Aligarh Journal of Statistics
Vol. 30 (2010), 39-54

GENERALIZED TWO PHASE SAMPLING ESTIMATORS FOR THE POPULATION MEAN IN THE PRESENCE OF NON RESPONSE

B.B. Khare and S. Srivastava

ABSTRACT

Two generalized two phase sampling estimators T_g and T'_g for the population mean in presence of non-response have been proposed and their properties have been studied under the finite population and super population approaches. The numerical illustration suggests that in the case of finite population approach, we see that for any value of ρ, T_g is preferred over T'_g for higher values of ρ_2. while T'_g is preferred over T_g for smaller values of ρ_2 and higher values of ρ, where ρ and ρ_2 denote the correlation coefficient between y and x for the entire population and for the non-responding part of the population. The empirical study shows that T_g is more efficient than T'_g in the case of fixed cost as well as for the fixed sample sizes (n, n') for different values of sub-sampling fraction $(1/K)$. However, T_g is always preferred over T'_g under superpopulation approach.

1. INTRODUCTION

The problem of non-response is very common in the field of research in Agriculture, Sociology, Economics and Medical Sciences. Hansen and Hurwitz (1946) have considered the method of sub-sampling from non–respondents in the sample to eradicate the problem of non-response in mail surveys. Srinath (1971) has suggested a different rule for sub-sampling from non-respondents. Bouza (1983) has suggested two sampling strategies for the difference of two population means in the presence of missing observations in two different situations. In the presence of non-response, the estimation of the population mean $\bar{Y}$, of the study character y using the auxiliary character x with known population mean $\bar{X}$, has been considered by Rao (1986, 90), Khare (1987) and Khare and Srivastava (1997, 2000). When $\bar{X}$ is unknown, Khare and Srivastava (1993, 1995) have proposed some two phase sampling ratio and product type estimates for the population mean and studied their properties under finite population approach.

In the present paper, we have proposed two generalized two phase sampling estimators T_g and T'_g for $\bar{Y}$ in the presence of non-response. The properties of these estimators have been studied in the case of finite population and the superpopulation approaches. A Comparative study of the proposed estimators with the other relevant estimators has also been conducted in the case of finite population as well as in the case of superpopulation approach. Under the finite population approach, the expressions for the optimum values of the first phase sample size (n'), the second phase sample size (n) and the sub-sampling fraction $(1/K)$ have been obtained for fixed cost $C \le C_0$. A numerical illustration has been given to study the performance of T_g and T'_g with respect to ρ and ρ_2. An empirical study is also conducted to study the performance of T_g and T'_g with respect to relevant estimators for fixed cost, $C \le C_0$.

2. THE ESTIMATORS

When $\bar{X}$ is unknown, we may replace $\bar{X}$ by its consistent estimate as the sample mean based on a larger sample of size (n'). So we first select a large sample of size n' from the finite population of size N by using simple random sampling without replacement $(SRSWOR)$ method of sampling and observe x for the selected n' units. It is assumed that there is no non-response in selecting n' units for x character. Further, we select a subsample of the size $n(\le n')$ from the selected n' units by using $SRSWOR$ method of sampling to observe y for these n selected units. We observe that from n selected units, only n_1 units respond and n_2 units do not respond. Further from remaining n_2 non-responding units, we again select a subsample of size $r = \frac{n_2}{K} (K > 1)$ by using $SRSWOR$ method of sampling and obtain the values on y after applying extra efforts. Now using the incomplete information on y and the corresponding incomplete/complete information on x from the sample of size n, the proposed estimators T_g and T'_g are given as follows:

$$T_g = \bar{y}^* \left(\frac{\bar{x}^*}{\bar{x}'} \right)^{\alpha} \tag{2.1}$$

$$T'_g = \bar{y}^* \left(\frac{\bar{x}}{\bar{x}'} \right)^{\alpha_0}, \tag{2.2}$$

where

$$\bar{y}^* = \frac{n_1}{n} \bar{y}_1 + \frac{n_2}{n} \bar{y}'_2, \quad \bar{x}^* = \frac{n_1}{n} \bar{x}_1 + \frac{n_2}{n} \bar{x}'_2, \quad \bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i \text{ and } \bar{x}' = \frac{1}{n'} \sum_{i=1}^{n'} x_i .$$

Here α, α_0 are constants which may take positive as well as negative values depending upon the correlation between y and x. It is easy to see that for $\alpha = 1, -1$, the estimators, T_g reduces to the conventional ratio and product estimators and for $\alpha_0 = 1, -1$, the estimator T'_g reduces to alternative ratio and product estimators [Khare and Srivastava (1995)]. Here $(\bar{x}_1, \bar{y}_1)$ and $(\bar{x}'_2, \bar{y}'_2)$ are the sample means of the characters (x, y) based on the sample of size n_1 and r respectively.

