Aligarh Journal of Statistics
Vol. 29 (2009), 91-106

TRANSFORMED TWO PHASE SAMPLING RATIO AND PRODUCT TYPE ESTIMATORS FOR POPULATION MEAN IN THE PRESENCE OF NON-RESPONSE

B.B. Khare and Sanjay Kumar

ABSTRACT

The proposed transformed two phase sampling ratio and product type estimators for population mean in the presence of non-response are found to be more efficient than corresponding two phase sampling ratio and product type estimators and are equally efficient to regression estimators. The empirical study shows that the proposed estimators are more efficient with respect to corresponding estimators in a wide range of the transformation constants. In the case of fixed cost and the specified precision, the conventional transformed two phase sampling ratio type estimator is found to be equally efficient to the regression estimator and has equal minimum cost, while it is found to be more efficient than the other relevant estimators and also has minimum cost in comparison to the other relevant estimators.

1. INTRODUCTION

The two phase sampling ratio and product type estimators are biased, so their biases can be reduced to zero up to the terms of the order n^{-1} without having any change in their mean square errors (MSE) by using the methods of bias reduction given by Tin (1965) and Quenouille (1956). However, the bias of the two phase sampling ratio type estimator can also be reduced to zero up to the terms of the order n^{-1} following Mohanty and Das (1971) method of transforming auxiliary character.

In case of non-response in sample survey, Hansen and Hurwitz (1946) suggested the method of sub sampling from non-respondents for estimating the population mean. Further improvement in the estimation of population mean in presence of non-response using auxiliary character with known / unknown population means have been proposed by Rao (1986, 90), Khare and Srivastava (1993, 95). Further Khare and srivastava (1996, 97) have proposed ratio and product type estimators for population mean using the transformed auxiliary character with known population mean in the presence of non-response.

In the present context, we have proposed two phase sampling ratio and product type estimators for population mean using transformed auxiliary character in the presence of non-response and their properties have been studied for fixed sample sizes (n', n), for fixed cost $C \le C_0$ and for specified precision $V = V_0$. A comparative study of the proposed transformed two phase sampling ratio and product type estimators for population mean in the presence of non-response has been made with the conventional and alternative two phase sampling ratio, product and regression type estimators for population mean in the presence of non-response. The optimum value of first phase sample size (n'), second phase sample size n and sub sampling fraction (K) have been determined for fixed cost $(C \le C_0)$ as well as for specified variance $V = V_0$. An empirical investigation has been made to study the performance of the proposed estimators with the relevant estimators for fixed sample sizes (n', n), for fixed cost $(C \le C_0)$ and for the specified precision $V = V_0$.

2. THE ESTIMATORS

Let $\bar{Y}$ and $\bar{X}$ be the population mean of the main character y and the auxiliary character x for the population $U:(U_1, U_2, \ldots, U_N)$. The population U is supposed to be composed of N_1 responding and N_2 non-responding units. From the population of size N, a sample of size n is selected by using *SRSWOR* method of sampling and it was observed that n_1 units respond and n_2 units don't respond. Further, by making extra effort, a sub sample of size $r = \frac{n_2}{K}$ $(K > 1)$ is drawn from n_2 non-responding units by using *SRSWOR* method of sampling. Hence, we have n_1 units from respondent group and r units from non-respondent group of the population in the sample for which the value of the y character is obtained. Hansen and Hurwitz (1946) proposed the estimator for $\bar{Y}$, which is given as follows:

$$\bar{y}^* = \frac{n_1}{n}\bar{y}_1 + \frac{n_2}{n}\bar{y}'_2, \tag{2.1}$$

where $\bar{y}_1$ and $\bar{y}'_2$ are the sample means based on n_1 and r units respectively.

The estimator $\bar{y}^*$ is unbiased and the $V(\bar{y}^*)$ is given by

$$V(\bar{y}^*) = \frac{f}{n}S_y^2 + \frac{W_2(K-1)}{n}S_{y_2}^2, \tag{2.2}$$

where $f = \frac{N-n}{N}$, $W_i = \frac{N_i}{N}$ $(i = 1, 2)$, S_y^2 and $S_{y_2}^2$ are the population mean square of the character y for the whole population and for the non-responding part of the population.

