RT&A, No.3 (74)
Volume 18, September 2023

TWO-CLASSES FOR REGRESSION TYPE OF ESTIMATORS FOR THE RATIO OF TWO POPULATION MEANS IN TWO-PHASE SAMPLING IN THE PRESENCE OF NON-RESPONSE FOR STRATIFIED POPULATION

Manish Mishra, B. B. Khare and Sachin Singh

•

Department of Statistics, Banaras Hindu University, Varanasi-221005, India.
manish.mishra10@bhu.ac.in, bbkhare56@yahoo.com and singhat619@gmail.com

Abstract

Utilizing the auxiliary information in stratified population, in the current study, we have discussed two classes for the regression type of estimators to estimate the ratio of two population means in the presence of non-response with the unknown population mean of the auxiliary variable. To estimate the unknown value of the population mean of auxiliary variable, we have used two-phase sampling method. For the suggested classes of estimators, we have considered two situations for the use of auxiliary information along with the non-response in the study variable such as incomplete information on the study variable and incomplete information on the corresponding units of the auxiliary variable and in another situation we have considered incomplete information on the study variable and complete information on the auxiliary variable. To estimate the non-response in study variable and auxiliary variables, we have used the Hansen and Hurwitz method of sub-sampling from the non-respondents. For the suggested classes of estimators, some members have been recognized. Using large sample approximation, the expressions for bias and mean square error have been derived for the suggested classes. The optimum values of the constants involving in the expression of mean square error have also been calculated. Mean square errors of the Suggested classes are found to be equal in theoretical study and real data study. An empirical study has been conducted with the help of a real data set (The Primary Census Abstract-2011 published by the Office of the Registrar General & Census Commissioner, India.) in order to compare the proposed classes of estimators with the conventional estimator for the different rates of non-response and different choices of sub-sampling fraction. The Suggested classes are found to be most efficient with respect to the conventional estimator for the different rates of non-response and different choices of sub-sampling fraction in empirical study.

Keywords: Ratio of two population means, Regression type estimator, Two-phase sampling, Auxiliary variable, Non-response, Mean square error

1. Introduction

In the literature of sample surveys, the ratio of two population means plays a crucial role. In this context to have a better understanding, we have several examples in the field of scientific and Socio-economic studies such as:

- **Agricultural surveys**: The crop production per acre in a crop survey, agriculture labor per cultivator and the ratio of production of corn acres to wheat acres, etc.
- **Industrial surveys**: the outlay of total expenses per employee, Proportion of liquid to total asset, etc.

- **Medical surveys**: In the estimation of growth index, the estimate of ratio using the measurements on weight to height. The skull or chest circumference may be used as an auxiliary variable.

In the case of finite population, the estimate for the ratio of two population means by using known and unknown population mean of the auxiliary variable have been studied by several authors such as Singh [13], Tripathi [17], Khare [3], Upadhyaya et al. [18], Singh and Naqvi [14]and Kumar and Srivastava [10]. In a recent study, Ahuja et al.[16] have suggested a generalized two-phase sampling estimator for ratio of two population means.

The occurrence of non-response is very common in the field of sample surveys. Hansen and Hurwitz [2] have suggested a technique of sub-sampling from the non-respondents to treat the problem of non-response. Further, El-Badry [1] has made some improvements to reduce the effect of non-response. In case of finite population in the presence of non-response, the estimation of ratio of two population means using known and unknown population mean of auxiliary variable have been studied by Khare and Pandey [4], Khare and Sinha [[5],[6]], Khare and Sinha [7] and Khare et al. [8].

For the stratified population in the presence of non-response, Khare and Jha [9] and Singh et al. [12] have suggested the classes of estimators for estimating the population mean utilizing auxiliary information with known and unknown population mean of auxiliary variable.

Following the research work of Singh et al. [12], we have made an effort by suggesting two classes of regression type of estimators for the ratio of two population means utilizing two-phase sampling method for the estimation of unknown population mean of auxiliary variable in the presence of non-response for the stratified population. Some members of the suggested classes of estimators have been recognized. The properties of the suggested classes have been obtained. To support the effectiveness of the proposed classes of estimators with respect to the relevant estimator, an empirical study is conducted with the help of a real data set.

2. Notations and Sampling Procedure

We have a heterogeneous population of size η : $[\eta_1, \eta_2, \eta_3 ... \eta_N]$ study variables (y_1, y_2) and auxiliary variable x with respective population means $\overline{Y}_1, \overline{Y}_2$ and population mean $(\overline{X})$ of auxiliary variable is unknown, which is divided into L homogeneous strata. The population parameters used in this study are denoted as follows:

we have,

$$\overline{Y}_1 = \frac{1}{N}\sum_{i=1}^{N} y_{1i}, \overline{Y}_2 = \frac{1}{N}\sum_{i=1}^{N} y_{2i}, \overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i \text{ and } R = \frac{\bar{Y}_1}{\bar{Y}_2} \tag{1}$$

In this present study, we are dealing with the stratified population with unknown $\overline{X}$ in the presence of non-response. In such a situation, to estimate the unknown population mean of an auxiliary variable, we use the technique of two-phase sampling which is described as follows:

In the very first step, we have the population of size N divided in L homogeneous strata of sizes $N_1, N_2, N_3 ... N_L$. In the first-phase, using simple random sampling without replacement (SRSWOR), we draw a larger preliminary sample of size $n_i^{'}$ from i^{th} stratum of size N_i.

