STATISTICS IN TRANSITION new series, March 2026
Vol. 27, No. 1, pp. xx–XX, https://doi.org/10.59139/stattrans-2026-xxx
Received – xx.xx.xxxx; accepted – xx.xx.xxxx

Optimality of classical difference estimators of finite population variance under random non-response with comparative study

Mahamood Usman[1]

Abstract

In this study, we address the challenge of calculating the finite population variance when faced with random non-response. Such issues are commonly encountered in various fields like medical sciences, environmental sciences and business studies when dealing with data. Using the ranking of an auxiliary variable across three different methodologies of random non-response, we developed several novel difference-type estimators of population variance along with their optimal models. The strategies are shaped by using the varying levels of information available regarding the auxiliary variable. We have studied the properties of the proposed estimators under large sample approximations and determined their optimum situations in each strategy. The introduced estimators can be viewed as an advancement of traditional difference estimators. Within the associated methodologies, we conducted a comparative analysis based on some real datasets as well as simulated datasets, whereby the proposed estimators showed reduced variances when assessed in terms of the enhanced percentage relative efficiencies (PRE) compared to some standard ratio and difference-type estimators relevant to the respective methodologies.

Key words: study variable, population variance, dual use of auxiliary variable, percentage relative efficiency, random non-response.

AMS Subject Classification: 62D05.

1. Introduction

The measurement of variation provides a dynamic idea about the data. For example, a company sales representative may analyze the variations in sales records or the population of customers monthly to help them decide how to improve sales or customer satisfaction. Similarly, a marketing analyst of a company may be interested in analyzing the variability of company sales in a particular area over time to see which products the customers like most. To measure the variation of such kind of data, the survey practitioners often use the term variance. Variance measures the variability of data and is extensively used by analysts in various fields such as agriculture, forestry, medical science, politics, finance, population traits, etc. It plays an important role in the testing of hypotheses and the construction of confidence intervals for population parameters. The attention to the variance estimation techniques has been paid by researchers since long ago. Singh et al. (1973) have proposed

[1]Department of Mathematics (School of Advanced Sciences), Vellore Institute of Technology, Vellore-632014, India. E-mail: mahmoodu33@gmail.com. ORCID: https://orcid.org/0000-0003-1234-4401.

the estimator of population variance using the priori information about the population coefficient of kurtosis and compared it with the usual unbiased estimator. Das and Tripathi (1978) have introduced the estimator using the information on an auxiliary variable. Later on, ratio and regression-type estimators using the information on auxiliary variables have been discussed by various authors such as Isaki (1983), Singh et al. (1988), Upadhyay and Singh (2001), Yasmeen et al. (2019), Zaman and Bulut (2022), among many others. Belili et al. (2023) and Khodija et al. (2023) have presented some improved probability distributions and studied their mathematical properties. They have examined the efficiencies of the estimators through comparative studies. Ahmad et al. (2023), Zaman and Bulut (2024) and Daraz et al. (2025) have proposed some ratio and difference-type estimators of population mean and variance and investigated their optimal behaviors using real and simulated datasets. The above authors have studied the estimation procedures of population parameters in the presence of complete response.
When the non-response is observed in the sample, the problem of estimation of population variance has also been discussed by various authors in the context of random non-response, introduced by Rubin (1976). The authors, notably Singh and Joarder (1998), Kumar (2014), Sharma and Singh (2020), and Bhusan and Pandey (2021) have suggested improved estimators of finite population variance in the presence of random non-response using the information on single and multi-auxiliary variables. It is a well-known fact that the efficiency of the estimators may be increased by using the multiple auxiliary variables. When the information on multi-auxiliary variables is not available, the researchers like Yaqub et al. (2017), Hussain and Haq (2019), Irfan et al. (2020) and many others have published the higher efficient estimators just by recalling the dual or rank of an auxiliary variable. Singh and Usman (2022) have established improved estimators of population variance using the rank of an auxiliary variable in a customary way in the case of random non-response. Recently, the authors like Javed et al. (2023), Almulhim et al. (2024), Bhusan and Pandey (2025), etc., have suggested improved and optimal estimation procedures for estimation of population parameters in the related areas.
Inspired by the aforementioned researchers, the motivation of the present work can be stated as follows:

- Enhancing the efficiency of the classical difference estimator of population variance to the next level.
- Efficient utilization of the rank of an auxiliary variable in the construction of a new model.
- Investigation of the behavior of the new model in three distinct strategies of random non-response.
- Comparison of the new model with existing ones based on numerical and simulation studies.

In this study, we have developed some new models along with their optimal versions for the estimation of finite population variance under the missing at random (MAR) non-response mechanism. We have efficiently employed the rank of an auxiliary variable in the construction of new estimators in three distinct strategies of random non-response. The novelty of the present work may be stated as the extension of classical difference estimator by utilizing the rank of an auxiliary variable in order to achieve an enhanced level of efficiency. The role of the rank (dual) of an auxiliary variable in the formulation of newly suggested estimators may be easily recognized in terms of higher percentage relative efficiencies compared to existing estimators considered in this study.

The rest part of the paper is constituted as follows. In *Section 2*, the methodology and notations are presented and some customary estimators have been discussed in *Section 3*. The proposed estimators have been formulated in *Section 4*, and their optimal situations have been stated in *Section 5*. In *Section 6*, the properties of proposed estimators have been compared with some relevant existing estimators under a comparative study based on real and simulated datasets. Finally, the conclusions have been made in *Section 7*.

2. Methodology and Notations

Consider a finite population $\Omega = \{\Omega_1, \Omega_2, ..., \Omega_N\}$ of size N in which the study variable y and auxiliary variable x are properly correlated with an amount of correlation ρ_{yx}. Suppose that $Z_x = \{z_{x_1}, z_{x_2}, ..., z_{x_N}\}$ denote the ranks of corresponding values of variable $X = \{x_1, x_2, ..., x_N\}$ on which the information is already available in Ω. Draw a sample of size n from Ω using the simple random sampling without replacement (SRSWOR) technique where the information cannot be received on $m\{m = 0, 1, 2, ..., (n-2)\}$ units due to random non-response (MAR) for target variable y only. As a result, the (n-m) responding units that remain are treated as the sample based on the technique of simple random sampling. We assume that the information is missing for auxiliary variable x on corresponding units of y, as per the situations discussed in the present study. If the probability of non-response among the $(n-2)$ possible values of m non-responses is denoted by p, then m follows the distribution given by

$$P(m) = \frac{n-m}{nq+2p}\binom{n-2}{m}p^m q^{n-2-m}; \quad m = 0, 1, 2, ..., (n-2) \tag{1}$$

where $p+q=1$ (*instantly* see Singh et al., 2000). Here, p can be estimated using the maximum likelihood estimation method based on the distribution given in (2.1).

Singh and Joarder (1998) have obtained the maximum likelihood estimator of p as

$$\hat{p} = \frac{(n-1+m) - \sqrt{(n-1+m)^2 - \frac{4nm(n-3)}{(n-2)}}}{2(n-3)}$$

and therefore $\hat{q} = 1 - \hat{p}$.

Now, we define the following notations:

$\bar{Y} = \sum_{i=1}^{N} y_i/N$: Mean of y for entire population

$S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$: Variance of y for entire population.