In case, when $\bar{X}$ is unknown, then we select a larger sample of size $n'(n > n)$ from the population of size N by using *SRSWOR* method of sampling and we estimate $\bar{X}$ by $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$ and again draw a sub-sample of size n and observe y character. Now, it has been observed that n_1 units respond and n_2 units do not respond and then we select a sub sample of size $r = \frac{n_2}{K}$ $(K > 1)$ from n_2 non-responding units by using *SRSWOR* method of sampling and consequently the conventional and alternative ratio and product type estimators for population mean under scheme A (two phase sampling in the presence of non-response) are given by Khare and Srivastava (1993) which are given as follows:

$$T_1 = \frac{\bar{y}^*}{\bar{x}^*}\bar{x}', \quad T_2 = \frac{\bar{y}^*}{\bar{x}'}\bar{x}^*, \quad T_3 = \frac{\bar{y}^*}{\bar{x}}\bar{x}', \quad T_4 = \frac{\bar{y}^*}{\bar{x}'}\bar{x} \tag{2.3}$$

Now, we propose transformed two-phase sampling ratio and product type estimators for population mean $\bar{Y}$ in the presence of non-response, which are given as follows:

Case 1: $\bar{X}$ is unknown and we have incomplete information on y and x for the sample of the size n. Then the conventional transformed two phase sampling ratio and product type estimators for $\bar{Y}$ in the presence of non response are given by

$$t_1 = \frac{\bar{y}^*(\bar{x}' + D)}{(\bar{x}^* + D)} \text{ and } t_2 = \frac{\bar{y}^*(\bar{x}^* + D')}{(\bar{x}' + D')}, \tag{2.4}$$

where D and D' are the characterizing scalar having positive values and

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

Here $\bar{x}_1$ and $\bar{x}_2$ denote the sample means of responding and non-responding units in the sample of size n based on n_1 and n_2 units. We also denote $\bar{x}$ and $\bar{x}_2'$ by sample means of x character based on n and r units.

Case 2: $\bar{X}$ is unknown and incomplete information on y, but complete information on x is available for the sample of size n. Then the alternative transformed two phase sampling ratio and product type estimators for $\bar{Y}$ in the presence of non response are given by

$$t_3 = \frac{\bar{y}^*(\bar{x}' + D_1)}{(\bar{x} + D_1)} \text{ and } t_4 = \frac{\bar{y}^*(\bar{x} + D_1')}{(\bar{x}' + D_1')}, \tag{2.5}$$

where D_1 and D_1' are the characterizing scalar having positive values.

3. THE BIAS AND MEAN SQUARE ERROR (MSE) OF THE PROPOSED ESTIMATORS

The expressions for bias$(.)$ and $MSE(.)$ of the proposed estimators t_1, t_2, t_3 and t_4 up-to the terms of order n^{-1} are given as follows:

$$B(t_1)=\frac{f_1}{n(\bar{X}+D)}\left[\frac{\bar{Y}}{\bar{X}+D}S_x^2-S_{yx}\right]+\frac{W_2(K-1)}{n(\bar{X}+D)}\left[\frac{\bar{Y}}{\bar{X}+D}S_{x(2)}^2-S_{yx(2)}\right] \tag{3.1}$$

$$MSE(t_1)=V(\bar{y}^*)+\left(\frac{\bar{Y}}{\bar{X}+D}\right)^2\left[\frac{f_1}{n}S_x^2+\frac{W_2(K-1)}{n}S_{x(2)}^2\right]$$
$$-2\left(\frac{\bar{Y}}{\bar{X}+D}\right)\left[\frac{f_1}{n}S_{yx}+\frac{W_2(K-1)}{n}S_{yx(2)}\right] \tag{3.2}$$

$$B(t_2)=\frac{1}{(\bar{X}+D')}\left[\frac{f_1}{n}S_{yx}+\frac{W_2(K-1)}{n}S_{yx(2)}\right] \tag{3.3}$$