Further, in the second-phase, we draw a relatively small sample of size $n_i^{'}(n_i < n_i^{'})$ using SRSWOR from $n_i^{'}$ units of i^{th} stratum.

For the study (y_1, y_2)variables and auxiliary variable (x), a sample of size n_i, we draw from i^{th} stratum of size N_i. Due to the non-response in the population, we observe that in the sample of size n_i there are n_{i1} responding and n_{i2} non-responding units such that $[n_i = n_{i1} + n_{i2}]$.

In the next step, to get the estimate of these n_{i2}non-responding units, we draw a sub-sample of size $r_i[= \frac{n_{i2}}{l_i}, l_i > 1]$.

Hence, we have responding n_{i1} and r_i sub-sampled units for the i^{th} stratum which we use in defining the Hansen and Hurwitz [2] estimators. The population means of the i^{th} stratum for

(y_1, y_2) and x are given by $\overline{Y}_{1i}$,$\overline{Y}_{2i}$ and $\overline{X}_i$ and their estimators in the presence of non-response using Hansen and Hurwitz [2] sub-sampling method from non-respondents are given as follows:

$$\overline{y}^*_{1i} = \frac{n_{i1}}{n_i}\overline{y}_{1i(1)} + \frac{n_{i2}}{n_i}\overline{y}'_{1i(2)}, \overline{y}^*_{2i} = \frac{n_{i1}}{n_i}\overline{y}_{2i(1)} + \frac{n_{i2}}{n_i}\overline{y}'_{2i(2)} \text{ and } \overline{x}^*_i = \frac{n_{i1}}{n_i}\overline{x}_{i(1)} + \frac{n_{i2}}{n_i}\overline{x}'_{i(2)}. \quad (2)$$

where, $\overline{y}_{1i(1)}$, $\overline{y}_{2i(1)}$ and $\overline{x}_{i(1)}$ are the sample means for the variable (y_1, y_2) and x for the n_{i1} units in i^{th} stratum. The sample means based on r_i units sub-sampled from n_{i2} units in the i^{th} stratum are denoted by $\overline{y}'_{1i(2)}$, $\overline{y}'_{2i(2)}$ and $\overline{x}'_{i(2)}$ for study variables (y_1, y_2) and the auxiliary variable x.

In the presence of non-response, the stratified sample means for $\overline{Y}_1$,$\overline{Y}_2$ and $\overline{X}$ for the i^{th} stratum are given as follows:

$$\overline{y}^*_{1st} = \sum_{i=1}^{L} W_i\overline{y}^*_{1i}, \overline{y}^*_{2st} = \sum_{i=1}^{L} W_i\overline{y}^*_{2i} \text{ and } \overline{x}^*_{st} = \sum_{i=1}^{L} W_i\overline{x}^*_i. \quad (3)$$

For the auxiliary variable x,the stratified sample mean for estimating $\overline{X}$ based on first-phase sample of size n'_i is given as follows:

$$\overline{x}'_{st} = \sum_{i=1}^{L} W_i\overline{x}'_i. \quad (4)$$

where,$\overline{x}'_i$ is the sample mean based on first-phase sample of size n'_i,$(n'_i > n_i)$ drawn from N_i units of the i^{th} stratum is given as follows:

$$\overline{x}'_i = \frac{1}{n'_i}\sum_{i=1}^{n'_i} x_{ij} \quad (5)$$

.

The stratified sample mean based on n_i units in i^{th} stratum for auxiliary variable x is given as follows:

$$\overline{x}_{st} = \sum_{i=1}^{L} W_i\overline{x}_i. \quad (6)$$

.

The population variance and co-variance for i^{th} stratum and non-responding part of the i^{th} stratum used in this study are given as follows:

$$S^2_{y1i} = \frac{1}{(N_i - 1)}\sum_{j=1}^{N_i}(Y_{1i,j} - \overline{Y}_{1i})^2,$$

$$S^2_{y2i} = \frac{1}{(N_i - 1)}\sum_{j=1}^{N_i}(Y_{2i,j} - \overline{Y}_{2i})^2,$$

$$S^2_{xi} = \frac{1}{(N_i - 1)}\sum_{j=1}^{N_i}(X_{ij} - \overline{X}_{1i})^2,$$

$$S^2_{y1i(2)} = \frac{1}{(N_{i2} - 1)}\sum_{j=1}^{N_{i2}}(Y_{1ij(2)} - \overline{Y}_{1i(2)})^2,$$

$$S^2_{y2i(2)} = \frac{1}{(N_{i2} - 1)}\sum_{j=1}^{N_{i2}}(Y_{2ij(2)} - \overline{Y}_{2i(2)})^2,$$

$$S^2_{xi(2)} = \frac{1}{(N_{i2} - 1)}\sum_{j=1}^{N_{i2}}(X_{ij(2)} - \overline{X}_{i(2)})^2,$$