Further we may use regression type estimators using the auxiliary information for computing the estimate of the population mean in the presence of non-response. The two phase sampling regression type estimators in presence of non-response (Khare and Srivastava (1995)) are given as follows:

$$T_{lr} = \bar{y}^* + b^* (\bar{x}' - \bar{x}^*) \tag{2.3}$$

$$T'_{lr} = \bar{y}^* + b^{**} (\bar{x}' - \bar{x}), \tag{2.4}$$

where

$$b^* = \frac{\hat{S}_{yx}}{\hat{S}_x^2}, \quad ,b^{**} = \frac{\hat{S}_{yx}}{s_x^2} \quad \text{and} \quad s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 .$$

Following Rao (1990), the estimates $\hat{S}_{yx}$ and $\hat{S}_x^2$ are obtained on the basis of the available data under the given sampling design.

3. BIAS AND MEAN SQUARE ERROR OF THE PROPOSED ESTIMATORS (FINITE POPULATION APPROACH)

In the case of finite population approach, it is assumed that the total population of size N is divided into two strata having N_1 responding and N_2 non-responding units which are unknown. At the first stage, n' units are selected from N for x character and at the second stage; a sub-sample of size $n (< n')$ is selected from n' by using $SRSWOR$ method of sampling. The n_1 responding units from n units are supposed to be selected from N_1 units of respondent stratum and $r = \frac{n_2}{K} (K > 1)$ sub-sampled units from n_2 non-responding units are supposed to be selected from N_2 units of non-respondent stratum.

The expressions for bias and mean square error and T_g and T'_g up to the terms of order n^{-1} are given as follows:

$$B(T_g) = \frac{\alpha}{\bar{X}}\left[\frac{(\alpha-1)}{2} R\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2\right\} + \left(\frac{1}{n}-\frac{1}{n'}\right)S_{yx} + \frac{W_2(k-1)}{n} S_{yx(2)}\right] \quad (3.1)$$

$$B(T'_g) = \frac{\alpha_0}{\bar{X}}\left[\left(\frac{1}{n}-\frac{1}{n'}\right)\{R S_x^2 + S_{yx}\}\right] \quad (3.2)$$

$$MSE(T_g) = \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)[\alpha^2 R^2 S_x^2 + 2\alpha R S_{yx}] + \frac{W_2(k-1)}{n}[S_{y_2}^2 + \alpha^2 R^2 S_{x_2}^2 + 2\alpha R S_{yx(2)}] \quad (3.3)$$

$$MSE(T') = \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2 + \left(\frac{1}{n}-\frac{1}{n'}\right)[\alpha^2 R^2 S_x^2 + 2\alpha R S_{yx}] \quad (3.4)$$

The optimum values of α and α_0 which minimize $MSE(T_g)$ and $MSE(T'_g)$ are given as follows:

$$\alpha_{(opt)} = -\frac{\left(\frac{1}{n}-\frac{1}{n'}\right)S_{yx} + \frac{W_2(k-1)}{n} S_{yx(2)}}{R\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)S_y^2 + \frac{W_2(k-1)}{n} S_{x_2}^2\right\}} \quad (3.5)$$

$$\alpha_{0(opt)} = -\frac{S_{yx}}{R S_x^2} \quad (3.6)$$

The values of minimum mean square error of T_g and T'_g for $\alpha_{(opt)}$ and $\alpha_{0(opt)}$ are given as follows:

$$MSE(T_g)_{\min} = \left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2 - \frac{\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)S_{yx} + \frac{W_2(k-1)}{n} S_{yx(2)}\right\}^2}{\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2\right\}} \quad (3.7)$$

$$MSE(T'_g)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \frac{W_2(k-1)}{n}S_{y_2}^2 - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{S_{yx}^2}{S_{x_2}^2}, \qquad (3.8)$$

where $R = \frac{\bar{Y}}{\bar{X}}$, $W_2 = \frac{N_2}{N}$, the non-response rate and $(S_{y_2}^2, S_{x_2}^2, S_{yx(2)})$ denote the population mean square of y, x and population covariance between y and x for the entire population and for the non-responding group of the population respectively.