$$MSE(t_2)=V(\bar{y}^*)+\left(\frac{\bar{Y}}{\bar{X}+D}\right)^2\left[\frac{f_1}{n}S_x^2+\frac{W_2(K-1)}{n}S_{x(2)}^2\right]$$
$$+2\left(\frac{\bar{Y}}{\bar{X}+D}\right)\left[\frac{f_1}{n}S_{yx}+\frac{W_2(K-1)}{n}S_{yx(2)}\right] \tag{3.4}$$

$$B(t_3)=\frac{f_1}{n(\bar{X}+D_1)}\left[\frac{\bar{Y}}{\bar{X}+D_1}S_x^2-S_{yx}\right] \tag{3.5}$$

$$MSE(t_3)=V(\bar{y}^*)+\frac{f_1}{n}\left[\left(\frac{\bar{Y}}{\bar{X}+D_1}\right)^2S_x^2-2\left(\frac{\bar{Y}}{\bar{X}+D_1}\right)S_{yx}\right] \tag{3.6}$$

$$B(t_4)=\frac{f_1}{n(\bar{X}+D_1')}S_{yx} \tag{3.7}$$

and

$$MSE(t_4)=V(\bar{y}^*)+\frac{f_1}{n}\left[\left(\frac{\bar{Y}}{\bar{X}+D_1'}\right)^2S_x^2+2\left(\frac{\bar{Y}}{\bar{X}+D_1'}\right)S_{yx}\right], \tag{3.8}$$

where $f_1 = \dfrac{n' - n}{n'}$, $(S_x^2, S_{x(2)}^2)$ denote the population mean square of x character for the whole population and for the non-response group of the population and $(S_{yx}, S_{yx(2)})$ denote the covariance between y and x character for the whole population and for the non-response group of the population.

The expressions for the optimum values of D, D', D_1 and D_1', which minimize the $MSE(t_i)$ $i = 1, 2, 3, 4$ respectively are given as follows:

$$D_{opt} = \bar{X}\left[\frac{\dfrac{f_1}{n}S_x^2 + \dfrac{W_2(K-1)}{n}S_{x(2)}^2}{\dfrac{f_1}{n}S_{yx} + \dfrac{W_2(K-1)}{n}S_{yx(2)}}R - 1\right] \tag{3.9}$$

$$D'_{opt} = -\bar{X}\left[\frac{\dfrac{f_1}{n}S_x^2 + \dfrac{W_2(K-1)}{n}S_{x(2)}^2}{\dfrac{f_1}{n}S_{yx} + \dfrac{W_2(K-1)}{n}S_{yx(2)}}R + 1\right] \tag{3.10}$$

$$D_{1(opt)} = \bar{X}\left[\frac{S_x^2}{S_{yx}}R - 1\right] \tag{3.11}$$

$$D'_{1(opt)} = -\bar{X}\left[\frac{S_x^2}{S_{yx}}R + 1\right], \tag{3.12}$$

where $R = \dfrac{\bar{Y}}{\bar{X}}$.

If β is closed to $\beta_{(2)}$ and n' is large enough so that $\dfrac{1}{n'}$ is neglected, then in this case,

$$D_{opt} = \bar{X}\left[\frac{R}{\beta} - 1\right] \text{ and } D'_{opt} = -\bar{X}\left[\frac{R}{\beta} + 1\right], \tag{3.13}$$

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

Now, the expressions of $MSE(t_i)$ $i = 1, 2, 3, 4$ for the optimum values of D, D', D_1 and D_1' respectively are given as follows:

Table 1: The expressions of $MSE(.)_{opt}$ of the proposed estimators

Estimator	Expressions of $MSE(.)_{opt}$
t_1 or t_2	$MSE(t_i)_{opt} = V(\bar{y}^*) - \dfrac{\left[\dfrac{f_1}{n} S_{yx} + \dfrac{W_2(K-1)}{n} S_{yx(2)}\right]^2}{\left[\dfrac{f_1}{n} S_x^2 + \dfrac{W_2(K-1)}{n} S_{x(2)}^2\right]}$, $i = 1, 2$
t_3 or t_4	$MSE(t_i)_{opt} = V(\bar{y}^*) - \dfrac{\dfrac{f_1}{n} S_{yx}^2}{S_x^2}$, $i = 3, 4$

