ESTIMATION OF POPULATION MEAN USING RATIO TYPE IMPUTATION TECHNIQUE WITH LINEAR COMBINATION OF TWO AUXILIARY VARIABLE UNDER TWO-PHASE SAMPLING

Deepjan Gohain, North Eastern Regional Institute of Science and Technology
Krishnajyoti Nath, North Eastern Regional Institute of Science and Technology
Singh BK, North Eastern Regional Institute of Science and Technology

ABSTRACT

Present paper proposes four generalized classes of estimators for estimating population mean under the framework of two-phase sampling design by using auxiliary information and also the expressions for bias and mean square error are derived. These types of imputation techniques are used in different decision science related fields for better results. In addition, theoretical results showing the superiority of the proposed estimator over existing estimators from empirical studies based on different datasets from classical statistical literature are shown.

Keywords: Imputation; Bias; Mean Square Error (MSE); Missing Data; Large Sample Approximation; Simple Random Sampling without Replacement (SRSWOR).

INTRODUCTION

The sampling unit refuses to participate in the sample survey, cannot respond, cannot be contacted, or accidentally loses some of the information collected due to unexpected factors, resulting in incomplete survey responses. To deal with missing data effectively Kalton et al. (1981) and Sande (1979) suggested imputation methods that make an incomplete data set structurally complete and its analysis simple. Hyunshik Lee & Särndal (1994); and Lee et al. (1995) used the information on an auxiliary variable for the purpose of imputation. Later Singh and Horn (2000) introduced a compromised method of imputation based on auxiliary variables. Ahmed et al. (2006) discussed several new imputation based estimators that used the information on an auxiliary variate and compared their performance with the mean method of imputation.

Singh and Horn (2000); Wright & Capps (2011), Singh & Gogoi (2017); Singh & Nath (2018b; 2019) and Joyce et al. (2021) discussed designing mixed sampling plan based on IPD and some imputation methods of missing data for estimating the population mean using two-phase sampling scheme.

The objective of the present research work is to provide more efficient alternative estimators than the existing ones, when population parameter of auxiliary information is missing or unknown.

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

NOTATIONS

Let $\Omega = \{1, 2, 3, ..., N\}$ be a finite population of size N and Y is the study variable and X, Z are the auxiliary variable where $\bar{Y}, \bar{X}$ *and* $\bar{Z}$ are the population mean of the variable Y, X *and* Z respectively.

Consider a first phase sample S_1 of size n_1 drawn from the population Ω by using SRSWOR method and a second sample S of size $n(n < n_1)$ drawn from Ω *or* S_1.

Case-I: when second sample S is drawn from S_1 i.e. second sample S is depends on first sample S_1 (denoted by design I) as in Figure 1.

Case-II: when second sample S is drawn from Ω i.e. second sample S is independent of first sample S_1 (denoted by design II) as in Figure 1.

Let the second sample S contains $r(r < n)$ responding units forming a sub space R and $(n-r)$ non-responding units with sub space R^C, such that $S = R \cup R^C$. For every unit $i \in R$, y_i is observed available. For every unit $i \in R^C$ y_i values are missing and imputed values are computed. The i^{th} value x_i and z_i of auxiliary variables are used as a source of imputation for missing data when $i \in R^C$ assuming that in S and S_1 the data $\{(x_i, z_i), i \in S\}$ *and* $\{(x_i, z_i), i \in S_1\}$ are known.

$\bar{x}, \bar{y}, \bar{z}$: sample mean of X, Y *and* Z respectively.

$\bar{x}_r, \bar{y}_r, \bar{z}_r$: sample mean of X, Y *and* Z for corresponding responding units in S.

$\bar{x}_1, \bar{z}_1$: sample mean of X, Z for corresponding units in S_1

s_y^2, s_x^2, s_z^2 : mean square of Y, X *and* Z for corresponding units in S_1

$\rho_{XY}, \rho_{YZ}, \rho_{ZX}$: population correlation coefficient between X *and* Y, Y *and* Z & Z *and* X respectively.

C_X, C_Y, C_Z : the coefficient of variation of X, Y *and* Z respectively.

$$\delta_1 = \frac{1}{r} - \frac{1}{N} \quad \delta_2 = \frac{1}{n_1} - \frac{1}{N}$$

$$\delta_4 = \frac{1}{r} - \frac{1}{n} \quad \delta_3 = \delta_1 - \delta_2 = \frac{1}{r} - \frac{1}{n_1}$$

Let $\frac{\bar{y}_r}{\bar{Y}} = 1 + e_0$, $\frac{\bar{x}_r}{\bar{X}} = 1 + e_1$, $\frac{\bar{z}_r}{\bar{X}} = 1 + e_2$, $\frac{\bar{x}_1}{\bar{X}} = 1 + e_3$, $\frac{\bar{z}_1}{\bar{Z}} = 1 + e_4$

Such that $|e_i| < 1 \ \forall \ i = 0, 1, 2, 3, 4$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

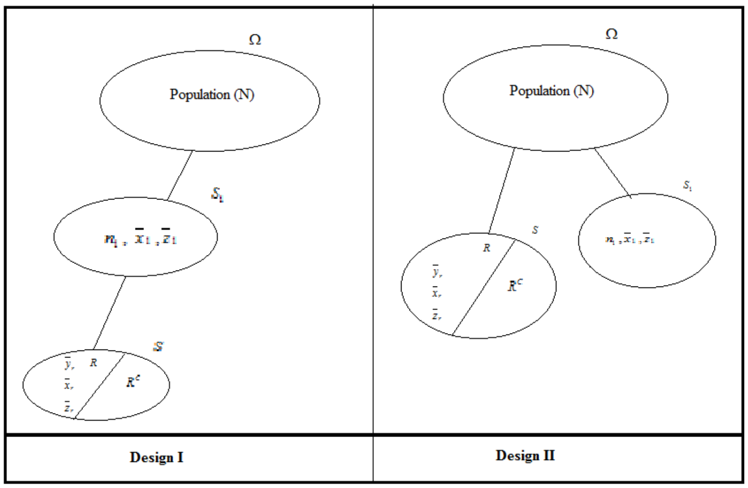

FIGURE 1
LARGE SAMPLE APPROXIMATION

Now using the concept of two-phase sampling and denoting E_1 and E_2 as the expectation over first phase and second phase respectively we have the following expected values.

