equality of variances, the optimum solution for p' is obtained as

$$p'^2 = \frac{1-\alpha}{\varrho' - \alpha}\left\{\frac{\sigma^2_{a(1)}(\xi_1 + \xi_2)}{\sigma^2_{ab(1)}(\xi_3 + \xi_4)}\right\}$$

with

$$\varrho' = \frac{C'_A}{C'_B}.$$

Using $N_{A1} = N_{a(1)} + N_{a(12)} + N_{ab(1)} + N_{ab(12)}$ and $N_{B1} = N_{ab(1)} + N_{ab(12)}$, ϱ' can be written as

$$\varrho' = \frac{C_1}{C_4}\alpha\frac{N_a(\xi_1 + \xi_2) + N_{ab}(\xi_3 + \xi_4)}{N_{ab}(\xi_3 + \xi_4)}$$

$$= \frac{C_1}{C_4}\alpha\left(\frac{1-\alpha}{\alpha}\frac{\xi_1 + \xi_2}{\xi_3 + \xi_4} + 1\right)$$

$$= \frac{\alpha}{K}\left(\frac{1}{\alpha_1^*} + 1\right)$$

where

$$\alpha_1^* = \frac{\alpha}{1-\alpha}\frac{\xi_3 + \xi_4}{\xi_1 + \xi_2}.$$

Then we have

$$p'^2 = \frac{1-\alpha}{\frac{\alpha}{K}\left(\frac{1}{\alpha_1^*} + 1\right) - \alpha}\frac{\xi_1 + \xi_2}{\xi_3 + \xi_4}\Phi_1'$$

$$= \frac{K\Phi_1'}{1 + \alpha_1^*(1-K)} \tag{A.1}$$

Define

$$\lambda_1 = (N_{a(1)} + N_{a(12)})\sigma^2_{a(1)} + p^2(N_{ab(1)} + N_{ab(12)})\sigma^2_{ab(1)},$$

$$\lambda_2 = q^2(N_{ab(1)} + N_{ab(12)})\sigma^2_{ab(1)},$$

$$\lambda_3 = (N_{a(1)} + N_{a(12)})\sigma^2_{a(1)} + p'^2(N_{ab(1)} + N_{ab(12)})\sigma^2_{ab(1)},$$

$$\lambda_4 = q'^2(N_{ab(1)} + N_{ab(12)})\sigma^2_{ab(1)}.$$

With the p' in (A.1), the optimum sample sizes will be

$$\frac{n^2_{AO}}{N_A} = \gamma' \frac{(N_{a(1)} + N_{a(12)})\sigma^2_{ab(1)} + p'^2(N_{ab(1)} + N_{ab(12)})\sigma^2_{ab(12)}}{C'_A}$$

$$= \gamma' \frac{\lambda_3}{C'_A}$$

$$\frac{n^2_{BO}}{N_B} = \gamma' \frac{q'^2(N_{ab(1)} + N_{ab(12)})\sigma^2_{ab(1)}}{C'_B} = \gamma' \frac{\lambda'_4}{C'_B}$$

with γ' determined wih respect to (18). From this we get

$$\frac{n_{BO}}{n_{AO}} = \frac{N_B}{N_A}\left(\frac{C_1 N_{A1}\lambda_4}{C_4 N_{B1}\lambda_3}\right)^{1/2}. \tag{A.2}$$

Also, the variances given by (5) and (17) at optimum sample sizes can be written as

$$V(\hat{Y}^{(1)}) = \frac{N_A}{n_A}\lambda_1 + \frac{N_B}{n_B}\lambda_2 \tag{A.3}$$

$$V(\hat{Y}^{*(1)}) = \frac{N_A}{n_{AO}}\lambda_3 + \frac{N_B}{n_{BO}}\lambda_4.$$

Equating the above variances and using (A.2), we obtain expression for n_{AO} and n_{BO} in terms of n_A and n_B as follows:

$$\frac{n_{AO}}{N_A} = \frac{\lambda_3 + \left(\dfrac{C_4\lambda_3\lambda_4 N_{B1}}{C_1 N_{A1}}\right)^{1/2}}{\dfrac{N_A}{n_A}\lambda_1 + \dfrac{N_B}{n_B}\lambda_2}$$

and

$$\frac{n_{BO}}{N_B} = \frac{\lambda_4 + \left(\dfrac{C_1\lambda_3\lambda_4 N_{A1}}{C_4 N_{B1}}\right)^{1/2}}{\dfrac{N_A}{n_A}\lambda_1 + \dfrac{N_B}{n_B}\lambda_2}.$$

Using these relationships, the cost C^* may be obtained as

$$C^* = \frac{(\xi_1 + \xi_2)\left[\left\{C_1(1 + \alpha_1^*)(\Phi_1' + \alpha_1^* p'^2)\right\}^{1/2} + \alpha_1^*(C_4 q'^2)^{1/2}\right]^2}{\dfrac{1}{1-\alpha}\left\{\dfrac{(\Phi_1' + \alpha_1^* p^2)}{n_A} + \dfrac{\alpha\alpha_1^* q^2}{n_B}\right\}}. \tag{A.4}$$

REFERENCES

ARMSTRONG, B. (1979). Test for multiple frames sampling technique for agricultural survey: New Brunswick, 1978. *Survey Methodology*, 5, 178-199.

BOSECKER, R.R., and FORD, B.L. (1976). Multiple frame estimation with stratified overlap domain. *American Statistical Association Proceedings of the Social Statistics Section*, 219-224.

HARTLEY, H.O. (1962). Multiple frame surveys. *American Statistical Association Proceedings of the Social Statistics Section*, 203-206.

HARTLEY, H.O. (1974). Multiple frame methodology and selected application. *Sankhya*, Series C, 36, 99-118.

LUND, R.E. (1968). Estimation in multiple frame surveys. *American Statistical Association Proceedings of the Social Statistics Section*, 282-288.

SERRURIER, D., and PHILLIPS, J. (1976). Double frame Ontario pilot hog surveys. *Survey Methodology*, 2, 138-170.

VOGEL, F.A. (1975). Surveys with overlapping frames, problems in application. *American Statistical Association Proceedings of the Social Statistics Section*, 695-699.