4. COMPARISION OF THE PROPOSED ESTIMATORS WITH THE RELEVANT ESTIMATORS

The conventional and alternative regression estimators T_{lr} and T'_{lr} for population mean under scheme A (Khare and Srivastava, 1995) are defined by

$$T_{lr} = \bar{y}^* + b^*(\bar{x}' - \bar{x}^*) \tag{4.1}$$

and

$$T'_{lr} = \bar{y}^* + b^{**}(\bar{x}' - \bar{x}), \tag{4.2}$$

where $b^* = \dfrac{\hat{S}_{yx}}{\hat{S}_x^2}$ and $b^{**} = \dfrac{\hat{S}_{yx}}{s_x^2}$. Also $\hat{S}_x^2$ and s_x^2 (Sample mean square) denote the estimates of S_x^2 based on $n_1 + r$ observations and n observations respectively.

In case of unknown $\bar{X}$ and incomplete information on y and x characters for the selected sample units, the regression estimator for population mean under scheme A (derived from the difference estimator by replacing the optimum value of the constant by its estimate) is given by

$$T_{lr(o)} = \bar{y}^* + \frac{\left[\dfrac{f_1}{n} \hat{S}_{yx} + \dfrac{W_2(K-1)}{n} \hat{S}_{yx(2)}\right]}{\left[\dfrac{f_1}{n} \hat{S}_x^2 + \dfrac{W_2(K-1)}{n} \hat{S}_{x(2)}^2\right]}(\bar{x}' - \bar{x}^*) \tag{4.3}$$

The MSE of T_{lr}, T'_{lr} and $T_{lr(o)}$ to terms of order n^{-1} and n'^{-1} are given as follows:

$$MSE(T_{lr}) = V(\bar{y}^*) + \beta^2\left[\frac{f_1}{n}S_x^2 + \frac{W_2(K-1)}{n}S_{x(2)}^2\right]$$

$$-2\beta\left[\frac{f_1}{n}S_{yx} + \frac{W_2(K-1)}{n}S_{yx(2)}\right] \quad (4.4)$$

$$MSE(T'_{lr}) = V(\bar{y}^*) + \frac{f_1}{n}\left[\beta^2 S_x^2 - 2\beta\, Syx\right] \quad (4.5)$$

and

$$MSE(T_{lr(o)}) = V(\bar{y}^*) - \frac{\left[\frac{f_1}{n}S_{yx} + \frac{W_2(K-1)}{n}S_{yx(2)}\right]^2}{\left[\frac{f_1}{n}S_x^2 + \frac{W_2(K-1)}{n}S_{x(2)}^2\right]} \quad , \quad (4.6)$$

where β and $\beta_{(2)}$ are the regression coefficients of y on x for the whole population and for the non-response part of the population.

For $D = D' = D_1 = D_1' = 0$, the estimators (t_1, t_3) and (t_2, t_4) reduces to the conventional and alternative two-phase sampling ratio (T_1, T_3) and product (T_2, T_4) estimators for the population mean in the presence of non-response (Khare and Srivastava, 1993).

For, $D = \frac{\bar{y}^*}{b^*} - \bar{x}^*$ and $D' = -\left(\frac{\bar{y}^*}{b^*} + \bar{x}'\right)$ the estimators t_1 and t_2 reduce to the estimator T_{lr} and for $D_1 = \frac{\bar{y}^*}{b^{**}} - \bar{x}$ and $D_1' = -\left(\frac{\bar{y}^*}{b^{**}} + \bar{x}'\right)$ the estimators t_3 and t_4 reduce to the estimator T'_{lr}.

It is observed that for positive values of D, D', D_1 and D_1', we have

$|B(t_1)| < |B(T_1)|$, $|B(t_2)| < |B(T_2)|$, $|B(t_3)| < |B(T_3)|$ and $|B(t_4)| < |B(T_4)|$

