Asian Journal of Statistical Sciences
Vol. 5, No-2, (2025), pp. 115-131

ARF INDIA
Academic Open Access Publishing
Journal homepage link:
http://arfjournals.com/ajss

Enhancing the Efficiency of the Estimators of Population Mean Using Double Sampling Technique in the Presence Of Nonresponse

[a]Bassey, Mbuotidem Okon, [b]Iseh, Matthew Joshua and [c]Enang, Ekaette Inyang

[a,b] *Department of Statistics, Akwa Ibom State University, Mkpat Enin, Nigeria*
[c]*Department of Statistics, University of Calabar, Calabar*

ARTICLE HISTORY
Compiled July 13, 2026

Received: 02 July 2024; Accepted: 13 April 2025

ABSTRACT
Estimating population parameters with a minimum error using survey data influenced by non-response is challenging. This has attracted the attention of researchers in this area of sampling. On the heels of the aforementioned problem, some classical estimators of the population mean utilizing the information on two auxiliary variables using a two-phase sampling scheme with sub-sampling of the non-respondents are proposed in this paper. Considered in this study is the case where the study and first auxiliary variables are affected by non-response while the supporting auxiliary variable is not affected by non-response. The expressions for the proposed estimators' bias and mean square error have been derived up to the first order of approximation. Efficiency comparison of the proposed estimator with that of the usual estimator and some well-known existing estimators of the population mean have been carried out. The theoretical results have also been validated through some empirical investigations, and it is evident from the results that the proposed estimators will be preferred in estimating population parameters from survey data with sensitive indicators that is inflicted with non-response.

KEYWORDS
Double sampling; sub-sampling; non-response; auxiliary variable

Introduction

In some practical situations, if the information on more than one auxiliary variable is readily accessible, the information on all such variables could be utilized to provide more precise estimators of the population parameter. [1] was the first to propose the multivariate ratio estimators of the population mean utilizing the information on several auxiliary variables. It was further extended by [2] and [3]. In addition, [4], [5], and [6] made useful contributions in this area. Numerous estimators of the population mean have been proposed by [7], [8], and others based on the data on two auxiliary variables.

In a sample survey, it is of great importance to deal with the problem of non-response. When the investigator fails to get information from the selected units of the population, non-response occurs. The non-response has no negative impact on the

CONTACT Bassey, Mbuotidem Okon Email: : mbuotistat@gmail.com

estimates' results if the traits of the non-responding units are strikingly comparable to those of the responding units. However, non-response becomes difficult to estimate in these circumstances since, in practice, it is always observed that responding units are not very similar to non-responding units. In order to lessen the impact of non-response error in sample surveys, [9] originally proposed the challenge of sub-sampling the non-respondents. The problem of sub-sampling the non-respondents was first suggested by [9] to reduce the effect of non-response error in sample survey. The authors obtained an unbiased estimator with the technique of sub-sampling the non-respondents, while the effect of the non-response was only noticed in the variance. [10] suggested the technique of estimating the non-response propensity. [11] and [12] extended the work by [10] to calibration technique with some outstanding results. Nonetheless, a number of authors have addressed the issue of nonresponse in great detail, either by expanding the method of sub-sampling the non-response by [9] or by suggesting different estimators of the population parameters. Some of which include [13], [14], [15], [16], [17], [18], [19], [20] and [21]. In addition, [22], [23], [24], etc, are among several authors who have discussed the problem of non-response in two-phase sampling scheme.

The problem of estimating the parameter in two-phase sampling becomes more complex when the non-response is observed on both study and auxiliary variables. This situation was first handled by [25] to obtain improved estimators. To account for the loss of information in both the study and auxiliary variables and in addition, obtain improved estimators of the population mean, this study seeks to utilize information on two auxiliary variables. The proposed estimators are formulated under the assumption that the additional (second) auxiliary variable is free from non-response. It is further assumed that the first
auxiliary variable has a high degree of correlation with the study variable than the second auxiliary variable. In order to validate the theoretical findings, two different sets of empirical data have been considered.

Sampling Strategy

Suppose a population comprises of N units $(U_1, U_2, \ldots, U_N)$. Let Y be the study variable with population mean $\bar{Y}$. Let X and Z be the two auxiliary variables with respective population means $\bar{X}$ and $\bar{Z}$. Let y_i, x_i and $z_i (i = 1, 2, \ldots..N)$ be observations on the i^{th} unit in the population for the variables, Y, X, and Z respectively. Assuming that non-response is observed on the study variable Y and the first auxiliary variable X whereas the support variable Z is free from the non-response. It is further assumed that the population mean $\bar{X}$ of the auxiliary variable X is unknown. In the present situation, we first estimate $\bar{X}$ adopting the concept of double sampling scheme and then use the estimate so obtained along with the information on the other support variable Z to get the estimate of $\bar{X}$. Let a larger sample of size 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 size 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 unit 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 is collected (see [9]). Thus, the estimators of $\bar{X}$ and $\bar{Z}$ at the first phase are respectively given as
$\bar{x}'^{*} = \frac{n'_1 \bar{x}'_{n_1} + n'_2 \bar{x}'_{h_2}}{n'}$
where $\bar{x}'_{n1}$ and $\bar{x}'_{h2}$ are respectively the means based on n'_1 responding units and h'_2

non-responding units for the auxiliary variable X . Also, the sample mean of the supplementary variable Z is given as $\bar{z}' = \frac{1}{n'}\sum_i^{n'} z_i$
The variance of the estimators $\bar{x}'^*$ and $\bar{z}'$ are respectively given by
$V(\bar{x}'^*) = \left(\frac{1}{n'} - \frac{1}{N}\right) S_X^2 + \frac{(L'-1)}{n} W_2 S_{X_2}^2$
$\mathrm{V}(\bar{z}') = \left(\frac{1}{n'} - \frac{1}{N}\right) S_Z^2$
where S_X^2 and $S_{X_2}^2$ are respectively the population mean squares of entire group and non-response of the auxiliary variable X. S_Z^2 is the population means square of the auxiliary variable Z and $W_2 = \frac{N_2}{N}$ represents the non-response rate in the population.

