A CLASS OF RATIO-CUM-PRODUCT TYPE ESTIMATORS UNDER DOUBLE SAMPLING IN THE PRESENCE OF NON-RESPONSE

Sunil Kumar[*][†], Housila P. Singh[‡], Sandeep Bhougal[§] and Rahul Gupta[*]

Abstract

In the present paper, we have proposed a class of ratio – type estimators for estimating the population mean $\bar{Y}$ of the study variate y under double sampling in the presence of non-response, where information on the auxiliary variable is not known. For an appropriate weight a and a good range of α values, it is found that the proposed class of estimators is more efficient than some estimators which are obtained by applying congenial values of a and α. Comparison of the proposed class of estimators with other estimators is also worked out. Also the good range of α is obtained empirically for different values of a and k.

1. Introduction

In most of the socio-economic studies, several variables are considered simultaneously. For example, while conducting a household survey, the investigator may be interested in studying characteristics such as number of wage earners, per capita income, land holding, number of illiterate persons, number of females etc, see Tripathi and Khare [15, p. 2255]. For the study of several variables generally, we assume that information in most cases is not obtained at the first attempt even after some call-backs. An estimate obtained from such incomplete data may be misleading, especially when the respondents differ from the

[*]Department of Statistics, University of Jammu, Jammu-180004, J & K, India.
E-mail: (S. Kumar) `sunilbhougal06@gmail.com` (R. Gupta) `rahulgupta68@yahoo.com`
[†]Corresponding Author.
[‡]School of studies in Statistics, Vikram University, Ujjain-456010, M. P., India.
E-mail: `hpsujn@rediffmail.com`
[§]School of Mathematics, Shri Mata Vaishno Devi University, Katra, Jammu, J & K, India.
E-mail: `sandeep.bhougal@smvdu.ac.in`

non-respondents because the estimates can be biased, see Okafor and Lee [7, p. 183]. Hansen and Hurwitz [2] suggested a technique for adjusting non-response to address the bias problem. Their idea is to take a sub sample from the non-respondents to get an estimate for the subpopulation represented by the non-respondents. Using the Hansen and Hurwitz [2] procedure, Cochran [1] suggested the ratio and regression estimators of the population mean of the study variable, in which information on the auxiliary variable is obtained from all the sample units, and the population mean of the auxiliary variable is known, while some sample units failed to supply information on the study variable. Later, various authors have paid their attention towards the estimation procedure for the population mean in the presence of non-response using an auxiliary character x, including Rao [8,9], Khare and Srivastava [3,5] and Singh and Kumar [11].

When the population mean $\bar{X}$ of the auxiliary character x is not known, Khare and Srivastava [4], Okafor and Lee [7], Singh and Kumar [10], Tabasum and Khan [13, 14] have applied the Hansen and Hurwitz [2] technique for treating the non-response to double sampling for ratio and regression estimation.

In this paper, for the case where the population mean $\bar{X}$ of the auxiliary character x is not known in advance, and motivated by Sisodia and Dwivedi [12], a class of ratio and product estimators is presented in the presence of non-response. The expressions for bias, mean squared error (MSE) and a condition for attaining minimum MSE of the suggested class of estimators have been obtained under a large sample approximation. We have also obtained the optimum values of the first and second phase sample and sub sampling fraction which minimize the survey cost for specified precision. Numerical illustration is given in support of the present study.

