STATISTICS IN TRANSITION new series, December 2022
Vol. 23, No. 4, pp. 91–111, DOI 10.2478/stattrans-2022-0044
Received – 15.01.2021; accepted – 28.02.2022

Missing data estimation based on the chaining technique in survey sampling

Narendra Singh Thakur[1], **Diwakar Shukla**[2]

ABSTRACT

Sample surveys are often affected by missing observations and non-response caused by the respondents' refusal or unwillingness to provide the requested information or due to their memory failure. In order to substitute the missing data, a procedure called imputation is applied, which uses the available data as a tool for the replacement of the missing values. Two auxiliary variables create a chain which is used to substitute the missing part of the sample. The aim of the paper is to present the application of the Chain-type factor estimator as a means of source imputation for the non-response units in an incomplete sample. The proposed strategies were found to be more efficient and bias-controllable than similar estimation procedures described in the relevant literature. These techniques could also be made nearly unbiased in relation to other selected parametric values. The findings are supported by a numerical study involving the use of a dataset, proving that the proposed techniques outperform other similar ones.

Key words: estimation, missing data, chaining, imputation, bias, mean squared error (MSE), factor type (F-T), chain type estimator, double sampling.

Mathematical Subject Code: 62D05

1. Introduction

In sample surveys, the auxiliary information is used to improve efficiency of the estimate [see, Cochran (2005), Sukhatme et al. (1984)]. The use of a ratio estimator is preferred when the population mean of auxiliary variate is known. However, when it is unknown then it is not possible to apply the ratio estimator directly and the concept of two-phase sampling is applied to get a sample-based estimate of population mean. Sometimes information on one more auxiliary variable highly correlated to earlier

[1] Govt. Adarsh Girls College, Sheopur (M.P.), India, Pin – 476337, Affiliation with Jiwaji University, Gwalior (M.P.), India. E-mail: nst_stats@yahoo.co.in. ORCID: https://orcid.org/0000-0001-9731-058X.

[2] Dr. Harisingh Gour Central University, Sagar (M.P.), India, Pin – 470003.
E-mail: diwakarshukla@rediffmail.com. ORCID: https://orcid.org/0000-0002-8694-0655.

auxiliary variate is available and easy to access at a lesser cost. This additional information could be intelligently utilized for obtaining efficient estimates. Chaining is one such technique, used by Chand (1975), Sukhatme and Chand (1977), which has a mechanism of combining wisely two auxiliary variates. Kiregyera (1980, 1984) proposed some chain type ratio and regression estimators whereas Singh et al. (1994) developed a class of chain type estimators under a double sample scheme. Al-Jrarha and Ahmed (2002) discussed the class of chain type estimators for population variance using double a sampling scheme. Some other useful contributions are Kumar and Bahl (2006), Pradhan (2005), Rao and Sitter (1995), Sharma and Tailor (2010), Shukla (2002), Singh and Espejo (2007), Singh et al. (2009), Singh et al. (1993), Srivastava and Jhajj (1980), etc.

The use of auxiliary information in the estimation of population values of the study variate is a common phenomenon in sampling theory of surveys. Auxiliary information is successfully utilized either at the planning stage or at the design stage or at the information stage to arrive at improved estimator compared to those not utilizing auxiliary information. The use of ratio and product strategies in survey sampling solely depends upon the knowledge of population mean $\overline{X}=N^{-1}\sum_{i=1}^{N}X_i$ of the auxiliary character X. In many situations of practical importance, the population mean $\overline{X}$ is unknown before the start of a survey. In such a situation, the usual thing to do is to estimate it by the sample mean $\bar{x}_m = m^{-1}\sum_{i=1}^{m}x_i$ based on a preliminary sample of size m of which n is a sub-sample $(n<m)$. If the population mean $\overline{Z}=N^{-1}\sum_{i=1}^{N}Z_i$ of another auxiliary variate Z, closely related to auxiliary variate X but compared to X remotely related to study variate Y is known, it is advisable to estimate $\overline{X}$ by $\overline{X}=\bar{x}_m\frac{\overline{Z}}{\bar{z}_m}$, where $\bar{z}_m = m^{-1}\sum_{i=1}^{m}z_i$, which would provide better estimate of $\overline{X}$ than $\bar{x}_m$ to the terms of order $o(n^{-1})$ if $\rho_{XZ}\frac{C_X}{C_Z}>0.5$ [see, Choudhury and Singh (2012)]. The symbol ρ_{XZ} is the coefficient of correlation between X and Z and C_X, C_Z are the coefficient of variation of X and Z respectively. Chand (1975) and Sukhatme and Chand (1977) proposed a technique of chaining of the available information on auxiliary characteristics with the main characteristic. Kiregyera (1980, 1984), Singh et al. (2006) also proposed some chain type ratio and regression estimators based on two auxiliary variables. Using prior information on parameters of auxiliary variate some useful contributions are Shukla et al. (1991), Bose (1943), Kadilar and Cingi (2003), Srivastava et al. (1990), Srivenkataramana (1980), etc.

According to Hietjan and Basu (1996), incompleteness in the form of missingness, censoring or grouping, is a troubling feature of several data sets. A key question is what one needs to assume to justify ignoring the incompleteness mechanism. Rubin (1976) addressed this question for Bayes/likelihood and frequentist inferences. Little and Rubin (1987) recognized for some time that failure to account for the stochastic nature of incompleteness can spoil inferences.

In brief, Rubin (1976) defined three key concepts: missing at random (MAR), observed at random (OAR) and Parameter Distinctness (PD). The data are MAR if the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the values of the unobserved data. The data are OAR if, for every possible value of the missing data, the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the values of the observed data. PD holds if there are no a priori ties between the parameters of the missingness model and those of the data model. For Bayesian inference this means that the parameters of the data model and missingness model are a priori independent. For direct likelihood inference it means that knowledge of one parameter's value does not place any constraints on the other parameter's value. Ignoring the missingness mechanism is justified for Bayes/likelihood inference if MAR and PD hold. The combination of MAR and OAR is called missing completely at random (MCAR). In what follows missing completely at random (MCAR) by Heitjan and Basu (1996) is used in this article. Some useful contributions available in the literature are Weeks (1999), Shukla et al. (2009), Seaman et al. (2013), Bhaskaran and Smeeth (2014), Pandey et al. (2015), Pandey et al. (2016), Doretti et al. (2018), etc. This manuscript presents the use of Chain-Type estimator as an imputation source for dealing with missing observations to estimate the population mean.

