JSR Publications

Available Online

J. Sci. Res. **8** (3), 321-339 (2016)

JOURNAL OF SCIENTIFIC RESEARCH

www.banglajol.info/index.php/JSR

Adapted Factor-Type Imputation Strategies

R. Pandey[1*], K. Yadav[1], N. S. Thakur[2]

[1]Department of Statistics, University of Delhi, New Delhi-110007, India

[2]Department of Mathematics and Statistics, Banasthali University, Rajasthan-304022, India

Received 19 May 2016, accepted in final revised form 18 June 2016

Abstract

The present paper provides alternative improved Factor-Type (F-T) estimators of population mean in presence of item non-response for the practitioners. The proposed estimators have been shown to be more efficient than the four existing estimators which are more efficient than the usual ratio and the mean estimators. Optimum conditions for minimum mean squared error are obtained for the new estimators. Empirical comparisons based on three different data sets establish that the proposed estimators record least mean squared error and hence a substantial gain in Percentage Relative Efficiency (P.R.E.), over these five contemporary estimators.

Keywords: Auxiliary variable; Imputation; Bias; Mean Squared Error (M.S.E.); Factor-Type (F-T) Estimator.

doi: http://dx.doi.org/10.3329/jsr.v8i3.27804 J. Sci. Res. **8** (3), 321-339 (2016)

1. Introduction

Often Surveys are accompanied by incomplete response or unavailable items. Analysis restricted to complete records when faced with an incomplete dataset may lead to biased inference [1]. The other alternative is to construct the estimates for the missing observations. Estimation of individual missing items based on survey response is called imputation. Missing patterns are classified as missing at random (MAR) and observed at random (OAR) [2]. 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. In other words, the data are missing only due to chance factors. The data are OAR, if for every possible value of the missing data, the probability of the observed missingness pattern, given the observed and the unobserved data, does not depend on the values of the observed data. The combination of MAR and OAR is called missing completely at random (MCAR) which means that the response propensity to provide information is constant for all the subjects. In this paper, we implicitly assume that the missing values are MCAR.

* *Corresponding author:* ranjitapandey111@gmail.com

Various techniques for mean estimation under non-response have been considered by several researchers [3-5], among others. Some well-known imputation methods in literature are deductive imputation, mean imputation overall (MO), random imputation overall (RO), mean imputation within classes (MC), hot deck, cold deck and so on [6]. Compromised imputation [7], imputation based on power transformation [8], modifications of ratio and regression methods of imputation [9], exponential ratio type imputation based on an auxiliary variable [10], Factor -Type estimator adapted as a tool of imputation [11], imputation based on modified Walsh estimator [12] have been considered in literature so far. Present work adapts strategy of [12] based on Factor-Type technique to provide better and more efficient means of imputing missing survey data. In the present paper, three *separate* strategies are proposed to impute missing data under (i) *low* non-response: a few missing data under item non-response-the most common missing data situation encountered in practice and under (ii) *high* non-response: large chunks of missing data under a variable- when a region is inaccessible for data collection or the information is invasive for a particular section of the target group.The present theoretical contribution proposes estimators with higher precision and substantially reduced M.S.E. as compared to the currently used five different imputation strategies. Numerical results based on three different types of populations A (moderate sized), B (small sized) and C (large sized) illustrate superiority of the estimatorsproposed where the variable to be imputed is highly positively correlated with the auxiliary variable or the covariate.

A finite population $\Omega = (1, 2, 3\ldots, N)$ is considered. A SRSWOR (Simple Random Sampling without Replacement), of size n is drawn from Ω to estimate $\overline{Y}$. The responding units out of sampled n units are denoted by r forming a set R and the set of non-responding units are denoted by R^c. The variable Y is of main interest and X is an auxiliary variable highly correlated with Y. For every unit $i \in R$, the value y_i is observed. For $i \in R^c$, y_i values are missing and imputed values are derived. The i^{th} value x_i of auxiliary variate X is used as a source of imputation for missing data when $i \in R^c$. For sample S, the data $x_S = \{x_i : i \in S\}$ are assumed to be known and $S = R \cup R^c$.