2. The suggested class of estimators

Let y and x be the main and the auxiliary characters with population means $\bar{Y}$ and $\bar{X}$ respectively. When non-response occurs, the sub sampling procedure of Hansen and Hurwitz [2] is an alternative to call-backs and similar procedures. In this approach, the population of size N is assumed to be composed of two strata of sizes N_1 and $N_2 = N - N_1$ of 'respondents' and 'non-respondents'. When the population mean $\bar{X}$ of the auxiliary character x is not known, then to furnish an estimate of the population mean $\bar{X}$ of the auxiliary character x, a large first phase sample of size n' is selected from a population of N units by the simple random sampling without replacement (SRSWOR) method of sampling and the auxiliary character x is measured. A smaller second phase sample of size n $(< n')$ is selected by the SRSWOR sampling scheme, and the character y is measured on it. Let us assume that in the first phase, all the n' units supplied information on the auxiliary character x. At the second phase, let n_1 units respond on the main character y and then n_2 units do not respond in the sample of size n. Using the Hansen and Hurwitz [2] approach to sub sampling, from the n_2 non-respondents a sub-sample, of size m units is selected using the SRSWOR sampling scheme, and enumerated by direct interview, such that $m = (n_2/k)$, $k > 1$, where k is the inverse sampling rate. Here we assume that response is obtained for all the m units. This method of double sampling can be applied in a household survey where the household size is used as an auxiliary variate x for the estimation of family expenditure. Information can be obtained completely on the family size, while there may be some non-response on the household expenditure, see Tabasum and Khan [13, p. 301].

Now, we have $(n_1 + m)$ responding units on y and consequently the estimator for the population mean $\bar{Y}$ of the study character y using the sub-sampling scheme suggested

by Hansen and Hurwitz [2] is given by

(1) $\bar{y}^* = w_1 \bar{y}_1 + w_2 \bar{y}_{2m},$

where $\bar{y}_1$ and $\bar{y}_{2m}$ denote the sample means of y based on n_1 and m units respectively, $w_1 = n_1/n$ and $w_2 = n_2/n$.

The estimator $\bar{y}^*$ is unbiased and has variance

(2) $\text{Var}\left(\bar{y}^*\right) = \left(\dfrac{1-f}{n}\right) S_y^2 + \dfrac{W_2\,(k-1)}{n} S_{y(2)}^2,$

where $f = n/N$, $W_2 = N_2/N$, S_y^2 and $S_{y(2)}^2$ are the population mean square of y for the entire population and for the non-responding part of the population, respectively.

Similarly, for estimating the population mean $\bar{X}$ of the auxiliary character x, the estimator $\bar{x}^*$ is given by

$\bar{x}^* = w_1 \bar{x}_1 + w_2 \bar{x}_{2m},$

where $\bar{x}_1$ and $\bar{x}_{2m}$ are the sample means of x based on n_1 and m units, respectively.

Now, we define a class of ratio-cum-product type estimators as

(3) $T_d^* = (1-a)\,\bar{y}^* + a\bar{y}^*\left(\bar{x}'/\bar{x}^*\right)^\alpha;\ a > 0$

where a and α are suitably chosen constants; $\bar{x}'$ is the sample mean based on a large preliminary sample of size n'.

For $\alpha > 0$, T_d^* is ratio-type estimator and for $\alpha < 0$, T_d^* is product-type estimator.

The following Table 1 gives some estimators of the population mean $\bar{Y}$ which can be obtained by suitable choice of the scalars a and α.

Table 1. Some members of the proposed class of ratio-cum-product estimators T_d^*

Estimator	Values of a and α	
	a	α
(i) Usual unbiased estimator: $t_{d1}^* = \bar{y}^*$	0	0
(ii) Ratio estimator $t_{d2}^* = \bar{y}^*\left(\bar{x}'/\bar{x}^*\right)$ by Khare and Srivastava [3], Okafor and Lee [7] and Tabasum and Khan [13]	1	1
(iii) The estimator: $t_{d3}^* = \bar{y}^*\left(\bar{x}'/\bar{x}^*\right)^\alpha$	1	–
(iv) The estimator: $t_{d4}^* = \bar{y}^*\left\{2 - \left(\bar{x}'/\bar{x}^*\right)^\alpha\right\}$	-1	-
(v) The estimator: $t_{d5}^* = \bar{y}^*\left\{(1-a) + a\left(\bar{x}'/\bar{x}^*\right)\right\}$	–	1

3. Bias and variance of T_d^*

The bias and variance of T_d^*, to the first degree of approximation are given by

(4)
$$B\left(T_d^*\right) = a\bar{Y}\alpha\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\frac{(\alpha+1)}{2} - C\right\}C_x^2 \right. $$
$$\left. + \frac{W_2(k-1)}{n}\left\{\frac{(\alpha+1)}{2} - C_{(2)}\right\}C_{x(2)}^2\right],$$