Furthermore, 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 study variable Y and auxiliary variable X , while the auxiliary variable Z is free from the non-response. Adopting the [9] technique of sub-sampling the non-respondents, we select a sub-sample of h_2 unit from n_2 non-responding unit 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 [9] unbiased estimators of $\bar{Y}$ and $\bar{X}$ at the second phase are respectively given by
$\bar{y}^* = \frac{n_1\bar{y}_{n1} + n_2\bar{y}_{h2}}{n}$ and $\bar{x}^* = \frac{n_1\bar{x}_{n1} + n_2\bar{x}_{h2}}{n}$
where $\bar{y}_{n1}$ and $\bar{y}_{h2}$ are the means based on n_1 responding units and h_2 non-responding units respectively for the study variable. $\bar{x}_{n1}$ and $\bar{x}_{h2}$ 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 unbiased estimators $\bar{y}^*$ and $\bar{x}^*$ are respectively given as:
$V(\bar{y}^*) = \left(\frac{1}{n} - \frac{1}{N}\right) S_Y^2 + \frac{(L-1)}{n} W_2 S_{Y2}^2$
$\mathrm{V}(\bar{x}^*) = \left(\frac{1}{n} - \frac{1}{N}\right) S_X^2 \frac{(L-1)}{n} W_2 S_{X2}^2$
where S_Y^2 and S_{Y2}^2 are the population mean squares of entire group and non-response respectively for the study variable.

Some Existing Estimators

- [17] proposed some ratio and regression-type estimators of population mean $\bar{Y}$ using the information on the auxiliary variables X under the condition that the non-response is observed on both study and auxiliary variables and the population mean $\bar{X}$ of the auxiliary variable X is not known. The estimators are respectively given as

$$T_{1\text{rat}}^* = \frac{\bar{y}^*}{\bar{x}^*}\bar{x}^*$$

$$T_{2\text{reg}}^* = \bar{y}^* + b^*\left(\bar{x}^* - \bar{x}^*\right)$$

where

$$b^* = \frac{s_{xy}^*}{s_x^{*2}}$$

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^{*2}_x are respectively the estimators of S_{XY} and S^2_X based on $(n_1 + h_2)$ units. Further,

$$S_{XY} = \rho_{XY} S_X S_Y,$$

and ρ_{XY} is the population correlation coefficient between Y and X.

The expression for the mean square error (MSE) of the estimators $T^*_{1\text{rat}}$ and $T^*_{2\text{reg}}$ up to the first order of approximation are respectively given as

$$\begin{aligned}\text{MSE}(T^*_{1\text{rat}}) = &\left(\frac{1}{n} - \frac{1}{N}\right) S^2_Y + \left(\frac{1}{n_1} - \frac{1}{n}\right)\left(S^2_Y + R^2 S^2_X - 2\rho_{XY} R S_X S_Y\right) \\ &+ \frac{(L-1)}{n} W_2 S^2_{Y2} + W_2\left(R^2 S^2_{X2} - 2\rho_{XY2} R S_{X2} S_{Y2}\right)\left[\frac{(L-1)}{n}\right],\end{aligned}$$

and

$$\begin{aligned}\text{MSE}(T^*_{2\text{reg}}) = &\left[\left(\frac{1}{n} - \frac{1}{N}\right) + \left(\frac{1}{n_1} - \frac{1}{n}\right)(1 - \rho^2_{XY})\right] S^2_Y \\ &+ \frac{(L-1)}{n} W_2 S^2_{Y2} + W_2\left(\beta^2_1 S^2_{X2} + 2\beta_1 \rho_{XY2} S_{X2} S_{Y2}\right)\left[\frac{(L-1)}{n} - \frac{(L-1)}{n'}\right].\end{aligned}$$

where ρ_{XY2} is the population correlation coefficient between Y and X for the non-response group and

$$R = \frac{\overline{Y}}{\overline{\overline{X}}}.$$

[18] proposed a product-ratio estimator under double sampling in the presence of non-response as follows

$$T_{ae} = \bar{y}^*_{st}\left(\frac{\bar{x}_{st} + \Phi}{\bar{x}_{st} - \Phi}\right),$$

where

$$\Phi = \sum_{i=1}^{k} C_{x_i}, \qquad \bar{y}^*_{st} = \sum_{i=1}^{k} p_i \bar{y}_i^*, \qquad \bar{y}_i^* = \frac{n_{1i}\bar{y}_{n1i} + n_{2i}\bar{y}_{n2i}}{n_i}, \qquad p_i = \frac{N_i}{N},$$

$$\bar{x}_{st} = \sum_{i=1}^{k} p_i \bar{x}_i, \qquad \bar{x}^1_{st} = \sum_{i=1}^{k} p_i \bar{x}_i^1,$$

