Available online at http://scik.org
J. Math. Comput. Sci. 11 (2021), No. 6, 7921-7938
https://doi.org/10.28919/jmcs/6671
ISSN: 1927-5307

STUDY OF IMPUTATION BASED GENERALIZED CLASSES OF DUAL TO PRODUCT CUM DUAL TO RATIO ESTIMATORS FOR MISSING DATA IN TWO-PHASE SAMPLING

KRISHNAJYOTHI NATH[1,*], DEEPJAN GOHAIN[1], NEELUTPOL GOGOI[2], B.K. SINGH[1]

[1]Department of Mathematics, North Eastern Regional Institute of Science and Technology (NERIST), Nirjuli, Arunachal Pradesh - 791109, India

[2]Department of Computer Science, Assam Down Town University, Guwahati-781026, Assam, India

Abstract: Present paper proposes three generalized classes of estimators for estimating population mean under the framework of two-phase sampling design in the presence of non-response in the study variable by using auxiliary information and also the expressions for bias and mean square error are derived. Further, theoretical results stating superiority of the proposed estimators, over the existing estimators by an empirical study based on different data sets from the classical statistical literature are shown.

Keywords: imputation; bias; mean square error (MSE); missing data; large sample approximation; simple random Sampling without replacement (SRSWOR); efficiency.

2010 AMS Subject Classification: 62D05.

---

*Corresponding author
E-mail address: krishnath10@gmail.com
Received August 20, 2021

1. INTRODUCTION

Sampling units may refuse to take part in the sample study or may not answer or not able to contact or unanticipated factors may accidently cause loss of some collected information leading to incomplete survey response. To deal with missing data effectively Kalton et al. (1981) [7] and Sande (1979) [14] suggested imputation methods that make an incomplete data set structurally complete and its analysis simple. Lee et al. (1994, 1995) [9, 10] used the information on an auxiliary variable for the purpose of imputation. Later Singh and Horn (2000) [20] introduced a compromised method of imputation based on auxiliary variables. Ahmed et al. (2006) [1] discussed several new imputation based estimators that used the information on an auxiliary variate and compared their performance with the mean method of imputation.

Shukla and Thakur (2008) [15], Rueda and Gonzalez (2008) [5, 13], Kadilar and Cingi (2008) [6], Shukla et al. (2009) [16], Shukla et al. (2009a) [17], Singh et al.(2010) [21], Diana and Francesco Perri (2010) [4], Baraldi and Enders (2010) [3], Thakur et al. (2011) [26], Shukla et al. (2011) [18], Thakur et al. (2013) [27], Shukla et al.(2013) [19], Singh et al. (2014) [22]. Singh et al. (2015, 2016,2017) [23, 24, 25], Thakur *et al.* (2016) [27] discussed 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.

2. NOTATIONS

Let $U=(U_1,U_2,U_3,\ldots,U_N)$ be a finite population of size N and character under study be y.

As usual, $\bar{Y}=N^{-1}\sum_{i=1}^{N}Y_i$, $\bar{X}=N^{-1}\sum_{i=1}^{N}X_i$ are population means, $\bar{X}$ is unknown and $\bar{Y}$ under investigation. Consider a preliminary large sample S' of size n' is drawn from population Ω by Simple Random Sampling without Replacement (SRSWOR) and a secondary sample S of size $n(n<n')$ is drawn in either as: a sub-sample from sample S' (denoted by design I) or as

independent to sample S' without replacing S' (denoted by design II).

Let sample size S of n units contains r responding units $(r<n)$ forming a sub-space R and $(n-r)$ non-responding with sub space R^C in $S=R\cup R^C$. For every $i\in R$, y_i is observed available. For $i\in R^C$, the y_i values are missing and imputed values are computed. The i^{th} value x_i of auxiliary variate is used as a source of imputation for missing data when $i\in R^C$. Assume for S, the data $x_s=\{x_i : i\in S\}$ and for $i'\in S'$, the data $\{x_i : i'\in S'\}$ are known with mean $\bar{x}=(n)^{-1}\sum_{i=1}^{n}x_i$ and $\bar{x}'=(n')^{-1}\sum_{i=1}^{n'}x_i$ respectively.

Remark 1: $\bar{y}_r=\bar{Y}(1+e_1)$; $\bar{x}_r=\bar{X}(1+e_2)$; $\bar{x}=\bar{X}(1+e_3)$; and $\bar{x}'=\bar{X}(1+e_3')$, which implies the results $e_1=\dfrac{\bar{y}_r}{\bar{Y}}-1$; $e_2=\dfrac{\bar{x}_r}{\bar{X}}-1$; $e_3=\dfrac{\bar{x}}{\bar{X}}-1$ and $e_3'=\dfrac{\bar{x}'}{\bar{X}}-1$. Now by using the concept of two-phase sampling, following Rao and Sitter (1995) [12] and the mechanism of Missing Completely at Random (MCAR), for given r,n and n', we have:

(i) Under design F1 [Case I]:

$$E(e_1)=E(e_2)=E(e_3)=E(e_3')=0;\quad E(e_1^2)=\delta_1 C_Y^2;\quad E(e_3^2)=\delta_2 C_X^2;\quad E(e_3'^2)=\delta_3 C_X^2;$$

$$E(e_1e_2)=\delta_1\rho C_Y C_X;\quad E(e_1e_3)=\delta_2\rho C_Y C_X;\quad E(e_1e_3')=\delta_3\rho C_Y C_X;\quad E(e_2e_3)=\delta_2 C_X^2;$$

$$E(e_2e_3')=\delta_3 C_X^2;\quad E(e_3e_3')=\delta_3 C_X^2;$$

(ii) Under design F2 [Case II]:

$$E(e_1)=E(e_2)=E(e_3)=E(e_3')=0;\quad E(e_1^2)=\delta_4 C_Y^2;\quad E(e_2^2)=\delta_4 C_X^2;\quad E(e_3^2)=\delta_5 C_X^2;$$

$$E(e_3'^2)=\delta_3 C_X^2;\quad E(e_1e_2)=\delta_4\rho C_Y C_X;\quad E(e_1e_3)=\delta_5\rho C_Y C_X;\quad E(e_1e_3')=0;\quad E(e_2e_3)=\delta_5 C_X^2;$$

$$E(e_2e_3')=0;\quad E(e_3e_3')=0$$

where, $\delta_1=\left(\dfrac{1}{r}-\dfrac{1}{n'}\right)$; $\delta_2=\left(\dfrac{1}{n}-\dfrac{1}{n'}\right)$; $\delta_3=\left(\dfrac{1}{n'}-\dfrac{1}{N}\right)$; $\delta_4=\left(\dfrac{1}{r}-\dfrac{1}{N-n'}\right)$; $\delta_5=\left(\dfrac{1}{n}-\dfrac{1}{N-n'}\right)$

3. REVIEWING EXISTING IMPUTATION METHOD AND CORRESPONDING ESTIMATORS

Let $\bar{Y} = N^{-1}\sum_{i=1}^{N} y_i$ be the mean of the finite population under consideration. A Simple Random Sampling Without Replacement (SRSWOR) S of size n is drawn from $\Omega = \{1,2,...,N\}$ to estimate the population mean $\bar{Y}$. Let the number of responding units out of sampled n units be denoted by $r(r < n),$ the set of responding units by $R,$ and that of non-responding units by R^C. For every unit $i \in R$ the value y_i is observed, but for the units $i \in R^C$, the observations y_i are missing and instead imputed values are derived. The i^{th} value x_i of auxiliary variate is used as a source of imputation for missing data when $i \in R^C$. Assume for S, the data $x_s = \{x_i : i \in S\}$ are known with mean $\bar{x} = (n)^{-1}\sum_{i'=1}^{n} x_i$. Under this setup, some well known imputation methods are given below.

