Original Research Article

ON THE EFFICIENCY OF ALMOST UNBIASED MEAN IMPUTATION WHEN POPULATION MEAN OF AUXILIARY VARIABLE IS UNKNOWN

Abstract

Human-assisted surveys, such as medical and social science surveys, are frequently plagued by non-response or missing observations. Several authors have devised different imputation algorithms to account for missing observations during analyses. Nonetheless, several of these imputation schemes' estimators are based on known population mean $\bar{X}$ of auxiliary variable. In this paper, a new class of almost unbiased imputation method that uses $\bar{x}_n$ as an estimate of $\bar{X}$ is suggested. Using the Taylor series expansion technique, the MSE of the class of estimators presented was derived up to first order approximation. Conditions were also specified for which the new estimators were more efficient than the other estimators studied in the study. The results of numerical examples through simulations revealed that the suggested class of estimators is more efficient.

Keywords: Estimators, Imputation Scheme, Population Mean, Study Variable

1.0 Introduction

Numerous studies in the field of sampling survey have estimators for estimating population parameters like population mean, population variance, standard deviation etc. under the assumption that complete information about sampling units is available. Some of these authors include Singh et al,[1], Sahai et al. [2], Srivastava et al. [3], Ahmed et al. [4], Audu et al. [5], Audu et al. [6], Muili et al. [7] have worked extensively in that direction. Authors like Singh and Tailor [8], Sisodia et al. [9], Khoshnevisan et al. [10], Singh et al. [11], Singh and Audu [12], Ahmed et al. [13] and Audu et al. [14] utilized coefficient of variation of auxiliary variable in the estimators' formulation and obtained highly efficient estimators. The estimators in the aforementioned literatures assumed that information on sampling units drawn from the population is completely available. However, this assumption is often violated due to non-response as a result of refusal to answer questions, inaccessibility to respondents, etc. In such situations, responses of non-respondents can be imputed using imputation techniques.
Imputation is the process of replacing missing data with substituted values. When substituting for a data point, it is known as "unit imputation"; when substituting for a component of a data point, it is known as "item imputation" (Singh [35]). There are three main problems that missing data causes. It can introduce a substantial amount of bias, make the handling and analysis of the data more arduous, and create reductions in efficiency (Barnard and Meng [23]). Missing data due to non-response can create problems for analyzing data and imputation is seen as a way to avoid pitfalls involved with likewise of cases that have missing values. That is to say, when one or more

values are missing for a case, most statistical packages default to discarding any case that has a missing value, which may introduce bias or affect the representativeness of the results. Imputation preserves all cases by replacing missing data with an estimated value based on other available information. Once all missing values have been imputed, the data set can then be analyze using standard techniques for complete data (Gelman and Jennifer [25]). There have been many theories embraced by scientists to account for missing data but the majority of them introduce bias.

Survey such as in medical and social science etc. conducted by human are often characterized by non-response. Hansen and Hurwitz [26] first discussed the issue of non-response and imputation methods to deal with non-response issues were suggested by several scholars like Singh and Horn [37], Singh and Deo [36], Ahmed *et al.* [40], Wang and Wang [39], Kadilar and Cingi [27], Toutenburg *et al*. [38], Singh (2009), Diana and Perri [24], Al-Omari *et al.* [17], Singh *et al.* [34], Mishra *et al.* [30], Singh and Gogoi [33], Singh et al. [32], Prasad [31], Audu *et al* ([19]-[22]) and Audu and Singh [18] are some of the most recent imputation methods. However, all the estimators of the schemes proposed by aforementioned authors are functions of population mean of auxiliary variable ($\bar{X}$) and if $\bar{X}$ is unknown, the schemes can not be applied to real life situations. This study, therefore, implored the concept of two-phase sampling in which a large sample of size $n\ (n<N)$ is taking to estimate $\bar{X}$ thereby addressing the problem of complete information about the auxiliary variable.

1.1 Notations

The following notations have been used.

Y: Study variable.

X: Auxiliary variable.

$\bar{X}, \bar{Y}$: The population mean of the variables X and Y respectively.

r: Response units size

n, N: Size of the sample, Population size.

$\bar{x}_n$: The sample mean of X based on sample of size n.

$\bar{x}_r$: The sample mean of X response units

$\bar{y}_r$: The sample mean of Y response units.

$S_x^2=(N-1)^{-1}\sum_{i=1}^{N}\left(x_i-\bar{X}\right)^2$ The population mean squares of X.

$S_Y^2=(N-1)^{-1}\sum_{i=1}^{N}\left(y_i-\bar{Y}\right)^2$: The population mean squares of Y.

$S_{YX}=\rho S_Y S_X$ is the covariance between variables Y and X.

$\psi_{r,N}=r^{-1}-N^{-1}$: Correction factor for response

$\psi_{r,n}=r^{-1}-n^{-1}$: Correction factor for non-response

$C_Y=S_Y/\bar{Y}$: Coefficient of variations Y

$C_X=S_X/\bar{X}$: Coefficient of variations Y

$R=\bar{Y}/\bar{X}$: Ratio Y to X

$$\rho_{YX} = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2 \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2}}:$$ Correlation coefficient of Y and X

2.0 Some Existing Imputation Schemes and their Estimators

Let Φ denotes the set of r units' response and Φ^c denotes the set of $n-r$ units' non-response or missing out of n units sampled without replacement from the N units' population Ω_N.

