Pakistan Journal of Statistics and Operation Research

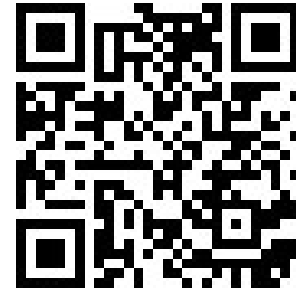# Some Improved
Estimators of Population Mean using Two-Phase
Sampling Scheme in the Presence of Non-Response

Manoj K. Chaudhary[1*], Amit Kumar[2],
Gautam K. Vishwakarma[3]

*Corresponding author

1. Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India, ritamanoj15@gmail.com
2. Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India, amit.stat40@gmail.com
3. Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India, gautamkr111@gmail.com

Abstract

In the present paper, we have proposed some improved estimators of the population mean utilizing the information on two auxiliary variables with the use of two-phase sampling scheme under non-response. We have discussed the situation in which study and first auxiliary variables suffer from the non-response while the second auxiliary variable is free from the non-response. The expressions for the bias and mean square error of the proposed estimators have been derived up to the first order of approximation. We have compared the efficiency of the proposed estimators with that of the usual mean estimator and some well known existing estimators of the population mean. The theoretical results have also been illustrated through some empirical data.

Key Words: Improved Estimators; Population Mean; Two-Phase Sampling; Auxiliary Information; Non-Response.

Mathematical Subject Classification: 60D05.

1. Introduction

In some practical situations, the information on more than one auxiliary variable is easily available. The information on all such variables can be utilized to provide more précised estimates of the population parameters. Olkin(1958) was the first who proposed multivariate ratio estimator of the population mean utilizing the information on a number of auxiliary variables. It was further extended by Srivastava(1966) ,Rao and Mudholkar(1967), Singh(1967) and Sahai(1980). Moreover, Chand(1975), Kiregyera(1980), Kadilar and Cingi(2005), Shukla and Thakur(2012), Kumar and Sharma(2020) and others have suggested a variety of estimators of the population mean using the information on two auxiliary variables.

It is of the great importance to deal with the problem of non-response in a sample survey. Non-response arises from the fact that the investigator fails to get the information from the expected units of the population. If the characteristics of the non-responding units are very similar to those of the responding units, then the non-response does not have an important role. But, practically, it is always seen that the responding units are not very similar to the non-responding units and hence the non-response plays an important role in such situations. The problem of non-response was first tackled by Hansen and Hurwitz(1946). They handled the problem by introducing a technique of sub-sampling of the non-respondents. Later, there are several authors who have dealt with the problem of non-response by proposing

various estimators of the population parameters. The authors such as Cochran(1977), Olkin(1986), Khare and Srivas-tava(1993),Khare and Srivastava (1995), Okafor and Lee(2000), Tabasum and Khan(2004), Tabasum and Khan(2006), Singh and Kumar(2007), Singh and Kumar(2009),Chaudhary and Smarandache(2014), Pal and Singh (2018),Khare and Sinha(2019) have discussed the problem of non-response in two-phase sampling scheme.

The problem of estimating the parameters in two-phase sampling would be more complex when the non-response is observed on both study and auxiliary variables. The problem of such situation was handled by Chaudhary and Ku-mar(2016). To get the improved estimators of the population mean under the present situation, we have utilized the information on two auxiliary variables. We have proposed the estimators under the assumption that the additional (second) auxiliary variable is free from the non-response. It is further assumed that the first auxiliary variable has a high degree of correlation with the study variable rather than the second auxiliary variable has with the study variable. In order to demonstrate the theoretical facts, we have considered two different sets of empirical data.

2. Sampling Strategy

Suppose a population comprises of N units $(U_1, U_2, ..., U_N)$. Let Y be the study variable with population mean $\overline{Y}$. Let X and Z be the two auxiliary variables with respective population means $\overline{X}$ and $\overline{Z}$. Let y_i, x_i and $z_i (i = 1, 2, ...N)$ be the observations on the ith unit in the population for the variables Y, X and Z respectively. Let us assume that the non-response is observed on the study variable Y and auxiliary variable X whereas the auxiliary variable Z is free from the non-response. It is further assumed that the population mean $\overline{X}$ of the auxiliary variable X is not known. In the present situation, we first estimate $\overline{X}$ adopting the idea of double (two-phase) sampling scheme and then use the estimate so obtained along with the information on another auxiliary variable Z to get the estimate of $\overline{Y}$. Let a larger sample of n' units be selected from N units using simple random sampling without replacement (SRSWOR) scheme at the first phase and then a smaller sub-sample of n units be selected from n' $(n < n')$ units using SRSWOR scheme at the second phase. At the first phase, out of n' units, n'_1 units respond and n'_2 units do not respond on the auxiliary variable X. Now, a sub-sample of h'_2 units is selected from the n'_2 units using SRSWOR scheme $(h'_2 = n'_2 / L', L' > 1)$ and the information from all the h'_2 units is collected [See Hansen and Hurwitz (1946)]. Thus, the estimators of $\overline{X}$ and $\overline{Z}$ at the first phase are respectively given as