4. PANDEY ET AL. (2015) SUGGESTED IMPUTATION STRATEGIES

For the case where y_{ji} denotes the i^{th} available observation for the j^{th} imputation method, the imputation method y_{1i} is given as follows:

(1) $$y_{1i}' = \begin{cases} y_i & if \quad i \in R \\ \dfrac{\bar{y}_r}{(1-f_1)}\left[\exp\left(\dfrac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right) - f_1\right] & if \;\; i \in R^C \end{cases} \qquad \text{Where } f_1 = \frac{r}{N};$$

Point estimator of $\bar{Y}$ is $t_1' = \bar{y}_r \exp\left(\dfrac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right)$

The Bias and MSE of t_1' under design I and Design II are given by

(i) $$B(t_1')_I = \frac{\bar{Y}}{8}\left[(3\delta_2 - 2\delta_3)C_x^2 + 4(\delta_3 - \delta_2)\rho C_y C_x\right]$$

(ii) $$B(t_1')_{II} = \frac{\bar{Y}}{8}\left[(3\delta_5 - \delta_3)C_x^2 + 4\delta_5 \rho C_y C_x\right]$$ and

(iii) $M\left(t_1^{'}\right)_I = \frac{\overline{Y}^2}{4}\left[4\delta_1 C_y^2 + \delta_2 C_x^2 - 4(\delta_3 - \delta_2)\rho C_y C_x\right]$

(iv) $M\left(t_1^{'}\right)_{II} = \frac{\overline{Y}^2}{4}\left[4\delta_5 C_y^2 + (\delta_3 + \delta_5)C_x^2 - 4\delta_5\rho C_y C_x\right]$

(2)
$$y_{2i}^{'} = \begin{cases} y_i & if \quad i \in R \\ \frac{\overline{y}_r}{(1-f_1)}\left[\exp\left(\frac{\overline{x}_r - \overline{x}}{\overline{x}_r + \overline{x}}\right) - f_1\right] & if \ i \in R^C \end{cases}$$

Point estimator of $\overline{Y}$ is $t_2^{'} = \overline{y}_r \exp\left(\frac{\overline{x}_r - \overline{x}}{\overline{x}_r + \overline{x}}\right)$

The Bias and MSE of $t_2^{'}$ under design I and Design II are given by

(i) $B(t_2^{'})_I = \frac{\overline{Y}}{2}\left[(\delta_1 - \delta_2)\rho C_y C_x - \frac{1}{4}(\delta_1 - 2\delta_2)C_x^2\right]$

(ii) $B(t_2^{'})_{II} = \frac{\overline{Y}}{2}\left[(\delta_4 - \delta_5)\rho C_y C_x - \frac{1}{4}(\delta_4 - 2\delta_5)C_x^2\right]$ and

(iii) $M\left(t_2^{'}\right)_I = \frac{\overline{Y}^2}{4}\left[4\delta_1 C_y^2 + \delta_1 C_x^2 + (\delta_1 - \delta_2)\rho C_y C_x\right]$

(iv) $M\left(t_2^{'}\right)_{II} = \frac{\overline{Y}^2}{4}\left[4\delta_4 C_y^2 + \delta_4 C_x^2 + (\delta_4 - \delta_5)\rho C_y C_x\right]$

(3)
$$y_{3i} = \begin{cases} y_i & if \quad i \in R \\ \frac{\overline{y}_r}{(1-f_1)}\left[\exp\left(\frac{\overline{x}^{'} - \overline{x}_r}{\overline{x}^{'} + \overline{x}_r}\right) - f_1\right] & if \ i \in R^C \end{cases}$$

Point estimator of $\overline{Y}$ is $t_3^{'} = \overline{y}_r \exp\left(\frac{\overline{x}^{'} - \overline{x}_r}{\overline{x}^{'} + \overline{x}_r}\right)$

The Bias and MSE of $t_3^{'}$ under design I and Design II are given by

(i) $B(t_3^{'})_I = \frac{\overline{Y}}{8}\left[(3\delta_1 - 2\delta_3)C_x^2 + 4(\delta_3 - \delta_1)\rho C_y C_x\right]$

(ii) $B(t_3^{'})_{II} = \frac{\overline{Y}}{8}\left[(3\delta_4 - \delta_3)C_x^2 + 4\delta_4\rho C_y C_x\right]$ and

(iii) $M\left(t_3^{'}\right)_I = \frac{\overline{Y}^2}{4}\left[4\delta_1 C_y^2 + \delta_1 C_x^2 - 4(\delta_3 - \delta_1)\rho C_y C_x\right]$

(iv) $M\left(t_3^{'}\right)_{II} = \frac{\overline{Y}^2}{4}\left[4\delta_4 C_y^2 + (\delta_3 + \delta_4)C_x^2 - 4\delta_4 \rho C_y C_x\right]$

5. SOME PROPOSED METHODS OF IMPUTATION AND THEIR ESTIMATORS

For estimating the population mean $\bar{y}$ of the study variate y, Singh and Ruiz Espejo (2003) considered an estimator of the ratio-product type given by

$$\bar{y}_{RP} = \bar{y}\left\{k\frac{\overline{X}}{\bar{x}} + (1-k)\frac{\bar{x}}{\overline{X}}\right\}$$

Where $\bar{y}$ and $\bar{x}$ are the sample means of y and x respectively based on a sample of size n out of the population of N units. $\overline{X}$ is the known population mean of x, and k is suitably chosen constant.