Case I: when S is drawn from S_1

$$E(e_0) = E_1\left[E_2(e_0)\right] = E_1\left[E_2\left(\frac{\bar{y}_r - \bar{Y}}{\bar{Y}}\right)\right] = E_1\left[\frac{\bar{y}_1 - \bar{Y}}{\bar{Y}}\right] = 0$$

Similarly $E(e_1) = E(e_2) = E(e_3) = E(e_4) = 0$

$$\begin{aligned}
E(e_0^2) &= E_1\left[E_2\left\{\frac{\bar{y}_r - \bar{Y}}{\bar{Y}}\right\}^2\right] \\
&= \frac{1}{\bar{Y}^2} E_1\left[E_2\left\{(\bar{y}_r - \bar{y}_1) + (\bar{y}_1 - \bar{Y})\right\}^2\right] \\
&= \frac{1}{\bar{Y}^2} E_1\left[E_2\left\{(\bar{y}_r - \bar{y}_1)^2 + (\bar{y}_1 - \bar{Y})^2\right\}\right] \\
&= \frac{1}{\bar{Y}^2} E_1\left[\left(\frac{1}{r} - \frac{1}{n_1}\right) s_y^2 + (\bar{y}_1 - \bar{Y})^2\right] \\
&= \frac{1}{\bar{Y}^2}\left[\left(\frac{1}{r} - \frac{1}{n_1}\right) S_Y^2 + \left(\frac{1}{n_1} - \frac{1}{N}\right) S_Y^2\right] \\
&= \left(\frac{1}{r} - \frac{1}{N}\right) C_Y^2
\end{aligned}$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

Similarly,

$$E\left(e_1^2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)C_X^2 \;,\; E\left(e_2^2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)C_Z^2$$

$$E\left(e_3^2\right)=E_1\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)^2=\frac{1}{\bar{X}^2}E_1\left(\bar{x}_1-\bar{X}\right)^2=\left(\frac{1}{n_1}-\frac{1}{N}\right)C_X^2$$

$$E\left(e_4^2\right)=E_1\left(\frac{\bar{z}_1-\bar{Z}}{\bar{Z}}\right)^2=\frac{1}{\bar{Z}^2}E_1\left(\bar{z}_1-\bar{Z}\right)^2=\left(\frac{1}{n_1}-\frac{1}{N}\right)C_Z^2$$

$$\begin{aligned}E\left(e_0e_1\right)&=\frac{1}{\bar{X}\,\bar{Y}}E_1\left[E_2\left(\bar{y}_r-\bar{Y}\right)\left(\bar{x}_r-\bar{X}\right)\right]\\&=\frac{1}{\bar{X}\,\bar{Y}}E_1\left[E_2\left(\bar{y}_r-\bar{y}_1\right)\left(\bar{x}_r-\bar{x}_1\right)+\left(\bar{y}_1-\bar{Y}\right)\left(\bar{x}_1-\bar{X}\right)\right]\\&=\frac{1}{\bar{X}\,\bar{Y}}E_1\left[\left(\frac{1}{r}-\frac{1}{n_1}\right)s_{yx}+\left(\bar{y}_1-\bar{Y}\right)\left(\bar{x}_1-\bar{X}\right)\right]\\&=\frac{1}{\bar{X}\,\bar{Y}}\left\{\left(\frac{1}{r}-\frac{1}{n_1}\right)S_{YX}+\left(\frac{1}{n_1}-\frac{1}{N}\right)S_{YX}\right\}\\&=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{YX}C_XC_Y\end{aligned}$$

Similarly,

$$E\left(e_0e_2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{ZX}C_ZC_X \;;\; E\left(e_1e_2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{YZ}C_YC$$

$$\begin{aligned}E\left(e_0e_3\right)&=E_1\left[E_2\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)\right]\\&=\frac{1}{\bar{X}\,\bar{Y}}E_1\left[\left(\bar{x}_1-\bar{X}\right)E_2\left(\bar{y}_r-\bar{Y}\right)\right]\\&=\frac{1}{\bar{X}\,\bar{Y}}E_1\left[\left(\bar{x}_1-\bar{X}\right)\left(\bar{y}_1-\bar{Y}\right)\right]\\&=\frac{1}{\bar{X}\,\bar{Y}}\left(\frac{1}{n_1}-\frac{1}{N}\right)S_{YX}\\&=\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_{YX}C_YC_X\end{aligned}$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

Similarly,

$$E\left(e_0e_4\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_{YZ}C_YC_Z\;;E\left(e_1e_4\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_{ZX}C_ZC_X\;;E\left(e_2e_4\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)C_Z^2$$

$$E\left(e_1e_3\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)C_X^2\;;E\left(e_2e_3\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_{ZX}C_ZC_X$$

$$E\left(e_3e_4\right)=E_1\left[\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)\left(\frac{\bar{z}_1-\bar{Z}}{\bar{Z}}\right)\right]=\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_{ZX}C_ZC_X$$

Case II: when S is drawn from Ω

$$E\left(e_0\right)=E_2\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)=\frac{\bar{Y}-\bar{Y}}{\bar{Y}}=0$$

Similarly $E\left(e_1\right)=E\left(e_2\right)=0$

$$E\left(e_3\right)=E_1\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)=0\quad;\quad E\left(e_4\right)=E_1\left(\frac{\bar{z}_1-\bar{Z}}{\bar{Z}}\right)=0$$

$$E\left(e_0e_1\right)=E_2\left[\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}_r-\bar{X}}{\bar{X}}\right)\right]=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{YX}C_YC_X$$

$$E\left(e_0e_2\right)=E_2\left[\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{z}_r-\bar{Z}}{\bar{Z}}\right)\right]=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{YZ}C_YC_Z$$

$$E\left(e_1e_2\right)=E_2\left[\left(\frac{\bar{x}_r-\bar{X}}{\bar{X}}\right)\left(\frac{\bar{z}_r-\bar{Z}}{\bar{Z}}\right)\right]=\left(\frac{1}{r}-\frac{1}{N}\right)\rho_{ZX}C_ZC_X$$

$$E\left(e_0e_3\right)=E_1\left[E_2\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)\right]=E_1\left[\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)E_2\left(\frac{\bar{y}_r-\bar{Y}}{\bar{Y}}\right)\right]=E_1\left[\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)\times 0\right]=0$$

Similarly,

$$E\left(e_1e_3\right)=E\left(e_1e_4\right)=E\left(e_0e_3\right)=E\left(e_0e_4\right)=E\left(e_2e_3\right)=E\left(e_2e_4\right)=0$$

$$E\left(e_3e_4\right)=E_1\left[\left(\frac{\bar{x}_1-\bar{X}}{\bar{X}}\right)\left(\frac{\bar{z}_1-\bar{Z}}{\bar{Z}}\right)\right]=\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_{ZX}C_ZC_X$$

Similarly,

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$E\left(e_0^2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)C_Y^2\ ;\ E\left(e_1^2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)C_X^2\ ;\ E\left(e_2^2\right)=\left(\frac{1}{r}-\frac{1}{N}\right)C_Z^2$$
$$E\left(e_3^2\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)C_X^2\ ;\ E\left(e_4^2\right)=\left(\frac{1}{n_1}-\frac{1}{N}\right)C_Z^2$$

SOME EXISTING IMPUTATION TECHNIQUES

Mean Method of Imputation

Under Mean method of imputation

$$y_{.i}=y_i \qquad ;\ i\in R$$
$$=\bar{y}_r \qquad ;\ i\in R^C$$

Using above the point estimator of population mean $\bar{Y}$ is $\bar{y}_m=\bar{y}_r$
The bias and Variance are given by

$$B\left(\bar{y}_m\right)=0$$
$$V\left(\bar{y}_m\right)=\delta_1 C_Y^2\bar{Y}^2$$