Define $e_1 = \left(\frac{\overline{y}_r}{\overline{\overline{Y}}} - 1\right)$, $e_2 = \left(\frac{\overline{x}_r}{\overline{\overline{X}}} - 1\right)$ and $e_3 = \left(\frac{\overline{x}_n}{\overline{\overline{X}}} - 1\right)$. Assuming MCAR for non-response with r and n known, we obtain: $E(e_1) = E(e_2) = E(e_3) = 0$, $E(e_1^2) = M_1 C_y^2$, $E(e_2^2) = M_1 C_x^2$, $E(e_3^2) = M_2 C_x^2$, $E(e_1 e_2) = M_1 \rho C_y C_x$

$E(e_2 e_3) = M_2 C_x^2$, $E(e_1 e_3) = M_2 \rho C_y C_x$ where $M_1 = \left(\frac{1}{r} - \frac{1}{N}\right)$, $M_2 = \left(\frac{1}{n} - \frac{1}{N}\right)$, $M_3 = M_1 - M_2$, and S_y^2, S_x^2 and S_{xy} have their usual meaning. $(y_i)_l$ denotes l^{th} imputation strategy for i^{th} observation.

Table 1. Some well-known imputation methods.

Imputation strategies	Point estimator of $\bar{Y}$
Mean $(y_i)_1 = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r & if \quad i \in R^c \end{cases}$	$t_1 = \bar{y}_r$ Its variance is $V(t_1) = M_1 S_y^2$
Ratio $(y_i)_2 = \begin{cases} y_i & if \quad i \in R \\ \hat{b}x_i & if \quad i \in R^c \end{cases}$	$t_2 = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right)$. The M.S.E. of t_2 are: $M(t_2) = M_2 S_y^2 + M_3 \left[S_y^2 + \Re^2 S_x^2 - 2\Re S_{xy} \right]$
Compromised $(y_i)_3 = \begin{cases} (\alpha' n/r) y_i + (1-\alpha')\hat{b}x_i & if \quad i \in R \\ (1-\alpha')\hat{b}x_i & if \quad i \in R^c \end{cases}$	$t_3 = \left[\alpha' \bar{y}_r + (1-\alpha')\bar{y}_r \frac{\bar{x}_n}{\bar{x}_r} \right]$ The minimum M.S.E. at $\alpha' = 1 - \rho \frac{C_Y}{C_X}$ of t_3 are given by $M(t_3)_{\min} = M(t_2) - M_1 \left(1 - \rho \frac{C_Y}{C_X} \right)^2 \bar{Y}^2 C_x^2$
[4] (a) $(y_i)_4 = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r \left[\frac{(n\text{-}r)\bar{x}_n + r\bar{x}_r}{n\bar{x}_n} \right] & if \quad i \in R^c \end{cases}$ (b) $(y_i)_5 = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r \left[\frac{(n-r)^2 \bar{x}_n + (2n-r) r \bar{x}_r}{(n-r) n\bar{x}_n + nr\bar{x}_r} \right] & if \quad i \in R^c \end{cases}$	(a) $t_4 = \bar{y}_r \left(\frac{\bar{x}_n}{\bar{x}_r} \right) = t_2$ (b) $t_5 = \bar{y}_r \frac{(n+r)\bar{x}_n - r\bar{x}_r}{n\bar{x}_n}$ The M.S.E of the estimator t_5 are given by $M(t_5) = M_2 S_y^2 + M_3 \left[S_y^2 + \left(\frac{r}{n} \right)^2 \Re^2 S_x^2 - 2\left(\frac{r}{n} \right) \Re S_{xy} \right]$
[12] $(y_i)_6 = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r \left[\frac{(n-r)\bar{x}_n + \beta\, r(\bar{x}_n - \bar{x}_r)}{\beta\, \bar{x}_r + (1-\beta)\bar{x}_n} \right] \frac{x_i}{\sum_{i \in R'} x_i} & if \quad i \in R^C \end{cases}$	$t_6 = \bar{y}_r \frac{\bar{x}_n}{\beta \bar{x}_r + (1-\beta)\bar{x}_n}$. The minimum M.S.E. of t_6 for the optimum value of $\beta = \rho \frac{C_y}{C_x}$ are $[M(t_6)]_{\min} = (M_1 - M_3 \rho^2) S_y^2$

*The estimator t_4 is similar to the usual ratio type estimator t_2 and therefore the expression of bias and M.S.E. are the same and $\Re = \bar{Y} / \bar{X}$.