with mean square error given as

$$
\begin{aligned}
\mathrm{MSE}(T_{ae}) = {} & b^2R^2\sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2S_{y_i}^2\right] - 2abR^2\sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2S_{x_i}^2\right] \\
& + a^2R^2\sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2S_{x_i}^2\right] - 2bR\sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\rho_iS_{x_i}S_{y_i}\right] \\
& + 2aR\sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2\rho_iS_{x_i}S_{y_i}\right] + \sum_{i=1}^{k}\left[\left(\frac{1}{n_i}-\frac{1}{N_i}\right)p_i^2S_{y_i}^2 + \frac{k_i-1}{n_i}p_i^2W_{i2}S_{y_i}^2\right].
\end{aligned}
$$

where

$$
R=\frac{\bar{Y}}{\bar{X}}, \qquad a=\frac{\bar{X}}{\bar{X}+\Phi}, \qquad b=\frac{\bar{X}}{\bar{X}-\Phi}.
$$

[25] proposed some improved estimators of population mean using two-phase sampling scheme in the presence of non-response as follows:

$$
T_{1rat}^{**}=\bar{y}^*\frac{\bar{x}_n^*}{\bar{x}^*}
$$

with

$$
\begin{aligned}
\mathrm{MSE}(T_{1rat}^{**}) = \bar{Y}^2\Bigg[& f^*\left(C_Y^2+C_X^2-2\rho_{YX}C_YC_X\right) \\
& + f'\left(C_Y^2+C_Z^2-2\rho_{YZ}C_YC_Z\right) \\
& + \frac{(L-1)}{n}W_2\frac{S_{Y2}^2}{\bar{Y}^2} \\
& + W_2\left(\frac{S_{X2}^2}{\bar{X}^2}-2\rho_{YX2}\frac{S_{X2}}{\bar{X}}\frac{S_{Y2}}{\bar{Y}}\right)\left[\frac{(L-1)}{n}-\left(\frac{L-1}{n'}\right)\right]\Bigg].
\end{aligned}
$$

where

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

$$
T_{2reg}^{**}=\bar{y}^*+b^*\left(\bar{x}_{1r}^*-\bar{x}^*\right),
$$

with

$$\begin{aligned}\operatorname{MSE}\left(T_{2reg}^{**}\right) = & f' S_Y^2 + \beta_1\beta_2 f'\left(\beta_1\beta_2 S_Z^2 - 2\rho_{YX} S_Y S_Z\right) \\ & + \beta_1 f^*\left(\beta_1 S_X^2 - 2\rho_{XY} S_Y S_X\right) \\ & + \frac{L-1}{n} W_2 \frac{S_{Y2}^2}{\bar{Y}^2} \\ & + W_2\left(\beta_1^2 S_{X2}^2 - 2\beta_1\rho_{YX2} S_{Y2} S_{X2}\right)\left[\frac{L-1}{n} - \frac{L-1}{n'}\right].\end{aligned}$$

$$\begin{aligned}\operatorname{MSE}(T_{ae1}) = & \Bigg[\left(\frac{\bar{X}}{\bar{X}-\rho_i}\right)^2 R^2 \sum_{i=1}^{K}\left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 S_{xi}^2 \\ & - \left(\frac{\bar{X}}{\bar{X}+\rho_i}\right)^2 R^2 \sum_{i=1}^{K}\left(\frac{1}{n_i^1} - \frac{1}{N_i}\right) p_i^2 S_{xi}^2 \\ & - 2R\left(\frac{\bar{X}}{\bar{X}+\rho_i}\right) \sum_{i=1}^{K}\left(\frac{1}{n_i} - \frac{1}{N_i}\right) p_i^2 \rho_i S_{xi} S_{yi} \\ & + 2\left(\frac{\bar{X}}{\bar{X}+\rho_i}\right) R \sum_{i=1}^{K}\left(\frac{1}{n_i^1} - \frac{1}{N_i}\right) p_i^2 \rho_i S_{xi} S_{yi} \\ & + \sum_{i=1}^{K}\left[\left(\frac{1}{n_i} - \frac{1}{N_i}\right)\left(\frac{k_i - 1}{n_i}\right) P_i^2 W_i S_{y2i}^2\right]\Bigg].\end{aligned}$$

Proposed estimators

In this study, new ratio and regression type estimators of the population mean $\bar{Y}$ using information on two auxiliary variables are proposed. Under the condition that the study variable Y and first auxiliary variable X are affected by non-response while the supplementary information Z is free from the non-response. It is further assumed that the population mean $\bar{X}$ is not known. The proposed ratio and regression type estimators under the given situation are respectively represented as follows

$$\hat{T}_{Rat} = \frac{\bar{y}^*}{\bar{x}_R^*}\bar{x}^* \tag{1}$$

$$\hat{T}_{Reg} = \bar{y}^* + b^*\left(\bar{x}^* - \bar{x}_{1r}^*\right) \tag{2}$$

where

$$\bar{x}_R^* = \frac{\bar{x}^*}{\bar{z}^*}\bar{Z}, \qquad \bar{x}_{1r}^* = \bar{x}^* + b^*\left(\bar{z}^* - \bar{Z}\right),$$

and

$$b^* = \frac{s^*_{xz}}{s^{2*}_z}$$

is an estimator of the population regression coefficient of X and Z, i.e.,

$$\beta_2 = \frac{S_{XZ}}{S^2_Z},$$

where s^*_{xz} and s^{2*}_z are respectively the unbiased estimators of S_{XZ} and S^2_Z, based on $(n'_1 + h_2)$ units.

