STATISTICS IN TRANSITION new series, Summer 2014
Vol. 15, No. 3, pp. 389–402

A CLASS OF TWO PHASE SAMPLING ESTIMATORS FOR RATIO OF TWO POPULATION MEANS USING MULTI-AUXILIARY CHARACTERS IN THE PRESENCE OF NON-RESPONSE

B. B. Khare[1], R. R. Sinha[2]

ABSTRACT

In this paper, a class of two phase sampling estimators for estimating the ratio of two population means using multi-auxiliary characters with unknown population means has been proposed in presence of non-response. The asymptotic bias, mean square error and minimum mean square error of the proposed class of estimators have been obtained. The optimum values of the sample at the first and the second phases along with the sub-sampling fraction of the non-responding group have been determined for the fixed cost and for the specified precision. The efficiency of the proposed class of estimators has also been shown through the theoretical and empirical studies.

Key words: two phase sampling, ratio of two means, bias, mean square error, auxiliary characters.

1. Introduction

The estimation of the ratio of two population means with known population mean of auxiliary character(s) has been discussed by Hartley and Ross (1954), Singh (1965), Tripathi (1970), Tripathi and Chaurvedi (1979) and Khare (1991). It has been well known that the ratio, product and regression types of estimators are used to increase the efficiency of the estimates when population mean of the auxiliary character is known in advance. But sometimes it has been observed in sample surveys that the population means of available auxiliary characters are not known in advance [sea Rao (1990)], in this condition it is customary to use two phase sampling for estimating the population means of the auxiliary characters. By introducing the two phase sampling scheme, Tripathi (1970), Singh (1982)

[1] Department of Statistics, Banaras Hindu University, Varanasi, India.
E-mail: bbkhare56@yahoo.com.
[2] Department of Mathematics, Dr. B. R. Ambedkar National Institute of Technology, Jalandhar, India. E-mail: raghawraman@gmail.com.

and Khare (1983, 91) have proposed the estimators for estimating the ratio of two population means $R = \bar{Y}_1/\bar{Y}_2$ using an auxiliary character with unknown population mean. Using two auxiliary characters with unknown population means, the estimators for estimating R have been proposed by Tripathi and Sinha (1976) and Srivasvata *et al.* (1988). Further Khare (1993) has proposed a class of estimators for R by using multi-auxiliary characters with unknown population means when the information is available on all selected units in the sample for main and auxiliary characters. But it has been observed in practice while conducting a sample survey related to human that we do not collect complete information for all the units selected in the sample due to the problem of non-response on study characters. Khare and Sinha (2002, 2004) have proposed classes of two phase sampling estimators for estimating the ratio of two population means using auxiliary character in presence of non-response while Khare and

Sinha (2012) have suggested the general classes of estimators using multi-auxiliary characters with subsampling the non-respondents.

In this paper, we have proposed a class of two phase sampling estimators for estimating the ratio of two population means $R = \bar{Y}_1/\bar{Y}_2$ of the study characters in presence of non-response using multi-auxiliary characters when their population means are not known. The expressions of bias, mean square error and minimum mean square error of the proposed class of estimators have been obtained. The optimum values of first phase sample, second phase sample and sub-sampling fraction of the non-responding group have been determined for the fixed cost and for the specified precision. The efficiency of the proposed class of estimators has also been shown through theoretical and empirical studies.

2. The proposed class of estimators

Consider a finite population which consists of N identifiable units $U_N = (u_1, u_2, \dots\dots\dots, u_N)$ in which (y_1, y_2) are the variables under study and $(x_1, x_2, \dots\dots, x_p)$ are the p auxiliary characters having population means $\bar{Y}_i$ $(i = 1, 2)$ of study characters and $\bar{X}_j$ $(j = 1, 2, \dots\dots, p)$ of auxiliary characters respectively. In many practical situations when the list of the sampling units is available but the population means of the auxiliary characters are not known then we use two phase sampling scheme to estimate the unknown population means of the auxiliary characters. In such situations, the estimate of population mean $\bar{X}_j$ $(j = 1, 2, \dots\dots, p)$ is furnished by taking a large first phase sample of size n' from the population of N units using simple random sampling without replacement (SRSWOR) method. Let the estimate of $\bar{X}_j$ $(j = 1, 2, \dots\dots, p)$ be the sample means $\bar{x}'_j$ $(j = 1, 2, \dots\dots, p)$ based on the information available on n' units. Again a second phase sample of size $n\,(< n')$ is drawn from the first phase selected units n' by SRSWOR method of sampling and collect the information on the study characters y_i $(i = 1, 2)$. We observe for the study characters y_i $(i = 1, 2)$ that only n_1 units are responding and $n_2 (= n - n_1)$ units are not responding in the sample of size n. In this case, it has been assumed that the whole population U_N is divided into two non-overlapping strata U_{N_1} and U_{N_2} of responding and non-responding soft-core groups; however they are not known in advance. The stratum weights of responding and non-responding groups are given by $P_1 = N_1/N$ and $P_2 = N_2/N$, and their estimates are respectively given by $\hat{P}_1 = p_1 = n_1/n$ and $\hat{P}_2 = p_2 = n_2/n$. Further, from the non-responding units n_2, we draw a subsample of size $r\,(= n_2\, k^{-1},\ k > 1)$ using SRSWOR technique of sampling and collect the information by the direct interview for yi $(i = 1, 2)$. Now using the approach of Hansen and Hurwitz (1946), the unbiased estimator for $\bar{Y}_i$ $(i = 1, 2)$ based on the information of $(n_1 + r)$ units is given by

$$\bar{y}_i^* = p_1 \bar{y}_{i1} + p_2 \bar{y}_{in_{(2r)}}, \quad i = 1, 2 \tag{2.1}$$

where $\bar{y}_{i1}$ and $\bar{y}_{in_{(2r)}}$ are the sample means of y_i based on n_1 and r units respectively.