$\bar{x}^* = \sum_{i=1}^{n-m} x_i/(n-m)$: Respondent mean of x

$\bar{x}=\sum_{i=1}^{n}x_i/n$: Mean of x for selected sample
$\bar{X}=\sum_{i=1}^{N}x_i/N$: Mean of x for entire population
$s_x^{*2}=\sum_{i=1}^{n-m}(x_i-\bar{x}^*)^2/(n-m-1)$: Respondent variance of x
$s_x^2=\sum_{i=1}^{n}(x_i-\bar{x})^2/(n-1)$: Variance of x for selected sample
$S_x^2=\sum_{i=1}^{N}(x_i-\bar{X})^2/(N-1)$: Variance of x for entire population
$S_{yx}=\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X})/(N-1)$: Population Covariance between y and x
$\bar{z}_x^*=\sum_{i=1}^{n-m}z_{x_i}/(n-m)$: Respondent mean of Z_x
$\bar{z}_x=\sum_{i=1}^{n}z_{x_i}/n$: Mean of Z_x for selected sample
$\bar{Z}_x=\sum_{i=1}^{N}z_{x_i}/N$: Mean of Z_x for entire population
$s_{z_x}^{*2}=\sum_{i=1}^{n-m}(z_{x_i}-\bar{z}_x^*)^2/(n-m-1)$: Respondent variance of Z_x
$s_{z_x}^2=\sum_{i=1}^{n}(z_{x_i}-\bar{z}_x)^2/(n-1)$: Variance of Z_x for selected sample
$S_{z_x}^2=\sum_{i=1}^{N}(z_{x_i}-\bar{Z}_x)^2/(N-1)$: Variance of Z_x for entire population
$S_{yz_x}=\sum_{i=1}^{N}(y_i-\bar{Y})(z_{x_i}-\bar{Z}_x)/(N-1)$: Population Covariance between y and Z_x
$S_{xz_x}=\sum_{i=1}^{N}(x_i-\bar{X})(z_{x_i}-\bar{Z}_x)/(N-1)$: Population covariance between x and Z_x.
$\rho_{yx}=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})(x_i-\bar{X})}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{Y})^2}\sqrt{\sum_{i=1}^{N}(x_i-\bar{X})^2}}$: Population Correlation between y and x
$\rho_{yz_x}=\frac{\sum_{i=1}^{N}(y_i-\bar{Y})(z_{x_i}-\bar{Z}_x)}{\sqrt{\sum_{i=1}^{N}(y_i-\bar{Y})^2}\sqrt{\sum_{i=1}^{N}(z_{x_i}-\bar{Z}_x)}}$: Population Correlation between y and z_x
$\rho_{xz_x}=\frac{\sum_{i=1}^{N}(x_i-\bar{X})(z_{x_i}-\bar{Z}_x)}{\sqrt{\sum_{i=1}^{N}(x_i-\bar{X})^2}\sqrt{\sum_{i=1}^{N}(z_{x_i}-\bar{Z}_x)}}$: Population Correlation between x and z_x

To obtain the biases and MSEs of the proposed estimators, we assume the following transformations in terms of errors:

$$s_y^{*2}=S_y^2(1+\varepsilon_0),$$
$$\bar{x}^*=\bar{X}(1+\varepsilon_1^*), \qquad \bar{x}=\bar{X}(1+\varepsilon_1),$$
$$\bar{z}_x^*=\bar{Z}_x(1+\varepsilon_2^*) \quad \text{and} \quad \bar{z}_x=\bar{Z}_x(1+\varepsilon_2)$$

such that
$E(\varepsilon_0)=E(\varepsilon_1^*)=E(\varepsilon_1)=E(\varepsilon_2^*)=E(\varepsilon_2)=0$ and $E(\varepsilon_0^2)=f_2(\lambda_{400}-1)=f_2\lambda_{400}^*$, $E(\varepsilon_1^{*2})=f_2C_x^2$, $E(\varepsilon_1^2)=f_1C_x^2$, $E(\varepsilon_2^{*2})=f_2C_{z_x}^2$, $E(\varepsilon_2^2)=f_1C_{z_x}^2$, $E(\varepsilon_0\varepsilon_1^*)=f_2\lambda_{210}C_x$, $E(\varepsilon_0\varepsilon_1)=f_1\lambda_{210}C_x$, $E(\varepsilon_0\varepsilon_2^*)=f_2\lambda_{201}C_{z_x}$, $E(\varepsilon_0\varepsilon_2)=f_1\lambda_{201}C_{z_x}$, $E(\varepsilon_1^*\varepsilon_2^*)=f_2C_{xz_x}$, $E(\varepsilon_1^*\varepsilon_2)=f_1C_{xz_x}$, $E(\varepsilon_1\varepsilon_2^*)=E(\varepsilon_1\varepsilon_2)=f_1C_{xz_x}$.

Here, $f_1=\left(\frac{1}{n}-\frac{1}{N}\right)$, $f_2=\left(\frac{1}{nq+2p}-\frac{1}{N}\right)$, $f_3=f_2-f_1=\left(\frac{1}{nq+2p}-\frac{1}{n}\right)$, $\lambda_{klm}=\frac{\mu_{klm}}{\mu_{200}^{k/2}\mu_{020}^{l/2}\mu_{002}^{m/2}}$, $\mu_{klm}=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^k(x_i-\bar{X})^l(z_{x_i}-\bar{Z}_x)^m$, $C_y=\frac{S_y}{\bar{Y}}$, $C_x=\frac{S_x}{\bar{X}}$, $C_{z_x}=\frac{S_{z_x}}{\bar{X}}$, $C_{yx}=\frac{S_{yx}}{\bar{Y}\bar{X}}=\rho_{yx}C_yC_x$, $C_{yz_x}=\frac{S_{yz_x}}{\bar{Y}\bar{Z}_x}=\rho_{yz_x}C_yC_{z_x}$, $C_{xz_x}=\frac{S_{xz_x}}{\bar{X}\bar{Z}_x}=\rho_{xz_x}C_xC_{z_x}$, $R_1=\frac{\bar{X}}{S_y^2}$ and $R_2=\frac{\bar{Z}_x}{S_y^2}$.

3. Classical Estimators Available in the Literature

In this part of the paper, we have discussed some already established estimation procedures of population variance S_y^2 for study variable along with their properties under the various strategies of random non-response given below.

Strategy I: When population mean $\bar{X}$ and sample mean $\bar{x}$ of auxiliary variable x are used. In this case, we assume that the information is missing at random only for y, and the population mean $\bar{X}$ is known.

Strategy II: When population mean $\bar{X}$ and respondent mean $\bar{x}^*$ of auxiliary variable x are used. In this case, we assume that the information is missing at random for y as well as the corresponding units of x, and the population mean $\bar{X}$ is known.

Strategy III: When sample mean $\bar{x}$ and respondent mean $\bar{x}^*$ of auxiliary variable x are used. In this case, we assume that the information is missing at random only for y, while the information for x is available on all the sampled units, but the population mean $\bar{X}$ is unknown.

The usual estimator of population variance in the case of random non-response, is given by

$$t_0 = s_y^{*2} \tag{2}$$

where $s_y^{*2} = \sum_{i=1}^{n-m}(y_i - \bar{y}^*)^2/(n-m-1)$ is the conditionally unbiased estimator of population variance respectively and where $\bar{y}^* = \sum_{i=1}^{n-m} y_i/(n-m)$ is the respondent mean of y (*see* Singh and Joarder, 1998).

The variance of the estimator t_0 to the first order approximation is given by

$$V(t_0) = S_y^4 f_2 \lambda_{400}^* \tag{3}$$

The classical ratio estimators on the lines of Upadhyay and Singh (2001) under the Strategies I, II and III, are respectively defined as

$$t_{m_1} = s_y^{*2}\left(\frac{\bar{X}}{\bar{x}}\right) \tag{4}$$

$$t_{m_2} = s_y^{*2}\left(\frac{\bar{X}}{\bar{x}^*}\right) \tag{5}$$

$$t_{m_3} = s_y^{*2}\left(\frac{\bar{x}}{\bar{x}^*}\right) \tag{6}$$

The MSEs of the estimators $t_{m_i} (i = 1,2,3)$ to the first order approximation are given by

$$MSE(t_{m_i}) = S_y^4[f_2\lambda_{400}^* + f_i C_x(C_x - 2\lambda_{210})] \tag{7}$$

Following Das (1978) and Upadhyay and Singh (2001), we define two sets of difference and ratio estimators under three strategies of random non-response, respectively given by difference-type estimators:

$$t_{m_{d1}} = s_y^{*2} + k_1^*(\bar{X} - \bar{x}) \quad (8)$$

$$t_{m_{d2}} = s_y^{*2} + k_2^*(\bar{X} - \bar{x}^*) \quad (9)$$

$$t_{m_{d3}} = s_y^{*2} + k_3^*(\bar{x} - \bar{x}^*) \quad (10)$$

Ratio type estimators:

$$t_{m_{r1}} = s_y^{*2}\left(\frac{\bar{X}}{\bar{x}}\right)^{\pi_1^*} \quad (11)$$

$$t_{m_{r2}} = s_y^{*2}\left(\frac{\bar{X}}{\bar{x}^*}\right)^{\pi_2^*} \quad (12)$$

$$t_{m_{r3}} = s_y^{*2}\left(\frac{\bar{x}}{\bar{x}^*}\right)^{\pi_3^*} \quad (13)$$

where k_i^* and π_i^* (i=1,2,3) are the unknown constants which are to be chosen such that the variances of the respective estimators is minimum.