$$\overline{x}'^* = \frac{n'_1 \overline{x}'_{n1} + n'_2 \overline{x}'_{h2}}{n'} \tag{1}$$

and

$$\overline{z}' = \frac{1}{n'} \sum_i^{n'} z_i \tag{2}$$

where $\overline{x}'_{n1}$ and $\overline{x}'_{h2}$ are respectively the means based on n'_1 responding units and h'_2 non-responding units for the auxiliary variable X.
The variances of the estimators $\overline{x}'^*$ and $\overline{z}'$ are respectively given by

$$V\left(\overline{x}'^*\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_X^2 + \frac{\left(L' - 1\right)}{n'} W_2 S_{X2}^2 \tag{3}$$

$$V\left(\overline{z}'\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_Z^2 \tag{4}$$

where S_X^2 and S_{X2}^2 are respectively the population mean squares of entire group and non-response group for the auxiliary variable X. S_Z^2 is the population mean square of the auxiliary variable Z and W_2 represents the non-response rate in the population.
Further, at the second phase, it is observed that there are n_1 responding units and n_2 non-responding units in the sample of n units for the study variable Y and auxiliary variable X. It is remembered that the auxiliary variable Z is free from the non-response. Adopting the Hansen and Hurwitz (1946) technique of sub-sampling of non-respondents,

we select a sub-sample of h_2 units from the n_2 non-responding units using SRSWOR scheme at the second phase ($h_2 = n_2/L, L > 1$) and gather the information from all the h_2 units. Thus, the Hansen and Hurwitz (1946) estimators of $\overline{Y}$ and $\overline{X}$ at the second phase are respectively given by

$$\overline{y}^* = \frac{n_1 \overline{y}_{n1} + n_2 \overline{y}_{h2}}{n} \tag{5}$$

$$\overline{x}^* = \frac{n_1 \overline{x}_{n1} + n_2 \overline{x}_{h2}}{n} \tag{6}$$

where $\overline{y}_{n1}$ and $\overline{y}_{h2}$ are the means based on n_1 responding units and h_2 non-responding units respectively for the study variable. $\overline{x}_{n1}$ and $\overline{x}_{h2}$ are respectively the means based on n_1 responding units and h_2 non-responding units for the auxiliary variable X.

The variances of the estimators $\overline{y}^*$ and $\overline{x}^*$ are respectively given as

$$V\left(\overline{y}^*\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^2 + \frac{(L-1)}{n} W_2 S_{Y2}^2 \tag{7}$$

$$V\left(\overline{x}^*\right) = \left(\frac{1}{n} - \frac{1}{N}\right) S_X^2 + \frac{(L-1)}{n} W_2 S_{X2}^2 \tag{8}$$

where S_Y^2 and S_{Y2}^2 are the population mean squares of entire group and non-response group respectively for the study variable.

Chaudhary and Kumar(2016) have proposed some ratio and regression-type estimators of the population mean $\overline{Y}$ using the information on the auxiliary variable X under the condition that the non-response is observed on both study and auxiliary variables and the population mean $\overline{X}$ of the auxiliary variable X is not known. The estimators are respectively given as

$$T_{1ratio}^* = \frac{\overline{y}^*}{\overline{x}^*} \overline{x}'^* \tag{9}$$

$$T_{2reg}^* = \overline{y}^* + b^* \left(\overline{x}'^* - \overline{x}^*\right) \tag{10}$$

where $b^* \left(= \frac{s_{xy}^*}{s_x^{2*}}\right)$ is an estimator of the population regression coefficient of Y on X i.e. $\beta_1 = \frac{S_{XY}}{S_X^2}$. s_{xy}^* and s_x^{2*} are respectively the unbiased estimators of S_{XY} and S_X^2, based on $(n_1 + h_2)$ units. $S_{XY} = \rho_{XY} S_X S_Y$ and ρ_{XY} is the population correlation coefficient between Y and X.

The expressions for the mean square error (MSE) of the estimators T_{1ratio}^* and T_{2reg}^* up to the first order of approximation are respectively given as

$$MSE\left(T_{1ratio}^*\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_Y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(S_Y^2 + R^2 S_X^2 - 2\rho_{XY} R S_X S_Y\right) + \frac{(L-1)}{n} W_2 S_{Y2}^2$$

$$+ W_2 \left(R^2 S_{X2}^2 - 2\rho_{XY2} R S_{X2} S_{Y2}\right) \left[\frac{(L-1)}{n} - \frac{(L'-1)}{n'}\right] \tag{11}$$

and, $$MSE\left(T_{2reg}^*\right) = \left[\left(\frac{1}{n'} - \frac{1}{N}\right) + \left(\frac{1}{n} - \frac{1}{n'}\right)\left(1 - \rho_{XY}^2\right)\right] S_Y^2 + \frac{(L-1)}{n} W_2 S_{Y2}^2$$

$$+ W_2 \left(\beta_1^2 S_{X2}^2 + 2\beta_1 \rho_{XY2} S_{X2} S_{Y2}\right) \left[\frac{(L-1)}{n} - \frac{(L'-1)}{n'}\right] \tag{12}$$

where ρ_{XY2} is the population correlation coefficient between Y and X for the non-response group. $R = \overline{Y}/\overline{X}$.