1.1. Some existing imputation strategies

A simple random sample *S* without replacement (SRSWOR), of size *n* is drawn from population $\Omega = \{1, 2, ..., N\}$ with Y_i as i^{th} unit of variable *Y* under study. Let $\overline{Y} = N^{-1}\sum_{i=1}^{N} Y_i$ be the mean of a finite population under estimation. The sample *S* of *n* units contains *r* responding units ($r < n$) forming a sub-space *R* and $(n - r)$ non-responding with the sub-space $(n - r)$ having symbol R^C in the space *S*. The sub-spaces *R* and R^C are disjoint and $R \cup R^C = S$. The variable *Y* is of main interest and *X* is auxiliary correlated with *Y*. For every unit $i \in R$, the value y_i is observed available. For units $i \in R^C$, the y_i values are missing and imputed values are to be derived. The i^{th} value x_i of *X* could be used as a source of imputation for y_i, $i \in R^C$. This is to assume for sample *S*, the data $x_s = \{x_i : i \in S\}$ is known and available completely. Responding units have missing data only for the study variable *Y*. Under this two variable set-up, some well-known imputation methods available in the literature are:

1.1.1. Ratio method of imputation

For y_i and x_i, define $y_{\bullet i}$ as

$$y_{\bullet i} = \begin{cases} y_i & if \quad i \in R \\ \hat{b}x_i & if \quad i \in R^C \end{cases} \tag{1.1}$$

Where $\hat{b} = \dfrac{\sum_{i\in R} y_i}{\sum_{i\in R} x_i} = \dfrac{\bar{y}_r}{\bar{x}_r}$

Using the above, the imputation-based estimator is:

$$\bar{y}_S = \frac{1}{n}\sum_{i\in S} y_{\bullet i} = \frac{1}{n}\left[\sum_{i\notin R} y_i + \hat{b}\sum_{i\notin R^C} x_i\right] = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) = \bar{y}_{RAT} \tag{1.2}$$

Where $\bar{y}_r = \frac{1}{r}\sum_{i\in R} y_i$, $\bar{x}_r = \frac{1}{r}\sum_{i\in R} x_i$ and $\bar{x}_n = \frac{1}{n}\sum_{i\in S} x_i$

1.1.2. Mean method of imputation

For y_i define $y_{\bullet i}$ as

$$y_{\bullet i} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r & if \quad i \in R^C \end{cases} \tag{1.3}$$

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

$$\bar{y}_m = \frac{1}{r}\sum_{i\in R} y_i = \bar{y}_r \tag{1.4}$$

1.1.3. Compromised method of imputation

Singh and Horn (2000) proposed a compromised imputation procedure:

$$y_{\bullet i} = \begin{cases} (\alpha n/r)y_i + (1-\alpha)\hat{b}x_i & if \quad i \in R \\ (1-\alpha)\hat{b}x_i & if \quad i \in R^C \end{cases} \tag{1.5}$$

Where α is a suitably chosen constant, such that the resultant variance of the estimator is minimum. The imputation-based estimator, for this case, is

$$\bar{y}_{COMP} = \left[\alpha\bar{y}_r + (1-\alpha)\bar{y}_r\frac{\bar{x}_n}{\bar{x}_r}\right] \tag{1.6}$$

1.1.4. Ahmed methods of imputation

For the case where y_{ji} denotes the i^{th} available observation for the j^{th} imputation method Ahmed et al. (2006) suggested:

(A)

$$y_{1i} = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[n\bar{y}_r\left(\dfrac{\bar{X}}{\bar{x}_n}\right)^{\beta_1} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{1.7}$$

Under this, the point estimator is:

$$t_1 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_1} \tag{1.8}$$

(B)

$$y_{2i} = \begin{cases} y_i & if \quad i \in R \\ \frac{1}{(n-r)}\left[n\bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_2} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{1.9}$$

The point estimator is under this set-up:

$$t_2 = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right)^{\beta_2} \tag{1.10}$$

(C)

$$y_{3i} = \begin{cases} y_i & if \quad i \in R \\ \frac{1}{(n-r)}\left[n\bar{y}_r\left(\frac{\bar{X}}{\bar{x}_n}\right)^{\beta_3} - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{1.11}$$

The point estimator is:

$$t_3 = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right)^{\beta_3} \tag{1.12}$$

Terms β_1, β_2 and β_3 are suitably chosen constants, so as to keep the variance of the resultant estimator minimum. As special cases, when

$$\beta_3 = 1,\ t_{Ratio} = \bar{y}_r\left(\frac{\bar{X}}{\bar{x}_r}\right) \tag{1.13}$$

and $$\beta_3 = \text{-}1,\ t_{\Pr oduct} = \bar{y}_r\left(\frac{\bar{x}_r}{\bar{X}}\right) \tag{1.14}$$

The last one (1.14) is natural analogue of the ratio estimator called the product estimator used when an auxiliary variate X has negative correlation with Y.

1.1.5. Factor type methods of imputation

Shukla and Thakur (2008) suggested factor-type imputation procedures as:

(D)

$$(y_{FT1})_i = \begin{cases} y_i & if \quad i \in R \\ \frac{\bar{y}_r}{(n-r)}\left[n\phi_1(k) - r\right] & if \quad i \in R^C \end{cases} \tag{1.15}$$

(E)

$$(y_{FT2})_i = \begin{cases} y_i & if \quad i \in R \\ \frac{\bar{y}_r}{(n-r)}\left[n\phi_2(k) - r\right] & if \quad i \in R^C \end{cases} \tag{1.16}$$

(F)
$$(y_{FT3})_i = \begin{cases} y_i & if \quad i \in R \\ \dfrac{\bar{y}_r}{(n-r)}\left[n\phi_3(k) - r\right] & if \quad i \in R^C \end{cases} \tag{1.17}$$

Where $\phi_1(k) = \dfrac{(A+C)\bar{X} + fB\bar{x}_n}{(A+fB)\bar{X} + C\bar{x}_n}$, $\phi_2(k) = \dfrac{(A+C)\bar{x}_n + fB\bar{x}_r}{(A+fB)\bar{x}_n + C\bar{x}_r}$, $\phi_3(k) = \dfrac{(A+C)\bar{X} + fB\bar{x}_r}{(A+fB)\bar{X} + C\bar{x}_r}$,

$A = (k-1)(k-2)$, $B = (k-1)(k-4)$, $C = (k-2)(k-3)(k-4)$, $f = \dfrac{n}{N}$, $0 < k < \infty$

Under (1.15), (1.16) and (1.17) point estimators are:

$$\left.\begin{aligned} T_{FT1} &= \bar{y}_r \phi_1(k) \\ T_{FT2} &= \bar{y}_r \phi_2(k) \\ T_{FT3} &= \bar{y}_r \phi_3(k) \end{aligned}\right\} \tag{1.18}$$

As special cases, when $k = 1, \beta_l = 1$ then $T_{FTl} = t_l$ when $k = 2, \beta_l = -1$ then $T_{FTl} = t_l$ when $k = 4, \beta_l = 0$ then $T_{FTl} = t_l = \bar{y}_r$; $(l = 1,2,3)$

2. Proposed imputation strategies

Consider a double sampling set-up with three variables Y, X and Z where Y is the main variable and X, Z are auxiliary variates. The correlation between X and Z is higher than other two. A specific way of combining X and Z is "chaining", which generates chain-type estimators in double sampling, and several authors have used this [see Singh and Singh (1991), Singh et al. (1994)] to get a series of alternative estimators for estimating population mean. Singh and Shukla (1987) discussed a family of factor-type ratio estimator for estimating population mean. In one more contribution, Singh and Shukla (1993) derived efficient factor-type estimator for estimating the same population parameter. Using the above contributions Singh et al. (1994) developed a factor-type-chain estimator, whose application as an imputation tool is the main source of motivation in this article.