Choudhury and Singh (2012) studied dual to product cum dual to ratio estimator. Further Swain (2013) generalized the estimator suggested by Choudhury and Singh (2012) by using transformation as

$$\overline{Y}_{DGR} = \bar{y}\left[\alpha\left(\frac{\overline{X}}{\bar{x}^*}\right)^g + (1-\alpha)\left(\frac{\bar{x}^*}{\overline{X}}\right)^h\right]^\delta$$

Where

$\bar{x}^* = \frac{N\overline{X} - n\bar{x}}{N-n}$, $\bar{y}\left(\frac{\overline{X}}{\bar{x}^*}\right)$ is the dual to product estimator and $\bar{y}\left(\frac{\bar{x}^*}{\overline{X}}\right)$ is the dual to ratio estimator; g, h and δ are the free parameters and α is chosen so as to minimize the approximate mean square error of $\bar{y}_{DGR}$ to $o\left(1/n\right)$.

Motivated by Choudhury and Singh (2012) and Swain (2013), we here propose the following three generalized dual to product cum dual to ratio methods of imputation under two-phase sampling.

Let y'_{ji} denote the i^{th} available observation for the j^{th} imputation method.

The ***first proposed*** method:

$$y'_{1i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{n-r}\left[\alpha\left(\dfrac{\bar{x}'}{\Psi_1}\right)^q + (1-\alpha)\left(\dfrac{\Psi_1}{\bar{x}'}\right)^h\right]^\delta & if \quad i \in R^c \end{cases}$$

Using above, the imputation-based estimator of population mean $\bar{Y}$ is

$$\bar{y}^d_{GDPR1} = \bar{y}_r\left[\alpha\left(\frac{\bar{x}'}{\Psi_1}\right)^q + (1-\alpha)\left(\frac{\Psi_1}{\bar{x}'}\right)^h\right]^\delta \tag{1}$$

Where $\Psi_1 = \dfrac{N\bar{x}' - n\bar{x}}{N-n}$, $\bar{y}_r\left(\dfrac{\bar{x}'}{\Psi_1}\right)$ is the dual to product estimator and $\bar{y}_r\left(\dfrac{\Psi_1}{\bar{x}'}\right)$ is the dual to ratio estimator; q, h and δ are the free parameters and α is chosen so as to minimize the approximate mean square error of $\bar{y}^d_{GDPR1}$.

Theorem 5.1:

The estimator, Bias, M.S.E. and minimum M.S.E of $\bar{y}^d_{GDPR1}$ in terms of e_i; i=1, 2, 3 and e'_3 under design $t = I, II$ upto first order of approximation are given by

$$\begin{aligned}\bar{y}^d_{GDPR1} = \bar{Y}\Big[&1 + e_1 + \delta q\alpha\big(ge_3 - ge'_3 + g^2 e'^2_3 + g^2 e^2_3 + 2ge'^2_3 - ge_3e'_3 - 2g^2 e_3e'_3 \\ &- ge'^2_3\big) + \frac{q(q-1)}{2}\delta\alpha g^2\left(e^2_3 + e'^2_3 - 2e_3e'_3\right) + \delta hg\left(e'_3 - e'^2_3 - e_3 + e_3e'_3\right) \\ &+ \frac{h(h-1)}{2}\delta g^2\left(e'^2_3 + e^2_3 - 2e_3e'_3\right) - \delta h\alpha g\left(e'_3 - e'^2_3 - e_3 + e_3e'_3\right) \\ &- \frac{h(h-1)}{2}\delta\alpha g^2\left(e'^2_3 + e^2_3 - 2e_3e'_3\right) + \frac{\delta(\delta-1)}{2}q^2\alpha^2 g^2\left(e^2_3 + e'^2_3 - 2e_3e'_3\right) \\ &+ \frac{\delta(\delta-1)}{2}h^2 g^2\left(e'^2_3 + e^2_3 - 2e_3e'_3\right) + \frac{\delta(\delta-1)}{2}h^2\alpha^2 g^2\left(e'^2_3 + e^2_3 - 2e_3e'_3\right) \\ &+ \delta(\delta-1)qh\alpha g^2\left(e_3e'_3 - e^2_3 - e'^2_3 + e_3e'_3\right) - \delta(\delta-1)qh\alpha^2 g^2\big(e_3e'_3 - e^2_3 \\ &- e'^2_3 + e_3e'_3\big) - \delta(\delta-1)h^2\alpha g^2\left(e'^2_3 - e_3e'_3 - e_3e'_3 + e^2_3\right) + \delta q\alpha g\left(e_1e_3 - e_1e'_3\right) \\ &+ \delta hg\left(e_1e'_3 - e_1e_3\right) - \delta h\alpha g\left(e_1e'_3 - e_1e_3\right)\Big]\end{aligned} \tag{2}$$

$$B\left(\bar{y}^{d}_{GDPR1}\right)_{I}=\bar{Y}\Big[(\delta_2-\delta_3)\Big\{\delta q\alpha g^2C_x^2+\frac{q(q-1)}{2}\delta\alpha g^2C_x^2+\frac{h(h-1)}{2}\delta g^2C_x^2+\frac{h(h-1)}{2}\delta g^2C_x^2-\frac{h(h-1)}{2}\delta\alpha g^2C_x^2+\frac{\delta(\delta-1)}{2}q^2\alpha^2g^2C_x^2+\frac{\delta(\delta-1)}{2}h^2g^2C_x^2+\frac{\delta(\delta-1)}{2}h^2\alpha^2g^2C_x^2-\delta(\delta-1)qh\alpha g^2C_x^2+\delta(\delta-1)qh\alpha^2g^2C_x^2-\delta(\delta-1)h^2\alpha g^2C_x^2+\delta q\alpha g\rho C_yC_x-\delta hg\rho C_yC_x+\delta h\alpha g\rho C_yC_x\Big\}\Big] \tag{3}$$

$$B\left(\bar{y}^{d}_{GDPR1}\right)_{II}=\bar{Y}\Big[(\delta_3+\delta_5)\Big\{\delta q\alpha g^2C_x^2+\frac{q(q-1)}{2}\delta\alpha g^2C_x^2+\frac{h(h-1)}{2}\delta gC_x^2+\frac{h(h-1)}{2}\delta\alpha g^2C_x^2+\frac{\delta(\delta-1)}{2}q^2\alpha^2g^2C_x^2+\frac{\delta(\delta-1)}{2}h^2g^2C_x^2+\frac{\delta(\delta-1)}{2}h^2\alpha^2g^2C_x^2-\delta(\delta-1)qh\alpha g^2C_x^2+\delta(\delta-1)qh\alpha^2g^2C_x^2-\delta(\delta-1)h^2\alpha g^2C_x^2\Big\}+\Big(\delta q\alpha g\delta_3C_x^2-\delta hg\delta_3C_x^2+2\delta\alpha g\delta_3C_x^2+\delta q\alpha g\delta_5\rho C_yC_x-\delta hg\delta_5\rho C_yC_x+\delta h\alpha g\delta_5\rho C_yC_x\Big)\Big] \tag{4}$$