2. Proposed Methods of Imputation

Under the proposed strategies for imputing missing observations, the data after imputation takes the following form:

$$(y_i)_{7j} = \begin{cases} y_i & if \quad i \in R \\ \bar{y}_r \left[\dfrac{(n-r)+\alpha r\left(1-\phi_j(\alpha)\right)}{1-\alpha\left(1-\phi_j(\alpha)\right)} \right] \dfrac{x_i}{\sum_{i\in R^C} x_i} & if \quad i \in R^c \end{cases} \quad \text{for } j = 1,2,3 \tag{1}$$

where $\phi_1(\alpha) = \left[\dfrac{(A+C)\bar{X} + fB\bar{x}_n}{(A+fB)\bar{X} + C\bar{x}_n} \right]$; $\phi_2(\alpha) = \left[\dfrac{(A+C)\bar{x}_n + fB\bar{x}_r}{(A+fB)\bar{x}_n + C\bar{x}_r} \right]$;

$\phi_3(\alpha) = \left[\dfrac{(A+C)\bar{X} + fB\bar{x}_r}{(A+fB)\bar{X} + C\bar{x}_r} \right]$; where, $A = (\alpha-1)(\alpha-2)$; $B = (\alpha-1)(\alpha-4)$;

$C = (\alpha-2)(\alpha-3)(\alpha-4)$; $f = \dfrac{n}{N}$, $0 < \alpha < \infty$ is a constant.

Under equation (1), point estimator of $\bar{Y}$ under this setup becomes

$$\bar{y}_{FTj} = \frac{\bar{y}_r}{1-\alpha\left(1-\phi_j(\alpha)\right)}; \quad j = 1,2,3 \tag{2}$$

Table 2. Some special transforms of $\bar{y}_{FTj}$ for various α.

α	$[\bar{y}_{FT1}]$	$[\bar{y}_{FT2}]$	$[\bar{y}_{FT3}]$
1	$\bar{y}_r\left(\dfrac{\bar{x}_n}{\bar{X}}\right)$	$\bar{y}_r\left(\dfrac{\bar{x}_r}{\bar{x}_n}\right)$	$\bar{y}_r\left(\dfrac{\bar{x}_r}{\bar{X}}\right)$
2	$\bar{y}_r\left(\dfrac{\bar{X}}{2\bar{x}_n - \bar{X}}\right)$	$\bar{y}_r\left(\dfrac{\bar{x}_n}{2\bar{x}_r - \bar{x}_n}\right)$	$\bar{y}_r\left(\dfrac{\bar{X}}{2\bar{x}_r - \bar{X}}\right)$
3	$\bar{y}_r\left(\dfrac{(1-f)\bar{X}}{\bar{X} + f(2\bar{X} - 3\bar{x}_n)}\right)$	$\bar{y}_r\left(\dfrac{(1-f)\bar{x}_n}{\bar{x}_n + f(2\bar{x}_n - 3\bar{x}_r)}\right)$	$\bar{y}_r\left(\dfrac{(1-f)\bar{X}}{\bar{X} + f(2\bar{X} - 3\bar{x}_r)}\right)$
4	$\bar{y}_r$	$\bar{y}_r$	$\bar{y}_r$

Remark 1: For $\alpha = \beta = 1$, $t_6 = \bar{y}_{FT2}$ i.e. both estimators are same.

The estimators $\bar{y}_{FTj}$, $j = 1,2,3$ are biased. The bias, M.S.E. and minimum M.S.E. are derived under large sample approximations upto first order of approximation in the following theorem.The expressions for Bias and M.S.E. are in fact equal to infinite Taylor series involving the terms which are functions of a variable. These functions are approximated to varying degrees by the partial sums of these series. For large sample size, $o\left(n^{-1}\right)$ are negligible, therefore, in the present paper first order approximation are considered.