2.1. Preliminaries

Typically, in double sampling, the population mean $\bar{X}$ of variable X is unknown. Hence, let S' be the preliminary sample drawn from $\Omega = \{1, 2, ..., N\}$ by *SRSWOR* containing m units with mean $\bar{x}_m$, $\bar{z}_m$ of X and Z. This implies $x_{s'} = \{x'_j : j \in S'\}$, $z_{s'} = \{z'_j : j \in S'\}$ are known data and at this stage data linked with variable Y are not known. A sub-sample S of n units $(n < m)$ is drawn from S' by *SRSWOR* having r responding units $(r < n)$ forming subspace R, having $(n-r)$ non-responding units with the sub-space R^C. Also, in S, $y_R = \{y_i, i \in R\}$, $x_s = \{x_i, i \in S\}$, $z_s = \{z_i, i \in S\}$ are available, whereas $y_{R^C} = \{y_i, i \in R^C\}$ is missing and needs to be estimated by an appropriate imputation technique. As discussed in previous section the sub-spaces R and R^C are disjoint and $R \cup R^C = S$.

Let us consider Ahmed et al. (2006) point estimator from equation (1.10) t_2 with $\beta_2 = 1$:

$$t_2^* = \bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \tag{*}$$

The term $\bar{x}_n$ could be improved by Chaining Technique as suggested by Chand (1975), Sukhatme and Chand (1977), Singh and Singh (1991) as:

$$t_2^{**} = \bar{y}_r \frac{\bar{x}_m}{\bar{x}_r} \frac{\bar{Z}}{\bar{z}_m} \qquad \text{(With } \bar{z}_m \text{ and } \bar{Z} \text{ known)} \tag{**}$$

Motivated from the above discussion, some proposed imputation strategies using Singh et al. (1994) are:

(G)
$$(y_{C1})_i = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[n\psi_1(k) - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{2.1}$$

(H)
$$(y_{C2})_i = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[n\psi_2(k) - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{2.2}$$

(I)
$$(y_{C3})_i = \begin{cases} y_i & if \quad i \in R \\ \dfrac{1}{(n-r)}\left[n\psi_3(k) - r\bar{y}_r\right] & if \quad i \in R^C \end{cases} \tag{2.3}$$

Where
$$\psi_1(k) = \bar{y}_r \frac{\bar{x}_m}{\bar{x}_r} \frac{(A+C)\bar{Z} + fB\bar{z}_m}{(A+fB)\bar{Z} + C\bar{z}_m} \tag{2.4}$$

$$\psi_2(k) = \bar{y}_r \frac{\bar{x}_m}{\bar{x}_r} \frac{(A+C)\bar{z}_m + fB\bar{z}_r}{(A+fB)\bar{z}_m + C\bar{z}_r} \tag{2.5}$$

$$\psi_3(k) = \bar{y}_r \frac{\bar{x}_m}{\bar{x}_r} \frac{(A+C)\bar{Z} + fB\bar{z}_r}{(A+fB)\bar{Z} + C\bar{z}_r} \tag{2.6}$$

Where $A = (k-1)(k-2); B = (k-1)(k-4); C = (k-2)(k-3)(k-4)$ and $0 < k < \infty$, is a constant. Also, $\bar{y}_r = \frac{1}{r}\sum_{i\in R} y_i$, $\bar{x}_r = \frac{1}{r}\sum_{i\in R} x_i$, $\bar{z}_r = \frac{1}{r}\sum_{i\in R} z_i$, $\bar{x}_m = \frac{1}{m}\sum_{i\in S'} x_i$, $\bar{z}_m = \frac{1}{m}\sum_{i\in S'} z_i$, $\bar{Z} = \frac{1}{N}\sum_{i\in\Omega} Z_i$.

Under strategies (2.1), (2.2) and (2.3) the point estimators of population mean of study variable $\overline{Y}$ are like (2.4), (2.5) and (2.6) respectively.

2.2. Special Cases:

(i) At $k=1$; $A=0$, $B=0$, $C=-6$

$$\psi_1(1)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\overline{Z}}{\bar{z}_m}\;;\qquad \psi_2(1)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\bar{z}_m}{\bar{z}_r};\qquad \psi_3(1)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\overline{Z}}{\bar{z}_r} \tag{2.7}$$

(ii) At $k=2$; $A=0$, $B=-2$, $C=0$

$$\psi_1(2)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\bar{z}_m}{\overline{Z}};\qquad \psi_2(2)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\bar{z}_r}{\bar{z}_m};\qquad \psi_3(2)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\bar{z}_r}{\overline{Z}} \tag{2.8}$$

(iii) At $k=3$; $A=2$, $B=-2$, $C=0$

$$\psi_1(3)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\overline{Z}-f\bar{z}_m}{(1-f)\overline{Z}};\quad \psi_2(3)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\bar{z}_m-f\bar{z}_r}{(1-f)\bar{z}_m};\quad \psi_3(3)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r}\frac{\overline{Z}-f\bar{z}_r}{(1-f)\overline{Z}} \tag{2.9}$$

(iv) At $k=4$; $A=6$, $B=0$, $C=0$

$$\psi_1(4)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r};\qquad \psi_2(4)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r};\qquad \psi_3(4)=\bar{y}_r\frac{\bar{x}_m}{\bar{x}_r} \tag{2.10}$$

3. Properties of the estimators under proposed strategies

Let $B(.)$ and $M(.)$ be the bias and mean squared error (MSE) of the estimators under a given sampling design respectively. Let the large sample approximations as $n\rightarrow N$ be: $\bar{y}_r=\overline{Y}(1+\delta_1)$; $\bar{x}_r=\overline{X}(1+\delta_2)$; $\bar{x}_m=\overline{X}(1+\delta_3)$; $\bar{z}_r=\overline{Z}(1+\delta_4)$ and $\bar{z}_m=\overline{Z}(1+\delta_5)$

Here, $|\delta_i|<1;\; i=1,2,3,4,5$.

Using the concept of two-phase sampling, following Rao and Sitter (1995) and using the mechanism of MCAR [Heitjan and Basu (1996)], for given r, n and m, we have:

$$E(\delta_i)=0;\; i=1,2,3,4,5;\; E(\delta_1^2)=M_1C_Y^2;\; E(\delta_2^2)=M_1C_X^2;\; E(\delta_3^2)=M_2C_X^2;\; E(\delta_4^2)=M_1C_Z^2;$$
$$E(\delta_5^2)=M_2C_Z^2;\; E(\delta_1\delta_2)=M_1\rho_{YX}C_Y\,C_X;\; E(\delta_1\delta_3)=M_2\rho_{YX}C_YC_X;\; E(\delta_1\delta_4)=M_1\rho_{YZ}C_Y\,C_Z;$$
$$E(\delta_1\delta_5)=M_2\rho_{YZ}C_Y\,C_Z;\; E(\delta_2\delta_3)=M_2\,C_X^2;\; E(\delta_2\delta_4)=M_1\rho_{XZ}C_X\,C_Z;\; E(\delta_2\delta_5)=M_2\rho_{XZ}C_X\,C_Z;$$
$$E(\delta_3\delta_4)=M_2\rho_{XZ}C_X\,C_Z;\; E(\delta_3\delta_5)=M_2\rho_{XZ}C_X\,C_Z;\; E(\delta_4\delta_5)=M_2\,C_Z^2$$

and $M_1=\frac{1}{r}-\frac{1}{N}$; $M_2=\frac{1}{m}-\frac{1}{N}$; $M_3=M_1-M_2=\frac{1}{r}-\frac{1}{m}$.