The minimum MSEs of the existing estimators $t_{md_i}(i = 1,2,3)$ and t_{mr_i}, are respectively given by

$$min.MSE(t_{md_i}) = min.MSE(t_{mr_i}) = S_y^4\left[f_2\lambda_{400}^* - f_i\lambda_{210}^2\right]. \quad (14)$$

The optimum values of k_i^* and π_i^* (i=1,2,3) are given by

$$k_{1(opt)}^* = k_{2(opt)}^* = k_{3(opt)}^* = S_y^2\frac{\lambda_{210}}{S_x} \qquad \text{and} \qquad \pi_{1(opt)}^* = \pi_{2(opt)}^* = \pi_{3(opt)}^* = \frac{\lambda_{210}}{S_x}.$$

In line with Singh et al. (1988), the optimal version of the estimators $t_{md_i}(i = 1,2,3)$ in three strategies are given by

$$t_{mD_1} = \kappa_1^* s_y^{*2} + d_1^*(\bar{X} - \bar{x}) \quad (15)$$

$$t_{mD_2} = \kappa_2^* s_y^{*2} + d_2^*(\bar{X} - \bar{x}^*) \quad (16)$$

$$t_{mD_3} = \kappa_3^* s_y^{*2} + d_3^*(\bar{x} - \bar{x}^*) \quad (17)$$

where κ_i^* and d_i^* $(i = 1,2,3)$ are the arbitrary constants to be chosen such that the MSEs of the respective estimators become minimum.

The optimum values of κ_i^* and d_i^* (i=1,2,3) are given by

$$\kappa_{i(opt)}^* = \frac{1}{\left[1 + f_2\lambda_{400}^* - f_i\lambda_{210}^2\right]} \qquad \text{and} \qquad d_{i(opt)} = S_y^2\frac{\lambda_{210}}{S_x}\kappa_{i(opt)}^*$$

The minimum MSEs of the estimators $t_{mD_i}(i = 1,2,3)$ are given by

$$min.MSE(t_{mD_i}) = \frac{S_y^4 MSE(t_{md_i})}{S_y^4 + MSE(t_{md_i})} \quad (18)$$

Thus the estimators $t_{mD_i}(i=1,2,3)$ are improvement over t_{md_i} as well as t_{mr_i} in the corresponding strategies.

4. Proposed Estimators

Here, we have suggested various novel difference-type estimators of finite population variance along with their optimal variants in case of random non-response by employing the rank of an auxiliary variable. The estimators are formulated in three different cases of random non-response, which are discussed in the following three strategies.

Strategy I: When the population means $(\bar{X},\bar{Z}_x)$ and corresponding estimates $(\bar{x},\bar{z}_x)$ from the sample are used.

Strategy II: When the population means $(\bar{X},\bar{Z}_x)$ and corresponding estimates $(\bar{x}^*,\bar{z}_x^*)$ from the respondents are used.

Strategy III: When the sample means $(\bar{x},\bar{z}_x)$ and corresponding estimates $(\bar{x}^*,\bar{z}_x^*)$ from the respondents are used.

The proposed estimators under the *Strategy I*, *Strategy II* and *Strategy III* are given as

$$t_{m_{dd1}} = s_y^{*2} + \phi_1^*(\bar{X}-\bar{x}) + \varphi_1^*(\bar{Z}_x-\bar{z}_x) \tag{19}$$

$$t_{m_{dd2}} = s_y^{*2} + \phi_2^*(\bar{X}-\bar{x}^*) + \varphi_2^*(\bar{Z}_x-\bar{z}_x^*) \tag{20}$$

$$t_{m_{dd3}} = s_y^{*2} + \phi_3^*(\bar{x}-\bar{x}^*) + \varphi_3^*(\bar{z}_x-\bar{z}_x^*) \tag{21}$$

where ϕ_i^* and φ_i^* (i=1,2,3) are the unknown constants to be chosen suitably. The optimum values of these constants are given later in *Appendix*.

The optimal versions of the proposed estimators $t_{m_{dd1}}$, $t_{m_{dd2}}$ and $t_{m_{dd3}}$ are respectively given by

$$t_{m_{dD1}} = \alpha_1^* s_y^{*2} + \beta_1^*(\bar{X}-\bar{x}) + \gamma_1^*(\bar{Z}_x-\bar{z}_x) \tag{22}$$

$$t_{m_{dD2}} = \alpha_2^* s_y^{*2} + \beta_2^*(\bar{X}-\bar{x}^*) + \gamma_2^*(\bar{Z}_x-\bar{z}_x^*) \tag{23}$$

$$t_{m_{dD3}} = \alpha_3^* s_y^{*2} + \beta_3^*(\bar{x}-\bar{x}^*) + \gamma_3^*(\bar{z}_x-\bar{z}_x^*) \tag{24}$$

where $\alpha_i^*(i=1,2,3)$, β_i^* and γ_i^* are the arbitrary chosen constants. The optimum values of these constants are given later in *Appendix*.

The difference-type estimators discussed in (3.7)-(3.9) and (3.14)-(3.16), are the special cases of the proposed difference-type estimators (4.1)-(4.3) and (4.4)-(4.6) respectively in the corresponding strategies.

Theorem 1: The biases of the estimators $t_{m_{ddi}}(i=1,2,3)$ and $t_{m_{dDi}}$ to the first degree of approximations are given by

$$B(t_{m_{ddi}}) = 0 \tag{25}$$

and

$$B(t_{m_{dDi}}) = S_y^2(\alpha_i^*-1) \tag{26}$$

Proof: *See* Appendix

Theorem 2: The minimum MSEs of the estimators $t_{m_{ddi}}(i=1,2,3)$ and $t_{m_{dDi}}$ to the first degree of approximations are given by

$$min.MSE(t_{m_{ddi}}) = S_y^4(f_2\lambda_{400}^* - f_i R_{y.xz_x}^{*2}) \tag{27}$$

and

$$min.MSE(t_{m_{dDi}}) = \frac{S_y^4 min.MSE(t_{m_{ddi}})}{S_y^4 + min.MSE(t_{m_{ddi}})} \tag{28}$$

where $R_{y.xz_x}^{*2} = \frac{\lambda_{201}^2+\lambda_{210}^2-2\rho_{xz_x}\lambda_{210}\lambda_{201}}{1-\rho_{xz_x}^2}$.

Proof: *See* Appendix.

5. Optimal Situations of Proposed Estimators

The optimal situations of the proposed estimators $t_{m_{ddi}}(i=1,2,3)$ and $t_{m_{dDi}}$ at which their variances are minimum are given as

$$t_{m_{dd1}}^* = s_y^{*2} + \phi_{1(opt)}^*(\bar{X}-\bar{x}) + \varphi_{1(opt)}^*(\bar{Z}_x - \bar{z}_x) \tag{29}$$

$$t_{m_{dd2}}^* = s_y^{*2} + \phi_{2(opt)}^*(\bar{X}-\bar{x}^*) + \varphi_{2(opt)}^*(\bar{Z}_x - \bar{z}_x^*) \tag{30}$$

$$t_{m_{dd3}}^* = s_y^{*2} + \phi_{3(opt)}^*(\bar{x}-\bar{x}^*) + \varphi_{3(opt)}^*(\bar{z}_x - \bar{z}_x^*) \tag{31}$$

$$t_{m_{dD1}}^* = \alpha_{1(opt)}^* s_y^{*2} + \beta_{1(opt)}^*(\bar{X}-\bar{x}) + \gamma_{1(opt)}^*(\bar{Z}_x - \bar{z}_x) \tag{32}$$

$$t_{m_{dD2}}^* = \alpha_{2(opt)}^* s_y^{*2} + \beta_{2(opt)}^*(\bar{X}-\bar{x}^*) + \gamma_{2(opt)}^*(\bar{Z}_x - \bar{z}_x^*) \tag{33}$$

and

$$t_{m_{dD3}}^* = \alpha_{3(opt)}^* s_y^{*2} + \beta_{3(opt)}^*(\bar{x}-\bar{x}^*) + \gamma_{3(opt)}^*(\bar{z}_x - \bar{z}_x^*) \tag{34}$$

where

$$\phi_{1(opt)}^* = \phi_{2(opt)}^* = \phi_{3(opt)}^* = \frac{\lambda_{210} - \lambda_{201}\rho_{xz_x}}{1-\rho_{xz_x}^2}\frac{S_y^2}{S_x}, \tag{35}$$

$$\varphi_{1(opt)}^* = \varphi_{2(opt)}^* = \varphi_{3(opt)}^* = \frac{\lambda_{201} - \lambda_{210}\rho_{xz_x}}{1-\rho_{xz_x}^2}\frac{S_y^2}{S_{z_x}}, \tag{36}$$

$$\alpha_{i(opt)}^* = \frac{1}{1+(f_2\lambda_{400}^* - f_i R_{y.xz_x}^{*2})}; \qquad (i=1,2,3) \tag{37}$$

$$\beta_{i(opt)}^* = \alpha_{i(opt)}^*\phi_{i(opt)}^*; \qquad (i=1,2,3) \tag{38}$$

$$\gamma_{i(opt)}^* = \alpha_{i(opt)}^*\varphi_{i(opt)}^*; \qquad (i=1,2,3) \tag{39}$$

6. Comparative Study

We have judged the merits of the estimators based on real and simulated data under an empirical study and a simulation study given as follows.

