An Improvement in Estimation of Population Mean using Two Auxiliary Variables and Two- Phase Sampling Scheme under Non-Response

Manoj K. Chaudhary[1] and Amit Kumar[2,*]

[1]*Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India*
[2]*Department of Statistics, University of Allahabad, Prayagraj, India*
E-mail: ritamanoj15@gmail.com; amit.stat40@gmail.com
*Corresponding Author

Abstract

In the present paper, we have proposed some improved ratio and regression-type estimators of the finite population mean utilizing the information on two auxiliary variables in the presence of non-response. The two-phase sampling scheme has been used to accomplish the job of estimating the desired parameter. The expressions for the basic properties such as bias and mean square error (MSE) of the proposed estimators have been derived up to the first order of approximation. A comparative study of the proposed estimators with some existing estimators has also been carried out through a real data set.

1 Introduction

There are several practical situations in which the survey (study) variable is highly correlated with more than one auxiliary variable. In such situations, the information on the auxiliary variables can be utilized to provide the improved estimates of the population parameters. Olkin (1958) proposed a multivariate ratio estimator utilizing the information on a number of auxiliary variables. Further, it has been extended by Srivastava (1966). Chand (1975) and Kiregyera (1980) have suggested some ratio-type estimators to resolve the problem of estimating the population mean with the help of two or more auxiliary variables.

The problem of non-response is very common in most of the sample surveys. The non-response arises from the fact that the surveyor fails to obtain the information from the respondents. It is of the great importance to deal with the problem of non-response through some innovative ideas or techniques. Hansen and Hurwitz (1946) resolved the problem of non-response by pioneering out a technique of sub-sampling of non-respondents. Besides it, there are several authors who have contributed a lot to deal with the problem of non-response. The authors such as Khare and Srivastava (1995), Singh and Kumar (2009), Singh et al. (2010), Shabbir and Nasir (2013), Verma et al. (2014) and Chaudhary and Kumar (2016) have tackled the problem of non-response in the case of two-phase sampling scheme.

In the subsequent sections, we have suggested some improved ratio and regression-type estimators of the finite population mean for the survey variable under non-response. We have utilized the information on two auxiliary variables with unknown means under the situation in which the first auxiliary variable is suffering from non-response, whereas the second (additional) auxiliary variable is free from the non-response. The basic properties of the proposed estimators have been discussed in detail. An empirical study has also been carried out to compare the efficiency of the suggested estimators with that of some existing estimators.

2 Sampling Strategy

Let $U = \{1, 2, \ldots, N\}$ be a population of N units. Let Y be the study variable with population mean $\overline{Y}$. Let X and Z be the auxiliary variables with respective population means $\overline{X}$ and $\overline{Z}$. Let y_i, x_i and z_i be the observations on the ith unit for the variables Y, X and Z respectively. The aim of the present work is to estimate the population mean $\overline{Y}$ using the information on

two auxiliary variables X and Z with unknown means $\overline{X}$ and $\overline{Z}$ under the assumption that the non-response occurs on study variable Y and auxiliary variable X whereas the auxiliary variable Z is free from the non-response. Adopting the concrete idea of two-phase (or double) sampling scheme, we select a first-phase sample of n' units using simple random sampling without replacement (SRSWOR) scheme and collect the information on the auxiliary variables X and Z. Now, a second-phase sample of n units is selected from the first-phase sample of n' units by the method of SRSWOR and the information is collected on Y. Out of n' units at the first-phase, it is noted that there are n'_1 responding units and n'_2 non-responding units for the auxiliary variable X. Further, we select a sub-sample of $h'_2(h'_2 = n'_2/L'; L' > 1)$ units from n'_2 non-responding units and collect the information on all h'_2 units. Following Hansen and Hurwitz (1946), the unbiased estimator of $\overline{X}$ at the first-phase is given by

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

The unbiased estimator of $\overline{Z}$ at the first-phase is given as

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

where $\overline{x}'_{n1}$ and $\overline{x}'_{h2}$ are the means based on n'_1 responding units and h'_2 non-responding units respectively for the auxiliary variable X.

The expressions for the variance of the estimators $\overline{x}'^*$ and $\overline{z}'$ are respectively given by

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

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

where S_X^2 and S_Z^2 are the population mean squares of the entire group for the auxiliary variables X and Z respectively. S_{X2}^2 is the population mean square of the non-response group for the auxiliary variable X. W_2 is the non-response rate in the population.