$$
\begin{aligned}
S^2_{y1iy2i} &= \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{1i,j}-\overline{Y}_{1i})(Y_{2i,j}-\overline{Y}_{2i}),\\
S^2_{y1ixi} &= \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{1i,j}-\overline{Y}_{1i})(X_{i,j}-\overline{X}_{i}),\\
S^2_{y2ixi} &= \frac{1}{(N_i-1)}\sum_{j=1}^{N_i}(Y_{2i,j}-\overline{Y}_{2i})(X_{i,j}-\overline{X}_{i}),\\
S^2_{y1iy2i(2)} &= \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{1i,j(2)}-\overline{Y}_{1i(2)})(Y_{2i,j(2)}-\overline{Y}_{2i(2)}),\\
S^2_{y1ixi(2)} &= \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{1i,j(2)}-\overline{Y}_{1i(2)})(X_{i,j(2)}-\overline{X}_{i(2)}),\\
S^2_{y2ixi(2)} &= \frac{1}{(N_{i2}-1)}\sum_{j=1}^{N_{i2}}(Y_{2i,j(2)}-\overline{Y}_{2i(2)})(X_{i,j(2)}-\overline{X}_{i(2)}).
\end{aligned}
\tag{7}
$$

where, $y_{1i,j}$: j^{th} value of y_1 in the i^{th} stratum, $y_{2i,j}$: j^{th} value of y_2 in the i^{th} stratum, $x_{i,j}$: j^{th} value of x in the i^{th} stratum, $y_{1i,j(2)}$: j^{th} value of y_1 for non-responding units in the i^{th} stratum, $y_{2i,j(2)}$: j^{th} value of y_2 for non-responding units in the i^{th} stratum and $x_{i,j(2)}$: j^{th} value of x for non-responding units in the i^{th} stratum.

We denote,$W_{i1} = \frac{N_{i1}}{N_i}$,$W_{i2} = \frac{N_{i2}}{N_i}$ such that $[N_i = N_{i1} + N_{i2}]$ $\forall$, $i = 1, 2, 3...L$. Where, N_{i1} and N_{i2} are the size of the responding and non-responding units in i^{th} stratum.

3. Proposed Classes of Regression Type of Estimators for R in Two-phase Sampling

In the case of unknown population mean of the auxiliary variable, we are considering two different situations of non-response i.e. incomplete information on y_1, y_2 and corresponding information on the auxiliary variable x and also we use complete information on auxiliary variable x. Here, we propose two classes of two-phase sampling regression type of estimators which are given as follows:

$$\hat{R}_{1st} = [\hat{R}_{st} + \alpha_1 \overline{x}'_{st}(u_2 - 1)]\phi_{(1)}(u_1) \tag{8}$$

$$\hat{R}_{2st} = [\hat{R}_{st} + \alpha_2 \overline{x}'_{st}(u_1 - 1)]\phi_{(1)}(u_2) \tag{9}$$

where, $\hat{R}_{st} = \frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}$, $u_1 = \frac{\overline{x}^*_{st}}{\overline{x}'_{st}}$ and $u_2 = \frac{\overline{x}_{st}}{\overline{x}'_{st}}$.

Such that,

$$\phi_1(1) = 1, \phi_1(2) = 1, \phi_{1(1)}(1) = \left(\frac{\delta}{\delta u_1}\phi_1(1)\right)_1 \text{ and } \phi_{1(2)}(1) = \left(\frac{\delta}{\delta u_2}\phi_2(1)\right)_1 \tag{10}$$

$\phi_{(1)}(u_1)$ and $\phi_{(2)}(u_2)$ are the function of u_1 and u_2 satisfy the regularity conditions given as follows:

- The functions of u_1 and u_1 assume values in a bounded closed convex subset U^* of the two-dimensional real line containing the point (1).
- In U^*, the function of u_1 and u_2 are continuous and bounded.
- The first-order and second-order partial derivatives of the given function of u_1 and u_1 exist and are continuous and bounded U^*.

The first and second-order partial derivatives of the function $\phi_{(1)}(u_1)$ w.r.to u_1 and $\phi_{(2)}(u_2)$ w.r.to u_2 are denoted by $[\phi_{1(1)}(u_1), \phi_{11(1)}(u_1)]$ and $[\phi_{1(2)}(u_2), \phi_{11(2)}(u_2)]$ respectively.

Now, using Taylor's series expansion we expand the function $\phi_{(1)}(u_1)$ and $\phi_{(2)}(u_2)$ upto the second-order partial derivative about the point (1), we have

$$\hat{R}_{1st} = [\hat{R}_{st} + \alpha_1 \overline{x}'_{st}(u_2 - 1)][\phi_{(1)}(1) + (u_1 - 1)\phi_{1(1)}(1) + \frac{1}{2}(u_1 - 1)^2 \phi_{11(1)}(1)] \tag{11}$$

$$\hat{R}_{2st} = [\hat{R}_{st} + \alpha_2 \overline{x}'_{st}(u_1 - 1)][\phi_{(2)}(1) + (u_2 - 1)\phi_{1(2)}(1) + \frac{1}{2}(u_2 - 1)^2 \phi_{11(2)}(1)] \tag{12}$$