6.1. Empirical Study

To exhibit the performances of the estimators, we have chosen 8 populations given as follows.

Population-1: [Cochran (1977); p-182]:
Y: Number of placebo children.
X: Number of paralytic polio cases in the placebo group.
The description of the required parameters is as follows:
$N = 34$, $\bar{Y} = 2.588234$, $\bar{X} = 4.923528$, $C_y = 1.233279$, $C_x = 1.023332$, $C_z = 0.5687384$, $\rho_{yx} = 0.7328234$, $\rho_{yz_x} = 0.6571886$, $\rho_{xz_x} = 0.8165118$. Here, $n = 12$ and $m = 8$.

Population-2: [Anderson (1958); p-110]:
Y: Sepal Width of Iris flower. X: Sepal Length of Iris flower.
The description for this data is as follows:
$N = 150$, $\bar{Y} = 3.057334$, $\bar{X} = 5.843334$, $C_y = 0.1425641$, $C_x = 0.1417114$, $C_z = 0.5749112$, $\rho_{yx} = -0.1175699$, $\rho_{yz_x} = -0.1404247$, $\rho_{xz_x} = 0.9871834$. Here, $n = 50$ and $m = 12$.

Population-3: [Madala (1992); p-108]:
Y: salary (thousands of dollars)
X: years of experience (defined as years since receiving Ph D).
The description of the data is as follows:
$N = 32$, $\bar{Y} = 47.37813$, $\bar{X} = 18.376$, $C_y = 0.1819514$, $C_x = 0.4548527$, $C_z = 0.5677533$, $\rho_{yx} = 0.4245115$, $\rho_{yz_x} = 0.3367752$, $\rho_{xz_x} = 0.9447146$. Here, $n = 12$ and $m = 8$.

Population-4: [Anderson (1958); p-110]:
Y: Sepal Length. X: Petal Length.
The details of the required parameters are as follows:
$N = 150$, $\bar{Y} = 5.843334$, $\bar{X} = 3.7581$, $C_y = 0.1417112$, $C_x = 0.4697442$, $C_z = 0.5748254$, $\rho_{yx} = 0.8717537$, $\rho_{yz_x} = 0.8792952$, $\rho_{xz_x} = 0.9684332$. Here, $n = 50$ and $m = 10$.

Population-5: [Satici and Kadilar (2011)]:
Y: number of successful students. X: number of teachers.
The summary of the data is:
$N = 261$, $\bar{Y} = 222.5825$, $\bar{X} = 306.44831$, $C_y = 1.86541$, $C_x = 1.7596$, $C_z = 0.576241$, $\rho_{yx} = 0.9706$, $\rho_{yz_x} = 0.6372$, $\rho_{xz_x} = 0.6264$. Here, $n = 90$ and $m = 70$.

Population-6: [Singh (2003); p-1111]:
Y: amount (in \$000) of non-real estate farm loans in different states during 1997.
X: amount (in \$000) of real estate farm loans in different states during 1997.
The summary of the data is:
$N = 50$, $\bar{Y} = 878.1627$, $\bar{X} = 555.4346$, $C_y = 1.235166$, $C_x = 1.052917$, $C_z = 0.571663$, $\rho_{yx} = 0.8039$, $\rho_{yz_x} = 0.7462$, $\rho_{xz_x} = 0.9237$. Here, $n = 20$ and $m = 8$.

Population-7: [Anderson (1958); p-110]:
Y: Petal Length of Iris setosa. X: Sepal Length of Iris setosa.
The description of the data is given as:
$N = 50$, $\bar{Y} = 1.463$, $\bar{X} = 5.0061$, $C_y = 0.1187853$, $C_x = 0.07041345$, $C_z = 0.5683722$, $\rho_{yx} = 0.2671757$, $\rho_{yz_x} = 0.2687847$, $\rho_{xz_x} = 0.9797012$. Here, $n = 20$ and $m = 5$.

Population-8: [Mc Nill (1977)]:
Y: speed of cars. X: distances taken to stop.
The details of parameters for this data are as follows:
$N = 50$, $\bar{Y} = 15.4$, $\bar{X} = 42.981$, $C_y = 0.3433534$, $C_x = 0.5995668$, $C_z = 0.571306$, $\rho_{yx} = 0.8068948$, $\rho_{yz_x} = 0.8341367$, $\rho_{xz_x} = 0.9605414$. Here, $n = 20$ and $m = 5$.
We have estimated the percentage relative efficiencies (PREs) of the different estimators with respect to usual unbiased estimator s_y^{*2}. To compute the PREs of different estimators $(t_\bullet)$ we use the formula, given by

$$PRE(t_\bullet) = \frac{V(s_y^{*2})}{MSE(t_\bullet)} \times 100.$$

The results are shown in Table 1.

6.2. Simulation Study

We have conducted a simulation study based on artificially generated data using the R programming language. To generate the data, we have considered two statistical probability distributions: (i) Gamma distribution and (ii) Normal distribution, where the performances of the estimators are appraised for different amounts of correlation, 0.6-0.9, with a step of 0.1 between the study variable and auxiliary variable. The distributions are discussed below.

Gamma distribution

Following Singh and Horn (1998), we use the transformations to generate the study and auxiliary variables, which are given as follows:

$$y_i = \mu_y + \sqrt{(1-\rho_{yx}^2)}y_i^* + \rho_{yx}\frac{S_y}{S_x}x_i^* \tag{40}$$

Table 1. *PREs* of the various estimators with respect to s_y^{*2}.

Estimators	Populations 1	2	3	4	5	6	7	8
Strategy I								
t_0	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
t_{m_1}	125.37	97.34	89.07	94.72	159.79	123.98	99.26	88.34
$t_{md_1}=t_{mr_1}$	125.54	101.80	101.07	100.73	204.03	124.03	100.88	100.16
t_{mD_1}	157.89	106.08	116.76	103.34	222.89	142.70	113.92	106.73
t_{mdd_1}	128.33	103.10	109.51	105.07	230.60	129.88	101.06	113.04
t_{mdD_1}	160.68	107.38	125.19	107.67	249.45	148.55	114.11	119.61
Strategy II								
t_{m_2}	160.83	96.10	81.09	92.84	217.26	151.88	98.83	82.63
$t_{md_2}=t_{mr_2}$	161.36	102.70	102.06	101.02	377.99	152.01	101.41	100.25
t_{mD_2}	193.71	106.98	117.75	103.62	396.85	170.69	114.45	106.82
t_{mdd_2}	170.25	104.67	119.75	107.15	546.07	168.44	101.71	122.52
t_{mdD_2}	202.60	108.95	135.44	109.76	564.93	187.12	114.75	129.09
Strategy III								
t_{m_3}	121.34	98.69	90.05	97.90	119.84	117.40	99.56	92.74
$t_{md_3}=t_{mr_3}$	121.48	100.87	100.97	100.28	129.13	117.43	100.52	100.09
t_{mD_3}	153.83	105.15	116.65	102.88	147.98	136.11	113.56	106.66
t_{mdd_3}	123.74	101.48	108.47	101.89	133.43	121.40	100.63	107.35
t_{mdD_3}	156.09	105.76	124.16	104.63	152.28	140.07	113.67	113.92

and

$$x_i = \mu_x + x_i^* \tag{41}$$

where $y_i^* \sim G(a_y, b_y)$ and $x_i^* \sim G(a_x, b_x)$ are the independent gamma variables generated using R programming language. Here, (a_y, b_y) and (a_x, b_x) are the shape and scale parameters for y_i^* and x_i^*. Moreover, $\mu_y = a_y b_y$, $\mu_x = a_x b_x$, $S_y^2 = a_y b_y^2$ and $S_x^2 = a_x b_x^2$. The size of data is $N = 5000$ and sample size $n = 1500$. We have taken $m = 300$.