If $\frac{S_{yx}}{S_x^2} = \frac{S_{yx(2)}}{S_{x_2}^2}$ then $\dfrac{\left(\frac{1}{n} - \frac{1}{n'}\right)S_{yx} + \frac{W_2(k-1)}{n}S_{yx(2)}}{\left(\frac{1}{n} - \frac{1}{n'}\right)S_x^2 + \frac{W_2(k-1)}{n}S_{x_2}^2}$ will be equal to $\frac{S_{yx}}{S_x^2}$ and in this case, we have,

$$MSE(T_g)_{\min} = \left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2(1-\rho^2) + \frac{W_2(k-1)}{n}(1-\rho_2^2)S_{y_2}^2 + \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 \qquad (3.9)$$

where, ρ and ρ_2 are the correlation coefficient between y and x for the entire population and for the non-responding group of the population.

It is important to notice here that the optimum values of the constants α and α_0 depend on the values of $\rho C_y / C_x$ and $\rho_2 C_{y_2} / C_{x_2}$, where (C_y, C_x, ρ) and $(C_{y_2}, C_{x_2}, \rho_2)$ denote the coefficient of variation of y, x and correlation coefficient between y and x for the entire population and for the non-responding group of the population respectively. Reddy (1978) has shown that the values of $\rho C_y / C_x$ and $\rho_2 C_{y_2} / C_{x_2}$ are not only stable over the time but also over different regions, so one can use the past data (Tripathi *et al.* (1993)). However, in the case of non-availability of any prior information about the required parametric values from any source, one may estimate these parameters on the basis of sample observations without having any loss in efficiency up to the order n^{-1} (Srivastava and Jhajj (1983)).

4. COMPARISON OF THE ESTIMATORS

The mean square error of the estimator T_{lr} and T'_{lr} are given by

$$MSE(T_{lr}) = \left(\frac{1}{n} - \frac{1}{n'}\right)S_y^2(1-\rho^2) + \left(\frac{1}{n'} - \frac{1}{N}\right)S_y^2 + \frac{W_2(k-1)}{n}(S_{y_2}^2 + \beta^2 S_{x_2}^2 - 2\beta S_{yx(2)}) \qquad (4.1)$$

$$MSE(T'_{lr}) = \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^2 (1-\rho^2) + \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2 . \qquad (4.2)$$

In the present case, a two phase sampling difference estimator for $\bar{Y}$ is given by

$$t_d = \bar{y}^* + d_0 (\bar{x}' - \bar{x}^*), \qquad (4.3)$$

where d_0 is a constant.

The estimator t_d is unbiased and its minimum value of the variance for

$$d_{0(opt)} = \frac{\left(\frac{1}{n} - \frac{1}{n'}\right) S_{yx} + \frac{W_2(k-1)}{n} S_{yx(2)}}{\left(\frac{1}{n} - \frac{1}{n'}\right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2}$$ is given by

$$V(t_d)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2 + \frac{W_2(k-1)}{n} S_{y_2}^2 - \frac{\left\{\left(\frac{1}{n} - \frac{1}{n'}\right) S_{yx} + \frac{W_2(k-1)}{n} S_{yx(2)}\right\}^2}{\left(\frac{1}{n} - \frac{1}{n'}\right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2} \qquad (4.4)$$

Since the value of $d_{0(opt)}$ is not known in practice, so we replace $d_{0(opt)}$ by its estimate based on the sample units and hence we define estimators t'_d given by

$$t'_d = \bar{y}^* + \frac{\left\{\left(\frac{1}{n} - \frac{1}{n'}\right) \hat{S}_{yx} \frac{W_2(k-1)}{n} \hat{S}_{yx(2)}\right\}}{\left\{\left(\frac{1}{n} - \frac{1}{n'}\right) \hat{S}_x^2 + \frac{W_2(k-1)}{n} \hat{S}_{x_2}^2\right\}} (\bar{x}' - \bar{x}^*) . \qquad (4.5)$$

It is easy to see that up to terms of order n^{-1}, $MSE(t'_d)$ is the same as $V(t_d)_{\min}$. This is due to the fact that if we replace the parameter values by the sample values then up to the terms of order n^{-1}, the MSE of the estimator is not affected (Srivasvava and Jhajj (1983)).