3.   Proposed Estimators

We now propose some improved ratio and regression-type estimators of the population mean $\overline{Y}$ using the information on one more auxiliary variable Z under the situation in which study variable Y and auxiliary variable X are suffering from the non-response while the auxiliary variable Z is free from the non-response. It is further assumed that the population mean $\overline{X}$ is not known. The proposed ratio and regression-type estimators under the given situation are respectively represented as

$$T_{1ratio}^{**} = \bar{y}^* \frac{\bar{x}_R^{'*}}{\bar{x}^*} \tag{13}$$

$$T_{2reg}^{**} = \bar{y}^* + b^* \left(\bar{x}_{lr}^{'*} - \bar{x}^* \right) \tag{14}$$

where $\bar{x}_R^{'*} = \frac{\bar{x}^{'*}}{\bar{z}^{'}} \bar{Z}$, $\bar{x}_{lr}^{'*} = \bar{x}^{'*} + b^{'*} \left(\bar{Z} - \bar{z}^{'} \right)$ and $b^{'*} \left(= \frac{s_{xz}^{'*}}{s_z^{2'*}} \right)$ is an estimator of the population regression coefficient

of X on Z i.e. $\beta_2 = \frac{S_{XZ}}{S_Z^2}$. $s_{xz}^{'*}$ and $s_z^{2'*}$ are respectively the unbiased estimators of S_{XZ} and S_Z^2, based on $\left(n_1^{'} + h_2^{'} \right)$ units. $S_{XZ} = \rho_{XZ} S_X S_Z$ and ρ_{XZ} is the population correlation coefficient between X and Z.

To obtain the biases and mean square errors of the estimators T_{1ratio}^{**} and T_{2reg}^{**}, we adopt the theory of large sample approximations. Let us assume

$\bar{y}^* = \bar{Y} (1 + e_0), \bar{x}^* = \overline{X} (1 + e_1), \bar{x}^{'*} = \overline{X} \left(1 + e_1^{'} \right), \bar{z}^{'} = \bar{Z} \left(1 + e_2^{'} \right), s_{xy}^* = S_{XY} (1 + e_2), s_x^{2*} = S_X^2 (1 + e_3),$

$s_{xz}^{'*} = S_{XZ} (1 + e_4)$ and $s_z^{2'*} = S_Z^2 (1 + e_5)$ such that

$$E(e_0) = E(e_1) = E\left(e_1^{'}\right) = E\left(e_2^{'}\right) = E(e_2) = E(e_3) = E(e_4) = E(e_5) = 0,$$

$$E\left(e_0^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right) C_Y^2 + \frac{(L-1)}{n} W_2 \frac{S_{Y2}^2}{\overline{Y}^2}, E\left(e_1^2\right) = \left(\frac{1}{n} - \frac{1}{N}\right) C_X^2 + \frac{(L-1)}{n} W_2 \frac{S_{X2}^2}{\overline{X}^2},$$

$$E\left(e_1^{'2}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) C_X^2 + \frac{\left(L'-1\right)}{n'} W_2 \frac{S_{X2}^2}{\overline{X}^2}, E\left(e_2^{'2}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) C_Z^2,$$

$$E(e_0 e_1) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{XY} C_X C_Y + \frac{(L-1)}{n} W_2 \rho_{XY2} \frac{S_{X2}}{\overline{X}} \frac{S_{Y2}}{\overline{Y}},$$

$$E\left(e_0 e_1^{'}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{XY} C_X C_Y + \frac{\left(L'-1\right)}{n'} W_2 \rho_{XY2} \frac{S_{X2}}{\overline{X}} \frac{S_{Y2}}{\overline{Y}},$$

$$E\left(e_0 e_2^{'}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{YZ} C_Y C_Z, E\left(e_1 e_1^{'}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) C_X^2 + \frac{\left(L'-1\right)}{n'} W_2 \frac{S_{X2}^2}{\overline{X}^2},$$

$$E\left(e_1 e_2^{'}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{XZ} C_X C_Z, E\left(e_1^{'} e_2^{'}\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{XZ} C_X C_Z,$$

$$E(e_1 e_2) = \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{21}}{n \bar{X} S_{XY}} + W_2 \frac{(L-1)}{n} \frac{\mu_{21(2)}}{\bar{X} S_{XY}},$$

$$E\left(e_1^{'} e_2\right) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{21}}{n' \bar{X} S_{XY}} + W_2 \frac{\left(L'-1\right)}{n'} \frac{\mu_{21(2)}}{\bar{X} S_{XY}},$$

$$E(e_1 e_3) = \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{30}}{n \bar{X} S_X^2} + W_2 \frac{(L-1)}{n} \frac{\mu_{30(2)}}{\bar{X} S_X^2},$$