Normal distribution

We have considered the bivariate normal distribution as $(Y,X) \sim N(9,9,\rho_{yx},20^2,20^2)$ for the correlations (ρ_{yx}). We have chosen $N = 5000$, $n = 1500$ and $m = 300$.
The complete simulation process is as follows. Draw a sample of size n focusing on a variable of interest which is properly correlated with an auxiliary characteristic from a population of size N. Set the value of m and drop m units randomly from the sample. Now, compute the relevant statistics based on the information available on $(n-m)$ units. Repeat the whole procedure 50,000 times.

We have computed the simulated percentage relative efficiencies (PREs) of different estimators considered in this study with respect to usual estimator s_y^{*2} based on their simulated MSE values by using the formulae given as

$$V(s_y^{*2})_{simulated} = \frac{1}{50,000} \sum_{j=1}^{50,000} ((s_y^{*2})_j - S_y^2)^2;$$

$$MSE(t_\bullet)_{simulated} = \frac{1}{50,000} \sum_{j=1}^{50,000} ((t_\bullet)_j - S_y^2)^2; \quad PRE(t_\bullet)_{simulated} = \frac{V(s_y^{*2})_{simulated}}{MSE(t_\bullet)_{simulated}} \times 100.$$

The results are shown in Table 2.

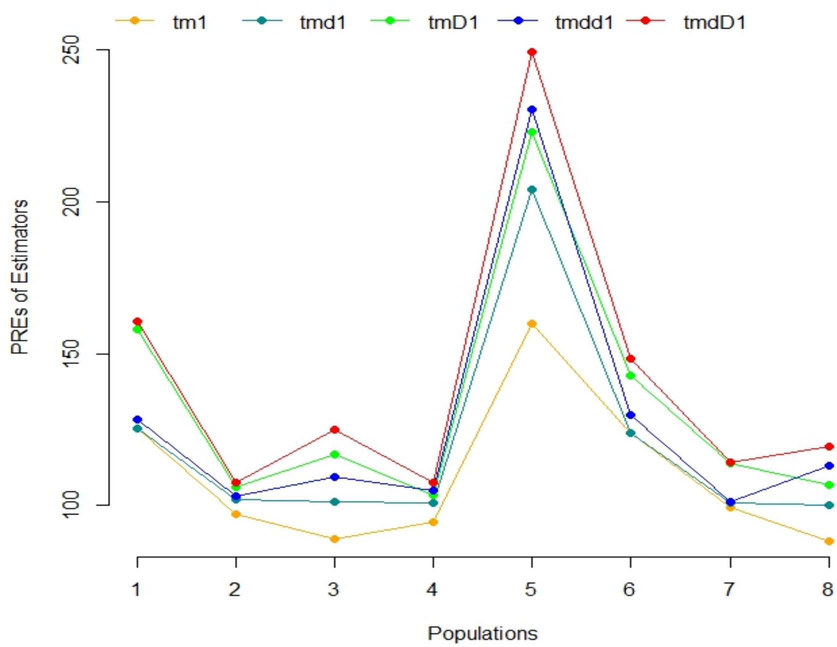

Figure 1. Comparison of PREs of different estimators for Populations 1-8 under ***Strategy-I***

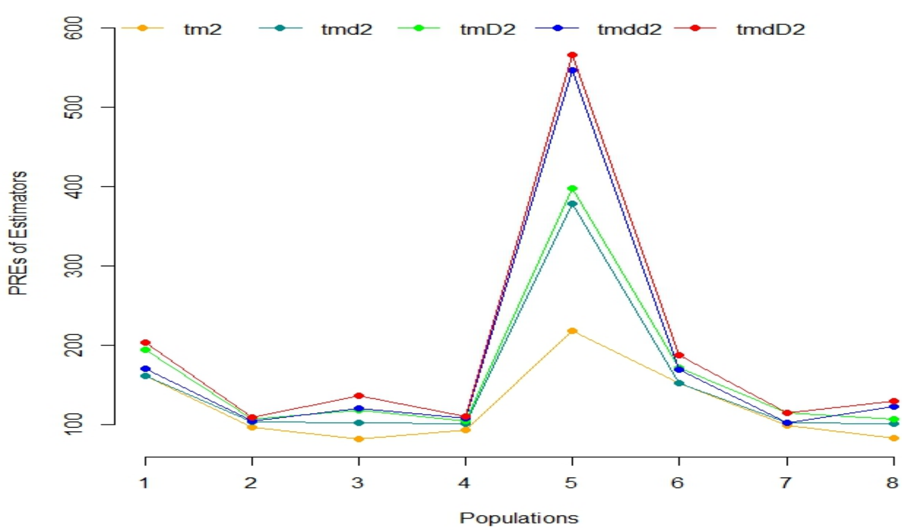

Figure 2. Comparison of PREs of different estimators for Populations 1-8 under ***Strategy-II***

Table 2. *PREs* of the different estimators with respect to s_y^{*2}.

Estimators	*Gamma distribution* Value of ρ_{yx} 0.6	0.7	0.8	0.9	*Normal distribution* Value of ρ_{yx} 0.6	0.7	0.8	0.9
Strategy I								
t_0	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
t_{m_1}	100.02	100.02	100.02	100.03	80.66	88.57	100.73	101.73
t_{md_1}	107.55	108.20	110.76	110.90	105.02	108.57	110.18	111.28
t_{mr_1}	107.56	108.23	110.77	110.89	105.03	108.59	110.19	111.29
t_{mD_1}	107.72	108.38	110.94	111.06	105.16	108.72	110.35	111.45
t_{mdd_1}	140.53	150.67	170.91	181.14	124.38	144.89	167.84	183.15
t_{mdD_1}	140.55	150.80	171.09	181.31	124.52	145.04	168.01	183.32
Strategy II								
t_{m_2}	100.03	100.03	100.03	100.03	83.79	97.62	101.68	102.32
t_{md_2}	110.50	111.41	115.22	115.38	106.94	112.00	114.34	115.96
t_{mr_2}	110.55	111.52	115.20	115.38	106.95	112.03	114.36	115.97
t_{mD_2}	110.71	111.66	115.37	115.55	107.08	112.15	114.51	116.12
t_{mdd_2}	164.09	183.90	228.93	255.13	136.26	172.59	221.56	260.62
t_{mdD_2}	165.25	184.11	229.11	255.30	136.40	172.74	221.73	260.78
Strategy III								
t_{m_3}	100.01	100.01	100.01	100.01	77.82	86.72	100.24	101.60
t_{md_3}	102.55	102.80	103.60	103.64	101.72	102.90	103.42	103.76
t_{mr_3}	102.58	102.80	103.60	103.64	101.74	102.92	103.43	103.77
t_{mD_3}	102.74	102.94	103.78	103.81	101.88	103.05	103.59	103.93
t_{mdd_3}	111.46	113.66	117.41	119.06	107.53	112.45	116.89	119.37
t_{mdD_3}	111.62	113.80	117.59	119.23	107.68	112.61	117.06	119.54

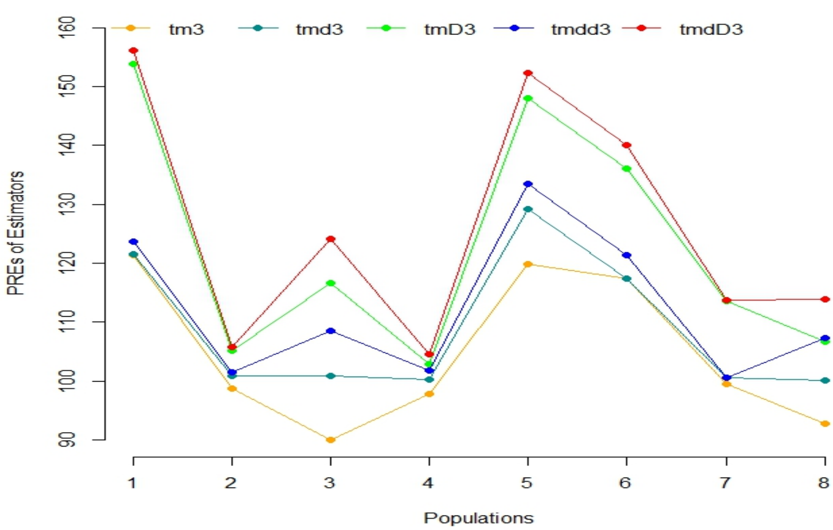

Figure 3. Comparison of PREs of different estimators for Populations 1-8 under ***Strategy-III***