Further,

$$S_{XZ} = \rho_{XZ} S_X S_Z,$$

and ρ_{XZ} is the population correlation coefficient between X and Z.

To obtain the bias and MSEs of $\hat{T}_{Rat}$ and $\hat{T}_{Reg}$, we adopt the theory of large sample approximations.

Let

$$\bar{y}^* = \bar{Y}(1+e_0), \qquad \bar{x}^* = \bar{X}(1+e_1), \qquad \bar{x}^{*'} = \bar{X}(1+e'_1),$$

$$\bar{z}^* = \bar{Z}(1+e_2), \qquad s^*_{xy} = S_{XZ}(1+e_2), \qquad s^{2*}_x = S^2_X(1+e_3),$$

$$s^*_z = S_{XZ}(1+e_4), \qquad s^{2*}_z = S^2_Z(1+e_5).$$

Taking expectation of the error terms, the following are obtained

$$E(e_0) = E(e_1) = E(e_2) = E(e_3) = E(e_4) = E(e_5) = E(e'_1) = E(e'_2) = 0$$

$$E(e_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C^2_Y + \frac{L-1}{n} W_2 \frac{S^2_{Y2}}{\bar{Y}^2},$$

$$E(e_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right) C^2_X + \frac{L-1}{n} W_2 \frac{S^2_{X2}}{\bar{X}^2},$$

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

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

$$E(e_0e_1) = \left(\frac{1}{n} - \frac{1}{N}\right) \rho_{XY} C_X C_Y + \frac{L-1}{n} W_2 \rho_{XY2} \frac{S_{X2}}{\bar{X}} \frac{S_{Y2}}{\bar{Y}},$$

$$E(e_0e_1') = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{XY} C_X C_Y + \frac{L'-1}{n'} W_2 \rho_{XY2} \frac{S_{X2}}{\bar{X}} \frac{S_{Y2}}{\bar{Y}},$$

$$E(e_0e_2) = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{YZ} C_Y C_Z,$$

$$E(e_1e_1') = \left(\frac{1}{n'} - \frac{1}{N}\right) C_X^2 + \frac{L'-1}{n'} W_2 \frac{S_{X2}^2}{\bar{X}^2},$$

$$E(e_1e_2) = \left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{XY} C_X C_Y,$$

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

$$E(e_1e_2) = \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{21}}{n\bar{X}S_Y} + W_2 \frac{L-1}{n} \frac{\mu_{21(2)}}{\bar{X}S_Y},$$

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

$$E(e_1e_3) = \frac{N(N-n)}{(N-1)(N-2)} \frac{\mu_{30}}{n\bar{X}S_X^2} + W_2 \frac{L-1}{n} \frac{\mu_{30(2)}}{\bar{X}S_{X2}^2},$$

$$E(e_1'e_3) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{30}}{n'\bar{X}S_X^2} + W_2\frac{L'-1}{n'}\frac{\mu_{30(2)}}{\bar{X}S_{X2}^2},$$

$$E(e_2'e_2) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{111}}{n'\bar{Z}S_Y},$$

$$E(e_2'e_3) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{21}'}{n'\bar{Z}S_X^2}.$$

$$E(e_2'e_4) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{12}}{n'\bar{Z}S_{XY}}, \qquad E(e_2'e_5) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{03}'}{n'\bar{Z}S_Z}.$$

$$E(e_1'e_4) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{31}}{n'\bar{X}S_{XZ}} + W_2\frac{L'-1}{n'}\frac{\mu_{31(2)}}{\bar{X}S_{XZ}},$$

$$E(e_1'e_5) = \frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{41}}{n'\bar{X}S_Z^2} + W_2\frac{L'-1}{n'}\frac{\mu_{41(2)}}{\bar{X}S_Z^2}.$$

where

$$\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{X})^r(y_i-\bar{Y})^s, \qquad \mu_{rs(2)} = \frac{1}{N_2}\sum_{i=1}^{N_2}(x_i-\bar{X}_2)^r(y_i-\bar{Y}_2)^s,$$

$$\mu_{rs}' = \frac{1}{N}\sum_{i=1}^{N}(x_i'-\bar{X})^r(y_i-\bar{Y})^s, \qquad \mu_{rst} = \frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{X})^r(y_i-\bar{Y})^s(z_i-\bar{Z})^t.$$

$$C_Y = \frac{S_Y}{\bar{Y}}, \qquad C_X = \frac{S_X}{\bar{X}}, \qquad C_Z = \frac{S_Z}{\bar{Z}},$$

$$\bar{X}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2}x_i, \qquad \bar{Y}_2 = \frac{1}{N_2}\sum_{i=1}^{N_2}y_i,$$

and ρ_{XY} is the population correlation coefficient between X and Y.
Recall Eq. (1),

$$\hat{T}_{Rat} = \frac{\bar{y}^*}{\bar{x}_R^*}\bar{x}^*,$$

$$= \bar{Y}(1+e_0)(1+e_1)(1+e_1')^{-1}(1+e_2')^{-1},$$

$$= \bar{Y}\left[1 - e_1'^2 - e_2' + e_1'e_2' + e_2'^2 + e_1 - e_1'e_1 + e_0 - e_0e_1' + e_0e_2'\right]. \tag{3}$$