Remark 3.1: Define the symbols

$$\phi_1=\frac{fB}{A+fB+C};\phi_2=\frac{C}{A+fB+C};\phi_3=\frac{A+C}{A+fB+C};\phi_4=\frac{A+fB}{A+fB+C};\;(\phi_1+\phi_3)=(\phi_2+\phi_4)=1$$

$$\phi=(\phi_1-\phi_2)=-(\phi_3-\phi_4);\,K_{YX}=\rho_{YX}\frac{C_Y}{C_X};K_{YZ}=\rho_{YZ}\frac{C_Y}{C_Z};K_{XZ}=\rho_{XZ}\frac{C_X}{C_Z}$$

Theorem 3.1:

[a₁] The estimator $\psi_1(k)$ in terms of $\delta_i;\ i=1,2,3,4,5$ up to the first order of approximation is:

$$\psi_1(k)=\overline{Y}\left[1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\left(\delta_5+\delta_1\delta_5-\delta_2\delta_5+\delta_3\delta_5-\phi_2\delta_5^2\right)\right] \quad (3.1)$$

[a₂] Bias of $\psi_1(k)$:

$$B[\psi_1(k)]=\overline{Y}\left[M_3C_X^2(1-K_{YX})-\phi M_2C_Z^2(\phi_2-K_{YZ})\right] \quad (3.2)$$

[a₃] Mean squared error of $\psi_1(k)$:

$$M[\psi_1(k)]=\overline{Y}^2\left[M_1C_Y^2+M_3C_X^2(1-2K_{YX})-\phi M_2C_Z^2(\phi+2K_{YZ})\right] \quad (3.3)$$

[a₄] Minimum MSE of the estimator $\psi_1(k)$ is when $\phi=-K_{YZ}$ holds and the expression is:

$$M[\psi_1(k)]_{\min}=\overline{Y}^2\left[M_1C_Y^2+M_3C_X^2(1-2K_{YX})+M_2K_{YZ}^2C_Z^2\right] \quad (3.4)$$

Proof:

[a₁]

$$\psi_1(k)=\overline{y}_r\frac{\overline{x}_m}{\overline{x}_r}\left[\frac{(A+C)\overline{Z}+fB\overline{z}_m}{(A+fB)\overline{Z}+C\overline{z}_m}\right]$$
$$=\overline{Y}(1+\delta_1)(1+\delta_2)^{-1}(1+\delta_3)(1+\phi_1\delta_5)(1+\phi_2\delta_5)^{-1}$$
$$=\overline{Y}\left[1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\left(\delta_5+\delta_1\delta_5-\delta_2\delta_5+\delta_3\delta_5-\phi_2\delta_5^2\right)\right]$$

[a₂] $B[\psi_1(k)]=E\left[\psi_1(k)-\overline{Y}\right]=\left[E[\psi_1(k)]-\overline{Y}\right]$

Using (3.1) and taking expectation both sides

$$E[\psi_1(k)]=\overline{Y}E\left[1-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\left(\delta_1\delta_5-\delta_2\delta_5+\delta_3\delta_5-\phi_2\delta_5^2\right)\right]$$
$$=\overline{Y}\left[1+M_3C_X^2(1-K_{YX})-\phi M_2C_Z^2(\phi_2-K_{YZ})\right]$$
$$B[\psi_1(k)]=\overline{Y}\left[M_3C_X^2(1-K_{YX})-\phi M_2C_Z^2(\phi_2-K_{YZ})\right]$$

[a₃] $M[\psi_1(k)]=E\left[\psi_1(k)-\overline{Y}\right]^2$

$$=E\left[\overline{Y}\left\{1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\left(\delta_5+\delta_1\delta_5-\delta_2\delta_5+\delta_3\delta_5-\phi_2\delta_5^2\right)\right\}-\overline{Y}\right]^2$$

[Using (3.1)]

$$=\overline{Y}^2\left[M_1C_Y^2+M_3C_X^2(1-2K_{YX})-\phi M_2C_Z^2(\phi+2K_{YZ})\right]$$

[a₄] To obtain minimum *MSE*, let

$$\frac{d}{d\phi}M[\psi_1(k)]=0 \quad \Rightarrow\overline{Y}^2\left[M_2C_Z^2(2\phi+2K_{YZ})\right]=0 \ \Rightarrow\phi=-K_{YZ}$$

$$M[\psi_1(k)]_{\min}=\overline{Y}^2\left[M_1C_Y^2+M_3C_X^2(1-2K_{YX})+M_2K_{YZ}^2C_Z^2\right]$$

Theorem 3.2:

[a5] The estimator $\psi_2(k)$ in terms of $\delta_i;\ i=1,2,3,4,5$ up to the first order of approximation is:

$$\psi_2(k)=\overline{Y}\left[1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi(\delta_4-\delta_5+\delta_1\delta_4-\delta_1\delta_5-\delta_2\delta_4+\delta_2\delta_5\right.$$
$$\left.+\delta_3\delta_4-\delta_3\delta_5+(\phi_2-\phi_4)\delta_4\delta_5-\phi_2\delta_4^2+\phi_4\delta_5^2)\right] \quad (3.5)$$

[a6] Bias of the estimator $\psi_2(k)$:

$$B[\psi_2(k)]=\overline{Y}M_3\left[C_X^2(1-K_{YX})-\phi C_Z^2(\phi_2-K_{YZ}+K_{XZ})\right] \quad (3.6)$$

[a7] Mean squared error of $\psi_2(k)$:

$$M[\psi_2(k)]=\overline{Y}^2\left[M_1C_Y^2+M_3\left\{C_X^2(1-2K_{YX})+\phi C_Z^2(\phi+2K_{YZ}-2K_{XZ})\right\}\right] \quad (3.7)$$

[a8] Minimum MSE of $\psi_2(k)$ is at $\phi=(-K_{YZ}+K_{XZ})$:

$$M[\psi_2(k)]_{\min}=\overline{Y}^2\left[M_1C_Y^2+M_3\left\{(1-2K_{YX})C_X^2-(K_{YZ}-K_{XZ})^2C_Z^2\right\}\right] \quad (3.8)$$

Proof:

[a5]

$$\psi_2(k)=\bar{y}_r\left(\frac{\bar{x}_m}{\bar{x}_r}\right)\left[\frac{(A+C)\bar{z}_m+fB\bar{z}_r}{(A+fB)\bar{z}_m+C\bar{z}_r}\right]$$
$$=\overline{Y}(1+\delta_1)(1+\delta_2)^{-1}(1+\delta_3)(1+\phi_1\delta_4+\phi_3\delta_5)(1+\phi_2\delta_4+\phi_4\delta_5)^{-1}$$
$$=\overline{Y}\left[1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\{\delta_4-\delta_5+\delta_1\delta_4-\delta_1\delta_5\right.$$
$$\left.-\delta_2\delta_4+\delta_2\delta_5+\delta_3\delta_4-\delta_3\delta_5+(\phi_2-\phi_4)\delta_4\delta_5-\phi_2\delta_4^2+\phi_4\delta_5^2\}\right]$$