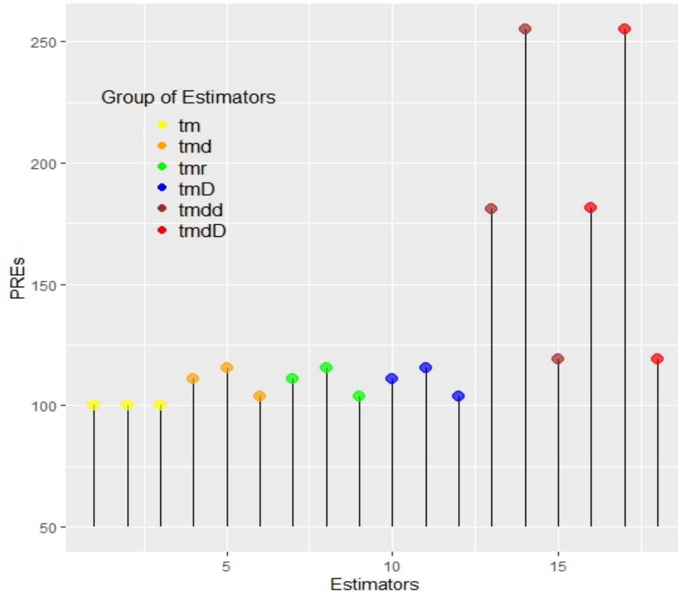

Figure 4. Comparison of PREs of different estimators in all the strategies based on Gamma distribution when $\rho_{yx} = 0.9$

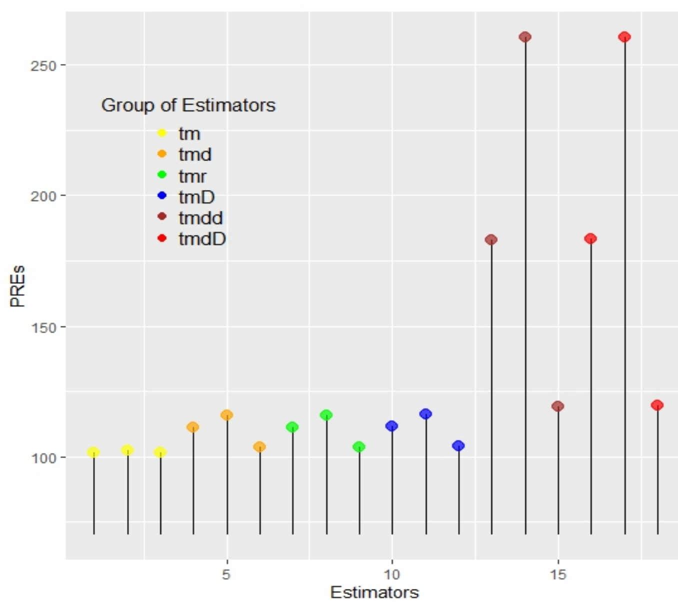

Figure 5. Comparison of PREs of different estimators in all the strategies based on Normal distribution when $\rho_{yx} = 0.9$

Interpretation of the results:

From Table 1, we report that:

(i) All the estimators, excluding the usual ratio estimators $t_{m_i}(i = 1,2,3)$, perform well in all the Populations 1-8 under each strategy. We see that the performances of usual ratio estimators t_{m_i} are good in Populations 1, 5, & 6, where the condition $\lambda^*_{210} > \frac{C_x}{2}$ holds. On the other hand, the performances of t_{m_i} are poor in Populations 2, 3, 4, 7 & 8 because the condition $\lambda^*_{210} > \frac{C_x}{2}$ does not hold in these populations.

(ii) The proposed estimators $t_{mdd_i}(i = 1,2,3)$ (constructed using the rank of an auxiliary variable) are paralleling more efficient than the existing estimators t_{md_i} or t_{mr_i}, respectively, which are formulated using the original information on an auxiliary variable. Thus, it is remarkable that the efficiency of usual difference-type estimators may be increased just by introducing the dual use of an auxiliary variable.

(iii) Similarly, the improved versions $t_{mdD_i}(i = 1,2,3)$ of the proposed estimators t_{mdd_i} also show their appreciable behaviors over the existing estimators t_{mD_i} respectively in terms of gain in percentage relative efficiencies. Thus, the efficiency of the optimal version of the usual difference estimator may also be increased using the rank of an auxiliary variable.

(iv) The proposed optimal estimators $t_{mdD_i}(i = 1,2,3)$ are the most efficient estimators among all the estimators discussed in Table 1 in the corresponding strategies.

(v) We see that the performances of all the estimators under *Strategy II* are superior to those of *Strategy I* and *Strategy III*. Thus, *Strategy II* is reasonably preferable over *Strategy I* and *Strategy III* when the information at different levels is available on an auxiliary variable.

(vi) In view of the arguments (iv) and (v), we can easily say that the efficiency of the proposed estimator t_{mdD_2} is highest among all the estimators considered in this study, which is evidently demonstrated in Table 1.

(vii) The merits of the proposed estimators based on the comparative results in Table 1 can be clearly visualized in Figures 1, 2, and 3 for Strategies I, II, and III, respectively.

Similar conclusions (as discussed above) for the proposed estimators can be drawn from the results in Table 2, which is based on the simulation study. An instant view of the results in Table 2 for gamma and normal distributions at the correlation value 0.9 can be obtained from the two scatter plots, which are displayed in Figures 4 and 5. Similar plots can be obtained for the rest of the correlation values for both distributions.

In Table 3 *(given in Appendix)*, we have demonstrated the values of estimates of all the estimators at their optimum situations considered in this study. These estimated values are based on the sample drawn from Population 6. It is observed that the estimates obtained

from proposed estimators $t_{mdd_i}(i = 1,2,3)$ and t_{mdD_i} based on the selected sample are very close to the true value of the parameter.
On the basis of the above arguments, we can easily say that the use of the rank of an auxiliary variable is capable enough to enhance the efficiency of the estimators to the next level, as the proposed estimators present considerable improvements over other existing estimators in estimating the population variance in the presence of random non-response under both empirical and simulation studies. Therefore, this comparative study may be appreciably extrapolated in general practice.

7. Conclusions

From the aforementioned results and discussions, it may be concluded that the efficiency of usual difference-type estimators may be easily increased without using any new (more than one) auxiliary variable, just by introducing the dual of an auxiliary variable. The proposed estimators, which are constructed using the rank (dual) of an auxiliary variable, are capable of providing increased efficiency in three different strategies of random non-response. The proposed difference-type estimators show better gain in terms of percentage relative efficiencies over the existing relevant estimators considered in this study for the corresponding situations of random non-response. The performances of the optimal versions of the proposed difference-type estimators are superior to all other estimators discussed in this study in respective situations at various amounts of correlations.

Hence, looking at their charming behaviors, they may be encouragingly recommended for real-life situations when faced with missing-at-random problems. The strengths of the proposed model are as follows: it may provide efficient results for both positive and negative correlations, it may be fruitfully appreciable for highly positively correlated datasets, and it may be highly preferable when the information is available only on a single auxiliary variable. On the other hand, the weakness may lie in the fact that the proposed model may not give an attractive result for the low-correlated datasets, and it may not be preferable when the information on multi-auxiliary variables is available. For future research, the present work may be extended to various sampling schemes such as successive sampling, two-phase sampling, stratified sampling, etc, for the estimation of mean, variance, and other population parameters.

Acknowledgement

Authors are highly thankful to Vellore Institute of Technology Vellore for providing 'VIT SEED GRANT (RGEMS)-Sanction Order No. SG20230035' for carrying the present research work.

Conflict of interest

There is no conflict of interest associated with the present article.

References

Ahmad, S., Adichwal, N. K., Aamir, M., Shabbir, J., Alsadat, N., Elgarhy, M. and Ahmad, H., (2023). An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling. *Scientific Reports*, 13(1), 21444.

Alam, S., Shabbir, J., (2020). Calibration estimation of mean by using double use of auxiliary information. *Commun. Stat. Simul. Comput.*, pp. 1–19.

Almulhim, F. A., Aljohani, H. M., Aldallal, R., Mustafa, M. S., Alsolmi, M. M., Elshenawy, A. and Alrashidi, A., (2024). Estimation of finite population mean using dual auxiliary information under non-response with simple random sampling. *Alexandria Engineering Journal*, 100, pp. 286–299.

Anderson, T. W., (1958). An Introduction to Multivariate Statistical Analysis. New York: *Wiley Series in Probability and Statistics*.