Under mean method of imputation, values found missing due to non-response are to be replaced by the mean of the rest of observed values (Kalton [28]). The study variable thereafter, takes the form given as,

$$y_{.i} = \begin{cases} y_i & \text{i} \in \Phi \\ \bar{y}_r & \text{i} \in \Phi^c \end{cases} \tag{2.1}$$

Under the method of imputation, sample mean denoted by $\hat{\mu}_0$ can be derived as

$$\hat{\mu}_0 = r^{-1}\sum_{i \in R} y_r \tag{2.2}$$

The bias and variance of $\hat{\mu}_0$ is given in (2.3) and (2.4)

$$Bias(\hat{\mu}_0) = 0 \tag{2.3}$$

$$Var(\hat{\mu}_0) = \psi_{r,N} S_Y^2 \tag{2.4}$$

where $\psi_{r,N} = r^{-1} - N^{-1}$, $S_Y^2 = \left(N-1\right)^{-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$, $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$

Lee *et al.*, [29] proposed ratio imputation method as given (2.5)

$$y_{.i} = \begin{cases} y_i & \text{i} \in \Phi \\ \hat{\beta}\text{x}_\text{i} & \text{i} \in \Phi^c \end{cases} \tag{2.5}$$

where $\hat{\beta} = \sum_{i=1}^{r} y_i / \sum_{i=1}^{r} x_i = \bar{y}_r / \bar{x}_r$

Under the method of imputation, estimator of population mean denoted by $\hat{\mu}_1$, as in (2.5)

$$\hat{\mu}_1 = \bar{y}_r \bar{x}_n / \bar{x}_r \tag{2.6}$$

The Bias and MSE of $\hat{\mu}_1$ up $\text{O}\left(n^{-1}\right)$ is given as:

$$Bias(\hat{\mu}_1) = \bar{Y}^{-1}\psi_{r,n}\left(R^2 S_x^2 - \rho_{YX} S_x S_y\right) \tag{2.7}$$

$$MSE(\hat{\mu}_1) = MSE\left(\bar{y}_r\right) + \left(\frac{1}{r} - \frac{1}{n}\right)\left(S_y^2 + R^2 S_X^2 - 2R\rho_{YX} S_Y S_X\right) \tag{2.8}$$

where $S_{YX} = \rho_{YX} S_Y S_X$, , $S_X^2 = \left(N-1\right)^{-1} \sum_{i=1}^{N} \left(x_i - \bar{X}\right)^2, \bar{X} = N^{-1} \sum_{i=1}^{N} x_i, \psi_{r,n} = r^{-1} - n^{-1}, R = \bar{Y} / \bar{X}$

Singh and Horn [37] utilized information from imputed values for responding and non-responding units as well, thereafter giving study variable the form given by (2.9).

$$y_{.i} = \begin{cases} \lambda \frac{n}{r} y_i + (1-\lambda)\hat{\beta} x_i, & \text{i} \in \Phi \\ (1-\lambda)\hat{\beta} x_i, & \text{i} \in \Phi^c \end{cases} \tag{2.9}$$

The estimator of population mean denoted by $\hat{\mu}_2$ as well as its bias and MSE are given as

$$\hat{\mu}_2 = \bar{y}_r \left(\lambda + \left(1-\lambda\right) \bar{x}_n \bar{x}_r^{-1}\right) \tag{2.10}$$

$$Bias(\hat{\mu}_2) = \bar{Y}^{-1}\left(1-\alpha\right)\psi_{r,n}\left(R^2 S_x^2 - \rho_{YX} S_x S_y\right) \tag{2.11}$$

$$MSE\left(\hat{\mu}_2\right) = MSE\left(\bar{y}_r\right) - \psi_{r,n}\left(S_y^2 + R^2 S_x^2 - 2RS_{xy}\right) - \psi_{r,n} R^2 \alpha^2 S_x^2 \tag{2.12}$$

where $\alpha = 1 - \rho_{YX} \frac{S_y}{RS_x}$

Ahmed *et al*. [40] proposed imputation scheme for population means estimators which is applicable when the study and auxiliary variables are either positively or negatively correlated, using power transformation.

$$\bar{y}_{.i} = \begin{cases} y_i & i \in \Phi \\ \frac{1}{n-r}\left[n\bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r}\right)^{\beta_1} - r\bar{y}_r\right] & i \in \Phi^C \end{cases} \tag{2.13}$$

Under the (2.13) method, the resultant estimator of the population mean $\bar{Y}$ as well as bias and MSE are given as

$$\hat{\mu}_3 = \bar{y}_r \left(\frac{\bar{X}}{\bar{x}_r}\right)^{\beta_1} \tag{2.14}$$

$$B\left(\hat{\mu}_3\right) = \bar{Y}^{-1}\psi_{n,N}\left(\frac{\beta_1\left(\beta_1+1\right)R^2 S_X^2}{2} - \beta_1 R \rho_{YX} S_Y S_X\right) \tag{2.15}$$

where $\beta = \rho_{YX} \frac{S_y}{RS_x}$

$$MSE\left(\hat{\mu}_3\right)_{\min} = \psi_{r,N}(1-\rho_{YX}^2)S_Y^2 \qquad (2.16)$$

Singh *et al*., [32] proposed Exponential-Type Compromised Imputation scheme to minimize the effect of distance between $\bar{X}$ and $\bar{x}_r$ on the efficiency of Ahmed *et al.* [40] as

$$y_{.i} = \begin{cases} v\dfrac{n}{r}y_i\left(1-v\right)\bar{y}_r \exp\left(\dfrac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) & if \;\; i \in \Phi \\ \left(1-v\right)\bar{y}_r \exp\left(\dfrac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r}\right) & if \;\; i \in \Phi^C \end{cases} \qquad (2.17)$$

The point estimator of population mean $\bar{Y}$ under the proposed method of imputation is:

$$\hat{\mu}_4 = v\bar{y}_r + (1-v)\bar{y}_r \exp\left(\left(\bar{X}-\bar{x}_r\right)\left(\bar{X}+\bar{x}_r\right)^{-1}\right) \qquad (2.18)$$

$$Bais\left(\hat{\mu}_4\right) = (1-v)\psi_{r,N}\bar{Y}^{-1}\left(\frac{3}{8}R^2S_x^2 - \frac{1}{2}\rho_{YX}S_xS_y\right) \qquad (2.19)$$

$$MSE\left(\hat{\mu}_4\right)_{\min} = \psi_{r,N}S_Y^2\left(1-\rho_{YX}^2\right) \qquad (2.20)$$