Now Eq. (3) can be expressed as

$$\hat{T}_{Rat} - \bar{Y} = \bar{Y}\left[1 - e_1'^2 - e_1' + e_1'e_2' + e_1'^2 + e_1 - e_1'e_1 + e_0 - e_0e_1' + e_0e_1 - 1\right]. \tag{4}$$

$$\begin{aligned} E(\hat{T}_{Rat} - \bar{Y}) &= \bar{Y}\Big[E(e_0) + E(e_1) - E(e_1') + E(e_1'^2) - E(e_1^2) \\ &\quad + E(e_1'e_2') - E(e_1'e_1) - E(e_0e_1') + E(e_0e_1)\Big] \\ &= \bar{Y}\Big[E(e_1'^2) - E(e_1^2) + E(e_1'e_2') - E(e_1'e_1) - E(e_0e_1') + E(e_0e_1)\Big]. \end{aligned}$$

$$\begin{aligned} \text{Bias}(\hat{T}_{Rat}) = \bar{Y}\Bigg[& f'C_X^2 + \frac{L'-1}{n'}W_2C_{X2}^2 - fC_X^2 - \frac{L-1}{n}W_2C_{X2}^2 \\ & + f'\rho_{YX}C_XC_Y - f'C_X^2 - \frac{L'-1}{n'}W_2C_{X2}^2 \\ & - f'\rho_{YX}C_XC_Y - \frac{L'-1}{n'}W_2\rho_{YX2}C_{X2}C_{Y2} \\ & - f'C_X^2 - \frac{L'-1}{n'}W_2C_{X2}^2f'\rho_{ZX}C_XC_Z\Bigg]. \end{aligned} \tag{5}$$

Squaring both sides of Eq. (4) and taking expectations,

$$E(\hat{T}_{Rat} - \bar{Y})^2 = \bar{Y}^2E\left\{e_0^2 + e_1^2 + e_1'^2 + 2e_0e_1 - 2e_0e_1' - 2e_1'e_1\right\}. \tag{6}$$

$$
\begin{aligned}
&= \bar{Y}^2 \Bigg\{ \left(\frac{1}{n} - \frac{1}{N}\right) C_Y^2 + \frac{L-1}{n} W_2 \frac{S_{Y2}^2}{\bar{Y}^2} \\
&\quad + \left(\frac{1}{n} - \frac{1}{N}\right) C_X^2 + \frac{L-1}{n} W_2 \frac{S_{X2}^2}{\bar{X}^2} \\
&\quad + \left(\frac{1}{n'} - \frac{1}{N}\right) C_X^2 + \frac{L'-1}{n'} W_2 \frac{S_{X2}^2}{\bar{X}^2} \\
&\quad + 2\left(\frac{1}{n} - \frac{1}{N}\right) \rho_{YX} C_X C_Y + 2\frac{L-1}{n} W_2 \rho_{YX2} \frac{S_{X2} S_{Y2}}{\bar{X}\bar{Y}} \\
&\quad - 2\left(\frac{1}{n'} - \frac{1}{N}\right) \rho_{YX} C_X C_Y - 2\frac{L'-1}{n'} W_2 \rho_{YX2} \frac{S_{X2} S_{Y2}}{\bar{X}\bar{Y}} \\
&\quad - 2\left(\frac{1}{n'} - \frac{1}{N}\right) C_X^2 - 2\frac{L'-1}{n'} W_2 \frac{S_{X2}^2}{\bar{X}^2} \Bigg\}.
\end{aligned}
$$

$$
\mathrm{MSE}(\hat{T}_{Rat}) = \bar{Y}^2 \left\{ f C_Y^2 + \frac{L-1}{n} W_2 C_{Y2}^2 + f^* \left(C_X^2 + 2f\rho_{YX2} C_X C_Y\right) - W_2 f_2 \left(C_{X2}^2 + 2\rho_{YX2} C_{X2} C_{Y2}\right) \right\}, \tag{7}
$$

where

$$
f^* = \left(\frac{1}{n} - \frac{1}{n'}\right), \qquad f_2 = \left[\frac{L-1}{n} - \frac{L'-1}{n'}\right].
$$

Consider Eq. (2),

$$
\begin{aligned}
\hat{T}_{Reg} &= \bar{y}^* + b^* \left(\bar{x}^* - \bar{x}_{lr}^*\right) \\
&= \bar{y}^* + b^* \left[\bar{x}^* - \left\{\bar{x}^* - b^{1*}(\bar{z}^* - \bar{Z})\right\}\right] \\
&= \bar{Y} + \bar{Y} e_0 + \beta_1 (1+e_2)(1+e_3)^{-1} \left[\bar{X} e_1 - \bar{X} e_1' + \beta_2 \left(1 - e_5 + e_5^2 + e_4 - e_4 e_5\right) (\bar{Z} e_2')\right].
\end{aligned} \tag{3}
$$

$$
\begin{aligned}
\hat{T}_{Reg} - \bar{Y} = \bar{Y} e_0 &+ \beta_1 \bar{X} \left[e_1 - e_1 e_3 + e_1 e_2 - e_1' + e_1' e_3 - e_1' e_2\right] \\
&+ \beta_1 \beta_2 \bar{Z} \left[e_2' - e_2' e_5 + e_2' e_4 - e_2' e_3 + e_2' e_2\right].
\end{aligned} \tag{8}
$$