$$M\left(\bar{y}^{d}_{GDPR1}\right)_{I}=\bar{Y}^2\Big[\delta_1C_y^2+(\delta_2-\delta_3)\Big\{\delta^2q^2\alpha^2g^2C_x^2+\delta^2h^2g^2C_x^2+\delta^2h^2\alpha^2g^2C_x^2+2\delta q\alpha g\rho C_yC_x-2\delta hg\rho C_yC_x+2\delta h\alpha g\rho C_yC_x-2\delta^2qh\alpha g^2C_x^2+2\delta^2qh\alpha^2g^2C_x^2-2\delta^2h^2\alpha g^2C_x^2\Big\}\Big] \tag{5}$$

$$M\left(\bar{y}^{d}_{GDPR1}\right)_{II}=\bar{Y}^2\Big[\delta_4C_y^2+(\delta_3+\delta_5)\Big\{\delta^2q^2\alpha^2g^2C_x^2+\delta^2h^2g^2C_x^2+\delta^2h^2\alpha^2g^2C_x^2-2\delta^2q\alpha g^2C_x^2+2\delta^2qh\alpha^2g^2C_x^2-2\delta^2h^2\alpha g^2\rho C_yC_x\Big\}+2\delta q\alpha g\delta_5\rho C_yC_x-2\delta hg\delta_5\rho C_yC_x\Big] \tag{6}$$

and $$\left\lfloor M\left(\bar{y}^{d}_{GDPR1}\right)_{I}\right\rfloor_{\min}=\bar{Y}^2\left[\delta_1-(\delta_2-\delta_3)\rho^2\right]C_y^2 \quad when\ q=\frac{\delta hgC_x(1-\alpha)-\rho C_y}{\delta\alpha gC_x} \tag{7}$$

$$\left\lfloor M\left(\bar{y}^{d}_{GDPR1}\right)_{II}\right\rfloor_{\min}=\bar{Y}^2\left[\delta_4-\delta_5^2(\delta_3+\delta_5)^{-1}\rho^2\right]C_y^2 \quad when\ q=\frac{(\delta_3+\delta_5)\delta hgC_x(1-\alpha)-\delta_5\rho C_y}{(\delta_3+\delta_5)\delta\alpha gC_x} \tag{8}$$

The ***second proposed*** method:

$$y'_{2i}=\begin{cases} y_i & if \quad i\in R \\ \dfrac{1}{n-r}\left[\alpha\left(\dfrac{\bar{x}}{\Psi_2}\right)^q+(1-\alpha)\left(\dfrac{\Psi_2}{\bar{x}}\right)^h\right]^\delta & if \quad i\in R^c \end{cases}$$

Using above, the imputation-based estimator of population mean $\bar{Y}$ is

$$\bar{y}^{d}_{GDPR2} = \bar{y}_r \left[\alpha \left(\frac{\bar{x}}{\Psi_2} \right)^q + (1-\alpha)\left(\frac{\Psi 2}{\bar{x}} \right)^h \right]^{\delta} \tag{9}$$

Where $\Psi_2 = \dfrac{N\bar{x} - n\bar{x}_r}{N-n}$, $\bar{y}_r\left(\dfrac{\bar{x}}{\Psi_2}\right)$ is the dual to product estimator and $\bar{y}_r\left(\dfrac{\Psi_2}{\bar{x}}\right)$ is the dual to ratio estimator; q, h and δ are the free parameters and α is chosen so as to minimize the approximate mean square error of $\bar{y}^{d}_{GDPR2}$.

Theorem 5.2:

The estimator, Bias, M.S.E. and minimum M.S.E of $\bar{y}^{d}_{GDPR2}$ in terms of e_i; i=1, 2, 3 and $e_3^{'}$ under design $t = I, II$ upto first order of approximation are given by

$$\begin{aligned}
\bar{y}^{d}_{GDPR2} = \bar{Y}\Big[& 1 + e_1 + \delta q\alpha\big(ge_2 - ge_3 + g^2e_3^2 + g^2e_2^2 + 2ge_3^{2'} - ge_2e_3 - 2g^2e_2e_3 \\
& - ge_3^2\big) + \frac{q(q-1)}{2}\delta\alpha g^2\left(e_2^2 + e_3^2 - 2e_2e_3\right) + \delta hg\left(e_3 - e_3^2 - e_2 + e_2e_3\right) \\
& + \frac{h(h-1)}{2}\delta g^2\left(e_3^2 + e_2^2 - 2e_2e_3\right) - \delta h\alpha g\left(e_3 - e_3^2 - e_2 + e_2e_3\right) \\
& - \frac{h(h-1)}{2}\delta\alpha g^2\left(e_3^2 + e_2^2 - 2e_2e_3\right) + \frac{\delta(\delta-1)}{2}q^2\alpha^2g^2\left(e_2^2 + e_3^{2'} - 2e_2e_3\right) \\
& + \frac{\delta(\delta-1)}{2}h^2g^2\left(e_3^2 + e_2^2 - 2e_2e_3\right) + \frac{\delta(\delta-1)}{2}h^2\alpha^2g^2\left(e_3^2 + e_2^2 - 2e_2e_3\right) \\
& + \delta(\delta-1)qh\alpha g^2\left(e_2e_3 - e_2^2 - e_3^2 + e_2e_3\right) - \delta(\delta-1)qh\alpha^2g^2\big(e_2e_3 - e_2^2 \\
& - e_3^2 + e_2e_3\big) - \delta(\delta-1)h^2\alpha g^2\left(e_3^2 - e_2e_3 - e_2e_3 + e_2^2\right) + \delta q\alpha g\left(e_1e_2 - e_1e_3\right) \\
& + \delta hg\left(e_1e_3 - e_1e_2\right) - \delta h\alpha g\left(e_1e_3 - e_1e_2\right)\Big]
\end{aligned} \tag{10}$$

$$\begin{aligned}
B\left(\bar{y}^{d}_{GDPR2}\right)_I = \bar{Y}\Big[& (\delta_1 - \delta_2)\Big\{\delta q\alpha g^2C_x^2 + \frac{q(q-1)}{2}\delta\alpha g^2C_x^2 + \frac{h(h-1)}{2}\delta g^2C_x^2 \\
& - \frac{h(h-1)}{2}\delta\alpha g^2C_x^2 + \frac{\delta(\delta-1)}{2}q^2\alpha^2g^2C_x^2 + \frac{\delta(\delta-1)}{2}h^2g^2C_x^2 \\
& + \frac{\delta(\delta-1)}{2}h^2\alpha^2g^2C_x^2 - \delta(\delta-1)qh\alpha g^2C_x^2 + \delta(\delta-1)qh\alpha^2g^2C_x^2 \\
& - \delta(\delta-1)h^2\alpha g^2C_x^2 + \delta q\alpha g\rho C_yC_x - \delta hg\rho C_yC_x + \delta h\alpha g\rho C_yC_x\Big]
\end{aligned} \tag{11}$$