Belili, M. C., Alshangiti, A. M., Gemeay, A. M., Zeghdoudi, H., Karakaya, K., Bakr, M. E., Balogun, O. S., Atchadé, M. N. and Hussam, E., (2023). Two-parameter family of distributions: Properties, estimation, and applications. *AIP Advances*, 13(10), Available from: https://doi.org/10.1063/5.0173532.

Bhushan S., Pandey, A. P., (2021). Optimal estimation of population variance in the presence of random non-response using simulation approach. *J Stat Comput Simul*. Available from: https://doi.org/10.1080/00949655.2021.1948547.

Bhushan, S., Pandey, S., (2025). Optimal random non-response framework for mean estimation on current occasion. Commun. *Stat. - Theory Methods*, 54(4), pp. 1205–1231.

Cochran, W. G., (1977). Sampling techniques, 3rd ed. New York: *John Wiley and Sons*.

Daraz, U., Wu, J., Agustiana and D., Emam, W., (2025). Finite population variance estimation using Monte Carlo simulation and real life application. *Symmetry*, 17(1), 84.

Das, A. K., (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya*, c, 40, pp. 139–148.

Das A. K., Tripathi T. P., (1978). Use of auxiliary information in estimating the finite population variance. *Sankhya* 34, 19.

Hussain, I., Haq, A., (2019). A New Family of Estimators for Population Mean with Dual Use of the Auxiliary Information. *J. Stat. Theory Pract.*, 13(1), 23.

Irfan, M., Javed, M., Bhatti, S. H., Raza, M. A. and Ahmad, T., (2020). Almost unbiased optimum estimators for population mean using dual auxiliary information. *Journal of King Saud University-Science*, 32(6), pp. 2835–2844.

Isaki, C. T., (1983). Variance estimation using auxiliary information. *Journal of the American Statistical Association*, 78(381), pp. 117–123.

Javed, S., Masood, S. and Shokri, A., (2023). Generalized Class of Finite Population Variance in the Presence of Random Non response Using Simulation Approach. *Complexity*, 2023(1), 6643435.

Khodja, N., Gemeay, A. M., Zeghdoudi, H., Karakaya, K., Alshangiti, A. M., Bakr, M. E., Balogun, O. S., Muse, A. H. and Hussam, E., (2023). Modeling voltage real data set by a new version of Lindley distribution. *IEEE Access*, 11, pp. 67220–67229.

Kumar, S., (2014). Variance estimation in presence of random non-response. *Journal of Reliability and Statistical Studies*, pp. 65–70.

Maddala, G. S., (1992). Introduction to econometrics (2nd Edit.). New York: *Macmillan*.

McNeil, D. R., (1977) Interactive Data Analysis. *Wiley*.

Rubin, D. B., (1976). Inference and missing data. *Biometrika*, 63(3), pp. 581–592.

Satici, E., Kadilar, C., (2011). Ratio Estimator for the Population Mean at the Current Occasion in the Presence of Non-Response in Successive Sampling. *Hacettepe Journal of Mathematics and Statistics*, 40(1), pp. 115–124.

Sharma, A. K., Singh, A. K., (2020). Estimation of population variance under an imputation method in two-phase sampling. Proceedings of the National Academy of Sciences, India Section A: *Physical Sciences*, 90, pp. 185–191.

Singh, G. N., Usman, M., (2022). Some Optimal Estimators of Finite Population Variance Using Dual of Auxiliary Variable in the Presence of Random Non-Response. *Journal of Statistical Theory and Practice*, 16(4), 65.

Singh, H. P., Upadhyaya, L. N. and Namjoshi, U. D., (1988). Estimation of finite population variance. *Current Science*, pp. 1331–1334.

Singh, J., Pandey, B.N. and Hirano, K., (1973). On the utilization of a known coefficient of kurtosis in the estimation procedure of variance. *Ann Inst Stat Math*, 25, pp. 51–55.

Singh, S., (2003). Advanced Sampling Theory With Applications. (Vol. 2). *Springer Science and Business Media*.

Singh, S., Horn, S., (1998). An alternative estimator in multi-character surveys. *Metrika*, pp. 99–107.

Singh, S., Joarder, A. H., (1998). Estimation of finite population variance using random non-response in survey sampling. *Metrika*, 47(1), pp. 241–249.

Singh S., Joarder A. H. and Tracy D. S., (2000). Regression type estimators for random non-response in survey. Sampling. *Statistica LX 1*, pp. 39–43.

Upadhyaya, L. N., Singh, H. P. (2001). Estimation of the population standard deviation using auxiliary information. *Am. J. Math. Manag. Sci*, 21(3-4), pp. 345–358.

Yaqub, M., Shabbir, J., and Gupta, S. N., (2017). Estimation of population mean based on dual use of auxiliary information in non-response. *Communications in Statistics-Theory and Methods*, 46(24), pp. 12130-12151.

Yasmeen, U., Noor-ul-Amin, M. and Hanif, M., (2019). Exponential estimators of finite population variance using transformed auxiliary variables. Proceedings of the National Academy of Sciences, India Section A: *Physical Sciences*, 89(1), pp. 185–191.

Zaman, T., Bulut, H., (2022). A new class of robust ratio estimators for finite population variance. *Scientia Iranica*.

Zaman, T., Bulut, H., (2024). A simulation study: Robust ratio double sampling estimator of finite population mean in the presence of outliers. *Scientia Iranica*, 31(15), pp. 1330–1341.

Appendix

Outline of the derivations of Theorem 1 and Theorem 2.

The proposed estimators $t_{mdd_i}(i=1,2,3)$ and t_{mdD_i} under the error transformations can be written as

$$t_{mdd_1} = S_y^2(1+\varepsilon_0) - \phi_1^*\bar{X}\varepsilon_1 - \varphi_1^*\bar{Z}_x\varepsilon_2 \tag{42}$$

$$t_{mdd_2} = S_y^2(1+\varepsilon_0) - \phi_2^*\bar{X}\varepsilon_1^* - \varphi_2^*\bar{Z}_x\varepsilon_2^* \tag{43}$$

$$t_{mdd_3} = S_y^2(1+\varepsilon_0) + \phi_3^*\bar{X}(\varepsilon_1 - \varepsilon_1) + \varphi_3^*\bar{Z}_x(\varepsilon_2 - \varepsilon_2^*) \tag{44}$$

$$t_{mdD_1} = \alpha_1^* S_y^2(1+\varepsilon_0) - \beta_1^*\bar{X}\varepsilon_1 - \gamma_1^*\bar{Z}_x\varepsilon_2 \tag{45}$$

$$t_{mdD_2} = \alpha_2^* S_y^2(1+\varepsilon_0) - \beta_2^*\bar{X}\varepsilon_1^* - \gamma_2^*\bar{Z}_x\varepsilon_2^* \tag{46}$$

$$t_{mdD_3} = \alpha_3^* S_y^2(1+\varepsilon_0) + \beta_3^*\bar{X}(\varepsilon_1 - \varepsilon_1^*) + \gamma_3^*\bar{Z}_x(\varepsilon_2 - \varepsilon_2^*) \tag{47}$$

The above equations can be rewritten as

$$tm_{dd_1} - S_y^2 = S_y^2\varepsilon_0 - \phi_1^*\bar{X}\varepsilon_1 - \varphi_1^*\bar{Z}_x\varepsilon_2 \tag{48}$$

$$tm_{dd_2} - S_y^2 = S_y^2\varepsilon_0 - \phi_2^*\bar{X}\varepsilon_1^* - \varphi_2^*\bar{Z}_x\varepsilon_2^* \tag{49}$$

$$tm_{dd_3} - S_y^2 = S_y^2\varepsilon_0 + \phi_3^*\bar{X}(\varepsilon_1^* - \varepsilon_1) + \varphi_3^*\bar{Z}_x(\varepsilon_2^* - \varepsilon_2) \tag{50}$$

$$t_{mdD_1} - S_y^2 = S_y^2\{\alpha_1^*(1+\varepsilon_0) - 1\} - \beta_1^*\bar{X}\varepsilon_1 - \gamma_1^*\bar{Z}_x\varepsilon_2 \tag{51}$$

$$t_{mdD_2} - S_y^2 = S_y^2\{\alpha_2^*(1+\varepsilon_0) - 1\} - \beta_2^*\bar{X}\varepsilon_1^* - \gamma_2^*\bar{Z}_x\varepsilon_2^* \tag{52}$$

$$t_{mdD_3} - S_y^2 = S_y^2\{\alpha_3^*(1+\varepsilon_0) - 1\} + \beta_3^*\bar{X}(\varepsilon_1^* - \varepsilon_1) + \gamma_3^*\bar{Z}_x(\varepsilon_2^* - \varepsilon_2) \tag{53}$$