$$
\begin{aligned}
E(\hat{T}_{Reg}-\bar{Y}) = \bar{Y}E(e_0) &+ \beta_1\bar{X}\Big[E(e_1)-E(e_1e_3)+E(e_1e_2)-E(e_1')\\
&\quad + E(e_1'e_3)-E(e_1'e_2)\Big]\\
&+\beta_1\beta_2\bar{Z}\Big[E(e_2')-E(e_2'e_5)+E(e_2'e_4)\\
&\quad -E(e_2'e_3)+E(e_2'e_2)\Big]\\
=\beta_1\bar{X}&\Big[-E(e_1e_3)+E(e_1e_2)+E(e_1'e_3)-E(e_1'e_2)\Big]\\
&+\beta_1\beta_2\bar{Z}\Big[-E(e_2'e_5)+E(e_2'e_4)-E(e_2'e_3)+E(e_2'e_2)\Big].
\end{aligned}
$$

$$
\begin{aligned}
=\beta_1\bar{X}\Bigg[&-\frac{N(N-n)}{(N-1)(N-2)}\frac{\mu_{30}}{n\bar{X}S_X^2}-W_2\frac{L-1}{n}\frac{\mu_{30(2)}}{\bar{X}S_{X2}^2}\\
&+\frac{N(N-n)}{(N-1)(N-2)}\frac{\mu_{21}}{n\bar{X}S_{XY}}+W_2\frac{L-1}{n}\frac{\mu_{21(2)}}{\bar{X}S_{XY}}\\
&+\frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{30}}{n'\bar{X}S_X^2}+W_2\frac{L'-1}{n'}\frac{\mu_{30(2)}}{\bar{X}S_{X2}^2}\\
&-\frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{21}}{n'\bar{X}S_{XY}}-W_2\frac{L'-1}{n'}\frac{\mu_{21(2)}}{\bar{X}S_{XY}}\Bigg]\\
+\beta_1\beta_2\bar{Z}\Bigg[&-\frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{41}}{n'\bar{Z}S_Z^2}-W_2\frac{L'-1}{n'}\frac{\mu_{41(2)}}{\bar{Z}S_Z^2}\\
&+\frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{31}}{n'\bar{Z}S_{XZ}}+W_2\frac{L'-1}{n'}\frac{\mu_{31(2)}}{\bar{Z}S_{XZ}}\\
&-\frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{30}}{n'\bar{X}S_X^2}-W_2\frac{L'-1}{n'}\frac{\mu_{30(2)}}{\bar{X}S_{X2}^2}\\
&+\frac{N(N-n')}{(N-1)(N-2)}\frac{\mu_{21}}{n'\bar{X}S_{XY}}+W_2\frac{L'-1}{n'}\frac{\mu_{21(2)}}{\bar{X}S_{XY}}\Bigg].
\end{aligned}
$$

$$
\begin{aligned}
\text{Bias}(\hat{T}_{Reg}) = \beta_1 \Bigg[& \frac{N^2}{(N-1)(N-2)} \left(\frac{n'-n}{nn'}\right) \left(\frac{\mu_{21}}{S_{XY}} - \frac{\mu_{30}}{S_X^2}\right) \\
& + W_2 \left(\frac{L'-1}{n'} - \frac{L-1}{n}\right) \left(\frac{\mu_{30(2)}}{S_X^2} - \frac{\mu_{21(2)}}{S_{XY}}\right) \\
& + \beta_1 P_2 \Bigg\{ \left(-\frac{N(N-n')}{(N-1)(N-2)n'}\right) \left(\frac{\mu_{41}}{S_Z^2} - \frac{\mu_{31(2)}}{S_{XZ}}\right) \\
& \quad + W_2 \frac{L'-1}{n'} \left(\frac{\mu_{31}}{S_{XZ}} - \frac{\mu_{41(2)}}{S_S^2}\right) \\
& \quad - R^* \Bigg\{ \frac{N(N-n')}{(N-1)(N-2)n'} \left(\frac{\mu_{21}}{S_{XY}} - \frac{\mu_{30}}{S_X^2}\right) \\
& \qquad + W_2 \frac{L'-1}{n'} \left(\frac{\mu_{21(2)}}{S_{XY}} - \frac{\mu_{30(2)}}{S_X^2}\right) \Bigg\} \Bigg\} \Bigg].
\end{aligned} \tag{9}
$$

where

$$
R^* = \frac{\bar{Z}}{\bar{X}}.
$$

$$
\begin{aligned}
E(\hat{T}_{Reg} - \bar{Y})^2 = & E\Big[\bar{Y}e_0 + \beta_1 \bar{X}\left(e_1 - e_1 e_3 + e_1 e_2 - e_1' + e_1' e_3 - e_1' e_2\right) \\
& + \beta_1 \beta_2 \bar{Z}\left(e_1' - e_1' e_5 + e_1' e_4 - e_1' e_3 + e_1 e_2\right)\Big]^2 \\
= & \Big[\bar{Y}^2 E(e_0^2) + 2\beta_1 \bar{Y}\bar{X}\left\{E(e_0 e_1) - E(e_0 e_1')\right\} + 2\beta_1\beta_2 \bar{Y}\bar{Z}E(e_0 e_1) \\
& + \beta_1^2 \bar{X}^2 \left\{E(e_1^2) - 2E(e_1 e_1') + E(e_1'^2)\right\} \\
& + 2\beta_1^2 \beta_2 \bar{X}\bar{Z}\left\{E(e_1 e_1') - E(e_1'^2)\right\} + \beta_1^2 \beta_2^2 \bar{Z}^2 E(e_1'^2)\Big].
\end{aligned} \tag{10}
$$

Empirical study

In order to examine the behaviour of the proposed estimators. It is essential to investigate the theoretical claims with empirical analysis. Therefore, two different data sets are considered with a view to validate the theoretical claims.