$$B\left(\overline{y}^{d}_{GDPR2}\right)_{II} = \overline{Y}\Big[(\delta_4-\delta_5)\Big\{\delta q\alpha g^2C_x^2 + \frac{q(q-1)}{2}\delta\alpha g^2C_x^2 + \frac{h(h-1)}{2}\delta g^2C_x^2 - \frac{h(h-1)}{2}\delta\alpha g^2C_x^2 + \frac{\delta(\delta-1)}{2}q^2\alpha^2g^2C_x^2 + \frac{\delta(\delta-1)}{2}h^2g^2C_x^2 + \frac{\delta(\delta-1)}{2}h^2\alpha^2g^2C_x^2 - \delta(\delta-1)qh\alpha g^2C_x^2 + \delta(\delta-1)qh\alpha^2g^2C_x^2 - \delta(\delta-1)h^2\alpha g^2C_x^2 + \delta q\alpha g\rho C_yC_x - \delta hg\rho C_yC_x + \delta h\alpha g\rho C_yC_x\Big]\tag{12}$$

$$M\left(\overline{y}^{d}_{GDPR2}\right)_{I} = \overline{Y}^2\Big[\delta_1C_y^2 + (\delta_1-\delta_2)\Big\{\delta^2\alpha^2q^2g^2C_x^2 + \delta^2h^2g^2C_x^2 + \delta^2h^2\alpha^2g^2C_x^2 - 2\delta^2qh\alpha g^2C_x^2 - 2\delta^2h^2\alpha g^2C_x^2 + 2\delta^2qh\alpha^2g^2C_x^2 + 2\delta q\alpha g\rho C_yC_x - 2\delta hg\rho C_yC_x + 2\delta h\alpha g\rho C_yC_x\Big]\tag{13}$$

$$M\left(\overline{y}^{d}_{GDPR2}\right)_{II} = \overline{Y}^2\Big[\delta_4C_y^2 + (\delta_4-\delta_5)\Big\{\delta^2\alpha^2q^2g^2C_x^2 + \delta^2h^2g^2C_x^2 + \delta^2h^2\alpha^2g^2C_x^2 - 2\delta^2qh\alpha g^2C_x^2 - 2\delta^2h^2\alpha g^2C_x^2 + 2\delta^2qh\alpha^2g^2C_x^2 + 2\delta q\alpha g\rho C_yC_x - 2\delta hg\rho C_yC_x + 2\delta h\alpha g\rho C_yC_x\Big]\tag{14}$$

and
$$\Big\lfloor M\left(\overline{y}^{d}_{GDPR2}\right)_{I}\Big\rfloor_{\min} = \overline{Y}^2\left[\delta_1 - (\delta_1-\delta_2)\rho^2\right]C_y^2 \;\; when\;\; q = \frac{\delta hgC_x(1-\alpha) - \rho C_y}{\delta\alpha gC_x}\tag{15}$$

$$\Big\lfloor M\left(\overline{y}^{d}_{GDPR2}\right)_{II}\Big\rfloor_{\min} = \overline{Y}^2\left[\delta_4 - (\delta_4-\delta_5)\rho^2\right]C_y^2 \qquad when\;\; q = \frac{\delta hgC_x(1-\alpha) - \rho C_y}{\delta\alpha gC_x}\tag{16}$$

The ***third proposed*** method:

$$y'_{3i} = \begin{cases} y_i & if \quad i\in R \\ \dfrac{1}{n-r}\left[\alpha\left(\dfrac{\overline{x}'}{\Psi_3}\right)^q + (1-\alpha)\left(\dfrac{\Psi_3}{\overline{x}'}\right)^h\right]^\delta & if \quad i\in R^c \end{cases}$$

Using above, the imputation-based estimator of population mean $\overline{Y}$ is

$$\overline{y}^{d}_{GDPR3} = \overline{y}_r\left[\alpha\left(\frac{\overline{x}'}{\Psi_3}\right)^q + (1-\alpha)\left(\frac{\Psi_3}{\overline{x}'}\right)^h\right]^\delta\tag{17}$$

Where $\Psi_3 = \dfrac{N\overline{x}' - n\overline{x}_r}{N-n}$, $\overline{y}_r\left(\dfrac{\overline{x}'}{\Psi_3}\right)$ is the dual to product estimator and $\overline{y}_r\left(\dfrac{\Psi_3}{\overline{x}'}\right)$ is the dual to ratio estimator; q, h and δ are the free parameters and α is chosen so as to minimize the approximate mean square error of $\overline{y}^{d}_{GDPR3}$.

Theorem 5.3:

The estimator, Bias, M.S.E. and minimum M.S.E of $\overline{y}^{d}_{GDPR3}$ in terms of e_i ; i=1, 2, 3 and $e_3^{'}$ under design $t = I, II$ upto first order of approximation are given by

$$\begin{aligned}
\overline{y}^{d}_{GDPR3} = \overline{Y}\Big[&1 + e_1 + \delta q\alpha\big(ge_2 - ge_3^{'} + g^2 e_3^{'2} + g^2 e_2^2 + 2ge_3^{'2} - ge_2e_3^{'} - 2g^2 e_2 e_3^{'} \\
&- ge_3^{'2}\big) + \frac{q(q-1)}{2}\delta\alpha g^2\left(e_2^2 + e_3^{'2} - 2e_2e_3^{'}\right) + \delta hg\left(e_3^{'} - e_3^{'2} - e_2 + e_2e_3^{'}\right) \\
&+ \frac{h(h-1)}{2}\delta g^2\left(e_3^{'2} + e_2^2 - 2e_2e_3^{'}\right) - \delta h\alpha g\left(e_3^{'} - e_3^{'2} - e_2 + e_2e_3^{'}\right) \\
&- \frac{h(h-1)}{2}\delta\alpha g^2\left(e_3^{'2} + e_2^2 - 2e_2e_3^{'}\right) + \frac{\delta(\delta-1)}{2}q^2\alpha^2 g^2\left(e_2^2 + e_3^{'2} - 2e_2e_3^{'}\right) \\
&+ \frac{\delta(\delta-1)}{2}h^2 g^2\left(e_3^{'2} + e_2^2 - 2e_2e_3^{'}\right) + \frac{\delta(\delta-1)}{2}h^2\alpha^2 g^2\left(e_3^{'2} + e_2^2 - 2e_2e_3^{'}\right) \\
&+ \delta(\delta-1)qh\alpha g^2\left(e_2e_3^{'} - e_2^2 - e_3^{'2} + e_2e_3^{'}\right) - \delta(\delta-1)qh\alpha^2 g^2\big(e_2e_3^{'} - e_2^2 \\
&- e_3^{'2} + e_2e_3^{'}\big) - \delta(\delta-1)h^2\alpha g^2\left(e_3^{'2} - e_2e_3^{'} - e_2e_3^{'} + e_2^2\right) + \delta q\alpha g\left(e_1e_2 - e_1e_3^{'}\right) \\
&+ \delta hg\left(e_1e_3^{'} - e_1e_2\right) - \delta h\alpha g\left(e_1e_3^{'} - e_1e_2\right)\Big)\Big]
\end{aligned} \tag{18}$$