[a6]

$$B[\psi_2(k)]=E\left[\psi_2(k)-\overline{Y}\right]=E[\psi_2(k)]-\overline{Y}$$

Using (3.5) and taking the expectation both sides,

$$E[\psi_2(k)]=\overline{Y}E\left[1-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\{\delta_1\delta_4-\delta_1\delta_5-\delta_2\delta_4+\delta_2\delta_5+\delta_3\delta_4-\delta_3\delta_5\right.$$
$$\left.+(\phi_2-\phi_4)\delta_4\delta_5-\phi_2\delta_4^2+\phi_4\delta_5^2\}\right]$$
$$=\overline{Y}\left[1+M_3\left\{C_X^2(1-K_{YX})-\phi C_Z^2(\phi_2-K_{YZ}+K_{XZ})\right\}\right]$$

$$B[\psi_2(k)]=E[\psi_2(k)]-\overline{Y}$$
$$=M_3\overline{Y}\left[C_X^2(1-K_{YX})-\phi C_Z^2(\phi_2-K_{YZ}+K_{XZ})\right]$$

[a7]

$$M[\psi_2(k)]=E\left[\psi_2(k)-\overline{Y}\right]^2$$
$$=\overline{Y}^2E\left[(\delta_1-\delta_2+\delta_3)+\phi(\delta_4-\delta_5)\right]^2 \quad \text{[Using (3.5)]}$$

$$M[\psi_2(k)]=\overline{Y}^2\left[M_1C_Y^2+M_3\left\{C_X^2(1-2K_{YX})+\phi C_Z^2(\phi+2K_{YZ}-2K_{XZ})\right\}\right]$$

[a8] To obtain minimum MSE, let

$$\frac{d}{d\phi} M\left[\psi_2(k)\right]=0 \quad \Rightarrow \quad \phi = K_{XZ} - K_{YZ}$$

and substitution provides

$$M\left[\psi_2(k)\right]_{\min} = \overline{Y}^2\left[M_1 C_Y^2 + M_3\left\{(1-2K_{YX})C_X^2 - (K_{YZ}-K_{XZ})^2 C_Z^2\right\}\right]$$

Theorem 3.3:

[a9] The estimator $\psi_3(k)$ in terms of $\delta_i;\ i=1,2,3,4,5$ up to the first order of approximation could be expressed as:

$$\psi_3(k) = \overline{Y}\left[1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2+\phi\left(\delta_4+\delta_1\delta_4-\delta_2\delta_4+\delta_3\delta_4-\phi_2\delta_4^2\right)\right] \tag{3.9}$$

[a10] Bias of $\psi_3(k)$:

$$B\left[\psi_3(k)\right] = \overline{Y}\left[M_3 C_X^2(1-K_{YX}) + \phi C_Z^2\left(M_1 K_{YZ} - M_3 K_{XZ} - M_1\phi_2\right)\right] \tag{3.10}$$

[a11] Mean squared error of $\psi_3(k)$:

$$M\left[\psi_3(k)\right] = \overline{Y}^2\left[M_1 C_Y^2 + M_3 C_X^2(1-2K_{YX}) + \phi C_Z^2\left(\phi M_1 + 2M_1K_{YZ} - 2M_3K_{XZ}\right)\right] \tag{3.11}$$

[a12] Minimum MSE of $\psi_3(k)$ is when $\phi = M_1^{-1}(M_3K_{XZ} - M_1K_{YZ})$ and the expression is:

$$M\left[\psi_3(k)\right]_{\min} = \overline{Y}^2\left[M_1C_Y^2 + M_3C_X^2(1-2K_{YX}) - (M_3K_{XZ}-M_1K_{YZ})^2 M_1^{-1} C_Z^2\right] \tag{3.12}$$

Proof:

[a9]

$$\psi_3(k) = \overline{y}_r\left(\frac{\overline{x}_m}{\overline{x}_r}\right)\left[\frac{(A+C)\overline{Z}+fB\overline{z}_r}{(A+fB)\overline{Z}+C\overline{z}_r}\right] = \overline{Y}(1+\delta_1)(1+\delta_2)^{-1}(1+\delta_3)(1+\phi_1\delta_4)(1+\phi_2\delta_4)^{-1}$$

$$= \overline{Y}(1+\delta_1)(1-\delta_2+\delta_2^2-\delta_3^3+\ldots..)(1+\delta_3)(1+\phi_1\delta_4)\left(1-\phi_2\delta_4+\phi_2^2\delta_4^2-\phi_2^2\delta_4^3+\ldots..\right)$$

$$= \overline{Y}\left[1+\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2 \quad +\phi\left(\delta_4+\delta_1\delta_4-\delta_2\delta_4+\delta_3\delta_4-\phi_2\delta_4^2\right)\right]$$

[a10]

$$B\left[\psi_3(k)\right] = E\left[\psi_3(k)-\overline{Y}\right]$$

$$= \overline{Y}E\left[\delta_1-\delta_2+\delta_3-\delta_1\delta_2+\delta_1\delta_3-\delta_2\delta_3+\delta_2^2 \quad +\phi\left(\delta_4+\delta_1\delta_4-\delta_2\delta_4+\delta_3\delta_4-\phi_2\delta_4^2\right)\right]$$

$$= \overline{Y}\left[M_3C_X^2(1-K_{YX}) + \phi C_Z^2\left(M_1K_{YZ} - M_3K_{XZ} - M_1\phi_2\right)\right]$$

[a11]

$$M\left[\psi_1(k)\right] = E\left[\psi_1(k)-\overline{Y}\right]^2 = \overline{Y}^2 E\left[\delta_1 - \delta_2 + \delta_3 + \phi\delta_4\right]^2$$

$$= \overline{Y}^2\left[M_1C_Y^2 + M_3C_X^2 - 2M_3\rho_{YX}C_YC_X + \phi^2 M_1 C_Z^2\right.$$
$$\left. + 2\phi\left(M_1\rho_{YZ}C_YC_Z - M_3\rho_{XZ}C_XC_Z\right)\right]$$

$$= \overline{Y}^2\left[M_1C_Y^2 + M_3C_X^2(1-2K_{YX}) \quad + \phi C_Z^2\left(\phi M_1 + 2M_1K_{YZ} - 2M_3K_{XZ}\right)\right]$$

[a12] To obtain minimum MSE, let

$$\frac{d}{d\phi} M\left[\psi_3(k)\right] = 0 \quad \Rightarrow \quad \phi = M_1^{-1}\left(M_3K_{XZ} - M_1K_{YZ}\right)$$

and substitution provides

$$M\left[\psi_3(k)\right]_{\min} = \overline{Y}^2\left[M_1C_Y^2 + M_3C_X^2(1-2K_{YX}) - (M_3K_{XZ}-M_1K_{YZ})^2 M_1^{-1}C_Z^2\right]$$

4. Comparison of the estimators under proposed imputation strategies

[b₁] $D_1 = M[\psi_1(k)]_{\min} - M[\psi_2(k)]_{\min}$

$$= \overline{Y}^2 C_Z^2 \left[M_3 (K_{YZ} - K_{XZ})^2 - M_2 K_{YZ}^2 \right] \quad (4.1)$$

$\psi_2(k)$ is better over $\psi_1(k)$ if $D_1 > 0$

$$\Rightarrow \quad \frac{K_{YZ} - K_{XZ}}{K_{YZ}} > \sqrt{\frac{M_2}{M_3}} \quad \Rightarrow \quad F_1 > F_2 \quad \text{(let)}$$

[b₂] $D_2 = M[\psi_1(k)]_{\min} - M[\psi_3(k)]_{\min}$

$$= \overline{Y}^2 C_Z^2 \left[(M_3 K_{XZ} - M_1 K_{YZ})^2 M_1^{-1} - M_2 K_{YZ}^2 \right] \quad (4.2)$$