Where $v = 1-2\rho_{YX}S_Y / RS_X$

Prasad [31] proposed ratio exponential imputation scheme given in (2.21) to address the problem of compromised in Singh et al [32] as

$$y_{.i} = \begin{cases} y_i & i \in \Phi \\ \dfrac{\bar{y}_r}{n-r}\left[n\eta \exp\left(\dfrac{\bar{X}-\bar{x}_r}{\bar{X}+\bar{x}_r+2\rho_{yx/\beta_2(x)}}\right) - r\right] & i \in \Phi^C \end{cases} \qquad (2.21)$$

Under this method, the resultant estimator of $\bar{Y}$ as well as the bias and MSE are given as

$$\hat{\mu}_5 = \eta\hat{t}_0 \exp\left(\left(\bar{X}-\bar{x}_r\right)\left(\bar{X}+\bar{x}+2\rho_{YX} / \beta_2\left(x\right)\right)^{-1}\right) \qquad (2.22)$$

$$Bais\left(\hat{\mu}_5\right) = \left[\left(\eta-1\right)\bar{Y} + \frac{\vartheta}{8}\psi_{r,N}\bar{Y}^{-1}\left(3\vartheta R^2S_X^2 - 4\rho_{yx}S_yS_x\right)\eta\right] \qquad (2.23)$$

$$MSE\left(\hat{\mu}_5\right)_{\min} = \frac{\bar{Y}^2\left(\psi_{r,N}\left(S_Y^2 + 0.25\vartheta^2R^2S_X^2 - \vartheta RS_{YX}\right)\right)}{\left(\bar{Y}^2 + \psi_{r,N}\left(S_Y^2 + 0.25\vartheta^2R^2S_X^2 - \vartheta RS_{YX}\right)\right)} \qquad (2.24)$$

where when $\eta = \bar{Y}^2 / \left(\bar{Y}^2 + \psi_{r,N}\left(S_y^2 + 0.25\vartheta^2R^2S_X^2 - \vartheta RS_{YX}\right)\right)$, $\vartheta = \beta_2\left(x\right)\bar{X} / \left(\beta_2\left(x\right)\bar{X} + \rho_{YX}\right)$.

Singh and Gogoi [33] Proposed imputation scheme which is applicable when X and Y are positively or negatively correlated, for population mean estimators using linear combination approach

$$y_{.i}=\begin{cases}\alpha\dfrac{n}{r}y_i\dfrac{\bar{X}}{\bar{x}_n}+(1-\alpha)\bar{y}_r\dfrac{\bar{x}_r}{\bar{X}} & i\in\Phi\\ (1-\alpha)\bar{y}_r\dfrac{\bar{x}_r}{\bar{X}} & i\in\Phi^C\end{cases} \tag{2.25}$$

The point estimator of population mean $\bar{Y}$ under proposed method of imputation is:

$$\hat{\mu}_6=\bar{y}_r\left\{\alpha\frac{\bar{X}}{\bar{x}_n}+(1-\alpha)\frac{\bar{x}_r}{\bar{X}}\right\} \tag{2.26}$$

where α_1 is an unknown parameter to be estimated.

The bias, mean square error and minimum mean square error are given by:

$$Bias(\hat{\mu}_6)=\bar{Y}^{-1}\psi_{n,N}\left\{\alpha R^2S_X^2+(1-2\alpha)\rho_{YX}S_YS_X\right\} \tag{2.27}$$

$$MSE(\hat{\mu}_6)_{\min}=\left(\psi_{r,N}-\psi_{n,N}\rho_{YX}^2\right)S_Y^2$$

Audu *et al* [19] proposed some new imputation schemes as in (2.28). They incorporated filtration parameters θ_i to obtain unbiased class of estimators.

$$y_{.i}=\begin{cases}\theta_1\dfrac{n}{r}y_i & \text{i}\in\Phi\\ \dfrac{n}{n-r}\bar{y}_r\left(\theta_2\left(\dfrac{\bar{X}}{\bar{x}_r}\right)^{\kappa_1}+\theta_3\exp\left(\dfrac{\kappa_2\left(\bar{X}-\bar{x}_r\right)}{\bar{X}+\bar{x}_r}\right)\right) & \text{i}\in\Phi^{\text{c}}\end{cases} \tag{2.28}$$

$$\left.\begin{aligned}\theta_3&=4\left(2^{-1}(\kappa_1+1)S_X-\rho_{XY}S_Y\right)\rho_{XY}S_Y/\kappa_2(\kappa_1-\kappa_2/2)RS_X^2\\ \theta_2&=-\left(2^{-1}(\kappa_2+2)S_X-2\rho_{XY}S_Y\right)\rho_{XY}S_Y/\kappa_1(\kappa_1-\kappa_2/2)RS_X^2\\ \theta_1&=1+\left(2\left(4^{-1}\kappa_2(\kappa_2+2)-\kappa_1(\kappa_1+1)\right)S_X-(\kappa_2-2\kappa_1)\rho_{XY}S_Y\right)\rho_{XY}S_Y/\kappa_1\kappa_2(\kappa_1-\kappa_2/2)RS_X^2\end{aligned}\right\} \tag{2.29}$$

where $\kappa_1,\kappa_2\in(1,-1)$

The point estimator of population mean $\bar{Y}$ under proposed method of imputation as well as bias and MSE are given as:

$$\hat{\mu}_7=\bar{y}_r\left(\theta_1+\theta_2\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\kappa_1}+\theta_3\exp\left(\frac{\kappa_2\left(\bar{X}-\bar{x}_r\right)}{\bar{X}+\bar{x}_r}\right)\right) \tag{2.30}$$

$$Bias(\hat{\mu}_7)=0 \tag{2.31}$$

$$MSE(\hat{\mu}_7)=\psi_{r,N}\left(S_Y^2+\zeta^2R^2S_X^2-2\zeta\rho_{XY}RS_XS_Y\right) \tag{2.32}$$

where $\zeta = \theta_2 \kappa_1 + \theta_3 \frac{\kappa_2}{2}$

3. Proposed Imputation Schemes

Having studied the work of Audu *et al.*[19], the following imputation scheme is proposed.