On comparing the $MSE(t_i)$ with $MSE(T_i)$ and $MSE(\bar{y}^*)$ for $\beta = \beta_{(2)}$, we see that

$$MSE(t_1) < MSE(T_1) < MSE(\bar{y}^*), \text{ if } \frac{1}{2}\frac{C_x}{C_y} < \rho < \frac{1}{2}\left(1 + \frac{\bar{X}}{\bar{X} + D}\right)\frac{C_x}{C_y} \quad (4.7)$$

$$MSE(t_2) < MSE(T_2) < MSE(\bar{y}^*), \text{ if } -\frac{1}{2}\left(1 + \frac{\bar{X}}{\bar{X} + D'}\right)\frac{C_x}{C_y} < \rho < -\frac{1}{2}\frac{C_x}{C_y} \quad (4.8)$$

$$MSE(t_3) < MSE(T_3) < MSE(\bar{y}^*) \text{, if } \frac{1}{2}\frac{C_x}{C_y} < \rho < \frac{1}{2}\left(1 + \frac{\bar{X}}{\bar{X}+D}\right)\frac{C_x}{C_y} \tag{4.9}$$

and

$$MSE(t_4) < MSE(T_4) < MSE(\bar{y}^*) \text{, if } -\frac{1}{2}\left(1 + \frac{\bar{X}}{\bar{X}+D'}\right)\frac{C_x}{C_y} < \rho < -\frac{1}{2}\frac{C_x}{C_y} \tag{4.10}$$

For the optimum value of D, D', D_1 and D_1', the amounts of bias in t_1, t_1, t_3 and t_4 are negligible and decreases as n increases and up-to the terms of order n^{-1}. The estimators t_1 and t_1 are equally efficient to T_{lr} if $\beta = \beta_{(2)}$ while t_3 and t_4 are always as equally efficient as T_{lr}'. The optimum values of D, D', D_1 and D_1' depend upon the unknown parameters, so for the practical purpose, one may use the prior information available from the past data [Reddy (1978), Tripathi *et. al* (1983)]. If there is no prior information or guess value available from the past data, then one may estimate the optimum values of D, D', D_1 and D_1' on the basis of sample observations available to the investigator without having any loss in the efficiency [Srivastava and Jhajj (1983)] of the estimators t_1, t_2, t_3 and t_4 up-to the terms of order n^{-1}.

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

Let C_0 be the total cost (fixed) of the survey apart from overhead cost.

The expected total of the survey apart from overhead cost is given by

$$C = C_1'n' + n\left(C_1 + C_2W_1 + C_3\frac{W_2}{K}\right), \tag{5.1}$$

where

C_1' = The cost per unit of identifying and observing auxiliary character at the first phase.

C_1 = The cost per unit of mailing questionnaire/visiting the unit at the second phase.

C_2 = The cost per unit of collecting and processing data from n_1 responding units.

C_3 = The cost per unit of obtaining and processing data after extra effort from the sub-sampled units.

$W_i = \dfrac{N_i}{N}$ $(i = 1, 2)$ denote the response and non-response rate in the population respectively.

The expression for $MSE(t_i)$, $i = 1, 2, 3, 4$ can be written as follows:

$$MSE(t_i) = \frac{M_{0i}}{n} + \frac{M_{1i}}{n'} + \frac{K}{n} M_{2i} + \text{terms independent from } n', \ n \text{ and } K,$$

where M_{0i}, M_{1i} and M_{2i} are the coefficients of the terms of $\dfrac{1}{n}$, $\dfrac{1}{n'}$ and $\dfrac{K}{n}$ respectively in the expression of $MSE(t_i)$, $i = 1, 2, 3, 4$.

Let us define a function ψ is given by

$$\psi = MSE(t_i) + \lambda_i \left\{ C_1' n' + n \left(C_1 + C_2 W_1 + C_3 \frac{W_2}{K} \right) \right\} \tag{5.2}$$

Now differentiating ψ with respect to n', n and K and equating them to zero, we get,

$$n' = \sqrt{\frac{M_{1i}}{\lambda_i C_1'}} \tag{5.3}$$

$$n = \sqrt{\frac{M_{0i} + K M_{2i}}{\lambda_i \left(C_1 + C_2 W_1 + C_3 \frac{W_2}{K} \right)}} \tag{5.4}$$

and

$$\frac{n}{K} = \sqrt{\frac{M_{2i}}{\lambda_i C_3 W_2}} \tag{5.5}$$

Since $n' > n$, so, we should have $C_1' < \dfrac{M_{1i} \left(C_1 + C_2 W_1 + \dfrac{C_3 W_2}{K_{opt}} \right)}{M_{0i} + K_{opt} M_{2i}}$.