From (3.7) and (4.4), we see that $MSE(t'_d) = MSE(T_g)_{\min}$. For $\beta = \beta_{(2)}$, we see that

$MSE(T_{lr}) = MSE(t'_d) = MSE(T_g)_{\min}$, since t'_d itself reduces to T_{lr}. We also observe that

$$MSE(T_{lr}) = MSE(T'_{lr}) > MSE(T_g)_{\min} \text{ for } \beta = 2\beta_{(2)} \tag{4.6}$$

$$MSE(T_g)_{\min} < MSE(T'_{lr}) < MSE(T_{lr}) \text{ for } \beta > 2\beta_{(2)} \tag{4.7}$$

and

$$MSE(T_g)_{\min} < MSE(T_{lr}) < MSE(T'_{lr}) \text{ for } \beta < 2\beta_{(2)} \tag{4.8}$$

where $\beta = \dfrac{S_{yx}}{S_x^2}$ and $\beta_{(2)} = \dfrac{S_{yx(2)}}{S_{x_2}^2}$.

For the optimum value of α and α_0, the relative efficiency of the estimator T_g with respect to T'_g is given by

$$RE(T_g) = \frac{MSE(T'_g)_{\min}}{MSE(T_g)_{\min}} = \frac{\left(1 - \dfrac{\rho^2\delta}{\Delta}\right)}{\dfrac{1-\rho^2(\delta + W_2(k-1)\Delta_0\Delta_1\Delta_2)}{\Delta(\delta + W_2(k-1)\Delta_1^2)}}, \tag{4.9}$$

where,

$$\delta = \frac{n'-n}{n},\ \Delta = \frac{N-n}{N} + W_2(k-1)\Delta_2^2,\ \Delta_0 = \frac{\rho_2}{\rho},\ \Delta_1 = \frac{S_{x_2}}{S_x} \text{ and } \Delta_2 = \frac{S_{y_2}}{S_y}.$$

Since the relative efficiency of T_g with respect to T'_g depends on the values of ρ, Δ_0, Δ_1, W_2 and k, so we have considered a numerical example in which $\Delta_1 = 1.1$, $\Delta_2 = 1.2$, $N = 500$, $n = 50$, $k = 4$ and $W_2 = 0.25$. The relative efficiency of T_g with respect to T'_g has been computed for different values of ρ and ρ_2.

Table 4.1: Relative efficiency of T_g with respect to T'_g (in %)

ρ	ρ_2								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.1	100.54	101.95	104.07	107.01	110.88	115.89	122.30	130.53	141.20
0.2	100.39	102.25	104.88	108.39	112.96	118.81	126.32	136.01	148.71
0.3	99.83	102.16	105.31	109.44	114.75	121.55	130.30	141.70	156.87
0.4	98.95	101.65	105.35	110.12	116.25	124.10	134.29	147.73	165.98
0.5	97.42	100.70	104.96	110.42	117.43	126.47	138.33	154.25	176.43
0.6	95.50	99.26	104.10	110.30	118.27	128.66	142.49	161.50	188.88
0.7	95.05	97.29	102.72	109.69	118.73	130.66	146.88	169.82	204.37
0.8	89.99	94.69	100.73	108.52	118.73	132.46	151.60	179.77	224.75
0.9	86.24	91.39	98.03	106.68	118.20	134.04	156.88	192.26	253.69

From Table 4.1, we see that for any given value of ρ, T_g is more efficient than T'_g for the higher values of ρ_2. However, T'_g is more efficient than T_g for smaller values ρ_2 and higher values of ρ.

In the case of two phase sampling, it is important to see whether deduction in variance would be worth the extra expenditure on the additional sample required to estimate the population mean of the auxiliary character. So one would like to choose that strategy which, for a fixed cost, can estimate $\overline{Y}$ with maximum precision. So we determine the size of the first and second phase sample which minimizes the variance of the estimator T_g and T'_g for the fixed cost.

5. DETERMINATION OF n', n AND k FOR FIXED COST $C \le C_0$

Let C_0 be the total cost apart from overhead cost. The expected total cost of the survey apart the overhead cost is given by

$$C = C'_1 n' + n\left\{C_1 + C_2 W_1 + C_3 \frac{W_2}{k}\right\}, \tag{5.1}$$

where

C' : Cost per unit for identifying and observing auxiliary character,

C_1: Cost per unit of mailing questionnaire at second phase,

C_2: Cost per unit of collecting and processing the data from responding units,

C_3: Cost per unit for obtaining and processing the data obtained from sub-sample units and $W_1 = \frac{N_1}{N}$ (response rate in the population).