Let Ω be a set of population with N units, $\Phi \subset \Omega$ with cardinality $|\Phi| = R$ and Φ^c be complement of Φ. Let $\bar{x}_n$ be an unbiased estimate of $\bar{X}$, the population mean of X based on the sample of size n, then,

$$y_{.i} = \begin{cases} f_1 \frac{n}{r} y_i & \text{i} \in \Phi \\ \frac{n}{n-r} \bar{y}_r \left(f_2 \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\tau_1} + f_3 \exp\left(\frac{\tau_2 \left(\bar{x}_n - \bar{x}_r \right)}{\bar{x}_n + \bar{x}_r} \right) \right) & \text{i} \in \Phi^{\text{c}} \end{cases} \tag{3.1}$$

where $\tau_1 \tau_2 \in (-1,1)$, $f_i, i = 1, 2, 3$ are filtration parameters, $\sum_{i=1}^{3} f_i = 1$.

3.1 Estimation Method/Procedure

The estimator of the proposed scheme is obtained as

$$\hat{t} = \frac{1}{n} \left(\sum_{i \in \Phi} y_{.i} + \sum_{i \in \Phi^c} y_{.i} \right) \tag{3.2}$$

$$\hat{t} = \frac{1}{n} \left(f_1 \frac{n}{r} \sum_{i=1}^{r} y_{.i} + \sum_{i=1}^{n-r} \bar{y}_r \left(f_2 \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\tau_1} + f_3 \exp\left(\tau_2 \left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r} \right) \right) \right) \right) \tag{3.3}$$

The estimator of the proposed scheme is given as;

$$\hat{t} = \bar{y}_r \left(f_1 + f_2 \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\tau_1} + f_3 \exp\left(\tau_2 \left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r} \right) \right) \right) \tag{3.4}$$

The Mean Square Errors of $(\hat{t})$ under Case I is defined as;

$$MSE(\hat{t})_I = \Delta \Sigma \Delta' \tag{3.5}$$

where, $\Delta = \begin{pmatrix} \frac{\partial \hat{t}}{\partial \bar{y}_r} & \frac{\partial \hat{t}}{\partial \bar{x}_n} & \frac{\partial \hat{t}}{\partial \bar{x}_r} \end{pmatrix}$ is a matrix of order 1×3, Δ' is its transpose and the variance-covariance matrix is defined as $\Sigma = \begin{pmatrix} \psi_{r,N} S_y^2 & \psi_{n,N} S_{yx} & \psi_{r,N} S_{yx} \\ \psi_{n,N} S_{yx} & \psi_{n,N} S_x^2 & \psi_{n,N} S_x^2 \\ \psi_{r,N} S_{yx} & \psi_{n,N} S_x^2 & \psi_{r,N} S_x^2 \end{pmatrix}$ is a 3×3 non-singular matrix.

On differentiating $\hat{t}$ with respect to $\bar{y}_r, \bar{x}_n, and\ \bar{x}_r$, to obtain (3.6), (3.7) and (3.8) respectively;

$$\frac{\partial \hat{t}}{\partial \bar{y}_r} = f_1 + f_2 \left(\frac{\bar{x}_n}{\bar{x}_r} \right)^{\tau_1} + f_3 \exp\left(\tau_2 \left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r} \right) \right) \tag{3.6}$$

$$\frac{\partial \hat{t}}{\partial \bar{x}_n} = \bar{y}_r \left(f_2 \tau_1 \frac{\bar{x}_n^{\tau_1 - 1}}{\bar{x}_r^{\tau_1}} + f_3 \tau_2 \left(\frac{2\bar{x}_r}{(\bar{x}_n + \bar{x}_r)^2} \right) \exp\left(\tau_2 \left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r} \right) \right) \right) \tag{3.7}$$

$$\frac{\partial \hat{t}}{\partial \bar{x}_r} = \bar{y}_r \left(-f_2 \tau_1 \frac{\bar{x}_n^{\tau_1}}{\bar{x}_r^{-(\tau_1+1)}} - f_3 \tau_2 \left(\frac{2\bar{x}_n}{(\bar{x}_n + \bar{x}_r)^2} \right) \exp\left(\tau_2 \left(\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r} \right) \right) \right) \tag{3.8}$$

By setting $\bar{x}_n = \bar{X}$, $\bar{x}_r = \bar{X}$, $\bar{y}_r = \bar{Y}$ in (3.6), (3.7) and (3.8), to get (3.9), (3.10) and (3.11) respectively,

$$\frac{\partial \hat{t}}{\partial \bar{y}_r} = f_1 + f_2 \left(\frac{\bar{X}}{\bar{X}} \right) + f_3 \exp\left(\frac{\bar{X} - \bar{X}}{\bar{X} - \bar{X}} \right) = f_1 + f_2 + f_3 = 1 \tag{3.9}$$

$$\frac{\partial \hat{t}}{\partial \bar{x}_n} = \frac{\bar{Y}}{\bar{X}} \left(f_2 \tau_1 + f_3 \frac{\tau_2}{2} \right) \tag{3.10}$$

$$\frac{\partial \hat{t}}{\partial \bar{x}_n} = -\frac{\bar{Y}}{\bar{X}} \left(f_2 \tau_1 + f_3 \frac{\tau_2}{2} \right) \tag{3.11}$$

Let $\theta = f_2 \tau_1 + f_3 \frac{\tau_2}{2}$

Substituting (3.9), (3.10), (3.11) into (3.5) to obtain the MSE of t_1 under case one as

$$MSE(\hat{t}) = \left(\psi_{r,N} S_y^2 - 2\theta (\psi_{r,N} - \psi_{n,N}) R \rho_{YX} S_y S_x + \theta^2 (\psi_{r,N} - \psi_{n,N}) R^2 S_x^2 \right) \tag{3.12}$$