Now putting the value of n in (5.5), we get,

$$K_{opt} = \sqrt{\frac{C_3 W_2 M_{0i}}{(C_1 + C_2 W_1) M_{2i}}} \tag{5.6}$$

For $K > 1$, the choice of C_3 should be such that $C_3 > \dfrac{(C_1 + C_2 W_1) M_{2i}}{W_2 M_{0i}}$.

Putting the value of n', n and K_{opt} from (5.4), (5.3) and (5.6) in (5.1), we get

$$\sqrt{\lambda_i} = \frac{1}{C_0}\left[\sqrt{C_1' M_{1i}} + \sqrt{\left\{C_1 + C_2 W_1 + \frac{C_3 W_2}{K_{opt}}\right\}\{M_{0i} + K_{opt} M_{2i}\}}\right] \tag{5.7}$$

It has also been seen that the determinant of the matrix of second order derivative of ψ with respect to n', n and K is negative for the optimum values of n', n and K, which shows that the solution for n', n given by (5.3) and (5.4) using (5.6), (5.7) and the optimum value of K for C for $C \le C_0$ minimizes the $V(t_i)$. The minimum value of $MSE(t_i)$ for the optimum value n', n and K are given by

$$MSE(t_i)_{\min} = \frac{1}{C_0}\left[\sqrt{C_1' M_{1i}} + \sqrt{(M_{0i} + K_{opt} M_{2i})\left(C_1 + C_2 W_1 + C_3 \frac{W_2}{K_{opt}}\right)}\right]^2 - \frac{S_y^2}{N}. \tag{5.8}$$

6. DETERMINATION OF n', n AND K FOR A SPECIFIED VARIANCE $V = V_0$

From (5.6), we see that the optimum value of K is independent of the total cost or specified precision. Let V_0 be the variance of the estimator t_i, $i = 1, 2, 3, 4$ fixed in advance, then we have,

$$V_0 = \frac{M_{0i}}{n} + \frac{M_{1i}}{n'} + \frac{K}{n} M_{2i} - \frac{S_y^2}{N} \tag{6.1}$$

After putting the value of n', n and K_{opt} in (6.1), we get

$$\sqrt{\frac{1}{\lambda_i}} = \frac{1}{\left(V_0 + \dfrac{S_y^2}{N}\right)}\left[\sqrt{C_1' M_{1i}} + \sqrt{\left\{C_1 + C_2 W_1 + \frac{C_3 W_2}{K_{opt}}\right\}\{M_{0i} + K_{opt} M_{2i}\}}\right] \tag{6.2}$$

Now, putting the value of $\sqrt{\frac{1}{\lambda_i}}$ from (6.2) and K_{opt} from (5.6) in (5.3) and (5.4), we can obtain the value of n' and n for which the estimator t_i, $i=1,2,3,4$ attains the variance V_0 with expected cost given by

$$C(t_i)=\frac{1}{\left(V_0+\frac{S_y^2}{N}\right)}\left[\sqrt{C_1' M_{1i}}+\sqrt{\left\{C_1+C_2W_1+\frac{C_3W_2}{K_{opt}}\right\}\{M_{0i}+K_{opt}M_{2i}\}}\right]^2 \tag{6.3}$$

7. EMPIRICAL STUDY

Data set I:

The data from the population of 100 villages of Satna district on un-irrigated area (y) in acres as a study character and total area of villages(in acres) as auxiliary character (x) have been taken from the Census book, 1971. The values of the parameters of the population are given as follows:

$\bar{X}=1555.77$, $S_x=1740.68$, $\bar{Y}=346.75$, $S_y=388.94$,

$\rho=0.54$, $S_{yx}=36415570$, $R=0.22$.

The non-response rate in the population is considered to be 25%, 20% and 15%. So, the values of the population parameters based on the non-responding parts, which are taken as the last 25%, 20% and 15% units of the population are given as follows:

Table 2: Values of unknown parameters of population at different values of W_2

	$W_2=0.25$	$W_2=0.20$	$W_2=0.15$
$\bar{X}_2$	1400.57	1462.10	1372.10
$\bar{Y}_2$	348.69	373.44	207.18
$S_{x(2)}$	1953.75	2121.52	2398.08
$S_{y(2)}$	321.91	344.46	213.71
$S_{yx(2)}$	342833.09	426156.23	399235.90
ρ_2	0.55	0.58	0.78

In the present empirical study, the data set I suggests the use of estimators t_1, t_3, T_1 and T_3 for the estimation of the population mean of un-irrigated area.