Date Set 1

Data used is from population I considered by [27], page 97. Here, the head length of the second son is considered as the study variable Y, whereas the head length of the first son and the head breadth of the first son are respectively considered as the auxiliary X and Z. The population details are given below:
$N = 25, n' = 10, n = 7, \bar{X} = 185.72, \bar{Y} = 183.84, \bar{Z} = 151.121, C_X = 0.0526, C_Y =$

0.0546,
$C_Z = 0.0488 S_X = 9.78872, S_Y = 10.03766, S_Z = 7.374656, S_{X2} = 6.512581,$
$S_{Y2} = 6.691776, S_{Z2} = 4.91647, \rho_{XY} = 0.7108, \rho_{XZ} = 0.6932, \rho_{YZ} = 0.7346,$
$\rho_{XY2} = 0.473867.$

Table 1. Table1a: Variance/MSE of the Estimators

W_2	$L = L'$	Variance / MSE								
0.2		$\bar{y}^*$	$T^*_{1\text{ rat}}$	$T^*_{2\text{ reg}}$	$T^{**}_{1\text{ rat}}$	$T^{**}_{2\text{ reg}}$	T_{ae}	T_{ae1}	$\widehat{T}_{\text{Rat}}$	$\widehat{T}_{\text{Reg}}$
	1.5	11.0	9.04	8.7	6.4	6.0	5.3	5.0	4.51	4.00
	2	11.6	9.68	9.4	7.1	6.6	5.9	5.1	4.58	4.02
	2.5	12.2	10.3	10.0	7.7	7.2	6.1	5.2	5.33	5.13
	3	12.9	10.9	10.6	8.3	7.8	6.7	5.3	5.87	5.32

Table 2. Tablelb: PRE of the Estimators

W_2	$L = L'$	PRE								
0.2		$\bar{y}^*$	$T^*_{1\text{ rat}}$	$T^*_{2\text{ reg}}$	$T^{**}_{1\text{ rat}}$	$T^{**}_{2\text{ reg}}$	T_{ae}	T_{ae1}	$\widehat{T}_{\text{Rat}}$	$\widehat{T}_{\text{Reg}}$
	1.5	100	121.8	125.2	170.2	181.5	190.5	193.3	195.1	211.6
	2	100	120.3	123.9	164.4	174.6	189.3	190.6	192.5	203.4
	2.5	100	118.0	122.8	159.0	168.9	182.7	187.4	188.9	200.9
	3	100	117.8	121.8	154.5	163.9	173.3	184.8	185.7	198.3

Data Set 2
Here, the data considered by [28] is used to demonstrate the theoretical results. In this data set, number of female is considered to be the study variable Y, whereas the number of female in service and number of educated female are considered to be auxiliary variables X and Z respectively. The characteristics of the population are given below:

$$N = 61, n' = 20, n = 10, \bar{X} = 5.31, \bar{Y} = 7.46, \bar{Z} = 179.00, C_X = 0.7587, C_Y = 0.71303,$$
$$C_Z = 0.2515 S_X = 4.028697, S_Y = 5.298838, S_Z = 45.0185, S_{X2} = 2.685798, S_{Y2} = 3.532559,$$
$$S_{Z2} = 30.01233, \rho_{XY} = 0.7737, \rho_{XZ} = 0.207, \rho_{YZ} = 0.0033, \rho_{XY2} = 0.5158.$$

Table 3. Table 2a: Variance/MSE of the Estimators

W_2	$L = L'$	Variance / MSE								
0.2		$\bar{y}^*$	$T^*_{1\text{ rat}}$	$T^*_{2\text{ reg}}$	$T^{**}_{1\text{ rat}}$	$T^{**}_{2\text{ reg}}$	T_{ae}	T_{ae1}	$\widehat{T}_{\text{Rat}}$	$\widehat{T}_{\text{Reg}}$
	1.5	2.4	2.4	1.6	2.0	1.6	2.5	2.3	1.9	1.2
	2	2.5	2.6	1.7	2.1	1.7	2.6	2.4	2.0	1.2
	2.5	2.7	2.6	1.8	2.2	1.8	2.7	2.6	2.0	1.3
	3	2.8	2.8	1.9	2.4	1.9	2.8	2.7	2.1	1.4

Table 4. Table 2b: PRE of the Estimators

W_2	$L = L'$	PRE								
0.2		$\bar{y}^*$	$T^*_{1\ \text{rat}}$	$T^*_{2\ \text{reg}}$	$T^{**}_{1\ \text{rat}}$	T^{**}_{2reg}	T_{ae}	T_{ae1}	$\widehat{T}_{\text{Rat}}$	$\widehat{T}_{\text{Reg}}$
	1.5	100	101.9	152.6	122.8	152.7	100.3	120.8	134.8	164.7
	2	100	101.7	149.9	121.3	150.0	100.1	199.4	133.4	163.2
	2.5	100	101.5	147.6	120.0	147.6	99.7	198.7	132.4	162.6
	3	100	101.4	145.5	118.8	145.5	98.5	198.2	133.1	161.3