Using the condition given in equation (10) and regularity conditions, the expression (11) and (12) can be written as:

$$\begin{aligned} \hat{R}_{1st} &= [\hat{R}_{st} + \alpha_1 \overline{x}'_{st}(u_2 - 1)][1 + (u_1 - 1)\phi_{1(1)}(1) + \frac{1}{2}(u_1 - 1)^2 \phi_{11(1)}(1)] \\ &= \hat{R}_{st} + \hat{R}_{st}(u_1 - 1)\phi_{1(1)}(1) + \hat{R}_{st}\frac{1}{2}(u_1 - 1)^2 \phi_{11(1)}(1) + \alpha_1 \overline{x}'_{st}(u_2 - 1) \\ &+ \alpha_1 \overline{x}'_{st}(u_1 - 1)(u_2 - 1)\phi_{1(1)}(1) + \frac{1}{2}\alpha_1 \overline{x}'_{st}(u_1 - 1)^2 (u_2 - 1)\phi_{11(1)}(1) \end{aligned} \tag{13}$$

$$\begin{aligned} \hat{R}_{2st} &= [\hat{R}_{st} + \alpha_2 \overline{x}'_{st}(u_1 - 1)][1 + (u_2 - 1)\phi_{1(2)}(1) + \frac{1}{2}(u_2 - 1)^2 \phi_{11(1)}(2)] \\ &= \hat{R}_{st} + \hat{R}_{st}(u_2 - 1)\phi_{1(2)}(1) + \hat{R}_{st}\frac{1}{2}(u_2 - 1)^2 \phi_{11(2)}(1) + \alpha_2 \overline{x}'_{st}(u_1 - 1) \\ &+ \alpha_2 \overline{x}'_{st}(u_1 - 1)(u_2 - 1)\phi_{1(2)}(1) + \frac{1}{2}\alpha_2 \overline{x}'_{st}(u_2 - 1)^2 (u_1 - 1)\phi_{11(2)}(1) \end{aligned} \tag{14}$$

4. Properties of the proposed classes of regression type of estimators

To obtain the expression for bias and MSE of the suggested classes of estimators, we define:

$$\begin{aligned} \overline{y}^*_{1st} &= \overline{Y}_1(1+\epsilon_0) \quad , \overline{y}^*_{2st} = \overline{Y}_2(1+\epsilon_1),\ \overline{x}_{st} = \overline{X}(1+\epsilon_2),\ \overline{x}'_{st} = \overline{X}(1+\epsilon_3) \\ \overline{x}^*_{st} &= \overline{X}_1(1+\epsilon_4) \end{aligned} \tag{15}$$

such that, $|\epsilon_i| < 1$ and $E(\epsilon_i) = 0\ \forall\ i = 1, 2, 3, 4$.

Here, we ignore the finite population correction term because the population size is large enough under consideration. i.e. by using large sample approximation, we have

$$\begin{aligned} E(\epsilon_0^2) &= \frac{V(\overline{y}^*_{1st})}{\overline{Y}_1^2} = \frac{1}{\overline{Y}_1^2}\sum_{i=1}^{L}[W_i^2 \lambda_i S^2_{y_{1i}} + \frac{(l_i - 1)}{n_i} W_{i2} S^2_{y_{1i}(2)}], \\ E(\epsilon_1^2) &= \frac{V(\overline{y}^*_{2st})}{\overline{Y}_2^2} = \frac{1}{\overline{Y}_2^2}\sum_{i=1}^{L}[W_i^2 \lambda_i S^2_{y_{2i}} + \frac{(l_i - 1)}{n_i} W_{i2} S^2_{y_{2i}(2)}], \\ E(\epsilon_2^2) &= E(\epsilon_2\epsilon_4) = \frac{V(\overline{x}_{st})}{\overline{X}^2} = \frac{1}{\overline{X}^2}\sum_{i=1}^{L}[W_i^2 \lambda_i S^2_{x_i}], \\ E(\epsilon_3^2) &= E(\epsilon_2\epsilon_3) = E(\epsilon_3\epsilon_4) = \frac{V(\overline{x}'_{st})}{\overline{X}^2} = \frac{1}{\overline{X}^2}\sum_{i=1}^{L}[W_i^2 \lambda'_i S^2_{x_i}], \\ E(\epsilon_4^2) &= \frac{V(\overline{x}^*_{st})}{\overline{X}^2} = \frac{1}{\overline{X}^2}\sum_{i=1}^{L}\left\{W_i^2 \lambda_i S^2_{x_i} + \frac{(l_i - 1)}{n_i} W_{i2} S^2_{x_i(2)}\right\}, \end{aligned} \tag{16}$$