Differentiating (3.12) with respect to θ, we have,

$$\theta = \frac{\rho_{YX} S_y}{R S_x} \tag{3.13}$$

Substituting (3.13) into (3.12), we obtain the minimum MSE of the estimator $\hat{t}$ under case I as

$$MSE\left(\hat{t}\right)_{I\min} = S_y^2\left(\psi_{r,N} - \left(\psi_{r,N} - \psi_{n,N}\right)\rho^2\right) \quad (3.14)$$

To obtain the expressions for $f_1,\ f_2,\ f_3$ the following system of equations are used

$$\left.\begin{aligned} & f_1 + f_2 + f_3 = 1 \\ & 0\,f_1 + \tau_1 f_2 + \frac{\tau_2}{2}\,f_3 = \frac{\rho_{YX} S_y}{RS_x} \\ & 0\,f_1 + \gamma_2 f_2 + \gamma_3 f_3 = 0 \end{aligned}\right\} \quad (3.15)$$

Solving (3.15), we obtained (3.16) as

$$\left.\begin{aligned} & f_1 = 1 - \left[\left(\gamma_3 - \gamma_2\right)\rho_{YX} S_y S_x^{-1} R^{-1}\right] / \left[\left(\tau_1\gamma_3 - 2^{-1}\tau_2\gamma_2\right)\right] \\ & f_2 = \gamma_3 \rho_{YX} S_y S_x^{-1} R^{-1} / \left(\tau_1\gamma_3 - 2^{-1}\tau_2\gamma_2\right) \\ & f_3 = -\gamma_2 \rho_{YX} S_y S_x^{-1} R^{-1} / \left(\tau_1\gamma_3 - 2^{-1}\tau_2\gamma_2\right) \end{aligned}\right\} \quad (3.16)$$

where γ_1, γ_2, and γ_3 are biases of the estimators combined in the proposed schemes define by:

$$\left.\begin{aligned} & \gamma_1 = Bias\left(\bar{y}_r\right) = 0 \\ & \gamma_2 = Bias\left(\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\tau_1}\right) = \frac{R}{\bar{\bar{Y}}}\left(\psi_{r,N}\left(\frac{\tau_1 + \tau_1^2}{2} RS_x^2 - \tau_1\rho_{YX} S_y S_x\right) - \psi_{n,N}\left(\tau_1\rho_{YX} S_y S_x - \tau_1^2 RS_x^2\right)\right) \\ & \gamma_3 = Bias\left(\bar{y}_r \exp\left(\tau_2\frac{\bar{x}_n - \bar{x}_r}{\bar{x}_n + \bar{x}_r}\right)\right) = \frac{R}{\bar{\bar{Y}}}\left(\left(\frac{\tau_2}{4} + \frac{\tau_2^2}{8}\right)\psi_{r,N}\left(RC_x^2 - \rho_{YX} S_y S_x\right) - \frac{\tau_2}{2}\psi_{n,N}\rho_{YX} S_y S_x\right) \end{aligned}\right\} \quad (3.17)$$

The Mean Square Errors of $\hat{t}$ under Case II is derived using as;

$$MSE\left(\hat{t}\right)_{II} = \Delta\Sigma^*\Delta' \quad (3.18)$$

where, $\Sigma^* = \begin{pmatrix} \psi_{r,N} S_y^2 & 0 & \psi_{r,N}\rho_{YX} S_y S_x \\ 0 & \psi_{n,N} S_x^2 & 0 \\ \psi_{r,N}\rho_{YX} S_y S_x & 0 & \psi_{r,N} S_x^2 \end{pmatrix}$ is a 3×3non-singular matrix.

Substituting (3.9), (3.10), (3.11) into (3.18) to obtain the MSE of t_1 under case II as

$$MSE\left(\hat{t}\right)_{II} = \left(\psi_{r,N} S_y^2 - 2\psi_{r,N}\theta R\rho_{YX} S_y S_x + \left(\psi_{r,N} + \psi_{n,N}\right)\theta^2 R^2 S_x^2\right) \quad (3.19)$$

Differentiating (3.19) with respect to θ , we have,

$$\theta = \frac{\psi_{r,N}\rho_{YX}S_y}{\left(\psi_{n,N}+\psi_{r,N}\right)RS_x} \tag{3.20}$$

Substitutes (3.20) into (3.19) to obtain minimum mean square error t_1 under case II as;

$$MSE\left(\hat{t}\right)_{II\min} = S_y^2\left(\psi_{r,N} - \frac{\psi_{r,N}^2\rho_{YX}^2}{\left(\psi_{n,N}+\psi_{r,N}\right)}\right) \tag{3.21}$$

To obtain the expression for f_1, f_2, f_3 the following system of equation are solved

$$\left.\begin{aligned} & f_1+f_2+f_3=1 \\ & 0\,f_1+\tau_1 f_2+\frac{\tau_2}{2}f_3=\frac{\psi_{r,N}\rho_{YX}S_y}{\left(\psi_{n,N}+\psi_{r,N}\right)RS_x} \\ & \vartheta_1 f_1+\vartheta_2 f_2+\vartheta_2 f_2=0 \end{aligned}\right\} \tag{3.22}$$