$$E\left(e_1^{'} e_3\right) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{30}}{n' \bar{X} S_X^2} + W_2 \frac{\left(L'-1\right)}{n'} \frac{\mu_{30(2)}}{\bar{X} S_X^2}, E\left(e_2^{'} e_2\right) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{111}}{n' \bar{Z} S_{XY}},$$

$$E\left(e_2' e_3\right) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{21}'}{n'\bar{Z}S_X^2}, \; E\left(e_2' e_4\right) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{12}'}{n'\bar{Z}S_{XZ}} \text{ and }$$

$$E\left(e_2' e_5\right) = \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{03}'}{n'\bar{Z}S_Z^2}$$

where $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^r\left(y_i - \bar{Y}\right)^s$, $\mu_{rs(??)} = \frac{1}{N_2}\sum_{i}^{N_2}\left(x_i - \bar{X}_2\right)^r\left(y_i - \bar{Y}_2\right)^s$, $\mu_{rs}' = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^r\left(z_i - \overline{Z}\right)^s$, $\mu_{rst} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^r\left(y_i - \bar{Y}\right)^s\left(z_i - \overline{Z}\right)^t$, $C_Y = \frac{S_Y}{\bar{Y}}$, $C_X = \frac{S_X}{\bar{X}}$, $C_Z = \frac{S_Z}{\bar{Z}}$, $\bar{X}_2 = \frac{1}{N_2}\sum_{i}^{N_2} x_i$, $\bar{Y}_2 = \frac{1}{N_2}\sum_{i}^{N_2} y_i$ and ρ_{YZ} is the population correlation coefficient between Y and Z. Now, we express the equation (13) in terms of e_0, e_1, e_1', e_2' and neglect the terms involving powers of e_0, e_1, e_1', e_2' higher than two. Thus, we have

$$T_{1ratio}^{**} - \bar{Y} = \bar{Y}\left[e_0 - e_1 + e_1' - e_2' + e_1^2 + e_2'^2 - e_0 e_1 + e_0 e_1' - e_0 e_2' - e_1 e_1' + e_1 e_2' - e_1' e_2'\right] \quad (15)$$

Taking expectation on both the sides of equation (15), we get

$$E\left[T_{1ratio}^{**} - \bar{Y}\right] = \bar{Y}\left[E\left(e_1^2\right) + E\left(e_2'^2\right) - E\left(e_0 e_1\right) + E\left(e_0 e_1'\right) - E\left(e_0 e_2'\right) - E\left(e_1 e_1'\right) + E\left(e_1 e_2'\right) - E\left(e_1' e_1^2\right)\right]$$

$$= \bar{Y}\left[fC_X^2 + \frac{(L-1)}{n}W_2\frac{S_{X2}^2}{\bar{X}^2} + f'C_Z^2 - f\rho_{XY}C_X C_Y - \frac{(L-1)}{n}W_2\rho_{XY2}\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}} + f'\rho_{XY}C_X C_Y\right.$$

$$\left. + \frac{\left(L'-1\right)}{n'}W_2\rho_{XY2}\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}} - f'\rho_{YZ}C_Y C_Z - f'C_X^2 - \frac{\left(L'-1\right)}{n'}W_2\frac{S_{X2}^2}{\overline{X}^2} + f'\rho_{XZ}C_X C_Z - f'\rho_{XZ}C_X C_Z\right]$$

where $f = \left(\frac{1}{n} - \frac{1}{N}\right)$ and $f' = \left(\frac{1}{n'} - \frac{1}{N}\right)$. Thus, the expression for the bias of T_{1ratio}^{**} up to the first order of approximation is given by

$$B\left(T_{1ratio}^{**}\right) = \bar{Y}\left[f^*\left(C_X^2 - \rho_{XY}C_X C_Y\right) + f'\left(C_Z^2 - \rho_{YZ}C_Y C_Z\right) + W_2\left(\frac{S_{X2}^2}{\overline{X}^2} - \rho_{XY2}\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}}\right)\left[\frac{(L-1)}{n} - \frac{\left(L'-1\right)}{n'}\right]\right]$$

(16)

where $f^* = \left(\frac{1}{n} - \frac{1}{n'}\right)$.

Squaring both the sides of the equation (15) and then taking expectation on neglecting the terms having powers of e_0, e_1, e_1' and e_2' greater than two, we get

$$E\left[T_{1ratio}^{**} - \bar{Y}\right]^2 = \bar{Y}^2\left[E\left(e_0^2\right) + E\left(e_1^2\right) + E\left(e_1'^2\right) + E\left(e_2'^2\right) - 2E\left(e_0 e_1\right) + 2E\left(e_0 e_1'\right)\right.$$

$$\left. - 2E\left(e_0 e_2'\right) - 2E\left(e_1 e_1'\right) + 2E\left(e_1 e_2'\right) - 2E\left(e_1' e_2'\right)\right]$$

$$= \bar{Y}^2\left[fC_Y^2 + \frac{(L-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2} + fC_X^2 + \frac{(L-1)}{n}W_2\frac{S_{X2}^2}{\overline{X}^2} + f'C_X^2 + \frac{\left(L'-1\right)}{n'}W_2\frac{S_{X2}^2}{\overline{X}^2} + f'C_Z^2\right.$$