$$E(\epsilon_0\epsilon_1) = \frac{Cov(\overline{y}^*_{1st}\overline{y}^*_{1st})}{\overline{Y}_1\overline{Y}_2} = \frac{1}{\overline{Y}_1\overline{Y}_2}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{1i}y_{2i}} + \frac{(l_i-1)}{n_i}W_{i2}S_{y_{1i}y_{2i}(2)}\right\},$$
$$E(\epsilon_0\epsilon_4) = \frac{Cov(\overline{y}^*_{1st}\overline{x}^*_{st})}{\overline{Y}_1\overline{X}} = \frac{1}{\overline{Y}_1\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{1i}x_i} + \frac{(l_i-1)}{n_i}W_{i2}S_{y_{1i}x_i(2)}\right\},$$
$$E(\epsilon_1\epsilon_4) = \frac{Cov(\overline{y}^*_{2st}\overline{x}^*_{st})}{\overline{Y}_2\overline{X}} = \frac{1}{\overline{Y}_2\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{2i}x_i} + \frac{(l_i-1)}{n_i}W_{i2}S_{y_{2i}x_i(2)}\right\},$$
$$E(\epsilon_0\epsilon_2) = \frac{Cov(\overline{y}^*_{1st}\overline{x}_{st})}{\overline{Y}_1\overline{X}} = \frac{1}{\overline{Y}_1\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_ix_i}\right\},$$
$$E(\epsilon_0\epsilon_3) = \frac{Cov(\overline{y}^*_{1st}\overline{x}'_{st})}{\overline{Y}_1\overline{X}} = \frac{1}{\overline{Y}_1\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda'_i S_{y_{1i}x_i}\right\},$$
$$E(\epsilon_1\epsilon_3) = \frac{Cov(\overline{y}^*_{2st}\overline{x}'_{st})}{\overline{Y}_2\overline{X}} = \frac{1}{\overline{Y}_2\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda'_i S_{y_{2i}x_i}\right\}$$
$$E(\epsilon_1\epsilon_2) = \frac{Cov(\overline{y}^*_{2st}\overline{x}_{st})}{\overline{Y}_2\overline{X}} = \frac{1}{\overline{Y}_2\overline{X}}\sum_{i=1}^{L} W_i^2\left\{\lambda_i S_{y_{2i}x_i}\right\} \quad (17)$$

Where, $\lambda_i = \left(\frac{1}{n_i} - \frac{1}{N_i}\right)$ and $\lambda'_i = \left(\frac{1}{n'_i} - \frac{1}{N_i}\right)$

Now, using the condition given in equations (15) and (16) on equations (13) and (14), The expressions for bias and MSE of suggested classes of estimators are given as follows:

$$Bias(\hat{R}_{1st}) = Bias(\hat{R}_{st}) + R[E(\epsilon_0\epsilon_4) - E(\epsilon_0\epsilon_3) - E(\epsilon_3\epsilon_4) + E(\epsilon_1\epsilon_4) + E(\epsilon_1\epsilon_3) + E(\epsilon_3^2)]\phi_{1(1)}(1) + \overline{X}[E(\epsilon_2\epsilon_4) - E(\epsilon_2\epsilon_3) - E(\epsilon_3\epsilon_4) - E(\epsilon_3^2)]\alpha_1\phi_{1(1)}(1) \quad (18)$$

$$MSE(\hat{R}_{1st}) = MSE(\hat{R}_{st}) + A\phi^2_{1(1)}(1) + B\alpha_1^2 + 2C\phi_{1(1)}(1) + 2D\alpha_1 + 2E\alpha_1\phi_{1(1)}(1) \quad (19)$$

$$Bias(\hat{R}_{2st}) = Bias(\hat{R}_{st}) + R[E(\epsilon_0\epsilon_2) - E(\epsilon_0\epsilon_3) - E(\epsilon_3\epsilon_2) + E(\epsilon_1\epsilon_2) + E(\epsilon_1\epsilon_3) + E(\epsilon_3^2)]\phi_{1(2)}(1) + \overline{X}[E(\epsilon_2\epsilon_4) - E(\epsilon_4\epsilon_3) - E(\epsilon_2\epsilon_3) - E(\epsilon_3^2)]\alpha_2\phi_{1(2)}(1) \quad (20)$$

$$MSE(\hat{R}_{2st}) = MSE(\hat{R}_{st}) + A_1\phi^2_{1(2)}(1) + B_1\alpha_2^2 + 2C_1\phi_{1(2)}(1) + 2D_1\alpha_2 + 2E_1\alpha_2\phi_{1(2)}(1) \quad (21)$$

The optimum values of $\phi_{1(1)}(1)$, $\phi_{1(2)}(1)$, α_1 and α_2 are obtained to get the minimum value of mean square error. These are given as follows:

$$\phi_{1(1)}(1) = \frac{(CE - AD)}{(AB - E^2)},\ \alpha_1 = \frac{(DE - BC)}{(AB - E^2)} \quad (22)$$

$$\phi_{1(2)}(1) = \frac{(C_1E_1 - A_1D_1)}{(A_1B_1 - E_1^2)},\ \alpha_2 = \frac{(D_1E_1 - B_1C_1)}{(A_1B_1 - E_1^2)} \quad (23)$$

The expressions for the minimum MSEs after substituting the optimum values of $\phi_{1(1)}(1)$, α_1, $\phi_{1(2)}(1)$ and α_2 from equations (21) and (22) in equations (18) and (20). we get,

$$MSE(\hat{R_{1st}}) = MSE(\hat{R_{st}}) - \left[\frac{BC^2 + AD^2 - 2CDE}{AB - E^2}\right] \quad (24)$$

$$MSE(\hat{R_{2st}}) = MSE(\hat{R_{st}}) - \left[\frac{B_1C_1^2 + A_1D_1^2 - 2C_1D_1E_1}{A_1B_1 - E_1^2}\right] \quad (25)$$

where,

$$A = R^2\left[E(\epsilon_3^2) + E(\epsilon_4^2) - 2E(\epsilon_3\epsilon_4)\right],$$