Using the $MSE(T_g)_{\min}$ and $MSE(T'_g)_{\min}$ from (3.9) and (3.8), the expression for $MSE(T_g)_{\min}$ and $MSE(T'_g)_{\min}$ can be written as

$$MSE(T_g) = \left(\frac{V_{01}}{n} + \frac{V_{11}}{n'} + \frac{K}{n}V_{21}\right) + \text{terms independent of } n,\ n' \text{ and } K, \tag{5.2}$$

where

V_{01}, V_{11} and V_{12} are the coefficients of $\frac{1}{n}$, $\frac{1}{n'}$ and $\frac{K}{n}$ in the expression of $MSE(T_g)$ and V_{02}, V_{12} and V_{22} are the coefficients of $\frac{1}{n}$, $\frac{1}{n'}$ and $\frac{K}{n}$ in the expression of $MSE(T'_g)$.

Let us define

$$\psi_1 = MSE(T_g) + \lambda_1\left\{C'_1 n' + n(C_1 + C_2 W_1 + C_3\frac{W_2}{K})\right\} \tag{5.3}$$

$$\psi_2 = MSE(T'_g) + \lambda_2\left\{C'_1 n' + n(C_1 + C_2 W_1 + C_3\frac{W_2}{K})\right\}. \tag{5.4}$$

Now differentiating ψ_i, $i = 1, 2$ with respect to n, n' and K and equating to zero, we have the optimum values of n, n' and K for $C \le C_0$, which are given as follows:

$$K_{opt} = \sqrt{\frac{C_3 W_2 V_{0i}}{(C_1 + C_2 W_1)V_{2i}}}, \qquad i = 1, 2 \tag{5.5}$$

$$n'_{opt} = \sqrt{\frac{V_{1i}}{\lambda_i C'_1}}, \qquad i = 1, 2 \tag{5.6}$$

and

$$n_{opt} = \sqrt{\frac{V_{0i} + K_{opt} V_{2i}}{\lambda_i \left(C_1 + C_2 W_1 + C_3 \dfrac{W_2}{K_{opt}} \right)}}, \quad i = 1, 2, \tag{5.7}$$

where

$$\sqrt{\lambda_i} = \frac{1}{C_0} \left[\sqrt{C_1' V_{1i}} + \sqrt{\left(C_1 + C_2 W_1 + \frac{C_3 W_2}{K_{opt}} \right) (V_{0i} + K_{opt} V_{2i})} \right], \quad i = 1, 2.$$

Here we also observe that for the optimum values of n, n' and K, the determinant of the matrix of second order derivatives of ψ_i with respect to n, n' and K is positive which shows that the minimum value of $MSE(T_g)_{\min}$ and $MSE(T_g')_{\min}$ will be attained for the optimum values of n, n' and K.

Theorem 5.1: The minimum value $MSE(T_g)_{\min}$ and $MSE(T_g')_{\min}$ for the optimum values of n, n' and K are given by

$$MSE(T_g)_{opt} = \frac{1}{C_0} \left[\sqrt{C_1' V_{11}} + \sqrt{\left(C_1 + C_2 W_1 + \frac{C_3 W_2}{K_{opt}} \right) (V_{01} + K_{opt} V_{21})} \right]^2 \tag{5.8}$$

and

$$MSE(T_g')_{opt} = \frac{1}{C_0} \left[\sqrt{C_1' V_{12}} + \sqrt{\left(C_1 + C_2 W_1 + \frac{C_3 W_2}{K_{opt}} \right) (V_{02} + K_{opt} V_{22})} \right]^2 \tag{5.9}$$

6. AN EMPERICAL STUDY

Physical growth characteristics of 95 school going male children of $6-7$ years, belonging to upper socio-economic group of Varanasi district (India) has been taken under consideration. The first 25% (i.e. 24 children) units have been considered as non-response units. The values of the parameters related to the study character y (the weight of the children aged $6-7$ years) and the auxiliary character x (the height of the children aged $6-7$ years) have been given as follows:

$\bar{Y} = 19.4968$kg, $\bar{X} = 115.9526$cm, $S_y = 3.0440$kg, $S_{y_2} = 2.342$kg,

$S_{x_2} = 5.1043\text{kg}, \ \ \rho = 0.713, \ \ S_x = 115.9526\text{cm}, \ \ \rho_2 = 0.678.$

The problem considered is to estimate the average weight of the children aged $6-7$ years using height of the children the auxiliary character.