$$\begin{aligned}
B\left(\overline{y}^{d}_{GDPR3}\right)_I = \overline{Y}\Big[&(\delta_1 - \delta_3)\Big\{\delta q\alpha g^2 C_x^2 + \frac{q(q-1)}{2}\delta\alpha g^2 C_x^2 + \frac{h(h-1)}{2}\delta g^2 C_x^2 \\
&+ \frac{h(h-1)}{2}\delta g^2 C_x^2 - \frac{h(h-1)}{2}\delta\alpha g^2 C_x^2 + \frac{\delta(\delta-1)}{2}q^2\alpha^2 g^2 C_x^2 \\
&+ \frac{\delta(\delta-1)}{2}h^2 g^2 C_x^2 + \frac{\delta(\delta-1)}{2}h^2\alpha^2 g^2 C_x^2 - \delta(\delta-1)qh\alpha g^2 C_x^2 \\
&+ \delta(\delta-1)qh\alpha^2 g^2 C_x^2 - \delta(\delta-1)h^2\alpha g^2 C_x^2 + \delta q\alpha g\rho C_y C_x \\
&- \delta hg\rho C_y C_x + \delta h\alpha g\rho C_y C_x\Big\}\Big]
\end{aligned} \tag{19}$$

$$\begin{aligned}
B\left(\overline{y}^{d}_{GDPR3}\right)_{II} = \overline{Y}\Big[&(\delta_3 + \delta_4)\Big\{\delta q\alpha g^2 C_x^2 + \frac{q(q-1)}{2}\delta\alpha g^2 C_x^2 + \frac{h(h-1)}{2}\delta g C_x^2 \\
&+ \frac{h(h-1)}{2}\delta\alpha g^2 C_x^2 + \frac{\delta(\delta-1)}{2}q^2\alpha^2 g^2 C_x^2 + \frac{\delta(\delta-1)}{2}h^2 g^2 C_x^2 \\
&+ \frac{\delta(\delta-1)}{2}h^2\alpha^2 g^2 C_x^2 - \delta(\delta-1)qh\alpha g^2 C_x^2 + \delta(\delta-1)qh\alpha^2 g^2 C_x^2 \\
&- \delta(\delta-1)h^2\alpha g^2 C_x^2\Big\} + \Big(\delta q\alpha g\delta_3 C_x^2 - \delta hg\delta_3 C_x^2 + 2\delta\alpha g\delta_3 C_x^2 \\
&+ \delta q\alpha g\delta_5\rho C_y C_x - \delta hg\delta_5\rho C_y C_x + \delta h\alpha g\delta_5\rho C_y C_x\Big)\Big]
\end{aligned} \tag{20}$$

$$M\left(\bar{y}^{d}_{GDPR3}\right)_I = \bar{Y}^2\Big[\delta_1 C_y^2 + (\delta_1-\delta_3)\Big\{\delta^2 q^2\alpha^2 g^2 C_x^2 + \delta^2 h^2 g^2 C_x^2 + \delta^2 h^2\alpha^2 g^2 C_x^2 + 2\delta q\alpha g\rho C_y C_x - 2\delta hg\rho C_y C_x + 2\delta h\alpha g\rho C_y C_x - 2\delta^2 qh\alpha g^2 C_x^2 + 2\delta^2 qh\alpha^2 g^2 C_x^2 - 2\delta^2 h^2\alpha g^2 C_x^2\Big\}\Big] \quad (21)$$

$$M\left(\bar{y}^{d}_{GDPR3}\right)_{II} = \bar{Y}^2\Big[\delta_4 C_y^2 + (\delta_3+\delta_4)\Big\{\delta^2 q^2\alpha^2 g^2 C_x^2 + \delta^2 h^2 g^2 C_x^2 + \delta^2 h^2\alpha^2 g^2 C_x^2 - 2\delta^2 q\alpha g^2 C_x^2 + 2\delta^2 qh\alpha^2 g^2 C_x^2 - 2\delta^2 h^2\alpha g^2\rho C_y C_x\Big\} + 2\delta q\alpha g\delta_4\rho C_y C_x - 2\delta hg\delta_4\rho C_y C_x\Big] \quad (22)$$

and $$\left\lfloor M\left(\bar{y}^{d}_{GDPR3}\right)_I\right\rfloor_{\min} = \bar{Y}^2\left[\delta_1 - (\delta_1-\delta_3)\rho^2\right]C_y^2 \quad when\ q = \frac{\delta hgC_x(1-\alpha) - \rho C_y}{\delta\alpha gC_x} \quad (23)$$

$$\left\lfloor M\left(\bar{y}^{d}_{GDPR3}\right)_{II}\right\rfloor_{\min} = \bar{Y}^2\left[\delta_4 - \delta_4^2(\delta_3+\delta_4)^{-1}\rho^2\right]C_y^2 \quad when\ q = \frac{(\delta_3+\delta_4)\delta hgC_x(1-\alpha) - \delta_5\rho C_y}{(\delta_3+\delta_4)\delta\alpha gC_x} \quad (24)$$

6. Comparison of Proposed and Existing Estimators Under Two-Phase Sampling Design

In this section, we derive the conditions under which the suggested estimators are superior to the existing estimators in design I and II.

(1) $$\Delta_1 = M\left(t_1'\right)_I - \left\lfloor M\left(\bar{y}^{d}_{GDPR1}\right)_I\right\rfloor_{\min}$$

$$= \bar{Y}^2\left[\left\{\left(\frac{1}{n}-\frac{1}{n'}\right)C_x^2 + \left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(\rho C_y C_x - \rho^2 C_y^2\right)\right\}\right]$$

$\left(\bar{y}^{d}_{GDPR2}\right)_I$ is better than t_1', when $\Delta_1 > 0$

This generates two conditions,

(i) When $\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right) > 0 \Rightarrow n < \frac{Nn'}{2N-n'}$

and (ii) When $\left(\rho C_Y C_X - \rho^2 C_Y^2\right) > 0 > 0 \quad \Rightarrow \rho < \frac{C_Y}{C_X}$

(2) $$\Delta_2 = M\left(t_1^{'}\right)_{II} - \left\lfloor M\left(\overline{y}^{d}_{GDPR1}\right)_{II}\right\rfloor_{\min}$$

$$=\overline{Y}^2\left[\frac{1}{4}\left(\frac{1}{n'}-\frac{1}{N}+\frac{1}{n}-\frac{1}{N-n'}\right)C_x^2-\left(\frac{1}{n}-\frac{1}{N-n'}\right)\rho C_y C_x+\frac{\left(\frac{1}{n}-\frac{1}{N-n'}\right)^2\rho^2 C_y^2}{\left[\left(\frac{1}{n'}-\frac{1}{N}\right)+\left(\frac{1}{n}-\frac{1}{N-n'}\right)\right]}\right]$$

. Thus $\left(\overline{y}^{d}_{GDPR1}\right)_{II}$ is better than $\left(t_1^{'}\right)_{II}$ when $\left(\frac{1}{n}-\frac{1}{N-n'}\right)<0$.