Solving (3.22), we obtained (3.23) as

$$\left.\begin{aligned} & f_1=1-\left[\left(\vartheta_3-\vartheta_2\right)\psi_{r,N}\left(\psi_{n,N}+\psi_{r,N}\right)^{-1}\rho C_y C_x^{-1}\right]/\left[\left(\tau_1\vartheta_3-2^{-1}\tau_2\gamma_2\right)\right] \\ & f_2=\vartheta_3\lambda\left(\psi_{n,N}+\psi_{r,N}\right)^{-1}\rho C_y C_x^{-1}/\left(\tau_1\vartheta_3-2^{-1}\tau_2\vartheta_2\right) \\ & f_3=-\vartheta_2\psi_{r,N}\left(\psi_{n,N}+\psi_{r,N}\right)^{-1}\rho C_y C_x^{-1}/\left(\tau_1\vartheta_3-2^{-1}\tau_2\vartheta_2\right) \end{aligned}\right\} \tag{3.23}$$

where ϑ_1, ϑ_2, and ϑ_3 are biases of the estimators combined in the proposed schemes define by

$$\left.\begin{aligned} & \vartheta_1 = Bias\left(\bar{y}_r\right)=0 \\ & \vartheta_2 = Bias\left(\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\tau_1}\right)=\bar{Y}^{-1}R\left(\psi_{r,N}\left(\frac{\tau_1+\tau_1^2}{2}RS_x^2-\tau_1\rho_{YX}S_yS_x\right)-\psi_{n,N}\left(\frac{\tau_1^2-\tau_1}{2}RS_x^2\right)\right) \\ & \vartheta_3 = Bias\left(\bar{y}_r\exp\left(\tau_2\frac{\bar{x}_n-\bar{x}_r}{\bar{x}_n+\bar{x}_r}\right)\right)=\frac{R}{\bar{\bar{Y}}}\left\{\psi_{r,N}\left(\left(\frac{\tau_2}{4}+\frac{\tau_2^2}{8}\right)RS_x^2-\frac{\tau_2}{2}\rho_{YX}S_yS_x\right)+\psi_{n,N}\left(\frac{\tau_2}{4}-\frac{\tau_2^2}{8}\right)RS_x^2\right\} \end{aligned}\right\} \tag{3.24}$$

3.2 Theoretical Efficiency Comparison

In this section, efficiency conditions of the proposed estimators over sample mean $\hat{\mu}_0$, Audu *et al.* [19] $\hat{\mu}_7$ were established.

i. Sample mean Vs Proposed Estimator

$$Var\left(\hat{\mu}_0\right)-MSE\left(\hat{t}_1\right)_I>0$$

$$S_y^2-S_y^2\left(\psi_{r,N}-\left(\psi_{r,N}-\psi_{n,N}\right)\rho_{xy}^2\right)>0 \quad \Rightarrow \quad \left|\rho_{YX}\right|>\sqrt{\frac{\psi_{r,N}-1}{\psi_{r,N}-\psi_{n,N}}} \tag{3.25}$$

$$Var\left(\hat{\mu}_0\right)-MSE\left(\hat{t}_1\right)_{II}>0$$

$$S_y^2-S_y^2\left(\psi_{r,N}-\frac{\psi_{r,N}\rho_{xy}^2}{\psi_{r,N}+\psi_{n,N}}\right)>0 \quad \Rightarrow \quad \left|\rho_{YX}\right|>0 \tag{3.26}$$

ii. Audu *et al.* [19] Vs Proposed Estimators

$$MSE\left(\hat{\mu}_7\right)-MSE\left(\hat{t}\right)_I>0 \quad \Rightarrow \quad \left|\rho_{YX}\right|>\frac{\zeta RS_x\left(\sqrt{\psi_{r,N}S_y^2-\bar{Y}^2\left(\psi_{r,N}-\psi_{n,N}\right)}+\psi_{r,N}S_y\right)}{\bar{Y}S_y\left(\psi_{r,N}-\psi_{n,N}\right)} \tag{3.27}$$

$$MSE\left(\hat{\mu}_7\right)-MSE\left(\hat{t}\right)_{II}>0 \quad \Rightarrow \quad \left|\rho_{YX}\right|>\frac{\zeta^2 RS_x\sqrt{\frac{\psi_{r,N}}{\psi_{r,N}+\psi_{n,N}}}}{S_y} \tag{3.28}$$

4. Empirical Study for Efficiency Comparison

In this section, simulation study was conducted to examine the superiority of the proposed estimators over other estimators considered in the study. Data of size 10000 units were generated for study population using function defined in Table 1 below. Samples of sizes 500 units from which 60 units were selected as respondents were randomly chosen 10,000 times by method Simple Random Sampling without Replacement (SRSWOR) The efficiency (MSEs) and efficiency gained (PREs) of the considered estimators were computed using (4.1) and (4.2) respectively.

$$MSE\left(\hat{\theta}_d\right)=\frac{1}{10000}\sum_{d=1}^{10000}\left(\hat{\theta}_d-\bar{Y}\right)^2 \tag{4.1}$$

$$PRE\left(\theta_l\right)=\left(\frac{MSE\left(\hat{\mu}_0\right)}{MSE\left(\theta\right)}\right)\times 100 \tag{4.2}$$

Table 1: Simulation data used for empirical study

Population	Study Variable y	Auxiliary variable X
1	$Y=3+0.4X+\varepsilon$	$X\sim unif(0.5,3)$
II		$X\sim Norm(5,0.3)$
	Where $\varepsilon\sim N(0,1)$	

Table 2: MSE and PRE of Proposed and Other Estimators using Population I

Estimators	MSE	PRE	Estimators	MSE	PRE
Mean $(\hat{\mu}_0)$	89.189	100.00	Singh and Gogoi [33] $(\hat{\mu}_6)$	72.200	123.53
Lee *et al*,. [29] $(\hat{\mu}_1)$	50.501	176.61	Audu et al. [19]		
Singh and Horn [37] $(\hat{\mu}_2)$	44.883	198.71	$(\hat{\mu}_{71})$	62.457	142.80
Ahmed et al. [40] $(\hat{\mu}_3)$	78.210	114.04	$(\hat{\mu}_{72})$	44.460	200.60
Singh et al. [32] $(\hat{\mu}_4)$	82.923	107.56	$(\hat{\mu}_{73})$	47.499	187.77
Prasad [31] $(\hat{\mu}_5)$	77.445	115.17	$(\hat{\mu}_{74})$	56.111	158.95
Proposed estimators					
Case I	**MSE**	**PRE**	**Case 1I**	**MSE**	**PRE**
$(\hat{t}_{11})_1$	15.70613	567.861	$(\hat{t}_{11})_{II}$	25.116	355.11
$(\hat{t}_{12})_1$	42.2845	210.926	$(\hat{t}_{12})_{II}$	21.334	418.07
$(\hat{t}_{13})_1$	19.65429	453.789	$(\hat{t}_{13})_{II}$	21.334	418.07
$(\hat{t}_{14})_1$	26.23558	339.954	$(\hat{t}_{14})_{II}$	25.252	353.20