$$-2f\rho_{XY}C_X C_Y - 2\frac{(L-1)}{n}W_2\rho_{XY2}\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}} + 2f'\rho_{XY}C_X C_Y + 2\frac{\left(L'-1\right)}{n'}W_2\rho_{XY2}\frac{S_{X2}}{\overline{X}}\frac{S_{Y2}}{\overline{Y}}$$

$$\left. -2f'\rho_{YZ}C_Y C_Z - 2f'C_X^2 - 2\frac{\left(L'-1\right)}{n'}W_2\frac{S_{X2}^2}{\overline{X}^2} + 2f'\rho_{XZ}C_X C_Z - 2f'\rho_{XZ}C_X C_Z\right].$$

Thus, the expression for the MSE of T_{1ratio}^{**} up to the first order of approximation is given as

$$MSE\left(T_{1ratio}^{**}\right) = \bar{Y}^2\left[f^*\left(C_Y^2 + C_X^2 - 2\rho_{XY}C_X C_Y\right) + f'\left(C_Y^2 + C_Z^2 - 2\rho_{YZ}C_Y C_Z\right) + \frac{(L-1)}{n}W_2\frac{S_{Y2}^2}{\overline{Y}^2}\right.$$

$$+ W_2 \left(\frac{S_{X2}^2}{\bar{X}^2} - 2\rho_{XY2} \frac{S_{X2}}{\bar{X}} \frac{S_{Y2}}{\bar{Y}} \right) \left[\frac{(L-1)}{n} - \frac{\left(L'-1\right)}{n'} \right] \right]] \tag{17}$$

Now, on expressing the equation (14) in terms of $e_0, e_1, e_1', e_2', e_2, e_3, e_4, e_5$ and then ignoring the items involving powers of $e_0, e_1, e_1', e_2', e_2, e_3, e_4, e_5$ greater than two, we get

$$T_{2reg}^{**} - \bar{Y} = \bar{Y} e_0 + \beta_1 \bar{X} \left(e_1' - e_1 - e_1 e_2 + e_1 e_3 + e_1' e_2 - e_1' e_3 \right) - \beta_1 \beta_2 \bar{Z} \left(e_2' + e_2' e_2 - e_2' e_3 + e_2' e_4 - e_2' e_5 \right) \tag{18}$$

Taking expectation on both the sides of equation (18), we get

$$E\left[T_{2reg}^{**} - \bar{Y} \right] = \beta_1 \bar{X} \left[-E(e_1 e_2) + E(e_1 e_3) + E(e_1' e_2) - E(e_1' e_3) \right]$$

$$-\beta_1 \beta_2 \bar{Z} \left[E(e_2' e_2) - E(e_2' e_3) + E(e_2' e_4) - E(e_2' e_5) \right]$$

$$= \beta_1 \bar{X} \left(-\frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{21}}{n\bar{X} S_{XY}} - W_2 \frac{(L-1)}{n} \frac{\mu_{21(2)}}{\bar{X} S_{XY}} + \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{30}}{n\bar{X} S_X^2} + W_2 \frac{(L-1)}{n} \frac{\mu_{30(2)}}{\bar{X} S_x^2} \right.$$

$$\left. + \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{21}}{n'\bar{X} S_{XY}} + W_2 \frac{\left(L'-1\right)}{n'} \frac{\mu_{21(2)}}{\bar{X} S_{XY}} - \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{30}}{n'\bar{X} S_X^2} - W_2 \frac{\left(L'-1\right)}{n'} \frac{\mu_{30(2)}}{\bar{X} S_X^2} \right)$$

$$-\beta_1 \beta_2 \bar{Z} \left(\frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{111}}{n'\bar{Z} S_{XY}} - \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{21}'}{n'\bar{Z} S_X^2} + \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{12}'}{n'\bar{Z} S_{XZ}} - \frac{N(N-n')}{(N-1)(N-2)} \frac{\mu_{03}'}{n'\bar{Z} S_Z^2} \right)$$

Therefore, the expression for the bias of T_{2reg}^{**} up to the first order of approximation is given as

$$Bias\left(T_{2reg}^{**} \right) = \beta_1 \left\{ \frac{N^2}{(N-1)(N-2)} \left(\frac{n-n'}{nn'} \right) \left(\frac{\mu_{21}}{S_{XY}} - \frac{\mu_{30}}{S_X^2} \right) + W_2 \left(\frac{\left(L'-1\right)}{n'} - \frac{(L-1)}{n} \right) \left(\frac{\mu_{21(2)}}{S_{XY}} - \frac{\mu_{30(2)}}{S_X^2} \right) \right\}$$

$$- \beta_1 \beta_2 \left\{ \frac{N\left(N-n'\right)}{(N-1)(N-2)n'} \left(\frac{\mu_{111}}{S_{XY}} - \frac{\mu_{21}'}{S_X^2} + \frac{\mu_{12}'}{S_{XZ}} - \frac{\mu_{03}'}{S_Z^2} \right) \right\} \tag{19}$$