$\psi_3(k)$ is better over $\psi_1(k)$ if $D_2 > 0$

$$\Rightarrow \quad \frac{K_{XZ}}{K_{YZ}} > \frac{M_1 + \sqrt{M_1 M_2}}{M_3} \quad \Rightarrow \quad F_3 > F_4 \quad \text{(let)}$$

[b₃] $D_3 = M[\psi_2(k)]_{\min} - M[\psi_3(k)]_{\min}$

$$= \overline{Y}^2 C_Z^2 \left[(M_3 K_{XZ} - M_1 K_{YZ})^2 M_1^{-1} - M_3 (K_{YZ} - K_{XZ})^2 \right] \quad (4.3)$$

$\psi_3(k)$ is better than $\psi_2(k)$ if $D_3 > 0$

$$\Rightarrow \quad \frac{K_{XZ}}{K_{YZ}} > \frac{M_1 + \sqrt{M_1 M_3}}{M_3 + \sqrt{M_1 M_3}} \quad \Rightarrow \quad F_3 > F_5 \quad \text{(let)}$$

5. Empirical study

For numerical study consider the data as attached in Appendix A, which is a generated artificial population of size $N = 200$ containing values of main variable Y and auxiliary variables X, Z. Parameters of this population are given below:

$\overline{Y} = 42.485$; $\overline{X} = 18.515$; $\overline{Z} = 20.52$; $S_Y^2 = 199.0598$; $S_X^2 = 48.5375$; $S_Z^2 = 45.7684$;

$\rho_{YX} = 0.8734$; $\rho_{YZ} = 0.8667$; $\rho_{XZ} = 0.9943$; $C_Y = 0.3287$; $C_X = 0.3755$; $C_Z = 0.3296$;

$K_{YZ} = 0.8643$; $K_{XZ} = 1.1326$; $K_{YX} = 0.7645$

Reddy (1978) proved that K_{YX}, K_{YZ}, K_{XZ} are ratio values and bear very small change over a span of time. It could be easily guessed or assumed to be known a priori. Using preliminary large sample of size $m = 80$ and sub-random sample of size $n = 30$ with the number of responding units $r = 22$ and $f = 0.15$ by *SRSWOR*. The optimum values of constants of different estimators at their optimal condition are $\alpha = 0.2354$, $\beta_1 = \beta_2 = \beta_3 = 0.7646$, $k_1' = 1.5206$, $k_2' = 2.4505$, $k_3' = 8.9456$ for compromised, Ahmed's methods and Factor Type F-T Estimators of imputation respectively. By simplifying optimum conditions of proposed estimators for minimum MSE, the cubic equations provide the values of constants k as shown in Table 5.1.

Table 5.1. Optimum *k*-values for minimum *MSE* of proposed estimators

Estimators	Condition for Optimum *MSE*	Three optimum Values of *k* on one condition		
$\psi_1(k)$	$\phi = -K_{YZ}$	k_1 = 1.3137	k_2 = 2.5180	k_3 = 13.5979
$\psi_2(k)$	$\phi = K_{XZ} - K_{YZ}$	k_4 = 1.9321	k_5 = -----	k_6 = -----
$\psi_3(k)$	$\phi = M_1^{-1}(M_3 K_{XZ} - M_1 K_{YZ})$	k_7 = 1.8759	k_8 = 3.2154	k_9 = 4.0919

Note: k_5, k_6 do not exist because the solution of cubic equations provided no real roots.

The formula for efficiency measurement is $e(\hat{T}) = \frac{MSE(\bar{y}_r)}{MSE(\hat{T})}$, where $\hat{T}$ is any estimator under consideration. The steps followed for the simulation procedure are:

Step 1: Draw a preliminary random sample *S'* of size *m* = 80 from the population of size 200.

Step 2: Again draw a random sub-sample of size *n* = 30 from *S'* drawn in step 1.

Step 3: Drop away 8 units randomly from each sample corresponding to variable *Y*.

Step 4: Compute and impute the dropped units of *Y* with the help of existing and proposed imputation methods.

Step 5: Obtain the estimates of the population mean for existing and proposed imputation methods.

Step 6: Repeat the above steps (1 to 5) 50,000 times, which provides multiple sample based estimates $\hat{T}_1, \hat{T}_2, \hat{T}_3, \ldots\ldots, \hat{T}_{50,000}$.

Step 7: The bias of $\hat{T}$ is obtained by $B(\hat{T}) = \frac{1}{50000}\sum_{i=1}^{50000}(\hat{T}_i - \bar{Y})$.

Step 8: The *MSE* of $\hat{T}$ is obtained by $MSE(\hat{T}) = \frac{1}{50000}\sum_{i=1}^{50000}(\hat{T}_i - \bar{Y})^2$.

Following the above procedure bias and *MSE* of the existing and proposed estimators are computed based on 50,000 repeated samples drawn by SRSWOR from population of *N* = 200. These computations and efficiencies with respect to $\bar{y}_r$ are given in Tables 5.2 and 5.3 respectively.

Table 5.2. Bias and *MSE* of existing estimators

Estimators	Optimum Value	Bias	*MSE*	Efficiency
$\bar{y}_r$	-----	-0.3123	9.7252	1
$\bar{y}_{RAT}$	-----	-0.0996	7.8457	1.2395
$\bar{y}_{COMP}$	$\alpha = 0.2354$	-0.0809	6.9649	1.3963
t_1	$\beta_1 = 0.7646$	-0.3983	5.8967	1.6492

Table 5.2. Bias and MSE of existing estimators (cont.)

Estimators	Optimum Value	Bias	*MSE*	Efficiency
t_2	$\beta_2 = 0.7646$	-0.1871	7.6655	1.2686
t_3	$\beta_3 = 0.7646$	-0.2151	3.2967	2.9499
T_{FT1}	$k_1' = 1.5206$	-0.3878	4.8327	2.0123
	$k_2' = 2.4505$	-0.3736	5.1655	1.8827
	$k_3' = 8.9456$	-0.3961	4.9454	1.9665
T_{FT2}	$k_1' = 1.5206$	-0.1071	6.3071	1.5419
	$k_2' = 2.4505$	-0.0329	6.1072	1.5924
	$k_3' = 8.9456$	-0.0980	6.0561	1.6058
T_{FT3}	$k_1' = 1.5206$	-0.1826	1.8399	5.2857
	$k_2' = 2.4505$	-0.1944	2.2685	4.2870
	$k_3' = 8.9456$	-0.1818	1.9894	4.8885

5.1. Numerical computation of proposed estimators

From Section 4.0 we get computational values of conditions on the population given in Appendix A. $F_1 = \frac{K_{YZ} - K_{XZ}}{K_{YZ}}$ = - 0.3104; $F_2 = \sqrt{\frac{M_2}{M_3}}$ = 0.4774; $F_3 = \frac{K_{XZ}}{K_{YZ}}$ = 1.3104; $F_4 = \frac{M_1 + \sqrt{M_1 M_2}}{M_3}$ = 1.7570 and $F_5 = \frac{M_1 + \sqrt{M_1 M_3}}{M_3 + \sqrt{M_1 M_3}}$ = 1.1082

Since $F_1 < F_2$ holds, $\psi_1(k)$ is better than $\psi_2(k)$ for this data set.
Again, $F_3 < F_4$, which implies $\psi_1(k)$ is better than $\psi_3(k)$ for the data set, and $F_3 > F_5$, which implies $\psi_3(k)$ is better than $\psi_2(k)$ for this data set. Overall $\psi_1(k)$ is the best estimator.