$$B = \overline{X}^2\left[E(\epsilon_2^2) + E(\epsilon_3^2) - 2E(\epsilon_2\epsilon_3)\right],$$

$$C = R^2\left[E(\epsilon_0\epsilon_4) - E(\epsilon_1\epsilon_4) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$D = R\overline{X}\left[E(\epsilon_0\epsilon_2) - E(\epsilon_1\epsilon_2) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$E = R\overline{X}\left[E(\epsilon_2\epsilon_4) - E(\epsilon_2\epsilon_3) - E(\epsilon_3\epsilon_4) + E(\epsilon_3^2)\right],$$

$$A_1 = R^2\left[E(\epsilon_2^2) + E(\epsilon_3^2) - 2E(\epsilon_2\epsilon_3)\right],$$

$$B_1 = \overline{X}^2\left[E(\epsilon_3^2) + E(\epsilon_4^2) - 2E(\epsilon_3\epsilon_4)\right],$$

$$C_1 = R^2\left[E(\epsilon_0\epsilon_2) - E(\epsilon_1\epsilon_2) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$D_1 = R\overline{X}\left[E(\epsilon_0\epsilon_4) - E(\epsilon_1\epsilon_4) - E(\epsilon_0\epsilon_3) + E(\epsilon_1\epsilon_3)\right],$$

$$E_1 = R\overline{X}\left[E(\epsilon_2\epsilon_4) - E(\epsilon_2\epsilon_3) - E(\epsilon_3\epsilon_4) + E(\epsilon_3^2)\right] \text{ and}$$

$$MSE(\hat{R}_{st}) = R^2[E(\epsilon_0^2) + E(\epsilon_1^2) - E(\epsilon_0\epsilon_1)] \tag{26}$$

Here, we observe that $A_1 = R^2\frac{B}{\overline{X}^2}$, $B_1 = \overline{X}^2\frac{A}{R^2}$, $C_1 = R\frac{D}{\overline{X}}$, $D_1 = \overline{X}\frac{C}{R}$ and $E_1 = E$.

After substituting these values in equation (24), we find that $MSE(\hat{R}_{1st})$ and $MSE(\hat{R}_{2st})$ are equal. i.e. $MSE(\hat{R}_{1st})$ =$MSE(\hat{R}_{2st})$ The optimum MSE of $\hat{R}_{1st}$ and $\hat{R}_{2st}$ are the same because they are utilizing the same information on y_1, y_2 and x in both cases.

5. Members of the Suggested classes

All the members of the suggested classes satisfy the conditions given in equation (10). Hence if the optimum values of the constants presented in suggested members are calculated by the expression given in equations (21) and (22) then all the members shown in table 1 will attain the minimum mean square error equal to the expression of MSE given in equation (23) and (24). The optimum values of constants are sometimes in the form of some unknown parameters and sometimes in the form of value of unknown constants. The optimum values of constants, in this situation, may be obtained from past data on the value (Reddy [11]), or by estimating the parameters included in the optimum value of the constant based on sample values. The minimum values of mean square error of the estimator up to the term of order $\frac{1}{n}$ are unchanged if we estimate the optimum values of the constants by using the sample values [Srivastava and Jhajj [15]]. If the condition given in equation (10) is satisfied by any parametric function $\phi_{(1)}(u_1)$ and $\phi_{(2)}(u_2)$ then they can generate a class of asymptotic estimators. Such classes have a large number of members. Some of them are given as follows:

Table 1: *Members of the classes*

Member of class $\hat{R}_{1st}$	Member of class $\hat{R}_{2st}$
$\hat{R}_{11st}=\left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}+\alpha_1\overline{x}'_{st}(u_2-1)\right](u_1)^{\gamma}$	$\hat{R}_{21st}=\left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}+\alpha_2\overline{x}'_{st}(u_1-1)\right](u_2)^{\gamma}$
$\hat{R}_{12st}=\left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}+\alpha_1\overline{x}'_{st}(u_2-1)\right]\left(\frac{\exp(u_1-1)}{\exp(u_1+1)}\right)$	$\hat{R}_{22st}=\left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}+\alpha_2\overline{x}'_{st}(u_1-1)\right]\left(\frac{\exp(u_2-1)}{\exp(u_2+1)}\right)$
$\hat{R}_{13st}=\left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}+\alpha_1\overline{x}'_{st}(u_2-1)\right](2-u_1{}^{\beta})$	$\hat{R}_{23st}=\left[\frac{\overline{y}^*_{1st}}{\overline{y}^*_{2st}}+\alpha_2\overline{x}'_{st}(u_1-1)\right](2-u_2{}^{\beta})$

Here, γ and β are the constants.