The variance of the estimator $\bar{y}_i^*$ up to the terms of order (n^{-1}) is given by

$$V(\bar{y}_i^*) = V_i = \theta\, S_{y_i}^2 + \theta_k\, S_{y_{i(2)}}^2, \tag{2.2}$$

where $S_{y_i}^2$ and $S_{y_{i(2)}}^2$ denote the population mean square of y_i for the entire and non-responding part of the population, and $\theta = \frac{N-n}{Nn}$, $\theta_k = \frac{P_2(k-1)}{n}$.

If the ratio of two population means is $R = \bar{Y}_1/\bar{Y}_2$ and we have incomplete information on the study characters (y_1, y_2), then the usual estimator for estimating R may be given by

$$\hat{R} = \frac{\bar{y}_1^*}{\bar{y}_2^*} \quad . \tag{2.3}$$

The bias and mean square error of $\hat{R}$ under SRSWOR up to the terms of order (n^{-1}) are given by

$$B(\hat{R}) = R\{\theta\ \nabla_{21} + \theta_k \nabla'_{21}\}, \tag{2.4}$$

$$M(\hat{R}) = R^2\{\theta\, \Delta_{12} + \theta_k \Delta'_{12}\}, \tag{2.5}$$

where $\nabla_{21} = \frac{S_{y_2}^2}{\bar{Y}_2^2} - \rho \frac{S_{y_1}}{\bar{Y}_1} \frac{S_{y_2}}{\bar{Y}_2}$, $\nabla'_{21} = \frac{S_{y_{2(2)}}^2}{\bar{Y}_2^2} - \rho_2 \frac{S_{y_{1(2)}}}{\bar{Y}_1} \frac{S_{y_{2(2)}}}{\bar{Y}_2}$, $\Delta_{12} = \frac{S_{y_1}^2}{\bar{Y}_1^2} + \frac{S_{y_2}^2}{\bar{Y}_2^2} - 2\rho \frac{S_{y_1}}{\bar{Y}_1} \frac{S_{y_2}}{\bar{Y}_2}$, $\Delta'_{12} = \frac{S_{y_{1(2)}}^2}{\bar{Y}_1^2} + \frac{S_{y_{2(2)}}^2}{\bar{Y}_2^2} - 2\rho_2 \frac{S_{y_{1(2)}}}{\bar{Y}_1} \frac{S_{y_{2(2)}}}{\bar{Y}_2}$, ρ and ρ_2 are the correlation coefficients between (y_1, y_2) for the entire and non-responding group of the population respectively.

Hence, when we have incomplete information on the study characters y_1, y_2 but complete information on the auxiliary characters $x_1, x_2, \ldots, x_p$ for the sample of size n [See Rao (1986) p. 220], we propose a class of two phase sampling estimators for estimating the ratio of two population means $R(=\bar{Y}_1/\bar{Y}_2)$ of study characters using multi-auxiliary characters in presence of non-response on study characters only as

$$T = f\left(\bar{y}_1^*/\bar{y}_2^*,\ \underline{z}'\right) = f\left(m,\ \underline{z}'\right) \tag{2.6}$$

such that $f\left(R,\ \underline{e}'\right) = R$, and $f_{1(R,\ \underline{e}')} = \left(\frac{\partial}{\partial m} f\left(m,\ \underline{z}'\right)\right)_{(R,\ \underline{e}')} = 1,$ (2.7)

where $\underline{z}$ and $\underline{e}$ are the column vectors of $\left(z_1, z_2, \ldots \ldots, z_p\right)'$ and $(1, 1, \ldots \ldots, 1)'$ respectively and $z_j = \frac{\bar{x}_j}{\bar{x}_j'}$, $(j = 1, 2, \ldots, p)$. Here we assume that the function $f\left(m,\ \underline{z}'\right)$ is continuous and bounded in $(p+1)$ dimensional real space S^* containing the point $(R,\ \underline{e}')$ and the first and second order partial derivatives of $f\left(m,\ \underline{z}'\right)$ exist and are continuous and bounded in S^*.

3. Bias and mean square error (MSE)

Let the conventional estimator of R be $\hat{R}(=\bar{y}_1^*/\bar{y}_2^*)$. Since the number of possible samples is finite, so the bias and mean square error of the estimator T may be obtained. Now, expanding the function $f\left(m,\ \underline{z}'\right)$ about the point $\left(R, \underline{e}'\right)$ in a second order Taylor's series and using the condition (2.7), we have

$$T = R + D + \underline{D}' f_{2\,(R,\ \underline{e}')} + D\underline{D}' f_{12(m^*,\ \underline{z}^{*\prime})} + \frac{1}{2}\left[D^2 f_{11(m^*,\ \underline{z}^{*\prime})} + \underline{D}' f_{22(m^*,\ \underline{z}^{*\prime})} \underline{D}\right], \tag{3.1}$$

where $D = (m - R)$, $\underline{D}' = (\underline{z} - \underline{e})'$, $m^* = R + \emptyset(m - R)$, $\underline{z}^* = \underline{e} + \underline{\emptyset}\,(\underline{z} - \underline{e})$ such that $0 < \emptyset, \emptyset_j < 1$; $j = 1, 2, \ldots, p$ and $\underline{\emptyset}$ is a $p \times p$ diagonal matrix having j^{th} diagonal elements ϕ_j.