Ratio Method of Imputation (Hyunshik Lee & Särndal, 1994)

Under Ratio Method of Imputation

$$y_{.i}=y_i \qquad ;\ i\in R$$
$$=\hat{b}x_i \qquad ;\ i\in R^C$$
$$where \qquad \hat{b}=\frac{\sum_{i\in R}y_i}{\sum_{i\in R}x_i}$$

Using above the point estimator of population mean $\bar{Y}$ is

$$\bar{y}_R=\bar{y}_r\,\frac{\bar{x}_n}{\bar{x}_r}$$

The bias and MSE are given by

$$B\left(\bar{y}_R\right)_{opt}=\delta_1\bar{Y}\left(C_X^2-\rho_{YX}C_YC_X\right)$$
$$MSE\left(\bar{y}_R\right)_{opt}=\bar{Y}^2\left[\ \delta_1C_Y^2+\delta_4\left(C_X^2-2\rho_{YX}C_YC_X\right)\right]$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

Compromised Method of Imputation (Singh & Horn, 2000)

Under this method of imputation

$$y_{\cdot i} = \beta \frac{n}{r} y_i + (1-\beta)\hat{b}x_i \qquad ; \quad i \in R$$
$$= (1-\beta)\hat{b}x_i \qquad ; \quad i \in R^C$$

Using the above the point estimator of population mean $\bar{Y}$ is

$$\bar{y}_{comp} = \beta \bar{y}_r + (1-\beta)\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r}$$

Where β is a constant to be determined such that MSE of $\bar{y}_{comp}$ is minimum.
The optimum Bias and MSE are-

$$B(\bar{y}_{comp}) = \bar{Y}\delta_4 (1-\rho_{yx})\rho_{yx} C_x C_Y$$
$$MSE(\bar{y}_{comp}) = \bar{Y}^2 (\delta_1 C_Y^2 - \delta_4 \rho_{YX}^2 C_Y^2)$$

Exponential Ratio Method of Imputation in two-phase sampling (Pandey et al., 2015)

Under this method of imputation

$$y_{0i} = y_i \qquad ; \quad i \in R$$
$$= \frac{1}{n-r}\left[n\bar{y}_r T - r\bar{y}_r\right] \qquad ; \quad i \in R^C$$

Where $T = \exp\left\{\frac{\bar{x}_1 - \bar{x}_r}{\bar{x}_1 + \bar{x}_r}\right\}$

Using above the point estimator of population mean $\bar{Y}$ is

$$\bar{y}_{ier}^d = \bar{y}_r \exp\left\{\frac{\bar{x}_1 - \bar{x}_r}{\bar{x}_1 - \bar{x}_r}\right\}$$

Where α is a drawn constant such that MSE of $\bar{y}_{ier}^d$ is minimum.
The optimum Bias and MSE are given by

$$B(\bar{y}_{ier}^d)_I = \bar{Y}.\frac{1}{8}\left[(3\delta_1 - 2\delta_2)C_X^2 - 4(\delta_1 - \delta_2)\rho_{YX} C_X C_Y\right]$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$B\left(\bar{y}_{ier}^{d}\right)_{II}=\bar{Y}.\frac{1}{8}\left[\left(3\delta_1-\delta_2\right)C_X^2-4\delta_1\rho_{YX}C_YC_X\right]$$

$$MSE\left(\bar{y}_{ier}^{d}\right)_{I}=\bar{Y}^2\left[\delta_1C_Y^2+\frac{\delta_1}{4}C_X^2-\delta_3\rho_{YX}C_YC_X\right]$$

$$MSE\left(\bar{y}_{ier}^{d}\right)_{II}=\bar{Y}^2\left[\delta_1C_Y^2+\frac{1}{4}\left(\delta_1+\delta_2\right)C_X^2-\delta_1\rho_{YX}C_YC_X\right]$$

Dual to Ratio Method of Imputation in two-phase sampling (Singh & Nath, 2018a)

Under this method of imputation

$$y_{3i}=y_i \qquad ; \quad i\in R$$
$$=\frac{1}{n-r}\left[n\bar{y}_r\phi-r\bar{y}_r\right] \qquad ; \quad i\in R^C$$

Where $\phi=\left\{\frac{n_1\bar{x}_1-n\bar{x}_r}{\left(n-n\right)\bar{x}_r}\right\}^{\beta}$

Using above the point estimator of population mean $\bar{Y}$ is

$$\bar{y}_{idr3}^{d}=\bar{y}_r\left\{\frac{n_1\bar{x}_1-n\bar{x}_r}{\left(n_1-n\right)\bar{x}_r}\right\}^{\beta}$$

Where β is a drawn constant such that MSE of $\bar{y}_{ier}^{d}$ is minimum.

The optimum Bias and MSE are given by

$$B\left(\bar{y}_{idr3}^{d}\right)_{I}=-\frac{\bar{Y}}{2}\delta_5\left(g_1C_X+\rho_{YX}C_Y\right)\rho_{YX}C_Y$$
$$B\left(\bar{y}_{idr3}^{d}\right)_{II}=-\frac{\bar{Y}}{2}\delta_1(\delta_1+\delta_2)^{-1}\left[(\delta_1+\delta_2)\,\mathrm{g}_1C_X+2\delta_2C_X+\delta_1\rho_{YX}C_Y\right]\rho_{YX}C_Y$$
$$MSE\left(\bar{y}_{idr3}^{d}\right)_{I}=\bar{Y}^2\left[\delta_1C_Y^2-\delta_3\rho_{YX}^2C_Y^2\right]$$
$$MSE\left(\bar{y}_{idr3}^{d}\right)_{II}=\bar{Y}^2\left[\delta_1C_Y^2-\delta_1^2\left(\delta_1+\delta_2\right)^{-1}\rho_{YX}^2C_Y^2\right]$$

where

$$g_1=\frac{n}{n_1-n}$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

PROPOSED IMPUTATION STRATEGIES

Motivating the above imputation methods of population mean, we have proposed the following Multivariate Ratio type imputation methods of population mean in two-phase sampling.

Imputation Method $\left(D, \bar{y}_{GSN1}\right)$

The imputation scheme is as follows:

$$y_{1i} = y_i \left(\frac{\alpha_1 \bar{x}_1 + \alpha_2 \bar{z}_1 + \rho_{xz}}{\alpha_1 \bar{x}_r + \alpha_2 \bar{z}_r + \rho_{xz}} \right) \quad ; \; i \in R$$

$$= \bar{y}_r \left(\frac{\alpha_1 \bar{x}_1 + \alpha_2 \bar{z}_1 + \rho_{xz}}{\alpha_1 \bar{x}_r + \alpha_2 \bar{z}_r + \rho_{xz}} \right) \quad ; \; i \in R^C$$