Conclusion

In this study, two estimators have been proposed, the ratio and regression-type estimators of the population mean using the information on two auxiliary variables in double sampling scheme with sub-sampling the non-respondents. A theoretical framework on the properties of the proposed estimators has been presented in details. A comparative study of the proposed estimators with the usual mean estimators and some other existing estimators of the population mean have also been presented. To support the theoretical results, empirical study has been considered with two different real data sets. From the Tables 1a, 1b, 2a and 2b, it is observed that the proposed estimators $\widehat{T}_{\text{Rat}}$ and $\widehat{T}_{\text{Reg}}$ performed better than the existing estimators $T^*_{1\ \text{rat}}$, $T^*_{2\ \text{reg}}$, $T^{**}_{1\ \text{rat}}$, $T^*_{2\ \text{reg}}$, T_{ae} and $T_{\text{ae}\ 1}$ considered in this work. This study has provided more reliable estimators as a solution to the challenges encountered during estimation of population parameters from survey data such as; health statistics, household income and expenditure, etc, which involve sensitive indicators that might attract non-response. For instance, in the study of household consumption, household income serves as the first auxiliary variable which is highly correlated with the study variable, and might be affected with non-response.(see [28]), In this case, the support auxiliary variable could be family size or geographic location (urban or rural) which will not be affected by non-response. It becomes necessary for researchers and users of statistics to adopt the proposed estimators for precise and efficient estimates of the population parameters.

Acknowledgement

We are thankful to the Chief editor of the journal and anonymous referees for their valuable comments and suggestions which lead to an improved version of the manuscript.

References

[1] Olkin, I. (1958). Multivariate ratio estimator for finite populations. *Biometrika*, 45, 51–65.

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

[3] Sahai, A., and C. R. M. A. K. (1980). An alternative multivariate product estimator. *Annals of Mathematical Statistics*, 32(2), 6–12.

[4] Kiregyera, B. (1980). A chain ratio-type estimator in finite population double sampling

using two auxiliary variables. *Metrika*, 31, 215–226.

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

[6] Iseh, M. J., and Enang, E. I. (2021). A calibration synthetic estimator of population mean in small area under stratified sampling design. *Statistics in Transition New Series*, 22(3), 15–30.

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

[8] Kumar, S., and Sharma, V. (2020). Improved chain ratio-product type estimators under double sampling scheme. *Journal of Statistics Application and Probability Letters*, 7(2), 87–95.

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

[10] Iseh, M. J., and Bassey, K. J. (2021a). A new calibration estimator of population mean for small area with non-response. *Asian Journal of Probability and Statistics*, 12(2), 41–51.

[11] Iseh, M. J., and Bassey, K. J. (2021b). Calibration estimator for population mean in small sample size with non-response. *European Journal of Statistics and Probability*, 9(1), 32–42.

[12] Olkin, I. (1986). Ratio estimator with sub-sampling the nonrespondent. *Survey Methodology*, 12, 217–230.

[13]

[13] conventional and alternative two-phase sampling ratio, product and regression estimators in the presence of non-response. *Proceedings of the National Academy of Sciences, India*, 65(A), 195–203.

[14]

[14] ratio and regression estimators with sub-sampling the nonrespondent. *Survey Methodology*, 26, 183–188.

[15]

[15] the population mean in the presence of nonresponse under double sampling using auxiliary information. *SORT*, 33(1), 71–78.

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

[17] Anieting, A. E., and Enang, E. I. (2020). Two-phase stratified sampling estimator for population mean in the presence of non-response using one auxiliary variable. *Mathematical Journal of Interdisciplinary Sciences*, 8(2), 49–56.

[18] Iseh, M. J., and Bassey, M. O. (2022). Calibration estimators for population mean with sub-sampling the nonrespondents under stratified sampling. *Science Journal of Applied Mathematics and Statistics*, 10(4), 45–56.

[19] Iseh, M. J., and Bassey, M. O. (2024). Smoothing of estimators of population mean using

calibration technique with sample errors. *Journal of Modern Applied Statistical Methods*, 23(1), 1–20.

[20] Ikot, E. E., and Iseh, M. J. (2024). Calibration for efficiency of ratio estimator in domains of study with sub-sampling the non-respondents. *Biometrics & Biostatistics International Journal*, 13(2), 42–50.

[21] Tabasan, R., and Khan, I. A. (2004). Double sampling for ratio estimation for non-response. *Journal of the Indian Society of Agricultural Statistics*, 58, 300–306.

[22] Pal, S. K., and Singh, H. P. (2018). Estimator of mean with two-parameter ratio-product estimation in double sampling using auxiliary information under non-response. *Journal of Modern Applied Statistical Methods*, 17(2), 2–32.

[23] Khare, B. B., and Sinha, R. R. (2019). Estimator of production of two populations by multi-auxiliary characters under double sampling of the nonrespondent. *Statistics in Transition New Series*, 20(3), 81–95.

[24] Chaudhary, M. K., Kumar, A., and Vishwarkarma, G. K. (2021). Some improved estimators of population mean using two-phase sampling scheme in the presence of non-response. *Pakistan Journal of Statistics and Operations Research*, 17(4), 911–919.

[25] Anieting, A. E., Enang, E. I., and Onwukwe, C. E. (2021). Efficient estimator for population mean in stratified double sampling in the presence of non-response using one auxiliary variable. *African Journal of Mathematics and Statistical Studies*, 4(2), 40–50.

[26] Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, New York.

[27] Singh, M. P. (1967). Multivariate product method of estimation for finite populations. *Philosophical Transactions of the Royal Society*, 19(2), 1–10.

[28] Iseh, M. J. (2023). Model formulation on efficiency for median estimation under fixed cost in survey sampling, Model Assisted Statistica and Applications. Vol 18(4): 373-385