Here, $f_{1(m, \underline{z}')}$ and $f_{2(m,\ \underline{z}')}$ denote the first partial derivatives of $f\left(m,\ \underline{z}'\right)$ with respect to m and $\underline{z}'$ respectively. The second partial derivative of $f\left(m,\ \underline{z}'\right)$ with respect to $\underline{z}'$ is denoted by $f_{22(m,\ \underline{z}')}$ and the first partial derivative of $f_{2(m,\ \underline{z}')}$ with respect to m is denoted by $f_{12(m,\ \underline{z}')}$.

The expressions for bias and mean square error of T for any sampling design up to the terms of order n^{-1} $[O(n^{-1})]$ are given by

$$B(T) = B\left(\hat{R}\right) + E\left(D\underline{D}'\right) f_{12(m^*, \underline{z}^{*\prime})} + \frac{1}{2}\, E\left(\underline{D}' f_{22(m^*, \underline{z}^{*\prime})} \underline{D}\right) \tag{3.2}$$

and

$$M(T) = M\left(\hat{R}\right) + 2E\left(D\underline{D}'\right) f_{2(R, \underline{e}')} + E\left(f_{2(R, \underline{e}')}\right)' \underline{D}\,\underline{D}' f_{2(R,\ \underline{e}')}. \tag{3.3}$$

The mean square error of T is minimized for

$$f_{2(R,\ \underline{e}')} = -[E(\underline{D}\ \underline{D}')]^{-1}E(D\underline{D}) \tag{3.4}$$

and the resulting minimum mean square error of R up to the terms of $O(n^{-1})$ is given by

$$M(T)_{min.} = M(\hat{R}) + E(D\underline{D}')[E(\underline{D}\ \underline{D}')]^{-1}E(D\underline{D}). \tag{3.5}$$

To find the bias and mean square error of T under SRSWOR, we use the large sample approximation by assuming

$\bar{y}_i^* = \bar{Y}_i(1+\in_{0i})$, $\bar{x}_j = \bar{X}_j(1+\in_j')$, $\bar{x}_j' = \bar{X}_j(1+\in_j'')$ with $E(\in_{0i}) = E(\in_j') = E(\in_j'') = 0$ and $|\in_{0i}| < 1$, $|\in_j'| < 1$, $|\in_j''| < 1$ $\forall\, i = 1,2; j = 1,2,\dots\dots,p$.

We also assume that the contribution of the terms involving the powers in $\in_{0i}$, $\in_j'$ and $\in_j''$ of order higher than two in the bias and mean square error are assumed to be negligible.

Let $\rho_{jj'}$, ρ_{ij}^* be the correlation coefficients between $(x_j, x_{j'})$ and (y_i, x_j) respectively for the entire population and $\rho_{jj'(2)}$, $\rho_{ij(2)}^*$ be the correlation coefficients between $(x_j, x_{j'})$ and (y_i, x_j) for the non-responding group of the population.

So, the expressions of bias and mean square error of R in SRSWOR method of sampling up to the terms of $O(n^{-1})$ are given by

$$B(T) = B(\hat{R}) + R(\theta-\theta')\underline{\mathbb{B}}' f_{12(m^*, \underline{z}^{*\prime})} + \frac{(\theta-\theta')}{2}\ trace\ \underline{M}\, f_{22(m^*, \underline{z}^{*\prime})} \tag{3.6}$$

and

$$M(T) = M(\hat{R}) + (\theta-\theta')\left(f_{2(R,\ \underline{e}')}\right)'\underline{M}\, f_{2(R, \underline{e}')} + 2R\,(\theta-\theta')\,\underline{\mathbb{B}}' f_{2(m^*, \underline{z}^{*\prime})}, \tag{3.7}$$

where $\theta' = \frac{N-n'}{Nn'}$, $\underline{\mathbb{B}} = (\mathbb{B}_1, \mathbb{B}_2, \dots\dots, \mathbb{B}_p)'$is a column vector of order $(p \times 1)$ having the j^{th} element $\mathbb{B}_j = \frac{S_{x_j}}{\bar{X}_j}\left(\rho_{1j}^* \frac{S_{y_1}}{\bar{Y}_1} - \rho_{2j}^* \frac{S_{y_2}}{\bar{Y}_2}\right)$, $\underline{M} = [m_{jj'}]_{p\times p}$ is a $(p \times p)$ positive definite matrix having $m_{jj'} = \rho_{jj'}\frac{S_{x_j}}{\bar{X}_j}\frac{S_{x_{j'}}}{\bar{X}_{j'}}$; $\forall\, j \neq j' = 1,2,\dots\dots,p$ and $S_{x_j}^2$ denotes the mean square error of x_j for the entire part of the population.