Imputation Method $\left(D, \bar{y}_{GSN2}\right)$

The imputation scheme is as follows

$$y_{2i} = y_i \left[\frac{\alpha_1 \left(\bar{x}_1 + \beta_1 \left(X \right) \right) + \alpha_2 \left(\bar{z}_1 + \beta_1 \left(Z \right) \right)}{\alpha_1 \left(\bar{x}_r + \beta_1 \left(X \right) \right) + \alpha_2 \left(\bar{z}_r + \beta_1 \left(Z \right) \right)} \right] \quad ; \; i \in R$$

$$= \bar{y}_r \left[\frac{\alpha_1 \left(\bar{x}_1 + \beta_1 \left(X \right) \right) + \alpha_2 \left(\bar{z}_1 + \beta_1 \left(Z \right) \right)}{\alpha_1 \left(\bar{x}_r + \beta_1 \left(X \right) \right) + \alpha_2 \left(\bar{z}_r + \beta_1 \left(Z \right) \right)} \right] \quad ; \; i \in R^C$$

Imputation Method $\left(D, \bar{y}_{GSN3}\right)$

The imputation scheme is as follows

$$y_{3i} = y_i \left[\frac{\alpha_1 \left(\beta_1 \left(X \right) \bar{x}_1 + \rho_{xz} \right) + \alpha_2 \left(\beta_1 \left(Z \right) \bar{z}_1 + \rho_{xz} \right)}{\alpha_1 \left(\beta_1 \left(X \right) \bar{x}_r + \rho_{xz} \right) + \alpha_2 \left(\beta_1 \left(Z \right) \bar{z}_r + \rho_{xz} \right)} \right] \quad ; \; i \in R$$

$$= \bar{y}_r \left[\frac{\alpha_1 \left(\beta_1 \left(X \right) \bar{x}_1 + \rho_{xz} \right) + \alpha_2 \left(\beta_1 \left(Z \right) \bar{z}_1 + \rho_{xz} \right)}{\alpha_1 \left(\beta_1 \left(X \right) \bar{x}_r + \rho_{xz} \right) + \alpha_2 \left(\beta_1 \left(Z \right) \bar{z}_r + \rho_{xz} \right)} \right] \quad ; \; i \in R^C$$

Imputation Method $\left(D, \bar{y}_{GSN4}\right)$

The imputation scheme is as follows

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$y_{4i} = y_i \left[\frac{\alpha_1\left(\rho_{xz}\overline{x}_1 + \beta_1(X)\right) + \alpha_2\left(\rho_{xz}\overline{z}_1 + \beta_1(Z)\right)}{\alpha_1\left(\rho_{xz}\overline{x}_r + \beta_1(X)\right) + \alpha_2\left(\rho_{xz}\overline{z}_r + \beta_1(Z)\right)} \right] \qquad ; i \in R$$

$$= \overline{y}_r \left[\frac{\alpha_1\left(\rho_{xz}\overline{x}_1 + \beta_1(X)\right) + \alpha_2\left(\rho_{xz}\overline{z}_1 + \beta_1(Z)\right)}{\alpha_1\left(\rho_{xz}\overline{x}_r + \beta_1(X)\right) + \alpha_2\left(\rho_{xz}\overline{z}_r + \beta_1(Z)\right)} \right] \qquad ; i \in R^C$$

Point estimators for population mean $\overline{Y}$ under the proposed four types of imputation methods $(D, \overline{y}_{GSN1})$, $(D, \overline{y}_{GSN2})$, $(G, \overline{y}_{GSN3})$ *and* $(D, \overline{y}_{GSN4})$ can easily be deduced. We have the point estimators-

$$\overline{y}^d_{GSN1} = \overline{y}_r \left(\frac{\alpha_1\overline{x}_1 + \alpha_2\overline{z}_1 + \rho_{xz}}{\alpha_1\overline{x}_r + \alpha_2\overline{z}_r + \rho_{xz}} \right)$$

$$\overline{y}^d_{GSN2} = \overline{y}_r \left[\frac{\alpha_1\left(\overline{x}_1 + \beta_1(X)\right) + \alpha_2\left(\overline{z}_1 + \beta_1(Z)\right)}{\alpha_1\left(\overline{x}_r + \beta_1(X)\right) + \alpha_2\left(\overline{z}_r + \beta_1(Z)\right)} \right]$$

$$\overline{y}^d_{GSN3} = \overline{y}_r \left[\frac{\alpha_1\left(\beta_1(X)\overline{x}_1 + \rho_{xz}\right) + \alpha_2\left(\beta_1(Z)\overline{z}_1 + \rho_{xz}\right)}{\alpha_1\left(\beta_1(X)\overline{x}_r + \rho_{xz}\right) + \alpha_2\left(\beta_1(Z)\overline{z}_r + \rho_{xz}\right)} \right]$$

$$\overline{y}^d_{GSN4} = \overline{y}_r \left[\frac{\alpha_1\left(\rho_{xz}\overline{x}_1 + \beta_1(X)\right) + \alpha_2\left(\rho_{xz}\overline{z}_1 + \beta_1(Z)\right)}{\alpha_1\left(\rho_{xz}\overline{x}_r + \beta_1(X)\right) + \alpha_2\left(\rho_{xz}\overline{z}_r + \beta_1(Z)\right)} \right]$$

In general the above four imputation method can be defined as $\left(D, \overline{y}_{GSN}\right)$

The imputation scheme is a follows

$$y_{.i} = y_i \left[\frac{\alpha_1\left(\overline{x}_1 + \lambda\right) + \alpha_2\left(\overline{z}_1 + \mu\right)}{\alpha_1\left(\overline{x}_r + \lambda\right) + \alpha_2\left(\overline{z}_r + \mu\right)} \right] \qquad ; i \in R$$

$$= \overline{y}_r \left[\frac{\alpha_1\left(\overline{x}_1 + \lambda\right) + \alpha_2\left(\overline{z}_1 + \mu\right)}{\alpha_1\left(\overline{x}_r + \lambda\right) + \alpha_2\left(\overline{z}_r + \mu\right)} \right] \qquad ; i \in R^C$$

Point estimator for population mean $\overline{Y}$ under $\left(D, \overline{y}_{GSN}\right)$ is-

$$\overline{y}^d_{GSN} = \overline{y}_r \left[\frac{\alpha_1\left(\overline{x}_1 + \lambda\right) + \alpha_2\left(\overline{z}_1 + \mu\right)}{\alpha_1\left(\overline{x}_r + \lambda\right) + \alpha_2\left(\overline{z}_r + \mu\right)} \right]$$

Where α_1 *and* α_2 are suitable chosen constants to be determined such that MSE of the point estimator has minimum and $\alpha_1 + \alpha_2 = 1$.