Taking the expectation of both sides of equations (8.7)-(8.12), we can easily get the biases of the proposed estimators. *Hence the proof of Theorem 1.*

Now, squaring both sides of the above equations and ignoring the terms of errors having power greater than two, we get

$$\begin{aligned}(t_{mdd_1} - S_y^2)^2 =& S_y^4\left[\varepsilon_0^2 + \phi_1^{*2}R_1^2\varepsilon_1^2 + \varphi_1^{*2}R_2^2\varepsilon_2^2 + 2\phi_1^*\varphi_1^*R_1R_2\varepsilon_1\varepsilon_2 - 2\phi_1^*R_1\varepsilon_0\varepsilon_1\right.\\ &\left.-2\varphi_1^*R_2\varepsilon_0\varepsilon_2\right]\end{aligned} \tag{54}$$

$$\begin{aligned}(t_{mdd_2} - S_y^2)^2 =& S_y^4\left[\varepsilon_0^2 + \phi_2^{*2}R_1^2\varepsilon_1^{*2} + \varphi_2^{*2}R_2^2\varepsilon_2^{*2} + 2\phi_2^*\varphi_2^*R_1R_2\varepsilon_1^*\varepsilon_2^* - 2\phi_2^*R_1\varepsilon_0\varepsilon_1^*\right.\\ &\left.-2\varphi_2^*R_2\varepsilon_0\varepsilon_2^*\right]\end{aligned} \tag{55}$$

$$\begin{aligned}(t_{mdd_3} - S_y^2)^2 =& S_y^4\left[\varepsilon_0^2 + \phi_3^{*2}R_1^2(\varepsilon_1^{*2} - \varepsilon_1^2) + \varphi_3^{*2}R_2^2(\varepsilon_2^{*2} - \varepsilon_2^2) + 2\phi_3^*\varphi_3^*R_1R_2(\varepsilon_1^*\varepsilon_2^* - \varepsilon_1\varepsilon_2)\right.\\ &\left.-2\phi_3^*R_1(\varepsilon_0\varepsilon_1^* - \varepsilon_0\varepsilon_1) - 2\varphi_3^*R_2(\varepsilon_0\varepsilon_2^*\varepsilon_0\varepsilon_2)\right]\end{aligned} \tag{56}$$

$$(t_{mdD_1} - S_y^2)^2 = S_y^4 \left[\alpha_1^{*2}(1+\varepsilon_0^2+2\varepsilon_0) + \beta_1^{*2}R_1^2\varepsilon_1^2 + \gamma_1^{*2}R_2^2\varepsilon_2^2 - 2\alpha_1^*\beta_1^*R_1(\varepsilon_1+\varepsilon_0\varepsilon_1)\right.$$
$$-2\alpha_1^*\gamma_1^*R_2(\varepsilon_2+\varepsilon_0\varepsilon_2) + 2\beta_1^*\gamma_1^*R_1R_2\varepsilon_1\varepsilon_2 - 2\alpha_1^*(1+\varepsilon_0)$$
$$\left.+2\beta_1^*R_1\varepsilon_1 + 2\gamma_1^*R_2\varepsilon_2 + 1\right] \quad (57)$$

$$(t_{mdD_2} - S_y^2)^2 = S_y^4 \left[\alpha_2^{*2}(1+\varepsilon_0^2+2\varepsilon_0) + \beta_2^{*2}R_1^2\varepsilon_1^{*2} + \gamma_2^{*2}R_2^2\varepsilon_2^{*2} - 2\alpha_2^*\beta_2^*R_1(\varepsilon_1^*+\varepsilon_0\varepsilon_1^*)\right.$$
$$-2\alpha_2^*\gamma_2^{*2}R_2(\varepsilon_2^*+\varepsilon_0\varepsilon_2^*) + 2\beta_2^*\gamma_2^*R_1R_2\varepsilon_1^*e_2^* - 2\alpha_2^*(1+\varepsilon_0)$$
$$\left.+2\beta_2^*R_1\varepsilon_1^* + 2\gamma_2^*R_2\varepsilon_2^* + 1\right] \quad (58)$$

$$(t_{mdD_3} - S_y^2)^2 = S_y^4 \left[\alpha_3^{*2}(1+\varepsilon_0^2+2\varepsilon_0) + \beta_3^{*2}R_1^2(\varepsilon_1-\varepsilon_1^*)^2 + \gamma_3^{*2}R_2^2(\varepsilon_2-\varepsilon_2^*)^2\right.$$
$$-2\alpha_3^*\beta_3^*R_1(1+\varepsilon_0)(\varepsilon_1-e_1^*) - 2\alpha_3^*\gamma_3^{*2}R_2^2(1+\varepsilon_0)(\varepsilon_2-\varepsilon_2^*)$$
$$+2\beta_3^*\gamma_3^*R_1R_2(\varepsilon_1-\varepsilon_1^*)(\varepsilon_2-\varepsilon_2^*) - 2\alpha_3^*(1+\varepsilon_0)$$
$$\left.+2\beta_3^*R_1(\varepsilon_1-\varepsilon_1^*) + 2\gamma_3^*R_2(\varepsilon_2-\varepsilon_2^*) + 1\right] \quad (59)$$

Taking expectations of both sides of equations (8.13)-(8.18), we can easily get the MSEs of the proposed estimators to the first order of approximations, are given as

$$MSE(t_{mdd_i}) = S_y^4 \left[f_2\lambda_{400}^* + \phi_i^{*2}R_1^2f_iC_x^2 + \varphi_i^{*2}R_2^2f_iC_{z_x}^2 + 2\phi_i^*\varphi_i^*R_1R_2f_iC_{xz_x}\right.$$
$$\left.-2\phi_i^*R_1f_i\lambda_{210}C_x - 2\varphi_i^*R_2f_i\lambda_{201}C_{z_x}\right] \quad (60)$$

$$MSE(t_{mdD_i}) = S_y^4 \left[\alpha_i^{*2}(1+f_2\lambda_{400}^*) + \beta_i^{*2}R_1^2f_iC_x^2 + \gamma_1^{*2}R_2^2f_iC_{z_x}^2 - 2\alpha_i^*\beta_i^*R_1f_i\lambda_{210}C_x\right.$$
$$\left.-2\alpha_i^*\gamma_i^*R_2f_i\lambda_{201}C_{r_x} + 2\beta_1^*\gamma_1^*R_1R_2f_iC_{xz_x} - 2\alpha_1^* + 1\right] \quad (61)$$

Now, differentiating partially the equations (8.19) and (8.20) with respect to the constants $\phi_i^*(i=1,2,3)$, φ_i^*, α_i^*, β_i^* and γ_i^* and equating the resultant equations to zero then solving them we can easily obtain their optimum values as given in equations (5.7)-(5.11).

Finally, by putting these optimum values in equations (8.19) and (8.20) appropriately, we can easily obtain the minimum MSEs of the proposed estimators as given in equations (4.9) and (4.10). *Hence the proof of Theorem 2.*

Table 3. Estimates of the various estimators based on a sample drawn from **Population 6** where the true value of the parameter is **1176526**

Sample		Respondents			
y	x	y	x	Estimators	Estimates
348.334	408.978	38.067	40.775	t_0	1230451
494.730	639.571	3520.361	1248.761	t_{m_1}	1027631
1692.817	413.777	57.684	139.628	t_{m_2}	1131690
43.229	42.808	440.518	323.028	t_{m_3}	1355048
298.351	756.169	571.487	114.899	t_{mr_1}	1230035
440.518	323.028	43.229	42.808	t_{mr_2}	1230258
197.244	56.908	635.774	870.720	t_{mr_3}	1230674
38.067	40.775	2610.572	2131.048	t_{md_1}	988258.4
571.487	114.899	494.730	639.571	t_{md_2}	1123362
557.656	1045.106	348.334	408.978	t_{md_3}	1365554
848.317	907.700	1372.439	1229.752	t_{mD_1}	858921.0
540.696	939.460	197.244	56.908 13	t_{mD_2}	1000448
3520.361	1248.761			t_{mD_3}	1178178
386.490	100.964			t_{mdd_1}	1161129
1372.439	1229.752			t_{mdd_2}	1186860
3585.406	1337.852			t_{mdd_3}	1256182
57.684	139.628			t_{mdD_1}	1015148
635.774	870.720			t_{mdD_2}	1068391
388.869	553.266			t_{mdD_3}	1088691
2610.572	2131.048				