Table 3: MSE and PRE of Proposed and Other Estimators using Population II

Estimators	MSE	PRE	Estimators	MSE	PRE
Mean $(\hat{\mu}_0)$	107.824	100	Singh and Gogoi [33] $(\hat{\mu}_6)$	52.577	205.08
Lee *et al*,. [29] $(\hat{\mu}_1)$	60.74127	177.513	Audu et al. [19]		
Singh and Horn [37] $(\hat{\mu}_2)$	16.876	638.919	$(\hat{\mu}_{71})$	32.602	330.73
Ahmed et al [40] $(\hat{\mu}_3)$	94.32676	114.309	$(\hat{\mu}_{72})$	87.618	123.06
Singh et al [32] $(\hat{\mu}_4)$	38.603	279.315	$(\hat{\mu}_{73})$	76.777	140.44
Prasad [31] $(\hat{\mu}_5)$	94.30049	114.341	$(\hat{\mu}_{74})$	89.893	119.95
Proposed estimators					
Case I	**MSE**	**PRE**	**Case 1I**	**MSE**	**PRE**
$(\hat{t}_{11})_1$	15.65151	688.905	$(\hat{t}_{11})_{II}$	31.815	338.91
$(\hat{t}_{12})_1$	41.60992	259.131	$(\hat{t}_{12})_{II}$	28.816	374.18

$\left(\hat{t}_{13}\right)_1$	19.59101	550.375	$\left(\hat{t}_{13}\right)_{II}$	28.816	374.18
$\left(\hat{t}_{14}\right)_1$	35.96499	299.803	$\left(\hat{t}_{14}\right)_{II}$	31.854	338.50

Tables 2 and 3 show the results of MSEs and PREs of the proposed and related existing estimators considered in this study using models I and II respectively from the simulation studies in Table 1. The result revealed that proposed class of estimators have minimum MSEs and higher PREs compared to that of conventional and other related estimators considered in the study. This implies that the class of proposed schemes using sample mean $\bar{x}_n$ as estimate of population mean $\bar{X}$ has enhanced the performance of imputation scheme and make it more efficient and less cost in estimation of missing values than other related estimators considered in this study.

5. Conclusion

Considering the results obtained from the empirical study on the efficiency of the proposed scheme's estimators over some existing related schemes' estimators considered in the study, it was obtained that the estimators of the proposed scheme have minimum MSE compared to other estimators considered in all the numerical computations carried out in the study, hence, the proposed estimators demonstrated high level of efficiency over other estimator considered in this study. The results revealed that the proposed scheme which utilized sample mean $\bar{x}_n$ based on sample size $n < N$ instead of population mean $\bar{X}$ which required N population units, provides more efficient estimators that minimize resources for collecting information in mail survey characterized with non-response

References

[1] Singh HP, Solanki RS. An efficient class of estimators for the population mean using auxiliary
information in systematic sampling. Journal of Statistical Theory and Practice. 2012;6(2):274-285.

[2] Sahai A, Ray SK. And efficient estimator using auxiliary information, Metrika. 1980; 27(4):271-275.

[3] Srivastava Sk, Jhajj HA. A class of estimators of the population mean in survey sampling using auxiliary information Biometrika. 1981;68(1):341-343.

[4] Ahmed A, Adewara AA, Singh RVK. Class of ratio estimators with known functions of auxiliary variable for estimating finite population variance. Asian Journal of Mathematics and Computer Research. 2016;12(1):63-70.

[5] Audu A, Ishaq OO, Muili JO, Abubakar A, Rashida A, Akintola KA, Isah U. Modified estimators of population mean using robust multiple regression methods. NIPES Journal of Science and Technology Research. 2020;2(4):12–20.

Available:https://doi.org/10.37933/nipes/2.4.2020.2

[6] Audu A, Adewara AA. Modified factor-type estimators under two-phase sampling. Punjab Journal of Mathematics. 2017;49(2):59-73. ISSN: 1016-2526.

[7] Muili JO, Agwamba EN, Erinola YA, Yunusa MA, Audu A, Hamzat MA. Modified ratio-cum-product estimators of finite population variance. International Journal of Advances in Engineering and Management. 2020;2(4):309-319. DOI: 10.35629/5252-0204309319.

[8] Singh HP, Tailor R. Estimation of finite population mean with known coefficient of variation of an auxiliary character. Statistica. 2005;65(3):301-313.

[9] Sisodia BVS, Dwivedi VK. Modified ratio estimator using coefficient of variation of auxiliary variable. Journal-Indian Society of Agricultural Statistics. 1981;33:13-18.

[10] Khoshnevisan M, Singh R, Chauhan P, Sawan N, Smarandache F. A general family of estimators for estimating population means using known value of some population parameter(s). Far East Journal of Theoretical Statistics. 2007;22(2):181-191.

[11] Singh R, Mishra P, Audu A, Khare S. Exponential Type Estimator for Estimating Finite Population Mean. Int. J. Comp. Theo. Stat. 2020;7(1):37-41.

[12] Singh RVK, Audu A. Efficiency of ratio estimators in stratified random sampling using information on auxiliary attribute. International Journal of Engineering Science and Innovative Technology. 2013;2(1):166-172.