(5)
$$\operatorname{Var}\left(T_d^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_y^2 + a\alpha\left(a\alpha - 2C\right)C_x^2\right\} \right.$$
$$\left. + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + a\alpha\left(a\alpha - 2C_{(2)}\right)C_{x(2)}^2\right\} + \left(\frac{1}{n'} - \frac{1}{N}\right)C_y^2\right],$$

For large N, the variance of T_d^* reduces to

(6)
$$\operatorname{Var}\left(T_d^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{a\alpha\left(a\alpha - 2C\right)C_x^2\right\} \right.$$
$$\left. + \frac{W_2(k-1)}{n}\left\{C_{y(2)}^2 + a\alpha\left(a\alpha - 2C_{(2)}\right)C_{x(2)}^2\right\} + \frac{C_y^2}{n}\right],$$
$$= \bar{Y}^2\left\{\frac{1}{n}C_y^2 + \frac{W_2(k-1)}{n}C_{y(2)}^2 + a^2\alpha\left(\alpha - 2\alpha_0\right)B\right\},$$

where

$$A = \left\{\left(\frac{1}{n} - \frac{1}{n'}\right)CC_x^2 + \frac{W_2(k-1)}{n}C_{(2)}C_{x(2)}^2\right\},$$

$$B = \left\{\left(\frac{1}{n} - \frac{1}{n'}\right)C_x^2 + \frac{W_2(k-1)}{n}C_{x(2)}^2\right\}, \quad C = \left(\rho_{yx}C_y\right)/C_x,$$

$$C_{(2)} = \left(\rho_{yx(2)}C_{y(2)}\right)/C_{x(2)}, \quad C_y^2 = S_y^2/\bar{Y}^2, \quad C_{y(2)}^2 = S_{y(2)}^2/\bar{Y}^2, \quad \alpha_0 = A/aB,$$

$$C_x^2 = S_x^2/\bar{X}^2, \quad C_{x(2)}^2 = S_{x(2)}^2/\bar{X}^2, \quad S_y^2 = \sum_{i=1}^{N}\left(y_i - \bar{Y}_1\right)/(N-1),$$

$$S_{y(2)}^2 = \sum_{i=1}^{N_2}\left(y_i - \bar{Y}_{(2)}\right)/(N_2-1), \quad S_x^2 = \sum_{i=1}^{N}\left(x_i - \bar{X}_1\right)/(N-1),$$

$$S_{x(2)}^2 = \sum_{i=1}^{N_2}\left(x_i - \bar{X}_{(2)}\right)/(N_2-1),$$

ρ_{yx} and $\rho_{yx(2)}$ are the correlation coefficients between y and x for the entire population and for the non-responding group of the population respectively.

The variance of T_d^* in (6) is minimized when

(7) $\alpha = A/(aB) = \alpha_o$ (say),

The optimum value of α cannot be uniquely determined, as is clear from (7). Thus, to obtain the optimum value of α, the value of a is fixed approximately in advance as being a weight. Thus, substituting the optimum value of α for a given value of a from (7) in

(6), we get

$$\min .\mathrm{Var}\left(T_d^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\frac{A\left(A - 2BC\right)}{B^2}\right\}C_x^2\right.$$

(8)
$$+ \frac{W_2\left(k-1\right)}{n}\left\{C_{y(2)}^2 + \frac{A\left(A - 2BC_{(2)}\right)}{B^2}C_{x(2)}^2\right\} + \left.\frac{C_y^2}{n}\right]$$

$$= \bar{Y}^2\left\{\frac{1}{n}C_y^2 + \frac{W_2\left(k-1\right)}{n}C_{y(2)}^2 - \frac{A^2}{B}\right\}.$$

Now we state the following theorem:

3.1. Theorem. *To the first degree of approximation,*

$$MSE\left(T_d^*\right) \geq \bar{Y}^2\left\{\frac{1}{n}C_y^2 + \frac{W_2\left(k-1\right)}{n}C_{y(2)}^2 - \frac{A^2}{B}\right\},$$

with equality holding if

$$\alpha = \alpha_o. \qquad\qquad \Box$$

The variance of t_{di}^*, $i = 1, 2\ldots, 5$ to the first degree of approximation, can be obtained by merely substituting suitable values of a and α into the variance of T_d^* as follows:

(9)
$$\mathrm{Var}\left(t_{d1}^*\right) = \bar{Y}^2\left\{\left(\frac{1}{n} - \frac{1}{n'}\right)C_y^2 + \frac{W_2\left(k-1\right)}{n}C_{y(2)}^2\right\},$$

(10)
$$\mathrm{Var}\left(t_{d2}^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_y^2 + \left(1 - 2C\right)C_x^2\right\}\right.$$
$$\left.+ \frac{W_2\left(k-1\right)}{n}\left\{C_{y(2)}^2 + \left(1 - 2C_{(2)}\right)C_{x(2)}^2\right\} + \left(\frac{1}{n'} - \frac{1}{N}\right)C_y^2\right],$$

(11)
$$\mathrm{Var}\left(t_{d3}^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_y^2 + \alpha\left(\alpha - 2C\right)C_x^2\right\}\right.$$
$$\left.+ \frac{W_2\left(k-1\right)}{n}\left\{C_{y(2)}^2 + \alpha\left(\alpha - 2C_{(2)}\right)C_{x(2)}^2\right\} + \left(\frac{1}{n'} - \frac{1}{N}\right)C_y^2\right],$$

(12)
$$\mathrm{Var}\left(t_{d4}^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_y^2 + \alpha\left(\alpha + 2C\right)C_x^2\right\}\right.$$
$$\left.+ \frac{W_2\left(k-1\right)}{n}\left\{C_{y(2)}^2 + \alpha\left(\alpha + 2C_{(2)}\right)C_{x(2)}^2\right\} + \left(\frac{1}{n'} - \frac{1}{N}\right)C_y^2\right],$$

(13)
$$\mathrm{Var}\left(t_{d5}^*\right) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{C_y^2 + a\left(a - 2C\right)C_x^2\right\}\right.$$
$$\left.+ \frac{W_2\left(k-1\right)}{n}\left\{C_{y(2)}^2 + a\left(a - 2C_{(2)}\right)C_{x(2)}^2\right\} + \left(\frac{1}{n'} - \frac{1}{N}\right)C_y^2\right].$$

The expression for $\min .\mathrm{Var}\left(T_d^*\right)$ in (8) implies that the estimator T_d^* will always be better than the estimators t_{di}^*, $i = 1, 2, \ldots, 5$ in α_o. Nevertheless, we shall further determine a range of α-values for which the estimator T_d^* will always have smaller variance than that of the estimators t_{di}^*, $i = 1, 2, \ldots, 5$.

4. Comparison of T_d^* with the other estimators

The range of α-values are derived in this section for which the variance of the estimator T_d^* is always less than that of the estimators t_{di}^*, $i = 1, 2, \ldots, 5$. The comparison of $\mathrm{Var}\left(T_d^*\right)$ *Vs* $\mathrm{Var}\left(t_{di}^*\right)$, $i = 1, 2, \ldots, 5$, are given below:

(i) $\text{Var}\,(T_d^*) < \text{Var}\,(t_{d1}^*)$ if $|\alpha - \alpha_0| < |\alpha_0|$, or equivalently

(14) $\min.\,(0,\,2\alpha_0) < \alpha < \max.\,(0,\,2\alpha_0)\,.$

(ii) $\text{Var}\,(T_d^*) < \text{Var}\,(t_{d2}^*)$ if $|\alpha - \alpha_0| < \left|\alpha_0 - \frac{1}{a}\right|$, or equivalently

(15) $\min.\,\{0,\,(2\alpha_0 - 1/a)\} < \alpha < \max.\,\{0,\,(2\alpha_0 - 1/a)\}\,.$

(iii) $\text{Var}\,(T_d^*) < \text{Var}\,(t_{d3}^*)$ if $|\alpha - \alpha_0| < \left|\alpha_0 \frac{a}{(a+1)}\right|$, or equivalently