For $n'=50$, $n=25$, the performance of t_1 w.r.t. T_1 for the various values of W_2 (for the fixed value of K) has been computed as given in the table 3.

Table 3: $RE(t_1)$ in % w.r.t. T_1 for different values of D in relation to W_2 (for fixed $K=1.5$)

$W_2=0.25$		$W_2=0.20$		$W_2=0.15$	
D	$MSE(t_1)$ $RE(t_1)$	D	$MSE(t_1)$ $RE(t_1)$	D	$MSE(t_1)$ $RE(t_1)$
0.000	5004.02 (100.00)	0.000	4910.98 (100.00)	0.000	4762.05 (100.00)
1475.56*	4040.54 (123.85)	**1437.56***	3985.22 (123.23)	**1586.16***	3751.44 (126.94)
3725.56	4218.31 (118.63)	2937.56	4093.66 (119.97)	2336.13	3781.93 (125.92)
5975.56	4396.68 (113.81)	5187.56	4296.25 (114.31)	3586.16	3885.36 (122.56)
7725.56	4495.20 (111.32)	6687.56	4395.69 (111.72)	4836.16	3984.12 (119.53)
10225.60	4595.67 (108.89)	8887.56	4494.32 (109.27)	6586.16	4093.99 (116.32)
14225.60	4699.35 (106.48)	11687.60	4699.47 (104.50)	8836.16	4195.86 (113.49)
20975.60	4798.20 (104.29)	16687.60	4699.47 (104.50)	12086.20	4294.32 (110.89)
20975.60	4798.20 (104.29)	16687.60	4699.47 (104.50)	12086.20	4294.32 (110.89)
55725.60	4950.25 (101.09)	25937.60	4798.53 (102.34)	18086.20	4399.98 (108.23)
107726.0	4999.96	41937.60	4874.42	30086.20	4498.62

	(100.08)		(100.75)		(105.86)
113226.0	5002.62 (100.03)	51937.60	4899.52 (100.23)	74836.20	4599.76 (103.53)
116476.0	5004.08 (99.998)	58187.60	4911.06 (99.998)	91086.20	4612.75 (103.24)

(*: optimum value of D).

From table 3, we observe that the $MSE(t_1)$ w.r.t. $MSE(T_1)$ decreases as W_2 decreases. The estimators t_1 is more efficient than T_1 in a wide range of $D > 0$ for the different values of W_2. It is also to be noted that the optimum value of D is different for different value of W_2 (for fixed value of K) . For the data set I, we also observe that for the fixed value of W_2, the $MSE(t_1)$ is less than the $MSE(T_1)$ for the optimum value of D for the different values of K. However, t_1 is more efficient than T_1 in a wide range of D for the different values of K. It is also to be noted that the optimum value of D is different for different value of K (for fixed value of W_2).

Similarly for the data set I, we observe that the estimator t_3 has minimum value of MSE for the common optimum value of D_1 for different values of W_2 (for fixed K) and for different values of K (for fixed W_2). For the optimum value of D_1, the $RE(t_3)$ w.r.t. T_3 increases as W_2 decreases for the fixed value of K. However, the $RE(t_3)$ w.r.t. T_3 also increases as K increases for the fixed value of W_2. It has also been observed that the estimator t_3 is more efficient than T_3 in a wide range of D_1 for different values of K and W_2.