At the second-phase, n_1 units do respond and n_2 units do not respond on the study variable Y and auxiliary variable X such that $n_1 + n_2 = n$. A sub-sample of $h_2(h_2 = n_2/L; L > 1)$ units is now selected from the n_2

non-responding units and the information is collected from all h_2 units on the variables Y and X. Thus, the unbiased estimators of $\overline{Y}$, $\overline{X}$ and $\overline{Z}$ at the second-phase are respectively given as [See Hansen and Hurwitz (1946)]

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

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

$$\overline{z}^* = \frac{n_1 \overline{z}_{n1} + n_2 \overline{z}_{h2}}{n} \tag{7}$$

where $\overline{y}_{n1}$, $\overline{x}_{n1}$ and $\overline{z}_{n1}$ are the means based on n_1 responding units for the variables Y, X and Z respectively. $\overline{y}_{h2}$, $\overline{x}_{h2}$ and $\overline{z}_{h2}$ are respectively the means based on the sub-sample of h_2 non-responding units for the variables Y, X and Z.

Note: The functional form of the estimator $\overline{z}^*$ is similar to that of the estimators $\overline{y}^*$ and $\overline{x}^*$. The concept behind it is only to get the similar estimator, however the non-response does not occur on the auxiliary variable Z.

The expressions for the variance of the estimators $\overline{y}^*$, $\overline{x}^*$ and $\overline{z}^*$ are respectively given by

$$V(\overline{y}^*) = \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2 + \frac{(L-1)}{n} W_2 S_{Y2}^2 \tag{8}$$

$$V(\overline{x}^*) = \left(\frac{1}{n} - \frac{1}{N} \right) S_X^2 + \frac{(L-1)}{n} W_2 S_{X2}^2 \tag{9}$$

$$V(\overline{z}^*) = \left(\frac{1}{n} - \frac{1}{N} \right) S_Z^2 + \frac{(L-1)}{n} W_2 S_{Z2}^2 \tag{10}$$

where S_Y^2 and S_{Y2}^2 are respectively the population mean squares of the entire group and non-response group for the study variable Y. S_{Z2}^2 is the population mean square of the non-response group for the auxiliary variable Z.

Chaudhary and Kumar (2016) have suggested some ratio and regression-type estimators of the finite population mean $\overline{Y}$ utilizing the information on the auxiliary variable X with unknown mean under the situation in which both the variables Y and X are suffering from non-response as follows:

$$T_1^* = \frac{\overline{y}^*}{\overline{x}^*} \overline{x}'^* \tag{11}$$

$$T_2^* = \overline{y}^* + b^* (\overline{x}'^* - \overline{x}^*) \tag{12}$$

where $b^* = s^*_{xy}/s^{2*}_x$. s^*_{xy} and s^{*2}_x are the unbiased estimators based on $(n_1 + h_2)$ units for S_{XY} and S^2_X respectively. $S_{XY} = \rho_{XY}S_XS_Y$. ρ_{XY} is the population correlation coefficient between Y and X.

The expressions for the mean square error (MSE) of the estimators T^*_1 and T^*_2 up to the first order of approximation are respectively given as

$$MSE(T^*_1) = \left(\frac{1}{n'} - \frac{1}{N}\right)S^2_Y + \left(\frac{1}{n} - \frac{1}{n'}\right)(S^2_Y + R^2S^2_X - 2\rho_{XY}RS_XS_Y)$$

$$+ \frac{(L-1)}{n}W_2S^2_{Y2}$$

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

$$\tag{13}$$

$$MSE(T^*_2) = \left[\left(\frac{1}{n'} - \frac{1}{N}\right) + \left(\frac{1}{n} - \frac{1}{n'}\right)(1 - \rho^2_{XY})\right]S^2_Y$$

$$+ \frac{(L-1)}{n}W_2S^2_{Y2}$$

$$+ W_2(\beta^2_1S^2_{X2} + 2\beta_1\rho_{XY2}S_{X2}S_{Y2})\left[\frac{(L-1)}{n} - \frac{(L'-1)}{n'}\right]$$

$$\tag{14}$$

where β_1 is the population regression coefficient of Y on X and ρ_{XY2} is the population correlation coefficient between Y and X of the non-response group. $R = \overline{Y}/\overline{X}$.