Table 5.3. Bias and *MSE* of proposed chain type estimators

Estimator	*k-optimum*	Bias	*MSE*	Efficiency
$\psi_1(k)$	k_1 =1.3137	-0.0030	1.9169	5.0734
	k_2 =2.5180	0.0215	1.9328	5.0317
	k_3 =13.5979	-0.0038	1.9409	5.0106
$\psi_2(k)$	k_4 =1.9321	0.3534	9.0303	1.0769
	k_5 = -----	——	——	-----
	k_6 = -----	——	——	-----
$\psi_3(k)$	k_7 =1.8759	0.6036	8.6779	1.1206
	k_8 =3.2154	0.6215	8.6360	1.1261
	k_9 = 4.0919	0.5992	8.6621	1.1227

6. Almost unbiased imputation based chain type estimator

By expression (3.2), (3.6) and (3.10), bias of $\psi_i(k)$; i = 1, 2, 3 could be made zero up to the first order of approximation. This provides three equations:

$$M_3 C_X^2 (1 - K_{YX}) - \phi M_2 C_Z^2 (\phi_2 - K_{YZ}) = 0 \quad (6.1)$$

$$C_X^2 (1 - K_{YX}) - \phi C_Z^2 (\phi_2 - K_{YZ} + K_{XZ}) = 0 \quad (6.2)$$

and
$$M_3 C_X^2 (1 - K_{YX}) + \phi C_Z^2 (M_1 K_{YZ} - M_3 K_{XZ} - M_1 \phi_2) = 0 \quad (6.3)$$

These equations are cubic or more function of *k*-values to provide multiple values of *k* on which bias is zero. The best choice is to have lowest mean squared error. So, the proposed estimators bear property of reducing *MSE* along with being almost unbiased also. Many similar estimators existing in the literature do not control both bias and *MSE* at their optimal level but the proposed estimators have this property. For equation (6.1), we get two real values k_1'' = 0.3829 and k_2'' =6.5038 and from (6.2) and (6.3) all values are imaginary, viz. there are no real roots. These results are obtained using the data set on which the empirical study was performed. The term almost unbiased is used because biases of proposed estimates $\psi_i(k)$ are obtained only up to the first order of approximation. The bias $B[\psi_2(k)] = 0$ holds approximately not completely, therefore mentioned almost unbiased.

Table 6.1. Almost unbiased comparison of chain type estimators

***k*-values**	$\psi_1(k)$		$\psi_2(k)$		$\psi_3(k)$	
	Bias	***MSE***	**Bias**	***MSE***	**Bias**	***MSE***
k_1'' = 0.3829	0.0005	4.4522	0.0002	15.4062	0.0002	14.4033
k_2'' =6.5038	0.0004	2.4831	0.0001	7.4559	0.0011	6.4898

7. Discussion and conclusions

In the present article some imputation procedures and their estimators of population mean are suggested and the expression of their bias, mean squared error and minimum mean squared error have been derived under large sample approximations up to the first order. An empirical study has been done over a data set and the bias and mean squared error have been calculated. Among the existing and proposed estimators, under Chain-based imputation strategies, i.e. $\psi_i(k)$; (i = 1, 2, 3), the estimator $\psi_1(k)$ is found best. The general perception regarding imputation of missing data is that it increases the bias of the estimate when *MSE* is optimized. In contrary, a key feature of $\psi_i(k)$; (i = 1, 2, 3) is that there are many values of the parameter *k* on which *MSE* is optimal. One can choose the value with the lowest bias. Therefore, suggested strategies are bias-controlled at the optimum level of *MSE.* Apart from this, estimators are almost unbiased also over multiple choices of *k*-values. The best selection is to have the lowest *MSE* by proposed strategies one can get almost unbiased estimator with lowest possible *MSE*. Thus, the suggested Chain-based imputation strategies $\psi_i(k)$; (i = 1, 2, 3) are useful and have advantage over other similar procedures.

Acknowledgement

Authors are thankful to the reviewers of this journal for their critical comments and useful suggestions, which has improved the quality of the manuscript.

References

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

Al-Jararha, J., Ahmed, M. S., (2002). The class of chain estimators for a finite population variance using double sampling. *Information and Management Sciences*, 13(2), pp. 13–18.

Bhaskaran, K., Smeeth, L., (2014). What is the difference between missing completely at random and missing at random? I*nternational Journal of Epidemiology*, 43(4), pp. 1336–1339.

Bose, C., (1943). Note on the sampling error in the method of double sampling. *Sankhya*, 6, 330.

Chand, L., (1975). Some ratio-type estimators based on two or more auxiliary variables unpublished Ph.D. Thesis, *IOWA State University*, Ames, Iowa, U.S.A.

Choudhury, S., Singh, B. K., (2012). A class of chain ratio-cum-dual to ratio type estimator with two auxiliary characters under double sampling in sample surveys. *Statistics in Transition-new series*, 13(3), pp. 519–536.

Cochran, W. G., (2005). Sampling Techniques. *John Wiley and Sons*, New York.

Doretti, M., Geneletti, S. and Stanghellini, E., (2018). Missing data: A unified taxonomy guided by conditional independence. *International Statistical Review*, 86(2), pp. 189–204.

Heitjan, D. F., Basu, S., (1996). Distinguishing 'missing at random' and 'missing completely at random'. *The American Statistician*, 50, pp. 207–213.

Kadilar, C., Cingi, H., (2003). A study on the chain ratio-type estimator. *Hacettepe Journal of Mathematics and Statistics*, 32, pp. 105–108.

Kiregyera, B., (1980). A chain ratio-type estimator in finite population double sampling using two auxiliary variables. *Metrika*, 27 (1), pp. 217–223.

Kiregyera, B., (1984). Regression type estimators using two auxiliary variables and the model of double sampling from finite population. *Metrika*, 31, pp. 215–226.

Kumar, M., Bahl, S., (2006). Class of dual to ratio estimators for double sampling. *Statistical Papers*, 47, pp. 319–326.

Little, R. J. A., Rubin, D. B., (1987). Statistical analysis with missing data, New York: *John Wiley* & *Sons, Inc.*

Pandey, R., Thakur, N. S. and Yadav, K., (2016). Adapted factor-type imputation strategies. *Journal of Scientific Research*, J. Sci. Res., 8(3), pp. 321–339.

Pandey, R., Thakur, N. S. and Yadav, K., (2015). Estimation of population mean using exponential ratio type imputation method under survey non-response. *Journal of the Indian Statistical Association*, Vol.53 No. 1 & 2, pp. 89–107.

Pradhan, B. K., (2005). A chain regression estimator in two phase sampling using multi-auxiliary information. *Bulletin of the Malaysian Mathematical Sciences Society (2),* 28(1), pp. 81–86.

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

Reddy, V. N., (1978). A study on the use of prior knowledge on certain population parameters in estimation. *Sankhya*, C, 40, pp. 29–37.

Rubin, D. B., (1976). Inference and missing data. *Biometrika*, 63, pp. 581–593.

Seaman, S., Galati, J., Jackson, D. and Carlin, J., (2013). *What is meant by "Missing at Random"? Statistical Science*, 28(2), pp. 257–268.