Theorem 1

(i) The proposed estimator $\bar{y}_{FT1}$ is $\bar{y}_{FT1} = \bar{Y}\left[1 + e_1 + P\left(e_3 + e_1 e_3 - \theta_2 e_3^2\right)\right]$ (3)

has bias $B\left(\bar{y}_{FT1}\right) = M_2 \psi_3(\alpha)$ (4)

and M.S.E $M\left(\bar{y}_{FT1}\right) = \bar{Y}^2\left[M_1 C_y^2 + \psi_4(\alpha) M_2\right]$ (5)

The minimum M.S.E. at $P = -\rho \frac{C_y}{C_x} = -V$ is $\left[M\left(\bar{y}_{FT1}\right)\right]_{\min} = \left(M_1 - M_2 \rho^2\right) S_y^2$ (6)

(ii) The estimator $\bar{y}_{FT2}$ in terms of e_1, e_2, e_3 is

$\bar{y}_{FT2} = \bar{Y}\left[1 + e_1 + P\left\{e_2 - e_3 + e_1 e_2 - e_1 e_3 + \left(\theta_2 - \theta_4\right) e_2 e_3 - \theta_2 e_2^2 + \theta_4 e_3^2\right\}\right]$ (7)

with the conditional bias $B\left(\bar{y}_{FT2}\right) = M_3 \psi_3(\alpha)$ (8)

and M.S.E. is $M\left(\bar{y}_{FT2}\right) = \bar{Y}^2\left[M_1 C_y^2 + M_3 \psi_4(\alpha)\right]$ (9)

The minimum M.S.E. at $P = -\rho \frac{C_y}{C_x} = -V$ is $\left[M\left(\bar{y}_{FT2}\right)\right]_{\min} = \left(M_1 - M_3 \rho^2\right) S_y^2$ (10)

(iii) The estimator $\bar{y}_{FT3}$ is $\bar{y}_{FT3} = \bar{Y}\left[1 + e_1 + P\left(e_2 + e_1 e_2 - \theta_2 e_2^2\right)\right]$ (11)

has the conditional bias $B\left(\bar{y}_{FT3}\right) = M_1 \psi_3(\alpha)$ (12)

and M.S.E. $M\left(\bar{y}_{FT3}\right) = M_1 \bar{Y}^2\left[C_y^2 + \psi_4(\alpha)\right]$ (13)

The minimum M.S.E. at $P = -\rho \frac{C_y}{C_x} = -V$ is $\left[M\left(\bar{y}_{FT3}\right)\right]_{\min} = \left(1 - \rho^2\right) M_1 S_y^2$ (14)

Proof: $\bar{y}_{FTj} = \frac{\bar{y}_r}{1 - \alpha\left(1 - \phi_j(\alpha)\right)}$; *j=1,2,3.*

Substituting the value of $\phi_j(\alpha)$; $j = 1,2,3$ and using the concept of large sample approximation, we get

$$\bar{y}_{FT1} = \frac{\bar{Y}\left(1 + e_1\right)\left[\left(A + fB + C\right) + C e_3\right]}{\left(A + fB + C\right) + \left[\alpha f B + (1-\alpha) C\right] e_3} = \bar{Y}\left(1 + e_1\right) \frac{\left(1 + \theta_1 e_3\right)}{\left(1 + \theta_2 e_3\right)}$$

$$\bar{y}_{FT2} = \frac{\bar{Y}\left(1 + e_1\right)\left[\left(A + fB + C\right) + \left(A + fB\right) e_3 + C e_2\right]}{\left(A + fB + C\right) + \left(A + (1-\alpha) fB + \alpha C\right) e_3 + \left[\alpha f B + (1-\alpha) C\right] e_2} = \bar{Y}\left(1 + e_1\right) \frac{1 + \theta_3 e_3 + \theta_1 e_2}{1 + \theta_4 e_3 + \theta_2 e_2}$$

$$\bar{y}_{FT3} = \frac{\bar{Y}\left(1 + e_1\right)\left[\left(A + fB + C\right) + C e_2\right]}{\left(A + fB + C\right) + \left(\alpha fB + (1-\alpha) C\right) e_2} = \bar{Y}\left(1 + e_1\right) \frac{\left(1 + \theta_1 e_2\right)}{\left(1 + \theta_2 e_2\right)}$$

Subsequent Taylor's expansion and ignoring terms of $o\left(n^{-1}\right)$ and higher order leads to equation (3), (7) and (11).