3 Proposed Estimators

We now propose some improved ratio and regression-type estimators of the finite population mean $\overline{Y}$ utilizing the information on two auxiliary variables X and Z with unknown means under the given situation of non-response as

$$T'^*_1 = \frac{\overline{y}^*}{\overline{x}^*}\overline{x}'^*_R \tag{15}$$

$$T'^*_2 = \overline{y}^* + b^*\left(\overline{x}'^*_{lr} - \overline{x}^*\right) \tag{16}$$

where $\overline{x}'^*_R = \frac{\overline{x}'^*}{\overline{z}^*}\overline{z}'$, $\overline{x}'^*_{lr} = \overrightarrow{x}'^* + b'^*(\overline{z}' - \overline{z}^*)$ and $b'^* = \frac{s'^*_{xz}}{s^{2'*}_z}$. s'^*_{xz} and s'^{*2}_z are respectively the unbiased estimators based on $(n'_1 + h'_2)$ units for S_{XZ}

and S_Z^2. $S_{XZ} = \rho_{XZ} S_X S_Z$. ρ_{XZ} is the population correlation coefficient between X and Z.

To obtain the biases and mean square errors of the proposed estimators $T_1'^*$ and $T_2'^*$, we use the concept of large sample approximations. Let

$$\bar{y}^* = \overline{Y}(1 + e_0), \bar{x}^* = \overline{X}(1 + e_1), \bar{z}^* = \overline{Z}(1 + e_2), \bar{x}'^* = \overline{X}(1 + e_1'),$$

$$\bar{z}' = \overline{Z}(1 + e_2'), s_{xy}^* = S_{XY}(1 + e_3),$$

$$s_x^{*2} = S_X^2(1 + e_4), s_{xz}' = S_{XY}(1 + e_5), s_z'^2 = S_Z^2(1 + e_6)$$

such that $E(e_0) = E(e_1) = E(e_2) = E(e_1') = E(e_2') = E(e_3) = E(e_4) = E(e_5) = E(e_6) = 0$,

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

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

$$E(e_2^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C_Z^2 + \frac{(L-1)}{n} W_2 \frac{S_{Z2}^2}{\overline{Z}^2}, E\left(e_1'^2\right) = E\left(e_1 e_1'\right)$$

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

$$E(e_2'^2) = E\left(e_2 e_2'\right) = \left(\frac{1}{n'} - \frac{1}{N}\right) C_Z^2, E\left(e_0 e_1\right)$$

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

$$E(e_0 e_2) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{YZ} C_Y C_Z + \frac{(L-1)}{n} W_2 \rho_{YZ2} \frac{S_{Y2}}{\overline{Y}} \frac{S_{Z2}}{\overline{Z}}, E\left(e_0 e_1'\right)$$

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

$$E(e_0 e_2') = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{YZ} C_Y C_Z, E\left(e_1 e_2\right)$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{XZ} C_X C_Z + \frac{(L-1)}{n} W_2 \rho_{XZ2} \frac{S_{X2}}{\overline{X}} \frac{S_{Z2}}{\overline{Z}},$$

$$E(e_1 e_2') = E\left(e_1' e_2'\right) = \left(\frac{1}{n'} - \frac{1}{N}\right)\rho_{XZ}C_X C_Z, E\left(e_1 e_3\right)$$

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

$$E(e_1 e_4) = \frac{N(N-n)}{(N-1)(N-2)}\frac{\mu_{30}}{n\bar{X}S_X^2} + W_2\frac{(L-1)}{n}\frac{\mu_{30(2)}}{\bar{X}S_X^2}, E\left(e_1 e_5\right)$$

$$= \frac{N(N-n)}{(N-1)(N-2)}\frac{\mu'_{21}}{n\bar{X}S_{XZ}} + W_2\frac{(L-1)}{n}\frac{\mu'_{21(2)}}{\bar{X}S_{XZ}},$$

$$E(e_2 e_1') = \left(\frac{1}{n'} - \frac{1}{N}\right)\rho_{XZ}C_X C_Z + \frac{(L'-1)}{n'}W_2\rho_{XZ2}\frac{S_{X2}}{\bar{X}}\frac{S_{Z2}}{\bar{Z}},$$

$$E(e_2 e_5) = E\left(e_2 e_4\right) = E\left(e_2 e_3\right) = E\left(e_1 e_5\right)$$