Using (3.5) and (3.6), the optimum values of the constants involved in T_g and T_g' are

$\alpha_{opt(K=4)} = -2.0034, \ \ \alpha_{opt(K=3)} = -2.0339, \ \ \alpha_{opt(K=2)} = -2.0809$ and

$\alpha_{0(opt)} = -2.1626.$

Table 6.1: Relative $RE(.)$ in % with respect of $\bar{y}^*$ for different values of K for $n' = 60$ and $n = 30$

Estimator	$1/k$		
	$1/4$	$1/3$	$1/2$
$\bar{y}^*$	100.00 (0.3509)*	100.00 (0.3044)	100.00 (0.2578)
T_{lr}	166.94 (0.2102)	165.08 (0.1844)	162.55 (0.1586)
T_g	167.98 (0.2089)	165.80 (0.1836)	162.96 (0.1582)
T_g'	128.87 (0.27233)	134.87 (0.2257)	143.86 (0.1792)

*Figures in parenthesis give $MSE(.)$.

From Table 6.1, we observe that for fixed value of n' and n, the estimators T_g and T_g' are more efficient $\bar{y}^*$ for different values of K. It has also been observed that the estimator T_g is more efficient than the estimator $T_g^{'}$ and the efficiency of T_{lr} is almost closer to the efficiency of T_g with respect to $\bar{y}^*$ for different value of K. The relative efficiency of $T_g^{'}$ with respect to $\bar{y}^*$ increases by increasing the value of sub-sampling fraction $1/K$. This is due to the reason that with increase in the value of sub-sampling fraction $1/K$, the variance of $\bar{y}^*$ decreased with a faster rate than the MSE of T_{lr} and T_g while the MSE of T_g' decreases with the faster rate than the variance of $\bar{y}^*$.

Table 6.2: Relative $RE(.)$ with respect to $\bar{y}^*$ in the case of fixed cost $C_0 = \text{Rs } 200.00$

Estimators	$C_1' = \text{Rs } 0.75$, $C_1' = \text{Rs } 0.75$, $C_1' = \text{Rs } 0.75$, $C_1' = \text{Rs } 0.75$			
	K_{opt}	n'_{opt} approx	n_{opt} approx	$RE(.)$ wrt $\bar{y}^*$ (in %)
$\bar{y}^*$	3.4147	–	34	100.00 (0.3715)*
T_{lr}	1.7237	65	23	123.59 (0.3006)
T_g	1.9085	65	23	125.38 (0.2963)
T'_g	1.1704	63	20	118.84 (0.3126)

*Figures in parenthesis give $MSE(.)$.

From Table 6.2, it has been observed that for the fixed cost, the relative efficiency of T_g with respect to $\bar{y}^*$ is higher than the relative efficiency of T_{lr} and T'_g with respect to $\bar{y}^*$.

7. COMPARISON THROUGH THE SUPER POPULATION MODEL

Let us assume that the population value $\underline{y} = (y_1, y_2, \ldots, y_N)$ is a realized outcome of a vector random variable $\underline{Y} = (Y_1, Y_2, \ldots, Y_N)$. We suppose that the whole population of size N is divided into N_1 (unknown) responding and N_2 (unknown) non-responding units such that $\underline{y} = (\underline{y}_{(1)}, \underline{y}_{(2)})$, where $\underline{y}_{(1)} = (y_1, y_2, \ldots, y_{N_1})$ and $\underline{y}_{(2)} = (y_{N_1+1}, y_{N_1+2}, \ldots, y_N)$. The joint distribution of $(Y_1, Y_2, \ldots, Y_N)$, $(Y_1, Y_2, \ldots, Y_{N_1})$ and $(Y_{N_1+1}, Y_{N_1+2}, \ldots, Y_N)$ are denoted by ξ, ξ_1 and ξ_2 respectively, so that $\xi = (\xi_1, \xi_2)$.