Now, taking expectation of both sides of equation (3), (7) and (11), we get

$$E\left(\bar{y}_{FT1} - \bar{Y}\right) = \bar{Y} E\left[e_1 + P\left(e_3 + e_1 e_3 - \theta_2 e_3^2\right)\right] \Rightarrow B\left(\bar{y}_{FT1}\right) = \bar{Y} P\left(M_2 \rho C_y C_x - \theta_2 M_2 C_x^2\right)$$

$$E\left(\bar{y}_{FT2} - \bar{Y}\right) = \bar{Y}\left[1 + e_1 + P\left\{e_2 - e_3 + e_1 e_2 - e_1 e_3 + \left(\theta_2 - \theta_4\right) e_2 e_3 - \theta_2 e_2^2 + \theta_4 e_3^2\right\}\right]$$

$$\Rightarrow B\left(\bar{y}_{FT2}\right) = -\bar{Y} P M_3\left(\theta_2 C_x^2 - \rho C_y C_x\right)$$

$E\left(\bar{y}_{FT3}-\bar{Y}\right)=\bar{Y}\left[1+e_1+P\left(e_2+e_1e_2-\theta_2e_2^2\right)\right] \Rightarrow B\left(\bar{y}_{FT3}\right)=-\bar{Y}\,PM_1\left(\theta_2C_x^2-\rho C_yC_x\right)$

Simplifying above three expressions, equations (4), (8) and (12) are obtained.

Squaring both sides of equations (3), (7) and (11), and neglecting terms of e's having power greater than two, we have

$\left[\left(\bar{y}_{FT1}-\bar{Y}\right)\right]^2=\bar{Y}^2\left[e_1^2+P^2e_3^2+2Pe_1e_3\right]^2$;

$\left[\left(\bar{y}_{FT2}-\bar{Y}\right)\right]^2=\bar{Y}^2\left[e_1^2+2Pe_1e_2-2Pe_1e_3+P^2e_2^2-2P^2e_2e_3+P^2e_3^2\right]^2$;

$\left[\left(\bar{y}_{FT3}-\bar{Y}\right)\right]^2=\bar{Y}^2\left[e_1^2+P^2e_2^2+2Pe_1e_2\right]^2$

Now taking expectation on both sides inall the above expressions, we get

$M\left(\bar{y}_{FT1}\right)=\bar{Y}^2\,E\left[e_1^2+P^2e_3^2+2Pe_1e_3\right]$;

$M\left(\bar{y}_{FT2}\right)=\bar{Y}^2E\left[e_1^2+2Pe_1e_2-2Pe_1e_3+P^2e_2^2-2P^2e_2e_3+P^2e_3^2\right]^2$;

$M\left(\bar{y}_{FT3}\right)=\bar{Y}^2\,E\left[e_1^2+P^2e_2^2+2Pe_1e_2\right]^2$

Substituting the expectations of e_1, e_2 and e_3 leads to (5), (9) and (13).

Now, differentiating with respect to P and then equating to zero gives

$$\frac{d}{dP}\left[M\left(\bar{y}_{FTj}\right), j=1,2,3\right]=0 \Rightarrow P=-\rho\frac{C_y}{C_x}=-V$$

Substituting the value of P in equations (5), (9) and (13) gives equations (6), (10) and (14).

where, $\psi_3(\alpha)=-P\bar{Y}\left(\theta_2C_x^2-\rho C_yC_x\right)$, $\psi_4(\alpha)=P\left(PC_x^2+2\rho C_yC_x\right)$, $\theta_1=\dfrac{C}{A+fB+C}$,

$\theta_2=\dfrac{\alpha fB+(1-\alpha)C}{A+fB+C}$, $\theta_3=\dfrac{A+fB}{A+fB+C}$, $\theta_4=\dfrac{A+(1-\alpha)fB+\alpha C}{A+fB+C}$ and

$P=\left(\theta_1-\theta_2\right)=-\left(\theta_3-\theta_4\right)$.

This completes the Proof.