Expanding $\overline{y}_{GSN}$ in terms of $e's$ retaining the terms upto first order approximate we have

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$\bar{y}^{d}_{GSN} = \bar{Y}\left(1+e_0\right)\left(1+\theta_1 e_3+\theta_2 e_4\right)\left(1+\theta_1 e_1+\theta_2 e_2\right)^{-1},$$

$$\bar{y}^{d}_{GSN} = \bar{Y}\left(1+e_0\right)\Big(1-\theta_1 e_1-\theta_2 e_2+\theta_1 e_3+\theta_2 e_4+\theta_1^2 e_1^2+\theta_2^2 e_2^2+2\theta_1\theta_2 e_1 e_2-\theta_1^2 e_1 e_3$$
$$-\theta_1\theta_2 e_2 e_3-\theta_1\theta_2 e_1 e_4-\theta_2^2 e_2 e_4\Big)$$

$$\Rightarrow \bar{y}^{d}_{GSN}-\bar{Y} = \bar{Y}\Big(e_0-\theta_1 e_1-\theta_2 e_2+\theta_1 e_3+\theta_2 e_4+\theta_1^2 e_1^2+\theta_2^2 e_2^2+2\theta_1\theta_2 e_1 e_2$$
$$-\theta_1^2 e_1 e_3-\theta_1\theta_2 e_2 e_3-\theta_1\theta_2 e_1 e_4-\theta_2^2 e_2 e_4-\theta_1 e_0 e_1-\theta_2 e_0 e_2$$
$$+\theta_1 e_0 e_3+\theta_2 e_0 e_4\Big) \qquad (1)$$

Where,

$$\theta_1 = \frac{\alpha_1 \bar{X}}{\alpha_1\left(\bar{X}+\lambda\right)+\alpha_2\left(\bar{Z}+\mu\right)} \qquad (2)$$

$$\theta_2 = \frac{\alpha_2 \bar{Z}}{\alpha_1\left(\bar{X}+\lambda\right)+\alpha_2\left(\bar{Z}+\mu\right)} \qquad (3)$$

PROPERTIES OF PROPOSED ESTIMATOR

The bias, MSE and min MSE of the proposed point estimators have been derived in the following theorems.

Theorem 1

Bias of the estimators $\bar{y}^{d}_{GSN}$ under design I and design II upto first order of approximation are as:

$$(i)\ B\left(\bar{y}^{d}_{GSN}\right)_I = \bar{Y}\delta_3\left(\theta_1^2 C_X^2+\theta_2^2 C_Z^2+2\theta_1\theta_2\rho_{XZ}C_X C_Z-\theta_1\rho_{YX}C_X C_Y-\theta_2\rho_{YZ}C_Y C_Z\right) \qquad (4)$$

$$(ii)\ B\left(\bar{y}^{d}_{GSN}\right)_{II} = \bar{Y}\delta_1\left(\theta_1^2 C_X^2+\theta_2 C_Z^2+2\theta_1\theta_2\rho_{XZ}C_X C_Z-\theta_1\rho_{YX}C_X C_Y-\theta_2\rho_{YZ}C_Y C_Z\right) \qquad (5)$$

Proof: Taking expectation on both sides of equation (1) we have

$$B\left(\bar{y}^{d}_{GSN}\right) = \bar{Y}E\Big[\theta_1^2\left(e_1^2-e_1 e_3\right)+\theta_2^2\left(e_2^2-e_2 e_4\right)+\theta_1\theta_2\left(2e_1 e_2-e_2 e_3-e_1 e_4\right)$$
$$-\theta_1\left(e_0 e_1-e_0 e_3\right)-\theta_2\left(e_0 e_2-e_0 e_4\right)\Big]$$

Putting the expected values under design I we have

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$(i)\ B\left(\bar{y}_{GSN}^{d}\right)_I=\bar{Y}\left(\delta_1-\delta_2\right)\left(\theta_1^2C_X^2+\theta_2^2C_Z^2+2\theta_1\theta_2\rho_{ZX}C_ZC_X-\theta_1\rho_{YX}C_YC_X-\theta_2\rho_{YZ}C_YC_Z\right)$$

$$\Rightarrow\ B\left(\bar{y}_{GSN}^{d}\right)_I=\bar{Y}\delta_3\left(\theta_1^2C_X^2+\theta_2^2C_Z^2+2\theta_1\theta_2\rho_{ZX}C_ZC_X-\theta_1\rho_{YX}C_YC_X-\theta_2\rho_{YZ}C_YC_Z\right)$$

Putting the expected values under design II we have

$$(ii)\,B\left(\bar{y}_{GSN}^{d}\right)_{II}=\bar{Y}\delta_1\left(\theta_1^2C_X^2+\theta_2^2C_Z^2+2\theta_1\theta_2\rho_{ZX}C_ZC_X-\theta_1\rho_{YX}C_YC_X-\theta_2\rho_{YZ}C_YC_Z\right)$$

Theorem 2

MSE of the estimators $\bar{y}^{d}{}_{GSN}$ under design I and design II upto first order approximation are as-

$$(i)\ \min\left[MSE\left(\bar{y}_{GSN}^{d}\right)_I\right]=\bar{Y}^2\left[\delta_1C_Y^2-\delta_3R_{Y.ZX}^2C_Y^2\right]$$

$$(ii)\ \min\left[MSE\left(\bar{y}_{GSN}^{d}\right)_{II}\right]=\bar{Y}^2\left[\delta_1C_Y^2-\frac{\delta_1^2}{\delta_1+\delta_2}R_{Y.ZX}^2C_Y^2\right]$$

Proof: Taking expectation after squaring the both sides of (3) we have

$$MSE\left(\bar{y}_{GSN}^{d}\right)=\bar{Y}^2E\left[e_0^2+\theta_1^2\left(e_1^2+e_3^2-2e_1e_3\right)+\theta_2^2\left(e_2^2+e_4^2-2e_2e_4\right)+2\theta_1\left(e_0e_3-e_0e_4\right)\right.$$
$$\left.+2\theta_2\left(e_0e_4-e_0e_2\right)+2\theta_1\theta_2\left(e_3e_4-e_1e_4-e_1e_3+e_1e_2\right)\right]$$

Putting the expected values under design I we have

$$MSE\left(\bar{y}_{GSN}^{d}\right)_I=\bar{Y}^2\left[\delta_1C_Y^2+\theta_1^2\left(\delta_1-\delta_2\right)C_X^2+\theta_2^2\left(\delta_1-\delta_2\right)C_Z^2+2\theta_1\left(\delta_2-\delta_1\right)\rho_{YX}C_XC_Y\right.$$
$$\left.+2\theta_2\left(\delta_2-\delta_1\right)\rho_{YZ}C_YC_Z+2\left(\delta_1-\delta_2\right)\rho_{XZ}C_XC_Z\right]$$
$$=\bar{Y}^2\left[\delta_1C_Y^2+\left(\delta_1-\delta_2\right)\left(\theta_1^2C_X^2+\theta_2^2C_Z^2-2\theta_1\rho_{YX}C_XC_Y-2\theta_2\rho_{YZ}C_YC_Z\right.\right.$$
$$\left.\left.+2\theta_1\theta_2\rho_{XZ}C_XC_Z\right)\right] \quad (6)$$

Putting the expected values under design II we have

$$MSE\left(\bar{y}_{GSN}^{d}\right)_{II}=\bar{Y}^2\left[\delta_1C_Y^2+\theta_1^2\left(\delta_1+\delta_2\right)C_X^2+\theta_2^2\left(\delta_1+\delta_2\right)C_Z^2-2\theta_1\delta_1\rho_{YX}C_YC_X\right.$$
$$\left.-2\theta_2\delta_1\rho_{YZ}C_YC_Z+2\theta_1\theta_2\left(\delta_1+\delta_2\right)\rho_{XZ}C_XC_Z\right] \quad (7)$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