The units of responding and non-responding groups which follows the superpopulation model ξ_1 and ξ_2 which are respectively represented by

$$y_{1i} = \alpha_1 + \beta_0 x_i + e_{1i}, \quad i = 1, 2, \ldots, N_1 \tag{7.1}$$

and

$$y_{2i} = \alpha_2 + \beta_0 x_i + e_{2i}, \quad i = 1, 2, \ldots, N_2 \tag{7.2}$$

with the assumptions that

$$E(e_{1i}/x_i)=E(e_{2i}/x_i)=0,\quad E(e_{1i}\,e_{1i'})=E(e_{2i}\,e_{2i'})=0,\quad \forall i\neq i'$$

$$V(e_{1i}/x_i)=v_1x_i^l \text{ and } V(e_{2i}/x_i)=v_2x_i^l \qquad (7.3)$$

for $0\le l<2$ and e_{1i} and e_{2i} are assumed to be uncorrelated.

The expressions for the expected bias and expected mean square error of any estimator T under the superpopulation model is given by

$$\xi B(T)=E_p\,E_\xi\,(T-\bar{Y}) \text{ and } \xi MSE(T)=E_p\,E_\xi\,(T-Y)^2 .$$

Let us denote

$$\alpha_W=W_1\alpha_1+W_2\alpha_2,\quad \bar{E}=\sum_{i=1}^{N}e_i/N,\quad \bar{e}_1=\sum_{i=1}^{n_1}e_{1i}/n_1,\quad \bar{e}_{2r}=\sum_{i=1}^{r}e_{2i}/r,$$

$$\bar{e}^*=w_1\bar{e}_1+w_2\bar{e}_{2r},$$

where

$$W_i=N_i/N \text{ and } w_i=n_i/n \text{ for } i=1,2.$$

Now, we have $\bar{Y}=\alpha_w+\beta_0\bar{X}+\bar{E}$ under the superpopulation model, the estimator $\bar{y}^*$ is unbiased and has expected mean square error

$$\xi MSE(\bar{y}^*)_l=V(\alpha_W)+\beta_0^2\,V(\bar{x}^*)+\xi E(\bar{e}^*-\bar{E})_{l'}^2, \qquad (7.4)$$

where

$$V(\alpha_W)=\frac{N-n}{(N-n)n}(\alpha_1-\alpha_2)^2W_1W_2,\quad V(\bar{x}^*)=\left(\frac{1}{n}-\frac{1}{N}\right)S_x^2+\frac{W_2(K-1)}{n}S_{x_2}^2,$$

$$\xi E(\bar{e}^*-\bar{E})_{l=0}^2=\left\{\left(\frac{1}{n}-\frac{1}{N}\right)W_1v_1+\left(\frac{K}{n}-\frac{1}{N}\right)W_2v_2\right\}$$

and

$$\xi E(\bar{e}^*-\bar{E})_{l=1}^2=\left\{\left(\frac{1}{n}-\frac{1}{N}\right)W_1\bar{X}_1v_1+\left(\frac{K}{n}-\frac{1}{N}\right)W_2\bar{X}_2v_2\right\}.$$

The expressions for the expected bias and expected mean square error of T_g and T_g' under the superpopulation model (7.1) and (7.2) up to the terms of order n^{-1} are given as follows:

$$\xi\, B(T_g) = \left\{ \frac{\alpha(\alpha-1)\alpha_w}{2\bar{X}^2} + \frac{\alpha(\alpha+1)\beta_0}{2\bar{X}} \right\} \left\{ \left(\frac{1}{n} - \frac{1}{n'} \right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2 \right\} \tag{7.5}$$

$$\xi\, B(T'_g) = \left\{ \frac{\alpha_0(\alpha_0-1)\alpha_w}{2\bar{X}^2} + \frac{\alpha_0(\alpha_0+1)\beta_0}{2\bar{X}} \right\} \left(\frac{1}{\text{n}} - \frac{1}{n'} \right) S_x^2 \tag{7.6}$$

$$\xi\, MSE(T_g)_l = \xi\, MSE(\bar{y}^*)_l + \left\{ \frac{\alpha^2 \alpha_w^2}{\bar{X}^2} + \frac{2\alpha(\alpha+1)\beta_0\alpha_w}{\bar{X}} + \beta_0^2(\alpha^2 + 2\alpha) \right\}$$

$$\times \left\{ \left(\frac{1}{n} - \frac{1}{n'} \right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2 \right\} \text{ for } l = 0,1 \tag{7.7}$$

and

$$\xi\, MSE(T'_g)_l = \xi\, MSE(\bar{y}^*)_l + \left\{ \frac{\alpha_0^2 \alpha_w^2}{\bar{X}^2} + \frac{2\alpha_0(\alpha_0+1)\beta_0\alpha_w}{\bar{X}} + \beta_0^2(\alpha_0^2 + 2\alpha_0) \right\}$$

$$\times \left(\frac{1}{n} - \frac{1}{n'} \right) S_x^2 \text{ for } l = 0,1. \tag{7.8}$$