6. An Empirical study

Table 2: *Population parameters for each stratum*

Parameters		(i)	(ii)	(iii)	(iv)	(v)	(vi)
N_i		152	112	85	82	109	96
n'_i		108	79	60	58	77	68
n_i		42	31	24	23	30	27
$\rho_{x_iy_{1i}}$		-0.3075	-0.0559	-0.0514	-0.2213	-0.6455	-0.5745
$\rho_{x_iy_{2i}}$		0.0739	0.6171	0.6025	0.1958	-0.245	0.5606
$\rho_{y_{1i}y_{2i}}$		0.2427	0.1695	0.4927	-0.0041	0.3483	-0.4074
$\overline{Y}_{1i}$		107.96	68.95	74.19	161.05	68.98	120.38
$\overline{Y}_{2i}$		203.1	190.41	282.74	222.39	251.98	250.28
$\overline{X}_i$		581.75	575.71	651.97	631.26	692.6	643.72
$S_{y_{1i}}$		52.07	37.79	61.75	80.31	50.22	57.27
$S_{y_{2i}}$		34.01	40.19	71.51	29.3	27.97	39.47
S_{x_i}		83.32	107.88	97.21	99.47	100.12	96.2
10% Non-resposne	$\rho_{x_iy_{1i}(2)}$	-0.6188	0.3031	0.535	-0.6709	-0.8781	-0.5204
	$\rho_{x_iy_{2i}(2)}$	0.3532	-0.2579	0.8087	-0.228	-0.3799	0.1083
	$\rho_{y_{1i}y_{2i}(2)}$	-0.3266	-0.0073	0.6923	-0.3547	0.4499	-0.2732
	$S_{y_{1i}(2)}$	61.7	22.09	79.44	65.24	68.36	29.71
	$S_{y_{2i}(2)}$	18.6	9.05	26.63	65.41	38.7	25.43
	$S_{x_i(2)}$	13.06	12.36	90.48	62.8	12.52	21.67
20 % Non-response	$\rho_{x_iy_{1i}(2)}$	-0.4829	-0.2723	0.0952	-0.7095	-0.8498	-0.263
	$\rho_{x_iy_{2i}(2)}$	0.1189	-0.653	0.7614	0.0035	0.0355	0.1717
	$\rho_{y_{1i}y_{2i}(2)}$	-0.0311	-0.3092	0.3305	-0.0823	-0.1392	-0.118
	$S_{y_{1i}(2)}$	67.53	52.33	69.71	77.85	59.14	60.05
	$S_{y_{2i}(2)}$	30.65	9.76	29.04	63.53	33.07	34.87
	$S_{x_i(2)}$	15.18	22.89	64.53	53.64	15.05	18.7
30% Non-response	$\rho_{x_iy_{1i}(2)}$	-0.4373	-0.1605	0.2785	-0.2861	-0.5	-0.3592
	$\rho_{x_iy_{2i}(2)}$	0.0884	-0.428	0.605	0.0022	0.2985	0.2831
	$\rho_{y_{1i}y_{2i}(2)}$	0.1563	0.2582	0.7877	-0.0932	-0.092	0.2379
	$S_{y_{1i}(2)}$	64.88	61.92	97.94	105.66	64.33	58.05
	$S_{y_{2i}(2)}$	31.78	10.76	72.99	76.89	32.25	46.2
	$S_{x_i(2)}$	20.39	20.51	88.09	43.81	15.46	26.26

The data used in the study has been taken from the Primary Census Abstract-2011 published by the Office of the Registrar General & Census Commissioner, India.

The number of cultivators=y_1 (Study variable)

The number of main workers=y_2 (Study variable)

The number of literate persons= x (auxiliary variable)

Here, we are considering the number of cultivators, main workers and literate persons per thousand population.

In the given population, we have six strata which are given as: Strata (i) = Central states, Strata (ii)= Eastern states, Strata (iii) = Northern states, Strata (iv)=North- East states, Strata (v) = Southern states, and Strata (vi) = Western states. And their parameters are given in table 2.

For the different choices of sub-sampling fraction i.e. $\frac{1}{l} = \frac{1}{2}$, $\frac{1}{l} = \frac{1}{3}$ and $\frac{1}{l} = \frac{1}{4}$ and for the different non-response rates 10%, 20% and 30% the optimum values of $\phi_{1(1)}(1)$,$\phi_{1(2)}(1)$, α_1 and α_2 are given in table 3.

Table 3: *Optimum values of $\phi_{1(1)}(1)$,$\phi_{1(2)}(1)$, α_1 and α_2*

Non-response rate(%)	Constants	$1/l = \frac{1}{2}$	$1/l = \frac{1}{3}$	$1/l = \frac{1}{4}$
10	$\phi_{1(1)}(1)$	1.1023	1.1985	1.2786
	α_1	0.00063	0.00059	0.00056
	$\phi_{1(2)}(1)$	0.92	0.8672	0.8233
	α_2	0.00075	0.00082	0.00087
20	$\phi_{1(1)}(1)$	1.1784	1.3019	1.3875
	α_1	0.0006	0.00055	0.00052
	$\phi_{1(2)}(1)$	0.8783	0.8105	0.7636
	α_2	0.00081	0.00089	0.00095
30	$\phi_{1(1)}(1)$	1.0054	1.0144	1.0204
	α_1	0.00067	0.00066	0.00066
	$\phi_{1(2)}(1)$	0.9732	0.9679	0.9649
	α_2	0.00069	0.00069	0.0007

Table 4: *The percentage relative efficiency (PRE) of $\hat{R}_{1st}$ and $\hat{R}_{1st}$ with $\hat{R}_{st}$*

Non-response rates(%)	**Estimators**	(N=636)		
		1/l		
		1/2	1/3	1/4
10	$\hat{R}_{st}$	**100(124.85)**	**100(134.11)**	**100(143.37)**
	$\hat{R}_{1st} = \hat{R}_{2st}$	**122.20(102.16)**	**122.53(109.45)**	**122.91(116.64)**
20	$\hat{R}_{st}$	**100(134.73)**	**100(153.88)**	**100(173.02)**
	$\hat{R}_{1st} = \hat{R}_{2st}$	**122.54(109.94)**	**123.26(124.83)**	**123.97(139.56)**
30	$\hat{R}_{st}$	**100(155.29)**	**100(194.99)**	**100(234.69)**
	$\hat{R}_{1st} = \hat{R}_{2st}$	**118.21(131.36)**	**116.08(167.97)**	**114.72(204.57)**

Note:Figures in parenthesis show MSE of the estimators in 10^{-6}.