$$= \frac{N(N-n)}{(N-1)(N-2)}\frac{\mu'_{21}}{n\bar{Z}S_{XZ}} + W_2\frac{(L-1)}{n}\frac{\mu'_{21(2)}}{\bar{Z}S_{XZ}},$$

$$E(e_2 e_6) = \frac{N(N-n)}{(N-1)(N-2)}\frac{\mu'_{30}}{n\bar{Z}S_Z^2} + W_2\frac{(L-1)}{n}\frac{\mu'_{30(2)}}{\bar{Z}S_Z^2},$$

$$E(e_1' e_3) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{21}}{n'\bar{X}S_{XY}} + W_2\frac{(L'-1)}{n'}\frac{\mu_{21(2)}}{\bar{X}S_{XY}},$$

$$E(e_1' e_4) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{30}}{n'\bar{X}S_X^2} + W_2\frac{(L'-1)}{n'}\frac{\mu_{30(2)}}{\bar{X}S_X^2},$$

$$E(e_2' e_3) = E\left(e_2' e_4\right) = E\left(e_2' e_5\right) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu'_{30}}{n'\bar{Z}S_Z^2},$$

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \bar{X}\right)^r\left(y_i - \bar{Y}\right)^s,$$

$$\mu_{rs(2)} = \frac{1}{N_2}\sum_{i}^{N_2}\left(x_i - \bar{X}_{(2)}\right)^r\left(y_i - \bar{Y}_{(2)}\right)^s,$$

$$\mu'_{rs} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i - \overline{X}\right)^r\left(z_i - \overline{Z}\right)^s,$$

$$\mu'_{rs(2)} = \frac{1}{N_2} \sum_{i=1}^{N_2} \left(x_i - \overline{X}_{(2)}\right)^r \left(z_i - \overline{Z}_{(2)}\right)^s,$$

$$\overline{Y}_{(2)} = \frac{1}{N_2} \sum_i^{N_2} y_i, \overline{X}_{(2)} = \frac{1}{N_2} \sum_i^{N_2} x_i, \overline{Z}_{(2)} = \frac{1}{N_2} \sum_i^{N_2} z_i,$$

$$C_Y = \frac{S_Y}{\overline{Y}}, C_X = \frac{S_X}{\overline{X}} \quad \text{and} \quad C_Z = \frac{S_Z}{\overline{Z}}$$

where ρ_{YZ} is the population correlation coefficient between Y and Z. ρ_{YX2} and ρ_{XZ2} are the population correlation coefficients between Y ,X and X, Z respectively for the non-response group. N_1 and N_2 are respectively the numbers of responding and non-responding units in the population such that $N = N_1 + N_2$.

Expressing the equation (15) in terms of e_0, e_1, e'_1, e_2, e'_2 and then removing the terms having powers of e_0, e_1, e'_1, e_2, e'_2 greater than two, we get

$$T'^*_1 - \overline{Y} = \overline{Y}(e_0 - e_1 - e_2 + e'_1 + e'_2 + e_0^2 + e_1^2 + e_1'^2 + e_2^2 + e_2'^2 - e_0 e_1$$
$$- e_0 e_2 + e_0 e'_1 + e_0 e'_2 + e_1 e_2 - e'_1 e_2 - e'_1 e_1 + e'_1 e'_2 - e_1 e'_2$$
$$- e'_2 e_2) \tag{17}$$

Now, taking expectation on both sides of the equation (17), we find the expression for the bias of T'^*_1 up to the first order of approximation as

$$Bias(T'^*_1) = \overline{Y} \left[\left(f + f'\right) C_X^2 + f C_Z^2 + \left(f - f'\right) \right.$$
$$\left(-\rho_{XY} C_X C_Y - \rho_{YZ} C_Y C_Z + \rho_{XZ} C_X C_Z\right)$$
$$- f' \rho_{XZ} C_X C_Z + \frac{(L-1)}{n} W_2 \frac{S_{Z2}^2}{\overline{Z}^2}$$
$$- \frac{(L-1)}{n} W_2 \rho_{YZ2} \frac{S_{Y2} S_{Z2}}{\overline{Y}\overline{Z}} - \frac{(L'-1)}{n'} W_2 \rho_{XZ2} \frac{S_{X2} S_{Z2}}{\overline{X}\overline{Z}}$$
$$+ \left(\frac{S_{X2}^2}{\overline{X}^2} - \rho_{XY2} \frac{S_{X2} S_{Y2}}{\overline{X}\overline{Y}} + \rho_{XZ2} \frac{S_{X2} S_{Z2}}{\overline{X}\overline{Z}} \right)$$
$$W_2 \left[\frac{(L-1)}{n} - \frac{(L'-1)}{n'} \right] \right] \tag{18}$$