Since the objective of this paper is to suggest a generalized class of estimators $T = f(m,\ \underline{z}')$ for estimating R and study its properties, so we may consider the following exponential, chain ratio and chain ratio cum regression types of estimators as members of T, which are as follows:

$$T_e = m\, e^{\sum_{j=1}^{p} a_j log z_j}, \tag{3.8}$$

$$T_r = m\left[\omega_1 z_1{}^{b_1/\omega_1} + \omega_2 z_2{}^{b_2/\omega_2} + \cdots \ldots \ldots + \omega_p z_p{}^{b_p/\omega_p}\right];\ \sum_{j=1}^{p} \omega_j = 1 \qquad (3.9)$$

and $$T_{crr} = \sum_{j=1}^{p}\{m + \varphi_j(z_j - 1)\}\left(\omega_j z_j{}^{c_j/\omega_j}\right), \qquad (3.10)$$

where a_j, b_j and c_j $(j = 1, 2, \ldots, p)$ are the scalar constants.

Now we state the following theorems:

Theorem 1. *Up to the terms of order $O(n^{-1})$ under SRSWOR, the mean square error of T is minimized for*

$$f_{2(R,\underline{e}')} = -R\underline{M}^{-1}\underline{\mathbb{B}} \qquad (3.11)$$

and minimum mean square error of T is given by

$$M(T)_{min.} = R^2\{(\theta\,\Delta_{12} + \theta_k \Delta'_{12}) - (\theta - \theta')(\underline{\mathbb{B}}'\underline{M}^{-1}\underline{\mathbb{B}})\}. \qquad (3.12)$$

Since the estimators T_e, T_r and T_{crr} are the members of T, so the values of the constants involved in them can be obtained by the condition (3.11) and their minimum mean square error will be equal to $M(T)_{min.}$. Sometimes this condition involves unknown parameters, so one may use the values of the parameters from past data or experience for obtaining the required value of the constants involved in (3.11). Reddy (1978) has shown that such values are stable not only over time but also over different regions. Srivastava and Jhajj (1983) have shown that the efficiency of such type of estimators does not decrease up to the terms of order $O(n^{-1})$ if we replace the optimum values of the constants by their estimates based on the sample values.

On comparing the proposed class of estimator T with $\hat{R}$ in terms of precision from (2.5) and (3.12), we have derived the following theorem:

Theorem 2. *Up to the terms of order $O(n^{-1})$,*

$$M(T) < M(\hat{R}) \text{ and } M(\hat{R}) - M(T) = R^2\{(\theta - \theta')(\underline{\mathbb{B}}'\underline{M}^{-1}\underline{\mathbb{B}})\} > 0.$$

Theorem 3. *Up to the terms of order $O(n^{-1})$,*

$$M(T) < M(\hat{R}) \text{ iff } -M(\hat{R}) < \left\{(\theta - \theta')\left(f_{2(R,\ \underline{e}')}\right)\right\}\left\{\left(f_{2(R,\ \underline{e}')}\right)'\underline{M} + 2R\underline{\mathbb{B}}'\right\} < 0.$$

If we compare the efficiency of proposed class of estimators (T) with the class of estimators suggested by Khare and Sinha (2012), we find that the Khare and Sinha (2012) estimator gives equal precision to T under the condition of known population mean of auxiliary characters.

It is also to be noted here that for $W_2 = 0$, i.e. when we have complete information on the study characters as well as on auxiliary characters for the sample of size n, then the proposed class of estimators T is equally efficient to the class of estimators for R as proposed by Khare (1993). Hence, it is clear that all the members of the proposed class of estimators T will attain minimum mean square error for one, two or p-auxiliary characters if the condition (3.11) is satisfied.

It is very important to know whether the reduction in variance would be worth the extra expenditure on the additional sample required to estimate the population mean of the auxiliary characters used in the case of two phase sampling. Hence, a rational approach is found by minimizing the mean square error of T for the fixed cost and obtaining the optimum values of n', n and k. Therefore, we determine the size of the first phase sample (n'), second phase sample (n) and the value of subsampling proportion (k^{-1}) which will minimize the mean square error of the proposed class of estimators T for the fixed cost $C \leq \boldsymbol{C_0}$.

4. Optimum sample size for the fixed cost C≤ C_0

The minimum value of the mean square error of T depends upon the values of n' n and k. Let the fixed total cost apart from overhead cost be $C \leq \boldsymbol{C_0}$. Let C_1' and C_1 are be the cost per unit of identifying and observing auxiliary characters and the cost per unit of mailing questionnaire/visiting the unit at the second phase respectively while C_2 and C_3 be the cost per unit of collecting/processing data for the study characters y_1, y_2 obtained from n_1 responding units and the cost per unit of obtaining and processing data for the study characters y_1, y_2 (after extra efforts) from the subsampled units. Now, the cost function under these assumptions is given by

$$C' = C_1'n' + C_1 n + C_2 n_1 + C_3 r. \tag{4.1}$$

Since C' will vary from sample to sample, so we consider the expected cost C to be incurred in the survey apart from overhead expenses, which is given by

$$C = E(C') = C_1'n' + n[C_1 + C_2P_1 + C_3P_2k^{-1}]. \tag{4.2}$$

Let $R^2\Psi_{0r}$, $R^2\Psi_{1r}$ and $R^2\Psi_{2r}$ be the coefficients of the terms n^{-1}, $(n')^{-1}$ and kn^{-1} respectively in the expressions of $M(T)$, then $M(T)$ can be expressed as

$$M(T) = (n^{-1})R^2\Psi_{0r} + (n')^{-1}R^2\Psi_{1r} + (kn^{-1})R^2\Psi_{2r} + I, \tag{4.3}$$

where I is the terms independent of n, n' and k in the expressions of $M(T)$.