The optimum value of θ_1, θ_2 is obtained by minimizing $MSE\left(\bar{y}^d{}_{GSN}\right)_I$ *and* $MSE\left(\bar{y}^d{}_{GSN}\right)_{II}$ given in equation (6) and (7) by using the method of maxima and minima we have-

$$(\theta_1)_I = \frac{C_Y}{C_X} R_{YX.Z} \quad , \quad (\theta_1)_{II} = \left(\frac{\delta_1}{\delta_1+\delta_2}\right)\frac{C_Y}{C_X} R_{YX.Z}$$

$$(\theta_2)_I = \frac{C_Y}{C_Z} R_{YZ.X} \quad , \quad (\theta_2)_{II} = \left(\frac{\delta_1}{\delta_1+\delta_2}\right)\frac{C_Y}{C_Z} R_{YZ.X}$$

$$where \quad R_{YX.Z} = \frac{\rho_{YX} - \rho_{XZ}\rho_{YZ}}{1-\rho_{XZ}^2} \quad and \quad R_{YZ.X} = \frac{\rho_{YZ} - \rho_{XZ}\rho_{YX}}{1-\rho_{XZ}^2}$$

Putting the optimum values of θ_1 *and* θ_2 under the design I and design II in equation (2) and (3) and solving for α_1 *and* α_2 we have

$$(\alpha_1)_{I,II} = \frac{R_{YX.Z} / C_X \overline{X}}{\left(R_{YX.Z} / C_X \overline{X}\right) + \left(R_{YZ.X} / C_Z \overline{Z}\right)}$$

$$(\alpha_2)_{I,II} = \frac{\left(R_{YZ.X} / C_Z \overline{X}\right)}{\left(R_{YX.Z} / C_X \overline{X}\right) + \left(R_{YZ.X} / C_Z \overline{Z}\right)}$$

Putting the optimum values of θ_1 *and* θ_2 under design I and II in equation (4) & (5) we have

$$(i) \min\left[MSE\left(\bar{y}^d_{GSN}\right)_I\right] = \bar{Y}^2\left[\delta_1 C_Y^2 - \delta_3 R_{Y.ZX}^2 C_Y^2\right]$$

$$(ii) \min\left[MSE\left(\bar{y}^d_{GSN}\right)_{II}\right] = \bar{Y}^2\left[\delta_1 C_Y^2 - \frac{\delta_1^2}{\delta_1+\delta_2} R_{Y.ZX}^2 C_Y^2\right]$$

$$where \quad R_{Y.ZX}^2 = \frac{\rho_{YX}^2 + \rho_{YZ}^2 - 2\rho_{YX}\rho_{YZ}\rho_{ZX}}{1-\rho_{XZ}^2}$$

Theorem 3

The estimator $\bar{y}^d{}_{GSN}$ is unbiased for optimum values of θ_1 *and* θ_2 under design I and II.

Proof: Putting the optimum values of θ_1 *and* θ_2 under design I in equation (6) we have

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$B\left(\bar{y}_{GSN}^{d}\right)_{I} = \bar{Y}\left(\delta_1-\delta_2\right)\left(R_{YX.Z}^2+R_{YZ.X}^2+2R_{YX.Z}R_{YZ.X}\rho_{XZ}-R_{YX.Z}\rho_{YX}-R_{YZ.X}\rho_{YZ}\right)$$

$$=\bar{Y}\delta_3\left\{\begin{array}{l}\left(\frac{\rho_{YX}-\rho_{XZ}\rho_{YZ}}{1-\rho_{XZ}}\right)^2+\left(\frac{\rho_{YZ}-\rho_{XZ}\rho_{XY}}{1-\rho_{XZ}^2}\right)^2+2\rho_{XZ}\left(\frac{\rho_{YX}-\rho_{XZ}\rho_{YZ}}{1-\rho_{XZ}^2}\right)\left(\frac{\rho_{YX}-\rho_{XZ}\rho_{YZ}}{1-\rho_{XZ}^2}\right)\\ -\left(\frac{\rho_{XY}^2-\rho_{XY}\rho_{YZ}\rho_{XZ}}{1-\rho_{XZ}^2}\right)-\left(\frac{\rho_{YZ}^2-\rho_{XY}\rho_{YZ}\rho_{XZ}}{1-\rho_{XZ}^2}\right)\end{array}\right\}$$

$$=\bar{Y}\delta_3\left\{\frac{\rho_{XY}^2+\rho_{YZ}^2-\rho_{XY}^2\rho_{XZ}^2-\rho_{YZ}^2\rho_{XZ}^2-2\rho_{XY}\rho_{YZ}\rho_{XZ}\left(1-\rho_{XZ}^2\right)}{\left(1-\rho_{XZ}^2\right)^2}-\frac{\rho_{XY}^2+\rho_{YZ}^2-2\rho_{XY}\rho_{YZ}\rho_{XZ}}{1-\rho_{XZ}^2}\right\}$$

$$=\bar{Y}\delta_3\left\{\frac{\rho_{XY}^2+\rho_{YZ}^2-2\rho_{XY}\rho_{YZ}\rho_{XZ}}{1-\rho_{XZ}^2}-\frac{\rho_{XY}^2+\rho_{YZ}^2-2\rho_{XY}\rho_{YZ}\rho_{XZ}}{1-\rho_{XZ}^2}\right\}$$

$$=0$$

Similarly put ting the optimum values of θ_1 *and* θ_2 under design II in equation (7) we have

$$B\left(\bar{y}_{GSN}^{d}\right)_{II}=0$$

COMPARISON

In this section we divide the conditions under which the suggested estimator is superior to the existing estimators in design I and design II. To compare the different estimators we use the following theorem of multiple correlation coefficients.

Comparison with Mean Method of Imputation

$$D_{11}=V\left(\bar{y}_m\right)-MSE\left(\bar{y}_{GSN}^{d}\right)_{I}=\bar{Y}^2\left[\delta_3R_{Y.ZX}^2\right]C_Y^2>0$$

$$D_{12}=V\left(\bar{y}_m\right)-MSE\left(\bar{y}_{GSN}^{d}\right)_{II}=\bar{Y}^2\left[\frac{\delta_1^2}{\delta_1+\delta_2}R_{Y.ZX}^2\right]C_Y^2>0$$

$\therefore \left(\bar{y}_{GSN}^{d}\right)$ is always efficient than $\bar{y}_m$ in design I and design II.

Comparison with Ratio Method of Imputation

$$D_{21}=MSE\left(\bar{y}_R\right)-MSE\left(\bar{y}_{GSN}^{d}\right)_{I}$$

$$=\bar{Y}\left[\delta_4\left(C_X^2-2\rho_{YX}C_YC_X\right)+\delta_3R_{Y.ZX}^2C_Y^2\right]>0$$

$$D_{22}=\bar{Y}^2\left[\delta_4\left(C_X^2-2\rho_{YX}C_YC_X\right)+\frac{\delta_1^2}{\delta_1+\delta_2}R_{Y.ZX}^2C_Y^2\right]>0$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$\therefore \bar{y}^{d}_{GSN}$ is always efficient than $\bar{y}_R$ in design I and design II.