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

Squaring both sides of the equation (17) and then taking expectation on avoiding the terms having powers of e_0, e_1, e'_1, e_2, e'_2 higher than two, we get

$$
E\left[T_1'^* - \overline{Y}\right]^2 = \overline{Y}^2[E(e_0^2) + E(e_1^2) + E(e_2^2) + E(e_1'^2) + E(e_2'^2)
$$
$$
- 2E(e_0 e_1) - 2E(e_0 e_2) + 2E(e_0 e_1') + 2E(e_0 e_2')
$$
$$
+ 2E(e_1 e_2) - 2E(e_1' e_2) - 2E(e_1' e_1) - 2E(e_1 e_2')
$$
$$
- 2E(e_2' e_2) + 2E(e_1' e_2')]
$$

Thus, the expression for the MSE of $T_1'^*$ up to the first order of approximation is given as

$$
MSE(T_1'^*) = \bar{Y}^2\left[\left(\frac{1}{n} - \frac{1}{N}\right)C_Y^2 + \left(\frac{1}{n} - \frac{1}{n'}\right)\right.
$$
$$
\left(C_X^2 + C_Z^2 - 2\rho_{XY}C_X C_Y - 2\rho_{YZ}C_Y C_Z\right.
$$
$$
+ 2\rho_{XZ}C_X C_Z) + W_2\frac{(L-1)}{n}
$$
$$
\left(\frac{S_{Y2}^2}{\bar{Y}^2} + \frac{S_{Z2}^2}{\bar{Z}^2} - 2\rho_{YZ2}\frac{S_{Y2}S_{Z2}}{\bar{Y}\bar{Z}}\right)
$$
$$
+ W_2\left(\frac{S_{X2}^2}{\bar{X}^2} - 2\rho_{XY2}\frac{S_{X2}S_{Y2}}{\bar{X}\bar{Y}} + 2\rho_{XZ2}\frac{S_{X2}S_{Z2}}{\bar{X}\bar{Z}}\right)
$$
$$
\left.\left[\frac{(L-1)}{n} - \frac{(L'-1)}{n'}\right]\right] \tag{19}
$$

Now, expressing the Equation (16) in terms of e_0, e_1, e'_1, e_2, e'_2, e_3, e_4, e_5, e_6 and then neglecting the terms involving powers of e_0, e_1, e'_1, e_2, e'_2 e_3, e_4, e_5, e_6 greater than two, we get

$$
T_2'^* - \overline{Y} = \overline{Y}e_0 + \beta_1\bar{X}(e_1' - e_1'e_4 + e_1'e_3 - e_1 + e_1 e_4 - e_1 e_3)
$$
$$
+ \beta_1\beta_2\bar{Z}(e_2' - e_2 - e_2'e_6 + e_2 e_6 - e_4 e_2' - e_4 e_2
$$
$$
+ e_3 e_2' - e_3 e_2 + e_2' e_5 - e_2 e_5) \tag{20}
$$

where β_2 is the population regression coefficient of X on Z.

The expression for the bias of $T_2'^*$ up to the first order of approximation can be obtained on taking expectation of both sides of the equation (20). Thus,

we have

$$Bias(T_2'^*)$$

$$= \beta_1 \left\{ \left(\frac{N(N-n')}{n'(N-1)(N-2)} - \frac{N(N-n)}{n(N-1)(N-2)} \right) \left(\frac{\mu_{21}}{S_{YX}} - \frac{\mu_{30}}{S_X^2} \right) \right.$$

$$- W_2 \left(\frac{\mu_{21(2)}}{S_{YX}} - \frac{\mu_{30(2)}}{S_X^2} \right) \left(\frac{(L-1)}{n} - \frac{(L'-1)}{n'} \right) \right\}$$

$$+ \beta_1 \beta_2 \left\{ \left(\frac{N(N-n')}{n'(N-1)(N-2)} \right) \left(\frac{\mu_{21}'}{S_{XZ}} - \frac{\mu_{30}'}{S_Z^2} \right) \right.$$

$$- \left(\frac{N(N-n)}{n(N-1)(N-2)} \right) \left(\frac{3\mu_{21}'}{S_{XZ}} - \frac{\mu_{30}'}{S_Z^2} \right)$$

$$\left. - W_2 \left(\frac{L-1}{n} \right) \left(\frac{3\mu_{21(2)}'}{S_{XZ}} - \frac{\mu_{30(2)}'}{S_Z^2} \right) \right\} \tag{21}$$