Now, let us define a function φ for minimizing the $M(T)$ for the fixed cost $C \leq \boldsymbol{C_0}$ and to obtain the optimum sample sizes as

$$\varphi = M(T) + \lambda_r\{C_1'n' + n(C_1 + C_2P_1 + C_3P_2k^{-1}) - \boldsymbol{C_0}\}, \tag{4.4}$$

where λ_r is a Lagrange's multiplier.

Differentiating φ with respect to n', n and k and equating to zero, we have

$$n' = R\sqrt{\frac{\Psi_{1r}}{\lambda_r C_1'}}, \tag{4.5}$$

$$n = R\sqrt{\frac{\Psi_{0r}+k\Psi_{2r}}{\lambda_r(C_1+C_2P_1+C_3P_2k^{-1})}} \tag{4.6}$$

and
$$k_{opt.} = R\sqrt{\frac{C_3P_2\Psi_{0r}}{(C_1+C_2P_1)\,\Psi_{2r}}}\ . \tag{4.7}$$

Now, putting the values of $n'$and n from (4.5) and (4.6) and using the value of $k_{opt.}$ from (4.7) in (4.2), we have

$$\sqrt{\lambda_r} = \frac{R}{\boldsymbol{C_0}}\left[\sqrt{\Psi_{1r}C_1'} + \sqrt{\left(\Psi_{0r} + k_{opt.}\Psi_{2r}\right)\left(C_1 + C_2P_1 + C_3P_2k_{opt.}^{-1}\right)}\right]. \tag{4.8}$$

It has also been observed that the determinant of the matrix of the second order derivative of φ with respect to n', n and k is positive for the optimum values of n', n and k, which shows that the solutions for n', n given by (4.5), (4.6) and the optimum value of k under the condition $C \leq \boldsymbol{C_0}$ minimize the variance of T. It is also important to note here that the subsampling fraction $k_{opt.}^{-1}$ will decrease as $\sqrt{C_3/(C_1 + C_2P_1)}$ increases.

Hence, for the optimum values of n', n and k, the minimum value of $M(T)$ is given by

$$M(T)_{min.} = \boldsymbol{C_0}\lambda_r - R^2\Delta_{12}N^{-1}. \tag{4.9}$$

5. Determination of sample sizes for the specified variance $\boldsymbol{M_0}$

Let $\boldsymbol{M_0}$ be the variance of the estimator T fixed in advance and we have

$$\boldsymbol{M_0} = (n^{-1})R^2\Psi_{0r} + (n')^{-1}R^2\Psi_{1r} + (kn^{-1})R^2\Psi_{2r} + R^2\Delta_{12}N^{-1}. \tag{5.1}$$

For minimizing the average total cost C for the specified variance of the estimator T (i.e. $M(T) = \boldsymbol{M_0}$), we define a function φ^* which is given as

$$\varphi^* = C_1'n' + n(C_1 + C_2P_1 + C_3P_2k^{-1}) - \mu(M(T) - M_0) \tag{5.2}$$

where μ is a Lagrange's multiplier.

Now, for obtaining the optimum values of n', n and k, differentiating φ^* with respect to n', n and k and equating to zero, we have

$$n' = R\sqrt{\frac{\mu\Psi_{1r}}{C_1'}} \tag{5.3}$$

$$n = R\sqrt{\frac{\mu(\Psi_{0r}+k\Psi_{2r})}{(C_1+C_2P_1+C_3P_2k^{-1})}} \tag{5.4}$$

and $$k_{opt.} = \sqrt{\frac{C_3 P_2 \Psi_{0r}}{(C_1+C_2P_1)\,\Psi_{2r}}}. \tag{5.5}$$

Again by putting the values of n' and n from (5.3) and (5.4) and utilizing the optimum value of k in (5.1), we get

$$\sqrt{\mu} = \frac{\left[\sqrt{\Psi_{1r}C_1'}+\sqrt{(\Psi_{0r}+k_{opt.}\Psi_{2r})(C_1+C_2P_1+C_3P_2k_{opt.}^{-1})}\right]}{[\boldsymbol{M_0}+R^2\Delta_{12}N^{-1}]}. \tag{5.6}$$

The minimum expected total cost incurred in attaining the specified variance $\boldsymbol{M_0}$ by the estimator T is then given by

$$C(T)_{min.} = \frac{\left[\sqrt{C_1'V_{11}}+\sqrt{(V_{01}+k_{opt.}V_{21})\left(C_1+C_2W_1+C_3\frac{W_2}{k_{opt.}}\right)}\right]^2}{[\boldsymbol{M_0}+R^2\Delta_{12}N^{-1}]}. \tag{5.7}$$

6. An empirical study

109 Village/Town/ward population of urban area under Police-station – Baria, Tahasil – Champua, Orissa has been taken under consideration from District Census Handbook, 1981, Orissa, published by Govt. of India. The last 25% villages (i.e. 27 villages) have been considered as non-response group of the population. Here we have considered the study characters and auxiliary characters given as follows:

y_1: Number of literate persons in the village,

y_2: Number of main workers in the village,

x_1: Number of non-workers in the village,

x_2: Total population of the village and

x_3: Number of cultivators in the village.