Sharma, B., Tailor, R., (2010). A new ratio-cum-dual to ratio estimator of finite population mean in simple random sampling. *Global Journal of Science Frontier Research*, 10(1), pp. 27–31.

Shukla, D., (2002). F-T estimator under two-phase sampling. *Metron*, 59, 1–2, pp. 253–263.

Shukla, D., Thakur, N. S., Pathak, S. and Rajput, D. S., (2009). Estimation of mean under imputation of missing data using factor type estimator in two-phase sampling. *Statistics in Transition*, Vol. 10, No. 3, pp. 397–414.

Shukla, D., Thakur, N. S., (2008). Estimation of mean with imputation of missing data Using Factor Type Estimator. *Statistics in Transition*, 9, 1, pp. 33–48

Shukla, D., Singh, V. K. and Singh, G. N., (1991). On the use of transformation in factor type estimator. *Metron*, 49(1-4), pp. 359–361.

Singh, H. P., Espejo, M. R., (2007). Double sampling ratio-product estimator of a finite population mean in sampling surveys. *Journal of Applied Statistics*, 34(1), pp. 71–85.

Singh, H. P., Mathur, N. and Chandra, P., (2009). A chain-type estimator for population variance using auxiliary variables in two-phase sampling. *Statistics in Transition-new series*, 10(1), pp. 75–84.

Singh, S., Horn, S., (2000). Compromised imputation in survey sampling. *Metrika*, 51, pp. 266–276.

Singh, S., Singh, H. P. and Upadhyaya, L. N., (2006). Chain ratio and regression type estimators for median estimation in survey sampling. *Statistical Papers*, 48, pp. 23–46.

Singh, V. K., Shukla, D., (1987). One parameter family of factor-type ratio estimator. *Metron*, 45, 1-2, pp. 273–283.

Singh, V. K., Shukla, D., (1993). An efficient one parameter family of factor – type estimator in sample survey. *Metron*, 51, 1–2, pp. 139–159.

Singh, V. K., Singh, G. N., (1991). Chain type estimator with two auxiliary variables under double sampling scheme. *Metron*, 49, pp. 279–289.

Singh, V. K., Singh, B. K. and Singh, G. N., (1993). An efficient class of dual to ratio estimators using two auxiliary characteristics. *Journal of Scientific Research*, 43, pp. 219–228.

Singh, V. K., Singh, G. N. and Shukla, D., (1994). A class of chain ratio estimator with two auxiliary variables under double sampling scheme. *Sankhya*, Ser. B., 46, 2, pp. 209–221.

Srivastava, S. K., Jhajj, H. S., (1980). A class of estimators using auxiliary information for estimating finite population variance. *Sankhya*, 42, pp. 87–96.

Srivastava, S. R., Khare, B. B. and Srivastava, S. R., (1990). A generalized chain ratio estimator for mean of finite population. *Journal of Indian Society of Agricultural Statistics,* 42(I), pp. 108–117.

Srivenkataramana, T., (1980). A dual to ratio estimator in sample surveys. *Biometrika*, 67(1), pp. 199–204.

Sukhatme, B. V., Chand, L., (1977). Multivariate ratio-type estimators, Proceeding of American Statistical Association. *Social Statistics Section,* pp. 927–931.

Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S. and Ashok, C., (1984). Sampling Theory of Surveys with Applications. *Iowa State University Press*, I.S.A.S. Publication, New Delhi.

Weeks, M., (1999). Methods of imputation for missing data (fifth draft), Faculty of Economics and Politics and Department of Applied Econometrics. *University of Cambridge*, Cambridge, UK.

Appendix

A. Population (N = 200)

Y_i	45	50	39	60	42	38	28	42	38	35
X_i	15	20	23	35	18	12	8	15	17	13
Z_i	16	22	26	37	19	14	11	17	18	15
Y_i	40	55	45	36	40	58	56	62	58	46
X_i	29	35	20	14	18	25	28	21	19	18
Z_i	30	37	23	15	19	27	30	22	21	21
Y_i	36	43	68	70	50	56	45	32	30	38
X_i	15	20	38	42	23	25	18	11	09	17
Z_i	18	22	39	44	25	26	19	13	12	20
Y_i	35	41	45	65	30	28	32	38	61	58
X_i	13	15	18	25	09	08	11	13	23	21
Z_i	16	17	19	27	12	10	13	14	24	23
Y_i	65	62	68	85	40	32	60	57	47	55
X_i	27	25	30	45	15	12	22	19	17	21
Z_i	28	26	33	46	17	15	23	20	19	23
Y_i	67	70	60	40	35	30	25	38	23	55
X_i	25	30	27	21	15	17	09	15	11	21
Z_i	26	32	30	23	17	18	12	18	14	24
Y_i	50	69	53	55	71	74	55	39	43	45
X_i	15	23	29	30	33	31	17	14	17	19
Z_i	17	24	30	33	35	32	19	16	19	21
Y_i	61	72	65	39	43	57	37	71	71	70
X_i	25	31	30	19	21	23	15	30	32	29
Z_i	27	33	32	21	23	25	17	32	33	32
Y_i	73	63	67	47	53	51	54	57	59	39
X_i	28	23	23	17	19	17	18	21	23	20
Z_i	30	25	24	20	22	20	21	23	26	22
Y_i	23	25	35	30	38	60	60	40	47	30
X_i	07	09	15	11	13	25	27	15	17	11
Z_i	10	11	18	14	14	26	29	18	20	14
Y_i	57	54	60	51	26	32	30	45	55	54
X_i	31	23	25	17	09	11	13	19	25	27
Z_i	32	25	27	19	12	13	14	20	27	28
Y_i	33	33	20	25	28	40	33	38	41	33
X_i	13	11	07	09	13	15	13	17	15	13
Z_i	16	14	9	10	14	17	14	20	17	15
Y_i	30	35	20	18	20	27	23	42	37	45
X_i	11	15	08	07	09	13	12	25	21	22
Z_i	13	18	11	8	12	16	14	26	24	23

Y_i	37	37	37	34	41	35	39	45	24	27
X_i	15	16	17	13	20	15	21	25	11	13
Z_i	16	18	19	16	22	18	23	26	14	14
Y_i	23	20	26	26	40	56	41	47	43	33
X_i	09	08	11	12	15	25	15	25	21	15
Z_i	11	10	14	15	17	26	17	27	22	17
Y_i	37	27	21	23	24	21	39	33	25	35
X_i	17	13	11	11	09	08	15	17	11	19
Z_i	19	16	13	12	12	11	17	20	13	20
Y_i	45	40	31	20	40	50	45	35	30	35
X_i	21	23	15	11	20	25	23	17	16	18
Z_i	22	25	18	13	21	27	26	19	17	19
Y_i	32	27	30	33	31	47	43	35	30	40
X_i	15	13	14	17	15	25	23	17	16	19
Z_i	17	16	16	14	17	28	25	18	18	22
Y_i	35	35	46	39	35	30	31	53	63	41
X_i	19	19	23	15	17	13	19	25	35	21
Z_i	22	21	24	17	20	15	22	26	36	23
Y_i	52	43	39	37	20	23	35	39	45	37
X_i	25	19	18	17	11	09	15	17	19	19
Z_i	26	20	20	19	13	12	17	18	21	22