Comparison with Compromised Method of Imputation

$$D_{31} = MSE\left(\bar{y}_{COMP}\right) - MSE\left(\bar{y}^{d}_{GSN}\right)_I$$
$$= \bar{Y}^2 \left[\delta_5 R^2_{Y.ZX} - \delta_4 \rho^2_{YX} \right] C^2_Y$$

$\therefore \left(\bar{y}^{d}_{GSN}\right)_I$ is more efficient than $\bar{y}_{COMP}$ if

$$\delta_5 R^2_{Y.ZX} > \delta_4 \rho^2_{YX}$$

which is always true as $\delta_5 > \delta_4$ *and* $R_{Y.ZX} > \rho_{YX}$

$$D_{32} = MSE\left(\bar{y}_{COMP}\right) - MSE\left(\bar{y}^{d}_{GSN}\right)_{II}$$
$$= \bar{Y}^2 \left[\frac{\delta_1^2}{\delta_1 + \delta_2} R^2_{Y.ZX} - \delta_4 \rho^2_{YX} \right] C^2_Y$$

$\therefore \left(\bar{y}^{d}_{GSN}\right)_{II}$ is more efficient than $\bar{y}_{COMP}$ if

$$\frac{\delta_1^2}{\delta_1 + \delta_2} R^2_{Y.ZX} > \delta_4 \rho^2_{YX}$$

Comparison with Exponential Ratio Method of Imputation

$$D_{41} = MSE\left(\bar{y}^{d}_{ier}\right)_I - MSE\left(\bar{y}^{d}_{GNS}\right)_I$$
$$= \bar{Y}^2 \left[\delta_3 \left(R^2_{Y.ZX} - \rho_{YX} \rho_{YX} \frac{C_X}{C_Y} \right) C^2_Y + \frac{\delta_1}{4} C^2_X \right]$$

$\therefore \left(\bar{y}^{d}_{GSN}\right)_I$ is more efficient than $\left(\bar{y}^{d}_{ier}\right)$ if

$$\delta_5 R^2_{Y.ZX} C^2_Y + \frac{\delta_1}{4} C^2_X > \delta_5 \rho_{YX} C_Y C_X$$

$$D_{42} = MSE\left(\bar{y}^{d}_{ier}\right)_{II} - MSE\left(\bar{y}^{d}_{GSN}\right)_{II}$$
$$= \bar{Y}^2 \left[\frac{1}{4} (\delta_1 + \delta_2) C^2_X + \frac{\delta_1^2}{\delta_1 + \delta_2} R^2_{Y.ZX} C^2_Y - \delta_1 \rho_{YX} C_Y C_X \right]$$

$\therefore \left(\bar{y}^{d}_{GSN}\right)_{II}$ is more efficient than $\left(\bar{y}^{d}_{ier}\right)_{II}$ if

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$$\frac{1}{4}(\delta_1+\delta_2)C_X^2+\frac{\delta_1^2}{\delta_1+\delta_2}R_{Y.ZX}^2C_Y^2>\delta_1\rho_{YX}C_YC_X$$

Comparison with dual to Ratio Method of Imputation

$$D_{51}=MSE\left(\bar{y}^d{}_{idr3}\right)_I-MSE\left(\bar{y}^d{}_{GSN}\right)_I$$
$$=\bar{Y}^2\delta_3\left(R_{Y.ZX}^2-\rho_{YX}^2\right)C_Y^2>0$$
$$D_{52}=MSE\left(\bar{y}^d{}_{idr3}\right)_{II}-MSE\left(\bar{y}^d{}_{GSN}\right)_{II}$$
$$=\frac{\delta_1^2}{\delta_1+\delta_2}\left(R_{Y.ZX}^2-\rho_{YX}^2\right)C_Y^2>0$$

$\therefore$ $\bar{y}^d_{GSN}$ is always efficient than $\bar{y}^d{}_{idr3}$ in design I and design II.

EMPIRICAL STUDY

To examine the performance of the proposed estimator of the population mean in two-phase sampling, we have considered the following three populations (Tables 1-10).

Population I (Cochran, 1977)

Y: Number of placebo children
X: Number of paralytic polio cases in the placebo group
Z: Number of paralytic polio cases in the 'not inoculated group

$$N=34,\quad n_1=22,\quad n=11,\quad r=9$$
$$\bar{Y}=4.92,\quad \bar{X}=2.59,\quad \bar{Z}=2.91$$
$$\rho_{YX}=0.7326,\quad \rho_{YZ}=0.6430,\quad \rho_{XZ}=0.6837$$
$$C_Y^2=1.0248,\quad C_x^2=1.5175,\quad C_z^2=1.1492$$

Population II (Murthy, 1967)

Y: Area under wheat in 1964
X: Area under wheat in 1963
Z: Cultivated area in 1961

$$N=34,\quad n_1=22,\quad n=10,\quad r=7$$
$$\bar{Y}=199.44\ acre,\quad \bar{X}=208.89\,acre,\quad \bar{Z}=747.59\,acre$$
$$\rho_{YX}=0.9801,\quad \rho_{YZ}=0.9043,\quad \rho_{XZ}=0.9097$$
$$C_Y^2=0.5673,\quad C_X^2=0.5191,\quad C_Z^2=0.3527$$

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

Population III (Anderson, 2003)

Y: Head length of second son
X : Head length of first son
Z :Head breathe of first son

$$N = 25, \quad n_1 = 14, \quad n = 9, \quad r = 7$$
$$\bar{Y} = 183.84, \quad \bar{X} = 185.72, \quad \bar{Z} = 151.16$$
$$\rho_{YX} = 0.7108, \quad \rho_{YZ} = 0.6932, \quad \rho_{XZ} = 0.7346$$
$$C_Y^2 = 0.002983, \quad C_X^2 = 0.002763, \quad C_Z^2 = 0.002329$$

Table 1
MSE OF THE DIFFERENT ESTIMATORS UNDER DESIGN I

Point Estimator	Population I	Population II	Population III
$\bar{y}_m$	2.026693	2559.906609	10.369737
$\bar{y}_R$	1.875253	1631.420886	.955755
$\bar{y}_{COMP}$	1.757726	1630.937426	8.752709
$\bar{y}^d_{ier}$	1.324985	1084.832174	7.845193
$\bar{y}^d_{idr3}$	1.152552	448.613009	6.731424
$\bar{y}^d_{GSN}$	1.090778	446.556316	6.273775

Table 2
MSE OF THE DIFFERENT ESTIMATORS UNDER DESIGN II

Point Estimator	Population I	Population II	Population III
$\bar{y}_m$	2.026693	2559.906609	10.369737
$\bar{y}_R$	1.875253	1631.420886	8.955755
$\bar{y}_{COMP}$	1.757726	1630.937426	8.752709
$\bar{y}^d_{ier}$	1.117529	828.239053	6.4115601
$\bar{y}^d_{idr3}$	1.117496	405.529871	6.356754
$\bar{y}^d_{GSN}$	1.053244	403.431209	5.851978