The values of the parameters of the population under study are as follows:

$\bar{Y}_1 = 145.3028$	$\bar{Y}_2 = 165.2661$	$\bar{X}_1 = 259.0826$	$\bar{X}_2 = 485.9174$	$\bar{X}_3 = 100.5505$
$S_{y_1} = 111.3891$	$S_{y_2} = 112.8437$	$S_{x_1} = 198.0687$	$S_{x_2} = 320.2197$	$S_{x_3} = 73.5426$
$S^2_{y_{1(2)}} = 100.2444$	$S^2_{y_{2(2)}} = 95.3420$	$\rho^*_{11} = 0.905$	$\rho^*_{12} = 0.905$	$\rho^*_{13} = 0.648$
$\rho = 0.816$	$\rho_2 = 0.787$	$\rho^*_{21} = 0.819$	$\rho^*_{22} = 0.908$	$\rho^*_{23} = 0.841$
	$\rho_{12} = 0.946$	$\rho_{13} = 0.732$	$\rho_{23} = 0.801$	

Let the costs at the different processing stages be $C_1' =$ Rs. 0.15, $C_1 =$ Rs. 5.00, $C_2 =$ Rs. 25.00 and $C_3 =$ Rs. 65.00.

To show the efficiency of the proposed class of estimators T for the ratio of two population means [i.e. $R = \bar{Y}_1/\bar{Y}_2$] using the auxiliary characters x_1, x_2 and x_3, we have considered $T_e = m\, e^{\sum_{j=1}^{p} a_j log z_j}$ as a member of the proposed class of estimators T.

The optimum values of the constants a_j, mean square error and the percentage relative efficiency (PRE) of T_e with respect to $\hat{R}$ for fixed sample sizes $n' = 80$, $n = 20$ and for the fixed cost $\boldsymbol{C_0} =$ Rs. 280 are shown in Table 1. The expected cost of $\hat{R}$ and T_e in case of specified precision $\boldsymbol{M_0} = 1250 \times 10^{-5}$ are also given in Table 1.

7. Conclusions

From Table 1 – see Appendix 2, it has been observed that the estimator T_e is more efficient than $\hat{R}$ for all the different values of the sub-sampling fraction k^{-1} and its efficiency increases as the value of sub-sampling fraction increases. The mean square error of the estimator T_e decreases while the relative efficiency of the estimator T_e with respect to $\hat{R}$ increases with the increase in the numbers of auxiliary characters used. Regarding the performance of the estimator T_e over $\hat{R}$ in case of fixed cost, we observe that the relative efficiency of T_e increases as the number of the auxiliary characters increases. We also observe that the values of $k_{opt.}$ and $n_{opt.}$ decrease while the value of $n'_{opt.}$ increases with the increase in the numbers of auxiliary characters used. Further, in case of specified variance, the expected cost incurred by T_e decreases with the increases in the numbers of auxiliary characters used. It has been also observed that $n'_{opt.}$increases while $n_{opt.}$ decreases by increasing the numbers of the auxiliary characters. Hence, on the basis of theoretical and empirical studies, we may recommend the proposed class of estimators T for the use in practice under its respective circumstances as discussed in the text.

Acknowledgements

Authors are grateful to the referee and the editor for their invaluable suggestions which helped in further improvement in the paper.

REFERENCES

HANSEN, M. H., HURWITZ, W. N., (1946). The problem of nonresponse in sample surveys, Jour. Amer. Statist. Assoc., 41, 517–529.

HARTLEY, H. O., ROSS, A., (1954). Unbiased ratio estimators, Nature, 174, 270–271.

KHARE, B. B., (1983). Some problems of estimation using auxiliary character, Ph.D. Thesis submitted to B.H.U., Varanasi, India.

KHARE, B. B., (1991). Determination of sample sizes for a class of two phase sampling estimators for ratio and product of two population means using auxiliary character, Metron, 49(1-4), 185–197.

KHARE, B. B., (1993). On a class of two phase sampling estimators for ratio of two population means using multi-auxiliary characters, Proc. Nat. Acad. Sci. India, 63(A) III, 513–519.

KHARE, B. B., SINHA, R. R., (2002). Estimation of the ratio of two population means using auxiliary character with unknown population mean in presence of non-response, Progress of Mathematic, 36, 337–348.

KHARE, B. B., SINHA, R. R., (2004). Estimation of finite population ratio using two phase sampling scheme in the presence of non-response, Aligarh J. Stat, 2004, 24, 43–56.

KHARE, B. B., SINHA, R. R., (2012). Improved classes of estimators for ratio of two means with double sampling the non respondents, Statistika, 49(3), 75–83.

REDDY, V. N., (1978). A study of use of prior knowledge on certain population parameters in estimation, Sankhaya, C, 40, 29–37.

RAO, P. S. R. S., (1986). Ratio estimation with subsampling the nonrespondents, Survey Methodology, 12(2), 217–230.

RAO, P. S. R. S., (1990). Regression estimators with subsampling of nonrespondents, In-Data Quality Control, Theory and Pragmatics, (Eds.) Gunar E. Liepins and V.R.R. Uppuluri, Marcel Dekker, New York, (1990), 191–208.

SINGH, M. P., (1965). On the estimation of ratio and product of population parameters, Sankhya, Ser.C, 27, 321–328.