(16) $\min.\,[0,\,\{2\alpha_0\,(a/a+1)\}] < \alpha < \max.\,[0,\,\{2\alpha_0\,(a/a+1)\}]\,.$

(iv) $\text{Var}\,(T_d^*) < \text{Var}\,(t_{d4}^*)$ if $|\alpha - \alpha_0| < \left|\alpha_0 \frac{a}{(a-1)}\right|$, or equivalently

(17) $\min.\,[0,\,\{2\alpha_0\,(a/a-1)\}] < \alpha < \max.\,[0,\,\{2\alpha_0\,(a/a-1)\}]\,.$

(vi) $\text{Var}\,(T_d^*) < \text{Var}\,(t_{d5}^*)$ if $|\alpha - \alpha_0| < |\alpha_0 - 1|$, or equivalently

(18) $\min.\,\{0,\,(2\alpha_0 - 1)\} < \alpha < \max.\,\{0,\,(2\alpha_0 - 1)\}\,.$

The difference (for a given value of a) between α and its optimum value α_o in the above inequalities provide quite a good range of α-values making the estimators T_d^* more efficient than t_{di}^*, $i = 1, 2, \ldots, 5$.

5. Determination of n', n and k

Let C^* be the total cost (fixed) of the survey apart from overhead cost. The expected total cost of the survey apart from overhead cost is given by

(19) $$C = c_1' n' + n\left(c_1 + c_2 W_1 + \frac{c_3 W_2}{k}\right)$$

where

$c_1' =$ the cost per unit of identifying and observing the auxiliary character,

$c_1 =$ 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 obtained from n_1 responding units,

$c_3 =$ the cost per unit of obtaining and processing data (after extra effort) from the sub sampled units, and

$W_1 = N_1/N$, $W_2 = N_2/N$ denote the response and non-response rate in the population.

The expressions $\text{Var}\,(t_{di}^*)$, $i = 0, 1, \ldots, 5$; $\text{Var}\,(t_{d0}^*) = Var\,(T_d^*)$ respectively, given by (6) and (9) to (13), can be written as

(20) $$\text{Var}\,(t_{di}^*) = \left\{\frac{1}{n}V_{0i} + \frac{1}{n'}V_{1i} + \frac{k}{n}V_{2i}\right\} + (\text{terms independent of } n', n \text{ and } k)$$

for $i = 0, 1, \ldots, 5$, where V_{0i}, V_{1i} and V_{2i} are respectively the coefficients of the terms $(1/n)$, $(1/n')$ and (k/n) in the expressions for $\text{Var}\,(t_{di}^*)$, $i = 0, 1, \ldots, 5$.

Let us define a function ϕ as follows:

(21) $$\phi = \text{Var}\,(t_{di}^*) + \lambda_i\left\{c_1' n' + n\left(c_1 + c_2 W_1 + \frac{c_3 W_2}{k}\right)\right\}.$$

Differentiating ϕ in (21) with respect to n', n and k, and equating to zero, we have

$$(22) \qquad n' = \sqrt{\frac{V_{1i}}{\lambda_i c_1'}},$$

$$(23) \qquad n = \sqrt{\frac{(V_{0i} + kV_{2i})}{\lambda_i \left(c_1 + c_2 W_1 + \frac{c_3 W_2}{k}\right)}},$$

$$(24) \qquad \frac{n}{k} = \sqrt{\frac{V_{2i}}{\lambda_i c_3 W_2}}.$$

Now putting the value of n from (23) in (24) we obtain the optimum value of k as

$$(25) \qquad k_{\text{opt}} = \sqrt{\frac{V_{0i} c_3 W_2}{(c_1 + c_2 W_1) V_{2i}}}.$$

Using the value of k_{opt} from (25), while putting the values of n' and n from (22) and (23) in (19), we have

$$(26) \qquad \sqrt{\lambda_i} = \frac{1}{C} \left\{ \sqrt{V_{1i} c_1'} + \sqrt{\left(V_{0i} + k_{\text{opt}} V_{2i}\right) \left(c_1 + c_2 W_1 + \frac{c_3 W_2}{k_{\text{opt}}}\right)} \right\}.$$