[13] Ahmed A, Singh RVK, Adewara AA. Ratio and product type exponential estimators of population variance under transformed sample information of study and supplementary variables. Asian Journal of Mathematics and Computer Research. 2016;11(3):175-183.

[14] Audu A, Yunusa MA, Ishaq OO, Lawal MK, Rashida A, Muhammad AH, et al. Difference-Cum-Ratio Estimators for Estimating Finite Population Coefficient of Variation in Simple Random Sampling. Asian Journal of Probability and Statistics. 2021;13(3):13-29

[15] Ishaq OO, Audu A, Ibrahim A, Abdulkadir HS, Tukur K. On the linear combination of sample variance, ratio, and product estimators of finite population variance under two-stage sampling. Science Forum Journal of Pure and Applied Science. 2020;20:307-315.
DOI: http://dx.doi.org/10.5455/sf.90563

[16] Audu A, Singh R, Khare S. Developing calibration estimators for population mean using robust measures of dispersion under stratified random. Statistics in Transition New Series. 2021;22(2):125–142. DOI: 10.21307/stattrans-2021-019

[17] Al-Omari, A. I., Bouza, C. N. and Herrera, C. (2013). Imputation methods of missing data for estimating the population mean using simple random sampling with known correlation coefficient. *Quality and Quantity*, 47, 353-365.

[18] Audu, A. and Singh, R. V. K. (2021): Exponential-type regression compromised imputation class of estimators, *Journal of Statistics and Management Systems*, 24(6), 1253-1266
DOI: 10.1080/09720510.2020.1814501

[19] Audu, A., Ishaq, O.O., Isah, U., Muhammed, S., Akintola, K. A., A. Rashida, A., and Abubakar. A. (2020) On the Class of Exponential-Type Imputation Estimators of Population Mean with Known Population Mean of Auxiliary Variable**.** *NIPES Journal of Science and Technology Research* 2(4) 2020 pp.1 - 11 pISSN-2682-5821, eISSN-2682-5821.

[20] Audu A, Singh R, Khare S. New regression-type compromised imputation class of estimators with known parameters of auxiliary variable. Communications in statistics – Simulation and Computation. 2021;1-13.
DOI: 10.1080/03610918.2021.1970182

[21] Audu A, Ishaq OO, Abubakar A, Akintola KA, Isah U, Rashida A, Muhammad S. Regression-type Imputation Class of Estimators using Auxiliary Attribute. Asian Research Journal of Mathematics. 2021;17(5):1-13.
DOI: 10.9734/ARJOM/2021/v17i530296

[22] Audu A, Ishaq OO, Muili J, Zakari Y, Ndatsu AM, Muhammed S. On the Efficiency of Imputation Estimators using Auxiliary Attribute. Continental J. Applied Sciences. 2020;15(1):1-13. DOI: 10.5281/zenodo.3721046

[23] Barnard, J. and Meng, X. L. (1999). "Applications of multiple imputation in medical studies: from AIDS to NHANES". *Statistical Methods in Medical Research*. **8**(1):17–36. doi:10.1177/096228029900800103.

[24] Diana, G. and Perri, P. F. (2010). Improved estimators of the population mean for missing data. *Communications in Statistics-Theory and Methods*, 39, pp.3245–3251.

[25] Gelman, Andrew, and Jennifer Hill (2006). Data analysis using regression and multilevel/ hierarchical models. Cambridge University Press, Ch.25.

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

[27] Kadilar, C. and Cingi, H. (2006). New Ratio Estimators using Correlation Coefficient. *Interstat, 4*:1-11

[28] Kalton, G. (1986) The treatment of Missing Survey Data, *Survey Methodology*, 12: 1-6

[29] Lee, H., Rancourt, E., and Sarndal, C.E. (1994). Experiments with variance estimation from survey data with imputed value. *Journal of Official Statistics, 10(3), 231-243*

[30] Mishra, P., Singh, P. and Singh, R. (2017). A Generalized Class of Estimators For Estimating Population Mean Using Imputation Technique, *Journal of Reliability and Statistical Studies*, Vol. 10 (2), pp.33-41.

[31] Prasad, S. (2017). A study on new methods of ratio exponential type imputation in sample surveys. *Hacettepe Journal of Mathematics and Statistics*, DOI: 10.15672/HJMS.2016.392.

[32] Singh, A.K., Singh, P. and Singh, V. K. (2014). Exponential-Type Compromised Imputation in Survey Sampling, *J. Stat. Appl*. Vol. 3 (2), pp.211-217.

[33] Singh, B.K. and Gogoi, U. (2018). Estimation of Population Mean using Ratio Cum Product Imputation Techniques in Sample Survey. *Res. Rev,:J.stat.,Vol.* 7, pp.38-49.

[34] Singh, G.N., Maurya, S., Khetan, M. and Kadilar, C. (2016). Some imputation methods for missing data in sample surveys. *Hacettepe Journal of Mathematics and Statistics*, Vol. 45 (6), pp.1865-1880.

[35] Singh, S. (2009). A new method of imputation in survey sampling. *Statistics*, 43, pp.499-511.

[36] Singh, S. and Deo, B. (2003). Imputation by power transformation. *Statistical Papers*, Vol. 44, pp.555-579.

[37] Singh, S. and Horn, S (2000). Compromised imputation in survey sampling. *Metrika*, Vol. 51, pp.267-276.

[38] Toutenburg, H., Srivastava, V.K. and Shalabh, A. (2008). Imputation versus imputation of missing values through ratio method in sample surveys. *Statistical Papers*, 49.

[39] Wang, L and Wang, Q. (2006). Empirical likelihood for parametric model under imputation for missing data. *Journal of Statistics and Management Systems*, Vol. 9 (1), pp.1-13.

[40] Ahmed, M. S., Al-Titi, Z., Al-Rawi, Z., Abu-Dayyeh, W. (2006). Estimation of a Population Mean using different Imputation Methods. *Trans.*, 7, 1247-1264