Squaring both sides of the equation (20) and then captivating expectation on neglecting the terms having powers of e_0, e_1, e_1', e_2, e_2' e_3, e_4, e_5, e_6 larger than two, we get

$$E(T_2'^* - \overline{Y})^2 = \overline{Y}^2 E(e_0^2) + \beta_1^2 \bar{X}^2 \left(E(e_1'^2) + E(e_1^2) - 2E(e_1 e_1') \right)$$

$$+ \beta_1^2 \beta_2^2 \bar{Z}^2 (E(e_2'^2) + E(e_2^2) - 2E(e_2 e_2')) + 2\bar{Y}\bar{X}\beta_1$$

$$\times (E(e_0 e_1') - E(e_0 e_1)) + 2\beta_1 \beta_2 \bar{Y}\bar{Z}(E(e_0 e_2') - E(e_0 e_2))$$

$$+ 2\beta_1^2 \beta_2 \bar{X}\bar{Z}(E(e_1' e_2') - E(e_1' e_2) - E(e_1 e_2') + E(e_1 e_2))$$

Therefore, the expression for the MSE of $T_2'^*$ up to the first order of approximation is given by

$$MSE\left(T_2'^*\right) = \left(\frac{1}{n} - \frac{1}{N} \right) S_Y^2 + \left(\frac{1}{n} - \frac{1}{n'} \right)$$

$$\times (\beta_1{}^2 S_X^2 + \beta_1^2 \beta_2^2 S_Z^2 - 2\beta_1 \rho_{XY} S_X S_Y$$

$$- 2\beta_1 \beta_2 \rho_{YZ} S_Y S_Z - 2\beta_1^2 \beta_2 \rho_{XZ} S_X S_Z)$$

$$+ \frac{(L-1)}{n} W_2 \left(S_{Y2}^2 + \beta_1^2 \beta_2^2 S_{Z2}^2 - 2\beta_1 \beta_2 \rho_{YZ2} S_{Y2} S_{Z2} \right)$$

$$+ W_2 \left({\beta_1}^2 S_{X2}^2 - 2\beta_1 \rho_{XY2} S_{X2} S_{Y2} + 2\beta_1^2 \beta_2 \rho_{XZ2} S_{X2} S_{Z2} \right)$$

$$\times \left[\frac{(L-1)}{n} - \frac{(L'-1)}{n'} \right] \tag{22}$$

4 Empirical Study

Here, the data considered by Khare and Sinha (2007) have been used to demonstrate the theoretical results. The data relate to the physical growth of upper socioeconomic group of 95 school children of Varanasi district under an ICMR study, Department of Pediatrics, B.H.U., during 1983–84. In this data set, we have $Y =$ Weight (in kg.) of the children, $X =$ Skull circumference (in cm.) of the children and $Z =$ Chest circumference (in cm.) of the children. The first 25% units have been considered as the non-responding units for all the variables Y, X and Z. The details are given below:

$$N = 95, \quad n' = 70, \quad n = 35, \bar{Y} = 19.4968, \quad \bar{X} = 51.1726$$

$$\bar{Z} = 55.8611, \quad C_Y = 0.15613, \quad C_X = 0.03006,$$

$$C_Z = 0.05860, \quad S_{Y2} = 2.3542, \quad S_{X2} = 1.2681, \quad S_{Z2} = 3.0176,$$

$$\rho_{YX} = 0.328, \quad \rho_{YX2} = 0.477,$$

$$\rho_{YZ} = 0.846, \quad \rho_{YZ2} = 0.729, \quad \rho_{XZ} = 0.297,$$

$$\rho_{XZ2} = 0.570, \quad W_2 = 0.25.$$

Table 1 shows the variance/MSE of the estimators $\bar{y}^*$, T_1^*, T_2^*, $T_1'^*$ and $T_2'^*$ along with their percentage relative efficiency (PRE). The PRE is computed with respect to $\bar{y}^*$.