SINGH, R. K., (1982). Generalized double sampling estimators for the ratio and product of population parameters, Jour. Ind. Statist. Assoc., 20, 39–49.

SRIVASTAVA, S. R., KHARE, B. B., SRIVASTAVA, S. R., (1988). On generalised chain estimator for ratio and product of two population means using auxiliary characters, Assam. Stat. Rev., 1, 21–29.

SRIVASTAVA, S. K., JHAJJ, H. S., (1983). A class of estimators of the population mean using multi-auxiliary information, Cal. Stat. Assoc. Bull., 32, 47–56.

TRIPATHI, T. P., (1970). Contribution to the sampling theory using multivariate information, Ph.D. Thesis submitted to Punjabi University, Patiyala, India.

TRIPATHI, T. P., CHATURVEDI, D. K., (1979). Use of multivariate auxiliary information in estimating the population ratio, Stat-Math. Tech. Report No. 24/79, I.S.I., Calcutta, India, Abs. J. Indian Soc. Agricultural Statist., bf 31.

TRIPATHI, T. P., SINHA, S. K. P., (1976). Estimation of ratio on successive occasions, Proceedings of Symposium on Recent Developments in Survey Methodology, held at I.S.I., Calcutta, March 1976.

APPENDIX 1

Expand the function $f(m,\ \underline{z}')$ given in (2.6) about the point $(R,\ \underline{e}')$ using Taylor's series up to the second order partial derivatives, we have

$$T = f(R, \underline{e}') + Df_{1\,(R,\ \underline{e}')} + \underline{D}'f_{2\,(R,\ \underline{e}')}$$
$$+\frac{1}{2}\left[D^2 f_{11(m^*,\ \underline{z}^{*\prime})} + 2D\underline{D}'f_{12(m^*,\ \underline{z}^{*\prime})} + \underline{D}'f_{22(m^*,\ \underline{z}^{*\prime})}\underline{D}\right]$$

Using condition (2.7), we get

$$T = R + D + \underline{D}'f_{2\,(R,\ \underline{e}')} + 2D\underline{D}'f_{12(m^*,\ \underline{z}^{*\prime})} + \frac{1}{2}\left[D^2 f_{11(m^*,\ \underline{z}^{*\prime})} + \underline{D}'f_{22(m^*,\ \underline{z}^{*\prime})}\underline{D}\right]$$

Now,

$$B(T) = E(T-R)$$
$$= B(\hat{R}) + E(D\underline{D}')f_{12(m^*,\underline{z}^{*\prime})} + \frac{1}{2}\,E(\underline{D}'f_{22(m^*,\underline{z}^{*\prime})}\underline{D})$$
$$M(T) = E(T-R)^2$$
$$= M(\hat{R}) + 2E(D\underline{D}')f_{2(R,\underline{e}')} + E(f_{2(R,\underline{e}')})'\underline{D}\,\underline{D}'f_{2(R,\ \underline{e}')}$$

Differentiating $M(T)$ with respect to $f_{2(R,\underline{e}')}$ and equating it to zero, we have

$$f_{2(R,\ \underline{e}')} = -\,[E(\underline{D}\,\underline{D}')]^{-1}E(D\underline{D})$$

Putting this $f_{2(R,\ \underline{e}')}$ in $M(T)$, we get

$$M(T)_{min.} = M(\hat{R}) + E(D\underline{D}')[E(\underline{D}\,\underline{D}')]^{-1}E(D\underline{D}).$$

Under simple random sampling without replacement (SRSWOR), we have obtained

$$E(\underline{D}\,\underline{D}') = E\left[(\underline{z}-\underline{e})(\underline{z}-\underline{e})'\right]$$

Consider

$$E[(z_1-1)(z_2-1)] = E\left[\left(\frac{\bar{x}_1}{\bar{x}_1'}-1\right)\left(\frac{\bar{x}_2}{\bar{x}_2'}-1\right)\right]$$

$$= E\left[\left(\frac{\bar{X}_1(1+\epsilon_1')}{\bar{X}_1(1+\epsilon_1'')}-1\right)\left(\frac{\bar{X}_2(1+\epsilon_2')}{\bar{X}_2(1+\epsilon_2'')}-1\right)\right]$$

$$= E\left[\left\{(1+\epsilon_1')(1+\epsilon_1'')^{-1}-1\right\}\left\{(1+\epsilon_2')(1+\epsilon_2'')^{-1}-1\right\}\right]$$

Neglecting the terms involving powers in $\epsilon_j', \epsilon_j''; j=1,2$ of order higher than two, we have

$$E[(z_1-1)(z_2-1)] = E\left[\epsilon_1''\epsilon_2'' - \epsilon_1''\epsilon_2' - \epsilon_1'\epsilon_2'' + \epsilon_1'\epsilon_2'\right]$$

Since $E[\epsilon_1''\epsilon_2''] = E[\epsilon_1''\epsilon_2'] = (\theta-\theta')\rho_{12}\frac{S_{x_1}}{\bar{X}_1}\frac{S_{x_2}}{\bar{X}_2}$

Therefore, $E[(z_1-1)(z_2-1)] = E[\epsilon_1'\epsilon_2'] - E[\epsilon_1'\epsilon_2'']$

$$= \theta\rho_{12}\frac{S_{x_1}}{\bar{X}_1}\frac{S_{x_2}}{\bar{X}_2} - \theta'\rho_{12}\frac{S_{x_1}}{\bar{X}_1}\frac{S_{x_2}}{\bar{X}_2}$$

$$= (\theta-\theta')\rho_{12}\frac{S_{x_1}}{\bar{X}_1}\frac{S_{x_2}}{\bar{X}_2} = (\theta-\theta')m_{12}$$

Similarly, we can define $m_{jj'} = \rho_{jj'}\frac{S_{x_j}}{\bar{X}_j}\frac{S_{x_{j'}}}{\bar{X}_{j'}}$.