Thus the minimum value of $\text{Var}\left(t_{di}^*\right)$, $i = 0, 1, \ldots, 5$ for the optimum values of n', n and k is given by

$$(27) \qquad \min . Var\left(t_{di}^*\right) = \left[\frac{1}{C} \left\{ \sqrt{V_{1i} c_1'} + \sqrt{\left(V_{0i} + k_{\text{opt}} V_{2i}\right) \left(c_1 + c_2 W_1 + \frac{c_3 W_2}{k_{\text{opt}}}\right)} \right\}^2 \right.$$
$$\left. - \frac{S_y^2}{N} \right].$$

6. Empirical study

To illustrate the findings we consider some natural population data earlier considered by Khare and Srivastava [4].

The population of 100 consecutive trips (after leaving 20 outlier values) measured by two fuel meters for a small family car in normal usage given by Lewisi et al [6] has been taken into consideration. The measurements of the turbine meter (in ml) are considered as the main (study) character y, and the measurements of the displacement meter (in cm^3) are considered as the auxiliary character x. The last 25% of the values are treated as non-response units.

The values of the parameters are as follows:

$$\bar{Y} = 3500.12, \ \bar{X} = 260.84, \ C_y = 0.5941, \ C_x = 0.5996, \ C_{y(2)} = 0.4931,$$
$$C_{x(2)} = 0.5151, \ C = 0.9759, \ C_{(2)} = 0.9525, \ \rho_{yx} = 0.985, \ \rho_{yx(2)} = 0.995,$$
$$W_2 = 0.25, \ N = 100, \ n' = 50, \ n = 30.$$

We have computed the optimum value of α and the ranges of α needed for the proposed estimator T_d^* to be more efficient than t_{di}^*, $i = 1, 2, \ldots, 5$ and these are shown in Tables 2 and 3 respectively for different values of a and k.

Table 2 The optimum value of α for the proposed estimator T_d^* to be more efficient than t_{di}^*, $i = 1, 2, \ldots, 7$ for different values of a and k

$k \rightarrow$	1/2	1/3	1/4	1/5
$a \downarrow$				
0.10	9.7143	9.6739	9.6491	9.6324
0.25	3.8857	3.8696	3.8596	3.8529
0.50	1.9429	1.9348	1.9298	1.9265
0.75	1.2952	1.2899	1.2865	1.2843
0.85	1.1429	1.1381	1.1352	1.1332
0.95	1.0226	1.0183	1.0157	1.0139

It is observed from Table 2 that:

(1) For a fixed value of a, the optimum value of α decreases as the value of k decreases, while for a fixed value of k, it decreases as the value of a increases,

(2) When a approaches unity, the optimum values of α are almost stable for different decreasing values of k; and

(3) When a is close to zero, the optimum value of α is larger compared to the value of a and approaches unity for different values of k.

Table 3. Ranges of α for the proposed estimator T_d^* to be more efficient than t_{di}^*, $i = 1, 2, \ldots, 5$ for different values of a and k

$a \downarrow$	$k = 1/2$				
	Ranges of α for the estimator T_d^* to be better than:				
	t_{d1}^*	t_{d2}^*	t_{d3}^*	t_{d4}^*	t_{d5}^*
0.10	(0,19.43)	(0,9.43)	(0, 1.77)	(0, -2.16)	(0, 18.43)
0.25	(0, 7.77)	(0,3.77)	(0, 1.55)	(0, -2.59)	(0, 6.77)
0.50	(0, 3.89)	(0, 1.89)	(0, 1.30)	(0, -3.89)	(0, 2.89)
0.75	(0, 2.59)	(0, 1.26)	(0, 1.11)	(0, -7.77)	(0, 1.59)
0.85	(0, 2.29)	(0, 1.11)	(0, 1.05)	(0, -12.95)	(0, 1.29)
0.95	(0, 2.05)	(0, 0.99)	(0, 1.00)	(0, -38.86)	(0, 1.05)
	$k = 1/3$				
0.10	(0,19.35)	(0, 9.35)	(0, 1.76)	(0, -2.15)	(0, 18.35)
0.25	(0, 7.74)	(0, 3.74)	(0, 1.55)	(0, -2.58)	(0, 6.74)
0.50	(0, 3.87)	(0, 1.87)	(0, 1.29)	(0, -3.87)	(0, 2.87)
0.75	(0, 2.58)	(0, 1.25)	(0, 1.11)	(0, -7.74)	(0, 1.58)
0.85	(0, 2.28)	(0, 1.10)	(0, 1.05)	(0, -12.90)	(0, 1.28)
0.95	(0, 2.04)	(0, 0.98)	(0, 1.00)	(0, -38.70)	(0, 1.04)