The optimum value of α and α_0 which minimize $\xi\, MSE(T_g)$ and $\xi\, MSE(T'_g)$ for $l = 0,1$ respectively is uniquely given by

$$\alpha_{opt} = \alpha_{0(opt)} = \frac{\beta_0 \bar{X}}{\alpha_W + \beta_0 \bar{X}} \tag{7.9}$$

and the minimum values of $\xi\, MSE(T_g)$ and $\xi\, MSE(T'_g)$ are given by

$$\xi\, MSE(T_g)_{l(\min)} = \xi\, MSE(\bar{y}^*)_l - \beta_0^2 \left(\left(\frac{1}{n} - \frac{1}{n'} \right) S_x^2 + \frac{W_2(k-1)}{n} S_{x_2}^2 \right) \tag{7.10}$$

$$\xi\, MSE(T'_g)_{l(\min)} = \xi\, MSE(\bar{y}^*)_l - \beta_0^2 \left(\frac{1}{n} - \frac{1}{n'} \right) S_x^2 \text{ for } l = 0,1. \tag{7.11}$$

It is important to notice here that,

$\xi\, MSE(T_g)_{l(\min)} < \xi\, MSE(T'_g)_{l(\min)}$, $\xi\, MSE(T_g)_{l(\min)} = \xi\, MSE(T_{lr})_{l(\min)}$

and

$\xi\, MSE(T'_g)_{l(\min)} = \xi\, MSE(T'_{lr})_{l(\min)}$ for $l = 0,1$.

Acknowledgement

The authors are grateful to the editor for giving very useful suggestions for the improvement of paper.

REFERENCES

Bouza, C. (1983): Estimation of a difference in finite populations with missing observations. *Biom. J.*, **25,** 123-128.

Hansen, M.H. and Hurwitz, W.N. (1946): The problem of non-response in sample surveys. *J. Amer. Statist. Assoc*., **41**, 517-529.

Khare, B.B. (1987): Allocation in stratified sampling in presence of non-response. *Metron*, **45**, 213-221.

Khare, B.B. and Srivastava, S. (1993): Estimation of population mean using auxiliary character in presence of non-response. *Nat. Acad. Sci. Lett.*, **16,** 111-114.

Khare, B.B. and Srivastava, S. (1995): Study of conventional and alternative two phase sampling ratio, product and regression estimators in presence of non-response. *Proc. Nat. Acad. Sci. India*, **65**, 195-203.

Khare, B.B. and Srivastava, S. (2000): Generalized estimator for population mean in presence of non-response. *Int. J. Math. Stat. Sci*., **9**, 75-87.

Khare B.B. and Srivastava, S. (1997): Transformed ratio type estimators for the population mean in the presence of non-response. *Comm. Statist. Theory Methods*, **26,** 1779-1791.

Rao, P.S.R.S. (1986): Ratio estimation with sub-sampling the non-respondents, *Survey Methodology*, **12,** 217-230.

Rao, P.S.R.S. (1990): Regression estimators with sub-sampling of non-respondents. *Data Quality Control Theory and Pragmatics*, (*Eds*.) Gunar E. Liepins and V.R.R. Uppuluri, *Marcel Dekker, New York*, 191-208.

Reddy, V.N. (1978): A study on the use of prior knowledge on certain population parameters in estimation. *Sankhyā*, *Ser C*, **4**, 29-37.

Srinath, K.P. (1971): Multiphase sampling in non-response problems. *J. Amer. Statist. Assoc.* **66**, 583-586.

Srivastava, S.K. and Jhajj, H.S. (1983): A class of estimators of the population mean using multi-auxiliary information. *Calcutta Statist. Assoc. Bull***.**, **32**, 47-56.

Tripathi, T.P., Maiti, P. and Sharma, S.D. (1983): Use of prior information on some parameters in estimating mean. . *Sankhyā*, *Ser A*, **45**, 372-376.

Received : 30-07-2008
Revised : 13-03-2010

B.B. Khare
Department Statistics,
Banaras Hindu University,
Varanasi- 221 005, India.
e-mail: bbkhare56@yahoo.com

S. Srivastava
NCERT, N. Delhi, India.