(3) $$\Delta_3 = M\left(t_2^{'}\right)_{I} - \left\lfloor M\left(\overline{y}^{d}_{GDPR2}\right)_{I}\right\rfloor_{\min} \quad =\overline{Y}^2\left[\left\{\left(\frac{1}{r}-\frac{1}{n}\right)\left(\rho C_y C_x-\rho^2 C_y^2\right)\right\}\right]$$

Thus $\left(\overline{y}^{d}_{GDPR2}\right)_{I}$ is better than $\left(t_2^{'}\right)_{I}$ when $\left(\rho C_Y C_X-\rho^2 C_Y^2\right)>0$ >0 $\Rightarrow \rho<\frac{C_Y}{C_X}$

(4) $$\Delta_4 = M\left(t_2^{'}\right)_{II} - \left\lfloor M\left(\overline{y}^{d}_{GDPR2}\right)_{II}\right\rfloor_{\min}$$

$$=\overline{Y}^2\left[\frac{1}{4}\left(\frac{1}{r}-\frac{1}{N-n'}\right)C_x^2+\frac{1}{4}\left(\frac{1}{r}-\frac{1}{n}\right)\rho C_y C_x+\frac{\left(\frac{1}{n}-\frac{1}{N-n'}\right)^2\rho^2 C_y^2}{\left[\left(\frac{1}{n'}-\frac{1}{N}\right)+\left(\frac{1}{n}-\frac{1}{N-n'}\right)\right]}\right]$$

Which is positive. Thus $\left(\overline{y}^{d}_{GDPR2}\right)_{II}$ is better than $\left(t_2^{'}\right)_{II}$

(5) $$\Delta_5 = M\left(t_3^{'}\right)_{I} - \left\lfloor M\left(\overline{y}^{d}_{GDPR3}\right)_{I}\right\rfloor_{\min}$$

$$=\overline{Y}^2\left[\left\{\frac{1}{4}\left(\frac{1}{r}-\frac{1}{n'}\right)C_x^2+\left(\frac{1}{r}-\frac{2}{n'}+\frac{1}{N}\right)\left(\rho C_y C_x-\rho^2 C_y^2\right)\right\}\right]$$

Thus $\left(\overline{y}^{d}_{GDPR3}\right)_{I}$ is better than $\left(t_3^{'}\right)_{I}$

This generates two conditions,

(i) When $\left(\frac{1}{r}-\frac{2}{n'}+\frac{1}{N}\right)>0 \Rightarrow r<\frac{Nn'}{2N-n'}$

and (ii) When $\left(\rho C_Y C_X-\rho^2 C_Y^2\right)>0$ >0 $\Rightarrow \rho<\frac{C_Y}{C_X}$

(6) $$\Delta_6 = M\left(t_3'\right)_{II} - \left\lfloor M\left(\bar{y}^{d}_{GDPR3}\right)_{II}\right\rfloor_{\min}$$

$$= \bar{Y}^2\left[\frac{1}{4}\left(\frac{1}{n'}-\frac{1}{N}+\frac{1}{r}-\frac{1}{N-n'}\right)C_x^2+\left(\frac{1}{r}-\frac{1}{N-n'}\right)\rho C_y C_x+\frac{\left(\frac{1}{r}-\frac{1}{N-n'}\right)^2\rho^2 C_y^2}{\left[\left(\frac{1}{n'}-\frac{1}{N}\right)+\left(\frac{1}{r}-\frac{1}{N-n'}\right)\right]}\right]$$

Which is positive. Thus $\left(\bar{y}^{d}_{GDPR3}\right)_{II}$ is better than $\left(t_3'\right)_{II}$

7. ILLUSTRATIVE EXAMPLES

Population **A [Source: Kadilar and Cingi (2006)]:** Y represents apple production and X represents number of apple trees. Other related information to population are:

$N = 106$ $n = 20$ $\bar{Y} = 2212.59$ $\bar{X} = 27421.70$ $\rho = 0.86$ $S_y = 11551.53$
$C_y = 5.22$ $S_x = 57460.61$ $C_x = 2.10$ $n' = 60$ $r = 15$

Population **B [Source: Koyuncu and Kadilar (2009)]:** Y is number of teachers and X is number of students. Other related information to population are:

$N = 923$ $n = 180$ $\bar{Y} = 436.4345$ $\bar{X} = 11440.498$ $\rho = 0.9543$ $S_y = 749.9394$
$C_y = 1.7183$ $S_x = 21331.131$ $C_x = 1.8645$ $n' = 360$ $r = 175$

Population **C [Source: Murthy (1967)]:** Y is area under winter paddy and X is corresponding geographical area. The following data is based on the given population:

$N = 108$ $n = 30$ $\bar{Y} = 172.3704$ $\bar{X} = 461.3981$ $\rho = 0.7896$ $S_y = 134.3567$
$C_y = 0.7795$ $S_x = 318.5022$ $C_x = 0.6903$ $n' = 70$ $r = 20$

The M.S.E.of suggested estimators $\bar{y}^{d}_{GDPR1}, \bar{y}^{d}_{GDPR2}$ *and* $\bar{y}^{d}_{GDPR3}$ and Pandey et al. (2015) under design I and II for population A, B and C are given in the table below.

Table 7.1: **MSE of suggested estimators for Population A, B and C under Two-Phase Sampling design**

Estimators	Population A		Population B		Population C	
	Under design-I	Under design-II	Under design-I	Under design-II	Under design-I	Under design-II
$\bar{y}^{d}_{GDPR1}$	4094721.103	3773113.28	1096.570493	878.2611311	486.9444334	381.5565279
$\bar{y}^{d}_{GDPR2}$	5025473.008	4348826.492	1570.161843	2133.437874	457.1674913	239.9844253
$\bar{y}^{d}_{GDPR3}$	2450392.738	2175247.21	1015.275195	804.5358235	299.3496978	207.6684095

Table 7.2: **MSE for Population A, B and C under Two-Phase Sampling design for Pandey et al. (2015) suggested estimators**

Estimators	Population A		Population B		Population C	
	Under design-I	Under design-II	Under design-I	Under design-II	Under design-I	Under design-II
t'_1	8054308.155	4880431.566	2742.222737	919.941012	889.1800587	381.6214812
t'_2	7131968.644	6427944.271	2160.675791	2889.753176	823.7714449	564.0080045
t'_3	4965874.345	4201188.065	1414.197776	853.7806146	383.7788822	230.218116