Table 3. Continued

$a \downarrow$	$k = 1/4$				
	Ranges of α for the estimator T_d^* to be better than:				
	t_{d1}^*	t_{d2}^*	t_{d3}^*	t_{d4}^*	t_{d5}^*
0.10	(0,19.30)	(0, 9.30)	(0, 1.75)	(0, -2.14)	(0, 18.30)
0.25	(0, 7.72)	(0,0.37)	(0, 1.54)	(0, -2.57)	(0, 6.72)
0.50	(0, 3.86)	(0, 1.86)	(0, 1.29)	(0, -3.86)	(0, 2.86)
0.75	(0, 2.57)	(0, 1.24)	(0, 1.10)	(0, -7.72)	(0, 1.57)
0.85	(0, 2.27)	(0, 1.09)	(0, 1.04)	(0, -12.87)	(0, 1.27)
0.95	(0, 2.03)	(0, 0.98)	(0, 0.99)	(0, -38.60)	(0, 1.03ss)
	$k = 1/5$				
0.10	(0,19.27)	(0, 9.27)	(0, 1.75)	(0, -2.14)	(0, 18.27)
0.25	(0, 7.71)	(0, 3.71)	(0, 1.54)	(0, -2.57)	(0, 6.71)
0.50	(0, 3.85)	(0, 1.85)	(0, 1.28)	(0, -3.85)	(0, 2.85)
0.75	(0, 2.57)	(0, 1.24)	(0, 1.10)	(0, -7.71)	(0, 1.57)
0.85	(0, 2.27)	(0, 1.09)	(0, 1.04)	(0, -12.84)	(0, 1.27)
0.95	(0, 2.03)	(0, 0.98)	(0, 0.99)	(0, -38.53)	(0, 1.03)

Table 3 exhibits the ranges of α for the estimator T_d^* to be more efficient than t_{di}^*, $i = 1, 2, \ldots, 5$ for different values of a and k. It is observed from Table 3 that for a fixed value of a, the range of α decreases as the value of k decreases, while for a fixed value of k, it decreases as the value of a increases for all the estimators t_{di}^*, $i = 1, 2, \ldots, 5$.

Thus we conclude that for fixed a and varying k, or for fixed k and varying a, there is enough scope of selecting the scalar α to obtain better estimators than t_{di}^*, $i = 1, 2, \ldots, 5$.

6.1. Remark. Suppose that complete information on the auxiliary variable x is available for both the first and second samples, and that incomplete information on the study variable y is available. So, in this case, we use information on the $(n_1 + m)$ responding units on the main character y, and complete information on the auxiliary variable x from the sample of size n. We suggest a class of ratio-cum-product-type estimators T_d for y in the presence of non-response, which is given by

$$(28) \qquad T_d = (1 - a) \, \bar{y}^* + a\bar{y}^* \left(\frac{\bar{x}'}{\bar{x}} \right)^\alpha,$$

where (a, α) are as defined above. A large number of estimators can be identified for suitable values of (a, α). For $a = 0$, the estimator T_d reduces to the conventional estimator $\bar{y}^*$, while for $\{(a, \alpha) = (1, 1)\}$, it reduces to the estimator $t_{rd} = \bar{y}^* \left(\frac{\bar{x}'}{\bar{x}} \right)$ envisaged by Khare and Srivastava [4], and revisited by Tabasum and Khan [14].