Data set II:

For studying the performance of t_1 and t_3 for fixed cost and for specified precision $V = V_0$, we are considering the data from the population of 96 villages of rural area under police station, Singur, District Hooghly obtained from 'District Census Handbook, 1981, published by government of India. The number of cultivators and the population of villages are considered as study character (y) and the auxiliary character (x). The values of the parameters are given as follows:

$$\bar{X} = 1807.23, \quad S_x = 1921.77, \quad \bar{Y} = 185.22, \quad S_y = 195.03, \quad \rho = 0.90,$$
$$S_{yx} = 33883588.$$

The non-response rate in the population of 96 villages is considered to be 25% (from 49th unit -72th unit). So, the values of the population parameters based on the non-responding parts, which are given as follows:

$\bar{X}_2 = 1571.71$, $S_{x(2)} = 1068.44$, $\bar{Y}_2 = 128.46$, $S_{y(2)} = 97.82$,
$\rho_2 = 0.90$, $S_{yx(2)} = 93560.01$.

Table 4: Relative efficiency of proposed estimators w.r.t. $\bar{y}^*$ for fixed cost $C = C_0$ and the expected cost for specified precision $V = V_0$

C_0 = Rs. 500.00 (fixed)					$V_0 = 2000$ cultivators		
Est.	K_{opt}	n'_{opt}	n_{opt}	$MSE(RE(.))$ w.r.t. $\bar{y}^*$	n'_{opt}	n_{opt}	Expected cost C (in Rs.)
$\bar{y}^*$	4.70	---	24	1577.97(100.00)	---	20	411.94
T_1	4.11	46	12	1014.28(155.58)	27	07	294.32
t_1	4.36	48	12	963.81(163.72)	27	07	283.78
T_3	1.76	43	10	1268.13(124.43)	30	07	347.28
t_3	1.68	43	9	1233.16(127.96)	30	06	399.99
$T_{lr(o)}$	4.36	48	12	963.81(163.72)	27	07	283.78

From the table 4, we observe that the estimator t_1 is more efficient than the estimators $\bar{y}^*$, T_1, t_3 and T_3. The estimator t_3 is also more efficient than $\bar{y}^*$ and T_3. However, the estimator t_1 is equally efficient to $T_{lr(o)}$ for fixed cost. For the fixed precision $V = V_0$, the estimator t_1 and $T_{lr(o)}$ have equal minimum cost. The estimator t_1 is found to be cheaper than t_3, T_1 and T_3. The estimator t_3 is also cheaper than $\bar{y}^*$ and T_3 for fixed precision $V = V_0$.

8. CONCLUDING REMARKS

The effective choice of transformation parameters D, D', D_1 and D'_1 for the proposed estimators t_1, t_2, t_3 and t_4 in comparison to the relevant estimators T_1, T_2, T_3 and T_4 respectively can be obtained from the equations (4.7), (4.8), (4.9) and (4.10). However, these values depend on the population parameters where prior information can be used (Reddy (1978), Tripathi *et. al* (1983)). In lack of prior information, the estimated values of the parameters based on sample values can be used (Srivastava and Jhajj (1983)). An empirical study has also been made on the basis of data set I for the choice of transformation parameter D for the proposed estimator t_1 w.r.t. T_1 and a discussion is also made for the choice of W_2 and K in the section 7.

The method of transforming auxiliary character can also used in the case of cluster sampling and PPS sampling with replacement. If the population is heterogeneous then we can also make use of transformation in each stratum and separate transformed ratio and combined transformed ratio estimators for population mean can be proposed under the scheme A. The proposed estimators may be used in the field of Medical, Agricultural and Biological Sciences where transformation of auxiliary variable can be used efficiently using prior information.

REFERENCES

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. and Srivastava, S. (1993): Estimation of population mean using auxiliary character in presence of non-response. *Nat. Acad. Sci. Lett. India,* **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 (A) II**, 195-203.

Khare, B.B. and Srivastava, S. (1996): Transformed product type estimators for population mean in presence of soft core observations. *Proc. Math. Soc.,* **12**, 29-34.

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.

Mohanty, S. and Das, M.N. (1971): Use of transformation in sampling. *J. Indian Soc. Agricultural Statist*., **23**, 83-87.

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

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.

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.

Tin, M. (1965): Comparison of some ratio estimators. *J. Amer. Statist. Assoc*., **60**, 294-307.

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.

Quenouille, M.H. (1956): Notes on bias in estimation. *Biometrika*, **43**, 353-360.

Received : 18-01-2009

Department of Statistics,
Banaras Hindu University, Varanasi-221005, India.
e-mail: bbkhare56@yahoo.com