Now, $E(D\underline{D}') = E(m-R)(\underline{z}-\underline{e})$

Consider

$$E(m-R)(z_1-1) = E\left[\left\{\frac{\bar{y}_1^*}{\bar{y}_2^*}-R\right\}\left\{\frac{\bar{x}_1}{\bar{x}_1'}-1\right\}\right]$$

$$= E\left[\left\{\frac{\bar{Y}_1(1+\epsilon_{01})}{\bar{Y}_2(1+\epsilon_{02})}-1\right\}\left\{\frac{\bar{X}_1(1+\epsilon_1')}{\bar{X}_1(1+\epsilon_1'')}-1\right\}\right]$$

$$= R\,E\left[\{(1+\epsilon_{01})(1+\epsilon_{02})^{-1}-1\}\left\{(1+\epsilon_1')(1+\epsilon_1'')^{-1}-1\right\}\right]$$

Neglecting the terms involving powers in $\epsilon_{01}, \epsilon_{02}, \epsilon_1'$ and ϵ_1'' of order higher than two, we have

$$E(m-R)(z_1-1) = R[\{E(\epsilon_{01}\epsilon_1') - E(\epsilon_{01}\epsilon_1'')\} - \{E(\epsilon_{02}\epsilon_1') - E(\epsilon_{02}\epsilon_1'')\}]$$

$$= R\left[\left\{\theta\rho_{11}^*\frac{S_{y_1}}{\bar{Y}_1}\frac{S_{x_1}}{\bar{X}_1} - \theta'\rho_{11}^*\frac{S_{y_1}}{\bar{Y}_1}\frac{S_{x_1}}{\bar{X}_1}\right\} - \left\{\theta\rho_{21}^*\frac{S_{y_2}}{\bar{Y}_2}\frac{S_{x_1}}{\bar{X}_1} - \theta'\rho_{21}^*\frac{S_{y_2}}{\bar{Y}_2}\frac{S_{x_1}}{\bar{X}_1}\right\}\right]$$

$$= R\left[(\theta-\theta')\rho_{11}^*\frac{S_{y_1}}{\bar{Y}_1}\frac{S_{x_1}}{\bar{X}_1} - (\theta-\theta')\rho_{21}^*\frac{S_{y_2}}{\bar{Y}_2}\frac{S_{x_1}}{\bar{X}_1}\right]$$

$$= R(\theta-\theta')\left[\frac{S_{x_1}}{\bar{X}_1}\left(\rho_{11}^*\frac{S_{y_1}}{\bar{Y}_1} - \rho_{21}^*\frac{S_{y_2}}{\bar{Y}_2}\right)\right]$$

$$= R(\theta-\theta')B_1$$

Similarly, we can define $B_j = \frac{S_{x_j}}{\bar{X}_j}\left(\rho_{1j}^*\frac{S_{y_1}}{\bar{Y}_1} - \rho_{2j}^*\frac{S_{y_2}}{\bar{Y}_2}\right)$

The expressions given in theorems can be obtained from (3.4) and (3.5).

APPENDIX 2

Table 1. Mean square error (MSE) and the percentage relative efficiency (PRE) of T_e with respect to $\hat{R}$ for fixed sample sizes, cost and variance

Estimators	Auxiliary character(s)	MSE in 10^{-5} and P.R.E. of T_e with respect to $\hat{R}$ for fixed $n' = 80$ and $n = 20$; k^{-1}: 4^{-1}	3^{-1}	2^{-1}	P.R.E. with respect to $\hat{R}$ for the fixed cost **C_0** = Rs. 280.00: $k_{opt.}$	$n'_{opt.}$ (approx.)	$n_{opt.}$ (approx.)	MSE(.) in 10^{-5}	P.R.E.(.) in %	Expected cost of the estimators $\hat{R}$ and T_e for the specified variance $\boldsymbol{M_0}$ = 1250 × 10^{-5}: $n'_{opt.}$ (approx.)	$n_{opt.}$ (approx.)	E.C. (.) in Rs.
$\hat{R}$	–	100.00 (1153)*	100.00 (979)	100.00 (804)	1.5199	-	8	1977	100.00	-	12	426.19
T_e	x_1; $a_1 = -0.1759$	104.72 (1101)	105.61 (927)	106.92 (752)	1.4280	36	8	1881	105.10	52	11	406.79
T_e	x_1, x_2; $a_1 = -0.8046$, $a_2 = 0.7711$	112.82 (1022)	115.45 (848)	119.47 (673)	1.2782	60	7	1650	119.82	78	10	360.47
T_e	x_1, x_2, x_3; $a_1 = -0.6947$, $a_2 = -0.8046$, $a_3 = 0.7711$	127.83 (902)	134.48 (728)	145.39 (553)	1.0072	95	7	1240	159.44	94	7	278.02

*Figures in parenthesis give the MSE(·).