To the first degree of approximation, the bias and MSE of T_d are respectively given by

$$(29) \qquad B\left(T_d\right) = \left(\frac{1}{n} - \frac{1}{n'}\right) \bar{Y} \left(\frac{a\alpha}{2}\right) C_x^2 \left(\alpha - 2C + 1\right),$$

$$(30) \qquad \text{MSE}\left(T_d\right) = \overline{Y}^2 \left[\left(\frac{1}{n} - \frac{1}{n'}\right) \left\{C_y^2 + a\alpha C_x^2 \left(a\alpha - 2C\right)\right\} + \left(\frac{1}{n'} - \frac{1}{N}\right) C_y^2 \right. \\ \left. + \frac{W_2\left(k-1\right)}{n} C_{y(2)}^2 \right].$$

The biases and MSEs of the estimators belonging to the class of estimators T_d defined by (28) can be obtained from (29) and (30) by giving suitable values to (a, α), respectively.

The MSE (T_d) in (30) is minimized for

$$(31) \qquad \alpha = \frac{C}{a} = \alpha_0^* \ \text{(say)}$$

Thus the resulting minimum MSE of T_d is given by

$$(32) \qquad \min .MSE\left(T_d\right) = \bar{Y}^2 \left[\left(\frac{1}{n} - \frac{1}{n'}\right) C_y^2 \left(1 - \rho^2\right) + \left(\frac{1}{n'} - \frac{1}{N}\right) C_y^2 \right. \\ \left. + \frac{W_2\left(k-1\right)}{n} C_{y(2)}^2 \right].$$

Following the procedure adopted in Section 5, cost aspects can also be discussed.

References

[1] Cochran, W. G. *Sampling Techniques* (3rd Edition, John Wiley & Sons, New York, 1977).

[2] Hansen, M. H. and Hurwitz, W. N. *The problem of non-response in sample surveys*, J. Amer. Statist. Assoc. **41**, 517–529, 1946.

[3] Khare, B. B. and Srivastava, S. *Estimation of population mean using auxiliary character in presence of non-response*, Nat. Acad. Sc. Letters, India, **16** (3), 111–114, 1993.

[4] Khare, B. B. and Srivastava, S. *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, 1995.

[5] Khare, B. B. and Srivastava, S. *Transformed ratio type estimators for the population mean in the presence of non-response*. Comm. Statist. - Theory Methods **26** (7), 1779–1791, 1997.

[6] Lewisi, P. A., Jones, P. W., Polak, J. W. and Tillotson, H. T. *The problem of conversion in method comparison studies*, Appl. Statist. **40** (1), 105–112, 1991.

[7] Okafor, F. C. and Lee, H. *Double sampling for ratio and regression estimation with sub sampling the non-respondent*. Survey Methodology **26**, 183–188, 2000.

[8] Rao, P. S. R. S. *Ratio Estimation with sub sampling the non-respondents*, Survey Methodology **12** (2), 217–230, 1986.

[9] Rao, P. S. R. S. *Ratio and regression estimates with sub-sampling the non respondents*, Paper presented at a special contributed session of the International Statistical Association Meetings, September, Tokyo, Japan, 2–16, 1987.

[10] Singh, H. P. and Kumar, S. *Estimation of mean in presence of non-response using two phase sampling scheme*, Statistical Papers, DOI 10.1007/s00362-008040-5, 2008.

[11] Singh, H. P. and Kumar, S. *A regression approach to the estimation of finite population mean in presence of non-response*, Aust. and New Zeal. J. Statist. **50** (4), 395–408, 2008.

[12] Sisodia, B. V. S. and Dwivedi, V. K. *A class of ratio cum product-type estimator in double sampling*, Biom. J. **24(4)**, 419-424, 1982.

[13] Tabasum, R. and Khan, I. A. *Double sampling for ratio estimation with non response*, J. Soc. Agril. Statist. **58** (3), 300–306, 2004.

[14] Tabasum, R. and Khan, I. A. *Double sampling ratio estimator for the population mean in presence of non-response*, Assam Statist. Review **20** (1), 73–83, 2006.

[15] Tripathi, T. P. and Khare, B. B. *Estimation of mean vector in presence of non response*, Commun. Statist.- Theory- Math. **26** (9), 2255–2269, 1997.