7. Discussion

Table 4 shows the PREs and MSEs (Figures in parenthesis) of the estimators $\hat{R}_{st}$, $\hat{R}_{1st}$ and $\hat{R}_{2st}$ for the different rates of non-response and choices of the sub-sampling fraction. We can see that in table 4, as we increase the non-response rates the MSE increases for each sub-sampling fractions.

The PRE for the 10% and 20% non-response rates shows almost same value. However, for 30% non-response rate the PRE decreases for each sub-sampling fraction.

By increasing the value of l the MSE increases and PRE for 10% and 20% non-response rates are almost same and PRE decreases slightly for 30% non-response rate.

The MSE and PRE of the suggested classes of estimators are found to be equal for different rates of non-response and different choices of sub-sampling fraction.

8. Conclusion

Hence, the findings on the basis of empirical study are justified that the proposed classes of estimators for estimating the ratio of two population means with unknown population mean of auxiliary variable in the presence of non-response for stratified population performed better than the usual estimator $\hat{R}_{st}$. So, we recommend the suggested classes of estimators for the use in practice.

References

[1] El-Badry, M. A. (1956). A sampling procedure for mailed questionnaires. *Journal of the American Statistical Association* , 51(274), 209-227.

[2] Hansen, M. H., Hurwitz, W. N. (1946). The problem of non-response in sample surveys. of the American Statistical Association, 41(236), 517-529.

[3] 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.

[4] Khare, B.B. Pandey, S.K. (2000). A class of estimators for ratio of two population means using auxiliary character in presence of non-response. *J. Sc. Res. B.H.U.*, 50, 115-124.

[5] Khare, B.B. Sinha, R.R. (2002). Estimators of the ratio of two population means using auxiliary character with unknown population mean in presence of non-response. *Prog. Maths. B.H.U.*, Vol.36, No.(1,2) 337-348.

[6] Khare, B. B., Sinha, R. R. (2004). Estimation of finite population ratio using two-phase sampling scheme in the presence of non-response. *Aligarh Journal of Statistics*, 24, 43-56.

[7] Khare, B. B., Sinha, R. R. (2007). Estimation of the ratio of the two population means using multi auxiliary characters in the presence of non-response. Statistical Techniques in Life Testing, *Reliability, Sampling Theory and Quality Control*, 1, 63-171.

[8] Khare, B. B., Jha, P. S., Kumar, K. (2014). Improved generalized chain estimators for ratio and product of two population means using two auxiliary characters in the presence of non-response. *International J. Stats Economics*, 13(1), 108-121.

[9] Khare, B. B., Jha, P. S. (2017). Classes of estimators for population mean using auxiliary variable in stratified population in the presence of non-response. *Communications in Statistics-Theory and Methods*, 46(13), 6579-6589.

[10] Kumar, K. Srivastava, U. (2018). Estimation of ratio and product of two population means using exponential type estimators in sample surveys. *International Journal of Mathematics and Statistics*, 19"3, pp. 102"109.

[11] Reddy, V. N. (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya C*, 40, 29-37.

[12] Singh, R. , Khare, S. , Khare, B.B. Jha , P.S. (2020). The general classes of estimators for population mean under stratified two phase random sampling in the presence of non-response. *Int. J. Agricult. Stat. Sci.*, Vol. 16, No. 2, pp. 557-565, 2020.

[13] Singh, M. P. (1965). On the estimation of ratio and product of the population parameters. *Sankhy´: The Indian Journal of Statistics*, Series B, 321-328.

[14] Singh, R. K. Naqvi, N. (2015). A generalized class of estimator of the ratio of two population means using auxiliary information. *Sri Lankan Journal of Applied Statistics*, 16"3, pp. 179"193.

[15] Srivastava, S. K., Jhajj, H. S. (1983). A class of estimators of the population mean using multiauxiliary information. *Calcutta Statistical Association Bulletin*, 32(1-2), 47-56.

[16] Ahuja, T. K. ., Misra, P. ., Belwal, O. K. . (2021). A Generalized Two Phase Sampling Estimator of Ratio of Population Means Using Auxiliary Information. *Journal of Reliability and Statistical Studies*, 14(01), 1"16.

[17] Tripathi, T.P. (1980). A general class of estimators for population ratio. *Sankhya*, Ser. C, 42, 63-75.

[18] Upadhyaya, L. N., Singh, G. N., Singh, H. P. (2000). Use of transformed auxiliary variable in the estimation of population ratio in sample survey. *Statistics in Transition*, 4(6),1019-102.