Squaring both the sides of the equation (18) and then taking expectation on neglecting the terms having powers of $e_0, e_1, e_1', e_2', e_2, e_3, e_4, e_5$ higher than two, we get

$$E\left(T_{2reg}^{**} - \bar{Y} \right)^2 = \bar{Y}^2 E(e_0^2) + \beta_1^2 \bar{X}^2 \left[E\left(e_1'^2 \right) + E\left(e_1^2 \right) - 2E\left(e_1' e_1 \right) \right] + \beta_1^2 \beta_2^2 \bar{Z}^2 E\left(e_2'^2 \right) + 2\beta_1 \bar{Y} \bar{X} \left[E\left(e_0 e_1' \right) - E\left(e_0 e_1 \right) \right]$$

$$-2\beta_1 \beta_2 \bar{Y} \bar{Z} E\left(e_0 e_2' \right) - 2\beta_1^2 \beta_2 \bar{X} \bar{Z} \left[E\left(e_1' e_2' \right) - E\left(e_1 e_2' \right) \right]$$

$$= \bar{Y}^2 \left(f C_Y^2 + \frac{(L-1)}{n} W_2 \frac{S_{Y2}^2}{\bar{Y}^2} \right) + \beta_1^2 \bar{X}^2 \left(f' C_X^2 + \frac{\left(L'-1\right)}{n'} W_2 \frac{S_{X2}^2}{\bar{X}^2} + f C_X^2 + \frac{(L-1)}{n} W_2 \frac{S_{X2}^2}{\bar{X}^2} \right.$$

$$-2f' C_X^2 - 2 \frac{\left(L'-1\right)}{n'} W_2 \frac{S_{X2}^2}{\bar{X}^2} \right) + \beta_1^2 \beta_2^2 \bar{Z}^2 f' C_Z^2 + 2\beta_1 \bar{Y} \bar{X} \left(f' \rho_{XY} C_X C_Y + \frac{\left(L'-1\right)}{n'} W_2 \rho_{XY2} \frac{S_{X2}}{\bar{X}} \frac{S_{Y2}}{\bar{Y}} \right.$$

$$\left. - f \rho_{XY} C_X C_Y - \frac{(L-1)}{n} W_2 \rho_{XY2} \frac{S_{X2}}{\bar{X}} \frac{S_{Y2}}{\bar{Y}} \right) - 2\beta_1 \beta_2 \bar{Y} \bar{Z} f' \rho_{YZ} C_Y C_Z - 2\beta_1^2 \beta_2 \bar{X} \bar{Z} \left(f' \rho_{XZ} C_X C_Z - f' \rho_{XZ} C_X C_Z \right)$$

Thus, the expression for the MSE of T_{2reg}^{**} up to the first order of approximation is given by

$$MSE\left(T_{2reg}^{**}\right) = fS_Y^2 + \beta_1\beta_2f'\left(\beta_1\beta_2S_Z^2 - 2\rho_{YZ}S_YS_Z\right) + \beta_1f^*\left(\beta_1S_X^2 - 2\rho_{XY}S_XS_Y\right) + \frac{(L-1)}{n}W_2S_{Y2}^2$$

$$+ W_2\left(\beta_1^2S_{X2}^2 - 2\beta_1\rho_{XY2}S_{X2}S_{Y2}\right)\left[\frac{(L-1)}{n} - \frac{\left(L'-1\right)}{n'}\right] \quad (20)$$

4. Efficiency Comparisons

In this section, we have tried to compare the efficiency of the proposed estimators with that of some existing estimators:

Case (i): A comparison of the equations (7) and (17) reveals that
$MSE\left(T_{1ratio}^{**}\right) - V\left(\bar{y}^*\right) < 0$ if

$$f^*\left(C_Y^2 + C_X^2 - 2\rho_{XY}C_XC_Y\right) + f'\left(C_Y^2 + C_Z^2 - 2\rho_{YZ}C_YC_Z\right) - f\frac{S_Y^2}{\bar{Y}^2}$$

$$+ W_2\left(\frac{S_{X2}^2}{\bar{X}^2} - 2\rho_{XY2}\frac{S_{X2}}{\bar{X}}\frac{S_{Y2}}{\bar{Y}}\right)\left[\frac{(L-1)}{n} - \frac{\left(L'-1\right)}{n'}\right] < 0.$$

Case (ii): From equations (11) and (17), we have
$MSE\left(T_{1ratio}^{**}\right) - MSE\left(T_{1ratio}^{*}\right) < 0$ if

$$f'\left(C_Z^2 - 2\rho_{YZ}S_YS_Z\right) < 0.$$

Case (iii): A comparison of the equations (7) and (20) provides
$MSE\left(T_{2reg}^{**}\right) - V\left(\bar{y}^*\right) < 0$ if

$$\beta_1\beta_2f'\left(\beta_1\beta_2S_Z^2 - 2\rho_{YZ}S_YS_Z\right) + \beta_1f^*\left(\beta_1S_X^2 - 2\rho_{XY}S_XS_Y\right)$$

$$+ W_2\left(\beta_1^2S_{X2}^2 - 2\beta_1\rho_{XY2}S_{X2}S_{Y2}\right)\left[\frac{(L-1)}{n} - \frac{\left(L'-1\right)}{n'}\right] < 0.$$