Table 1   Variance/MSE and PRE of $\bar{y}^*$, T_1^*, T_2^*, $T_1'^*$, $T_2'^*$

	Variance/MSE					PRE				
$L = L'$	$\bar{y}^*$	T_1^*	$T_1'^*$	T_2^*	$T_2'^*$	$\bar{y}^*$	T_1^*	$T_1'^*$	T_2^*	$T_2'^*$
1.5	0.187	0.171	0.150	0.146	0.106	100.00	109.57	124.67	127.65	176.36
2.0	0.207	0.188	0.165	0.165	0.116	100.00	109.78	125.27	125.51	177.66
2.5	0.226	0.206	0.180	0.183	0.127	100.00	109.96	125.77	123.79	178.74
3.0	0.246	0.224	0.195	0.201	0.137	100.00	110.10	126.19	122.39	179.65

From the above table, it is revealed that the MSE of the proposed ratio-type estimator $T_1'^*$ is much smaller than that of the existing ratio-type estimator T_1^* and hence the PRE of $T_1'^*$ is higher than that of T_1^*. The same pattern of results is seen in the case of the proposed regression-type estimator $T_2'^*$ and the existing regression-type estimator T_2^*. It is also revealed that the proposed estimators $T_1'^*$ and $T_2'^*$ perform very well than the usual mean estimator $\bar{y}^*$.

5 Concluding Remarks

We have proposed some improved ratio and regression-type estimators for estimating the finite population mean utilizing the information on two auxiliary variables in the presence of non-response. We have discussed the situation in which the knowledge about the population means of both auxiliary variables is not available and hence we have adopted the concrete idea of two-phase sampling scheme to get the desired information. A theoretical study along with the basic properties of the suggested estimators has been presented. To strengthen the theoretical results, we have carried out an empirical study by considering a natural data set. In Table 1, we have presented a comparative study of the proposed estimators with the usual mean estimator and some other existing estimators. The Table 1 shows that the proposed ratio and regression-type estimators $T_1'^*$ and $T_2'^*$ provide better estimates as compared to the usual mean estimator $\bar{y}^*$ and existing ratio and regression-type estimators T_1^* and T_2^*. Thus, the proposed estimators may be preferred over the usual mean estimator and existing estimators.

References

[1] Chand, L. (1975). Some ratio-type estimators based on two or more auxiliary variables, Unpublished Ph. D. Thesis, Iowa State University, Ames, Iowa, U.S.A.

[2] 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), Vol. 5, No. 2, 287–297.

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

[4] Khare, B. B. and Sinha, R. R. (2007). Estimation of the ratio of the two population means using multi auxiliary characteristics in the presence of non-response, In Statistical Techniques in Life Testing, Reliability, Sampling Theory and Quality Control, 163–171.

[5] 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, Nat. Acad. Sci. Lett. India, 65, 195–203.

[6] Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables, Metrika, 31, 215–226.

[7] Olkin, I. (1958). Multivariate ratio estimation for finite populations, Biometrika, 45, 154–165.

[8] Shabbir, J. and Nasir, S. (2013). On estimating the finite population mean using two auxiliary variables in two phase sampling in the presence of non response, Communication in Statistics-Theory and Methods, 42, 4127–4145.

[9] 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(2.1), 71–84.

[10] Singh, H. P., Kumar, S. and Kozak, M. (2010). Improved estimation of finite population mean using sub-sampling to deal with non response in two phase sampling scheme, Communication in Statistics-Theory and Method, 39(5), 791–802.

[11] Srivastava, S. K. (1966). A note on Olkin's multivariate ratio estimator, Journal of the Indian Statistical Association, 4, 202–208.

[12] Verma, H. K., Sharma, P. and Singh, R. (2014). Some ratio cum product type estimators for population mean under double sampling in the presence of non-response, J. Stat. Appl. Pro. 3, No. 3, 379–385.

Biographies

Manoj K. Chaudhary is a Professor (Associate) in the Department of Statistics at Banaras Hindu University, Varanasi, India. He has made significant contributions to the survey methodology and methods of estimation. He has published more than 45 research papers in internationally and nationally renowned journals.

Amit Kumar is a Guest Faculty in the Department of Statistics at University of Allahabad, Prayagraj, India. He has made significant contributions to the survey methodology and methods of estimation. He has published more than 8 research papers in internationally and nationally renowned journals.