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

Table 3
PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_R$ UNDER DESIGN I

Point Estimator	Population I	Population II	Population III
$\bar{y}_R$	100.000	100.000	100.000
$\bar{y}_{COMP}$	106.686	100.029	102.319
$\bar{y}_{ier}^{d}$	141.530	150.385	114.156
$\bar{y}_{idr3}^{d}$	162.704	363.659	133.044
$\bar{y}_{GSN}^{d}$	171.918	365.334	142.749

Table 4
PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_R$ UNDER DESIGN II

Point Estimator	Population I	Population II	Population III
$\bar{y}_R$	100.000	100.000	100.000
$\bar{y}_{COMP}$	106.686	100.029	102.319
$\bar{y}_{ier}^{d}$	167.803	196.975	139.681
$\bar{y}_{idr3}^{d}$	167.808	402.294	140.886
$\bar{y}_{GSN}^{d}$	178.045	404.386	153.038

Table 5
PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_{COMP}$ UNDER DESIGN I

Point Estimator	Population I	Population II	Population III
$\bar{y}_{COMP}$	100.000	100.000	100.000
$\bar{y}_{ier}^{d}$	132.660	150.340	111.568
$\bar{y}_{idr3}^{d}$	152.507	363.551	130.027
$\bar{y}_{GSN}^{d}$	161.144	365.225	139.513

Table 6
PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_{COMP}$ UNDER DESIGN II

Point Estimator	Population I	Population II	Population III
$\bar{y}_{COMP}$	100.000	100.000	100.000
$\bar{y}_{ier}^{d}$	157.287	196.916	136.514

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

$\bar{y}_{idr3}^{d}$	157.292	402.174	137.691
$\bar{y}_{GSN}^{d}$	166.887	404.267	149.568

Table 7

PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_{ier}$ UNDER DESIGN I

Point Estimator	Population I	Population II	Population III
$\bar{y}_{ier}^{d}$	100.000	100.000	100.000
$\bar{y}_{idr3}^{d}$	114.961	241.819	116.546
$\bar{y}_{GSN}^{d}$	121.472	242.933	125.047

Table 8

PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_{ier}$ UNDER DESIGN II

Point Estimator	Population I	Population II	Population III
$\bar{y}_{ier}^{d}$	100.000	100.000	100.000
$\bar{y}_{idr3}^{d}$	100.003	204.236	100.086
$\bar{y}_{GSN}^{d}$	106.103	205.299	109.562

Table 9

PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_{idr3}$ UNDER DESIGN I

Point Estimator	Population I	Population II	Population III
$\bar{y}_{idr3}^{d}$	100.000	100.000	100.000
$\bar{y}_{GSN}^{d}$	105.666	100.460	107.295

Table 10

PRE OF THE DIFFERENT ESTIMATORS WITH RESPECT TO $\bar{y}_{idr3}$ UNDER DESIGN II

Point Estimator	Population I	Population II	Population III
$\bar{y}_{idr3}^{d}$	100.000	100.000	100.000
$\bar{y}_{GSN}^{d}$	106.100	100.495	108.626

CONCLUSION

From the above tables, it is obvious that the suggested have smaller Mean Square Error (MSE) than the MSE's of the other existing estimators both theoretically as well as empirically under dependent and independent cases. Also the Bias of the proposed estimator vanishes at the

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.

optimum values of α_1 *and* α_2. Therefore it is concluded that the proposed estimator is preferable to use over other existing estimator.

REFERENCES

Ahmed, M.S., Al-Titi, O., Al-Rawi, Z., & Abu-Dayyeh, W. (2006). Estimation of a population mean using different imputation methods. *Statistics in Transition*, *7*(6), 1247-1264.

Anderson, T.W. (2003). An Introduction to Multivariate Statistical Analysis, 3rd edR John Wiley and Sons. *New York*.

Cochran, W.G. (1977). *Sampling techniques*. John Wiley & Sons.

Hyunshik Lee, E.R., & Särndal, C.E. (1994). Experiments with variance estimation from survey data with imputed values. *Journal of official Statistics*, *10*(3), 231-243.

Joyce, V.J., Merlin, G.S., Edna, K.R.J., & Fenella, S. (2021). Designing Mixed Sampling Plan Based on IPD. *Journal of Management Information and Decision Sciences*, *24*, 1-6.

Kalton, G., Kasprzyk, D., & Santos, R. (1981). Issues of nonresponse and imputation in the survey of income and program participation. In *Current topics in survey sampling* (pp. 455-480). Academic Press.

Lee, H., Rancourt, E., & Sarndal, C.E. (1995). Variance estimation in the presence of imputed data for the generalized estimation system. *Proc. of the American Statist. Assoc.(Social Survey Research Methods Section),* 384-389.

Murthy, M.N. (1967). *Sampling: Theory and methods*. Statistical Pub. Society.

Pandey, R., Thakur, N.S., & Yadav, K. (2015). Estimation of population mean using exponential ratio type imputation method under survey non-response. *Journal of the Indian Statistical Association*, *53*(1), 89-107.

Sande, I.G. (1979). A personal view of hot deck approach to automatic edit and imputation. *Journal Imputation Procedures. Survey Methodology*, *5*, 238-246.

Singh, B.K. & Nath K. (2018a). Estimation of population mean using ratio cum product compromised method of imputation in two-phase sampling scheme in sample survey. *Asian Journal of Mathematics & Statistics, 11*(1), 27-39.

Singh, B.K. & Nath K. (2019). Generalized class of dual to product cum dual to ratio estimator for population mean with imputation of missing data in two-phase sampling scheme. *International Journal of Mathematics and Statistics (IJMS), 21,* 86-96.

Singh, B.K., & Gogoi, U. (2017). Estimation of population mean using exponential dual to ratio type compromised imputation for missing data in survey sampling. *J Stat Appl Pro*, *3*, 515-522.

Singh, B.K., & Nath, K. (2018b). Some Imputation Methods in Two-Phase Sampling Scheme for Estimation of Population Mean. *Research & Reviews: Journal of Statistics (RRJoST), 7*(1), 1-16.

Singh, S., & Horn, S. (2000). Compromised imputation in survey sampling. *Metrika*, *51*(3), 267-276.

Wright, K., & Capps, C. (2011). A survey of information systems development project performance. *Academy of Information and Management Sciences Journal*, *14*(1), 87-105.

Received: 10-Feb-2022, Manuscript No. JMIDS-22-11218; **Editor assigned:** 15-Feb-2022, PreQC No. JMIDS-22-11218(PQ); **Reviewed:** 07-Mar-2022, QC No. JMIDS-22-11218; **Revised:** 29-Mar-2022, Manuscript No. JMIDS-22-11218 (R)**; Published:** 05-Apr-2022

Citation Information: Gohain, D., Nath, K., & Singh, B.K. (2022). Estimation of population mean using ratio type imputation technique with linear combination of two auxiliary variable under two-phase sampling. *Journal of Management Information and Decision Sciences, 25*(4), 1-20.