Case (iv): From equations (12) and (20), we get
$MSE\left(T_{2reg}^{**}\right) - MSE\left(T_{2reg}^{*}\right) < 0$ if

$$\rho_{XY}^2f^*S_Y^2 + \beta_1\beta_2f'\left(\beta_1\beta_2S_Z^2 - 2\rho_{YZ}S_YS_Z\right) + \beta_1f^*\left(\beta_1S_X^2 - 2\rho_{XY}S_XS_Y\right) < 0.$$

The cases (i), (ii), (iii) and (iv) provide the requisite conditions for which the proposed estimators T_{1ratio}^{**} and T_{2reg}^{**} would be preferred over the existing estimators $\bar{y}^*$, T_{1ratio}^{*} and T_{2reg}^{*}.

5. Empirical Study

In order to examine the behaviour of the proposed estimators, it is essential to illustrate the results, whatever obtained in the previous sections with some numerical data. Therefore, we have considered two different data sets with a view to support the theoretical results:

Data Set-1

We have used the data regarding the population I considered by Anderson(1958), p.97. Here, the head length of second son is considered as the study variable$Y$, whereas the head length of the first son and head breadth of the first son are

respectively considered as the auxiliary variables X and Z. The population details are given below:

$$N = 25, n' = 10, n = 7, \bar{X} = 185.72, \bar{Y} = 183.84, \bar{Z} = 151.12, C_X = 0.0526, C_Y = 0.0546, C_Z = 0.0488,$$

$$S_X = 9.768872, S_Y = 10.03766, S_Z = 7.374656, S_{X2} = 6.512581, S_{Y2} = 6.691776,$$

$$S_{Z2} = 4.916437, \rho_{XY} = 0.7108, \rho_{YZ} = 0.6932, \rho_{XZ} = 0.7346, \rho_{XY2} = 0.473867.$$

Table 1 shows the variance/MSE of the estimators $\bar{y}^*, T_{1ratio}^*, T_{1ratio}^{**}, T_{2reg}^*$ and T_{2reg}^{**} along with the percentage relative efficiency (PRE) of the estimators $\bar{y}^*, T_{1ratio}^*, T_{1ratio}^{**}, T_{2reg}^*$ and T_{2reg}^{**} with respect to the usual mean estimator $\bar{y}^*$.

Table 1: Variance/MSE and PRE of $\bar{y}^*, T_{1ratio}^*, T_{1ratio}^{**}, T_{2reg}^*, T_{2reg}^{**}$

W_2	$L = L'$	$Variance/MSE$					PRE				
		$\bar{y}^*$	T_{1ratio}^*	T_{1ratio}^{**}	T_{2reg}^*	T_{2reg}^{**}	$\bar{y}^*$	T_{1ratio}^*	T_{1ratio}^{**}	T_{2reg}^*	T_{2reg}^{**}
0.2	1.5	11.00	9.04	6.44	8.79	6.06	100.00	121.72	170.91	125.19	181.54
	2	11.64	9.68	7.08	9.40	6.67	100.00	120.25	164.43	123.91	174.60
	2.5	12.28	10.32	7.72	10.00	7.28	100.00	118.96	159.03	122.78	168.81
	3	12.92	10.97	8.37	10.61	7.88	100.00	117.83	154.46	121.78	163.92

Data Set-2

Here, the data considered by Singh(1967) have been used to demonstrate the theoretical results. In this data set, number of females employed is considered to be the study variable Y, whereas the number of females in service and number of educated females are considered to be the auxiliary variables X and Z respectively. The characteristics of the population are given below:

$$N = 61, n' = 20, n = 10, \bar{X} = 5.31, \bar{Y} = 7.46, \bar{Z} = 179.00, C_X = 0.7587, C_Y = 0.7103,$$

$$C_Z = 0.2515, S_X = 4.028697, S_Y = 5.298838, S_Z = 45.0185, S_{X2} = 2.685798, S_{Y2} = 3.532559,$$

$$S_{Z2} = 30.01233, \rho_{XY} = 0.7737, \rho_{YZ} = -0.207, \rho_{XZ} = -0.0033, \rho_{XY2} = 0.5158.$$

Table 2 also shows the variance/MSE of the estimators $\bar{y}^*, T_{1ratio}^*, T_{1ratio}^{**}, T_{2reg}^*$ and T_{2reg}^{**} along with the PRE of the estimators $\bar{y}^*, T_{1ratio}^*, T_{1ratio}^{**}, T_{2reg}^*$ and T_{2reg}^{**} with respect to the usual mean estimator $\bar{y}^*$.