The sampling efficiency of the suggested estimators under design I and II over Pandey et al. (2015) suggested estimator is defined as:

$$E_i = \frac{Opt\left[M\left(\bar{y}^{d}_{GDPRi}\right)_j\right]}{Opt\left[M\left(t'_i\right)\right]}; \quad i = 1, 2, 3; \quad j = I,\ II$$

Table 7.3: Efficiency of suggested estimators for population A, B and C over Pandey et al.(2015)

Estimators	Population A		Population B		Population C	
	Under design-I	Under design-II	Under design-I	Under design-II	Under design-I	Under design-II
E_1	0.50838893	0.77311058	0.399883816	0.954692877	0.547633102	0.999829797
E_2	0.704640368	0.676550124	0.72669942	0.738276851	0.55496885	0.425498261
E_3	0.493446384	0.51776954	0.717915989	0.942321493	0.780005653	0.902050686

8. CONCLUSION

Imputation is a strategy which provides nearest value corresponding to the missing unit in an item survey. The tables 7.1 and 7.2 are given for the MSE of the suggested estimators and the existing estimators for population A, B and C in two different designs. Based on these results, we notice in Table 7.3 that all the proposed estimators have higher efficiency than the Pandey et al (2015) proposed estimators. The present paper is therefore an important contribution to the field of study of missing data problem in estimation of population parameter in two phase sampling design.

CONFLICT OF INTERESTS

The author(s) declare that there is no conflict of interests.

REFERENCES

[1] M. S. Ahmed, O. Al-Titi, Z. Al-Rawi, W. Abu-Dayyeh, Estimation of a population mean using different imputation methods. Stat. Transit. 7(6) (2006), 1247-1264.

[2] R. Arnab, S. Singh, A new method for estimating variance from data imputed with ratio method of imputation. Stat. Probab. Lett. 76(5) (2006), 513-519.

[3] A. N. Baraldi, C. K. Enders, An introduction to modern missing data analyses. J. School Psychol. 48(1) (2010), 5-37.

[4] G. Diana, P. Francesco Perri, Improved estimators of the population mean for missing data. Commun. Stat. Theory Meth. 39(18) (2010), 3245-3251.

[5] S. Gonzalez, M. Rueda, A. Arcos, An improved estimator to analyse missing data. Stat. Papers, 49(4) (2008), 791-796.

[6] C. Kadilar, H. Cingi, Estimators for the population mean in the case of missing data. Commun. Stat. Theory Meth. 37(14) (2008), 2226-2236.

[7] G. Kalton, D. Kasprzyk, R. Santos, Issues of non-response and imputation of income and participation." Current topics in survey sampling. In D. Krevoski, R. Platek and J N. K. Rao (Eds.), (1981) pp. 455-480. New York: Acad. Press.

[8] M. Khoshnevisan, R. Singh, P. Chauhan, N. Sawan, F. Smarandache, A general family of estimators for estimating population mean using known value of some population parameter(s). Far East J. Theor. Stat. 22(2) (2007), 181-191.

[9] H. Lee, E. Rancourt, C. E. Sarndal, Experiments with variance estimation from survey data with imputed values. J. Official Stat. 10(3) (1994), 231-243.

[10] H. Lee, E. Rancourt, C. E. Sarndal, Variance estimation in the presence of imputed data for the generalized estimation system. Proc. Amer. Statist. Assoc. (Soc. Surv. Res. Meth. Sect.), (1995), p. 384-389.

[11] R. Pandey, N. S. Thakur, K. Yadav, Estimation of population mean using exponential ratio type imputation method under survey non-response. J. Indian Stat. Assoc. 52 (2015), 89-107.

[12] J. N. K. Rao, R. R. Sitter, Variance estimation under two-phase sampling with application to imputation for missing data, Biometrika, 82(2) (1995), 453-460.

[13] M. Rueda, S. González, A new ratio-type imputation with random disturbance. Appl. Math. Lett. 21(9) (2008), 978-982.

[14] I. G. Sande, A personal view of hot deck approach to automatic edit and imputation. J. Imput. Proc. Surv. Methodol. 5 (1979), 238-246.

[15] D. Shukla, N. S. Thakur, Estimation of mean with imputation of missing data using factor-type estimator. Stat. Transit. 9(1) (2008), 33-48.

[16] D. Shukla, N. S. Thakur, S. Pathak, D. S. Rajput, Estimation of mean with imputation of missing data using factor-type estimator in two-phase sampling. Stat. Transit. 10(3) (2009), 397-414.

[17] D. Shukla, N. S. Thakur, D. S. Thakur, Utilization of non-response auxiliary population mean in imputation for missing observations. J. Reliab. Stat. Stud. 2(1) (2009a), 28-40.

[18] D. Shukla, N. S. Thakur, D. S. Thakur, S. Pathak, Linear combination based imputation method for missing data in sample. Int. J. Mod. Eng. Res. 1(2) (2011), 580-596.

[19] D. Shukla, N. Thakur, S. Pathak, Some new aspects on imputation in sampling. Afr. J. Math. Computer Sci. Res. 6(1) (2013), 5-15.

[20] S. Singh, S. Horn, Compromised imputation in survey sampling. Metrika, 51(3) (2000), 267-276.

[21] G. Singh, K. Priyanka, J.-M. Kim, S. Singh, Estimation of population mean using imputation techniques in sample surveys. J. Korean Stat. Soc. 39(1) (2010), 67-74.

[22] A. K. Singh, P. Singh, V. K. Singh, Imputation methods of missing data for estimating the population mean using simple random sampling. Glob. J. Adv. Res. 1(2) (2014), 253-263.

[23] P. Singh, A. K. Singh, V. K. Singh, On the use of compromised imputation for missing data using Factor-Type estimators. J. Stat. Appl. Probab. Lett. 2(2) (2015), 1-9.

[24] A. K. Singh, P. Singh, V. K. Singh, Estimation of Mean under Imputation of Missing data using Exponential-type Estimator in Two-phase sampling. Int. J. Stat. Econ. 17(1) (2016), 73-81.

[25] P. Singh, A. K. Singh, V. K. Singh, On the Use of Chain –type estimator in Compromised Imputation under two phase sampling. Int. J. Stat. Econ. 18(1) (2017), 39-51

[26] N. S. Thakur, K. Yadav, Estimation of mean in presence of missing data under two-phase sampling scheme. J. Reliab. Stat. Stud. 4(2) (2011), 93-104.

[27] N. S. Thakur, K. Yadav, S. Pathak, On mean estimation with imputation in two-phase sampling design. Research J. Math. Stat. Sci. 1(3) (2013), 1-9.

[28] N. S. Thakur, K. Yadav, R. Pandey, A Family of Factor-type Estimators for estimation of population variance. J. Reliab. Stat. Stud. 9(2) (2016), 71-79.