Table 2: Variance/MSE and PRE of $\bar{y}^*, T_{1ratio}^*, T_{1ratio}^{**}, T_{2reg}^*, T_{2reg}^{**}$

W_2	$L = L'$	$Variance/MSE$					PRE				
		$\bar{y}^*$	T_{1ratio}^*	T_{1ratio}^{**}	T_{2reg}^*	T_{2reg}^{**}	$\bar{y}^*$	T_{1ratio}^*	T_{1ratio}^{**}	T_{2reg}^*	T_{2reg}^{**}
0.2	1.5	2.472	2.425	2.013	1.619	1.618	100.00	101.96	122.84	152.66	152.76
	2	2.597	2.552	2.140	1.732	1.731	100.00	101.77	121.37	149.96	150.05
	2.5	2.722	2.679	2.267	1.844	1.843	100.00	101.59	120.06	147.60	147.68
	3	2.847	2.806	2.394	1.956	1.955	100.00	101.43	118.89	145.50	145.57

6. Concluding Remarks

We have proposed some improved ratio and regression-type estimators of the population mean using the information on two auxiliary variables in double (two-phase) sampling scheme under non-response. A theoretical study on the properties of the proposed estimators has been presented in detail. A comparative study of the proposed estimators with the usual mean estimator and some other existing estimators of the population mean has also been presented. To support the theoretical results, we have presented an empirical study by considering two different real data sets. From the Tables 1 and 2, it is revealed that the proposed estimators T_{1ratio}^{**} and T_{2reg}^{**} provide much better estimates as compared to the existing estimators $\bar{y}^*, T_{1ratio}^*$ and T_{2reg}^*.

References

1.　Anderson, T. W. (1958). *An introduction to multivariate statistical analysis.* John Wiley & Sons, New York.

2.　Chand, L. (1975). Some ratio-type estimators based on two or more auxiliary variables. *Unpublished Ph. D. Thesis*.

3.　Chaudhary, M. K. and Kumar, A. (2016). Estimation of mean of finite population using double sampling scheme under non-response. *Journal of Statistics Application and Probability (JSAP)*, 5(2):287–297.

4.　Chaudhary, M. K., P. A. S. R. and Smarandache, R. (2014). Two-phase sampling in estimation of population mean in the presence of non-response. the efficient use of supplementary information in finite population sampling. *Education Publishing 1313 Chesapeake Avenue Columbus, Ohio 43212, USA*, pages 28–41.

5.　Cochran, W. G. (1977). *Sampling Techniques. 3rd ed.,*. John Wiley & Sons, New York, USA.

6.　Hansen, M. H. and Hurwitz, W. N. (1946). The problem of non-response in sample surveys. journal of the american statistical association. *Journal of Statistics Applications and Probability Letters*, 41:517–529.

7.　Kadilar, C. and Cingi, H. (2005). A new estimator using two auxiliary variables. *Applied Mathematics and Computation*, 162(2):901–908.

8.　Khare, B. B. and Sinha, R. R. (2019). Estimation of product of two population means by multi-auxiliary characters under double sampling the non-respondents. *Statistics in Transition New Series*, 20(3):81–95.

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

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

11.　Kiregyera, B. (1980). Achain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika*, 31:215–226.

12.　Kumar, S. and Sharma, V. (2020). Improved chain ratio-product type estimators under double sampling scheme. *Journal of Statistics Applications and Probability Letters*, 7(2):87–96.

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

14.　Olkin, I. (1958). Multivariate ratio estimation for finite populations. *Biometrika*, 45:154–165.

15.　Olkin, I. (1986). Ratio estimation with sub sampling the non-respondents. *Survey Methodology*, 12:217–230.

16.　Pal, S. K. and Singh, H. P. (2018). Estimation of mean with two-parameter ratio-product-ratio estimator in double sampling using auxiliary information under non-response. *Journal of Modern Applied Statistical Methods*, 17(2):02–32.

17.　Rao and Mudholkar (1967). Generalised multi-variate estimations for the mean of finite populations. *Journal of the American Statistical Association*, 62:1008–1012.

18.　Sahai, A., C. R. . M. A. K. (1980). An alternative multivariate product estimator. *Annals of Mathematical Statistics*, 32(2):06–12.

19.　Shukla, D., P. S. and Thakur, N. S. (2012). Estimation of population mean using two auxiliary sources in sample surveys. *Statistics in Transition New Series*, 13(1):21–36.

20.　Singh, H. P. and Kumar, S. (2007). Estimation of the population mean using difference cum ratio estimator with full response on the auxiliary character. *Res. J. Applied Sci.*, 2:769–772.

21.　Singh, H. P. and Kumar, S. (2009). A general procedure of estimating the population mean in the presence of non-response under double sampling using auxiliary information. *SORT*, 33(1):71–84.

22.　Singh, M. P. (1967). Multivariate product method of estimation for finite populations. *Philosophical Transactions of the Royal Society*, 19(2):01–10.

23.　Srivastava, S. K. (1966). A note on olkin's multivariate ratio estimator. *Journal of the Indian Statistical Association*, 4:202–208.

24.　Tabasum, R. and Khan, I. A. (2004). Double sampling for ratio estimation with non-response. *J. Ind. Soc. Agril. Statist.*, 58:300–306.

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