Remark 2 : Choices of α: The optimality condition $P=-\rho\dfrac{C_y}{C_x}=-V$ provides the equation

$$P=-V \Rightarrow AV+fB\left(V-\alpha\right)+C\left(V+\alpha\right)=0$$
$$\Rightarrow \alpha^4+\alpha^3\left(V-9-f\right)-\alpha^2\left[8V-26-\left(V+5\right)f\right]$$
$$+\alpha\left[23V-\left(5V+4\right)f-24\right]-22V+4Vf=0 \qquad (15)$$

The equation (15) is a polynomial of degree four and the pair $\left(f,V\right)$ can be treated as known. The four possible roots are denoted by α_1, α_2 α_3 and α_4 (some may be imaginary) for which optimum level of m.s.e. will attain minimum. The following algorithm yields bias control at the optimum level of m.s.e.

STEP I: Compute $\left|B\left(\bar{y}_{FTj}\right)_{\alpha_i}\right|$ for $i=1,2,3,4$; $j=1,2,3$. **STEP II**: For the given values of *i*, choose α_j as $\left|B\left(\bar{y}_{FTj}\right)_{\alpha_i}\right|=\min_{i=1,2,3,4}\left[\left|B\left(\bar{y}_{FTj}\right)_{\alpha_i}\right|\right]$. It is thus, observed that for any given pair of values of $(V, f), 0 < V < \infty; \ 0 < f < 1,$ one can generate a polynomial of degree four for α_1 , α_2, α_3 and α_4 so as to achieve the solution quickly.

Remark 3: The quantity V is stable over moderate length of time which would be initially known or could be guessed on the basis of the past data [13]. Also, equation (15) has only α in the power (of order four), while V and f are known in advance. Therefore, one can solve the equation (15) in order to obtain optimal values in the suggested class.

3. Comparisons

3.1. *Analytical comparison among the proposed estimators*

Using equations (9)-(11), the following identities are obtained:

(i) Let $D_1 = [M(\bar{y}_{FT1})_{\min}] - [M(\bar{y}_{FT2})_{\min}] = (M_3 - M_2)\rho^2 S_y^2$

The estimator $\bar{y}_{FT2}$ is better than $\bar{y}_{FT1}$, if $D_1 > 0 \Rightarrow r \langle \dfrac{n}{2-f}$ where $f = \dfrac{n}{N}$

Now, if $f = 0$ then $r \langle \dfrac{n}{2}$ and if $f = 1$ then $r \langle n$

Hence, $D_1 > 0$ occurs most of the time because $r \langle \dfrac{n}{2}$ always.

(ii) Let $D_2 = [M(\bar{y}_{FT1})_{\min}] - [M(\bar{y}_{FT3})_{\min}] = (M_1 - M_2)\rho^2 S_y^2$

The estimator $\bar{y}_{FT3}$ is better than $\bar{y}_{FT1}$, if $D_2 > 0 \Rightarrow r \langle n$ which is always true.

(iii) Let $D_3 = [M(\bar{y}_{FT2})_{\min}] - [M(\bar{y}_{FT3})_{\min}] = (M_1 - M_3)\rho^2 S_y^2$

The estimator $\bar{y}_{FT3}$ is better than $\bar{y}_{FT2}$, if $D_3 > 0 \Rightarrow \mathrm{N} > n$ which is always true.

Thus, $\bar{y}_{FT3}$ is always better than $\bar{y}_{FT2}$ until $\mathrm{N} = n$. Also, $\bar{y}_{FT2}$ is better than $\bar{y}_{FT1}$. Therefore, the proposed estimation strategy $\bar{y}_{FT3}$ is preferable over the other two proposed methods.

3.2. *Analytical comparison between proposed estimators and* [12] *strategy*

(iv) Let $D_4 = [M(t_6)_{\min}] - [M(\bar{y}_{FT1})_{\min}] = (M_2 - M_3)\rho^2 S_y^2 = 0$

The estimator $\bar{y}_{FT1}$ is better than t_6, if $D_3 > 0 \Rightarrow \mathrm{r} < \dfrac{n}{(2-f)}$

(v) Let $D_5 = [M(t_6)_{\min}] - [M(\bar{y}_{FT2})_{\min}] = (M_2 + M_3 - M_1)\rho^2 S_y^2 = 0$

Therefore, both are equally efficient.

(vi) Let $D_6 = [M(t_6)_{\min}] - [M(\bar{y}_{FT3})_{\min}] = (M_1 - M_3)\rho^2 S_y^2$

The estimator $\bar{y}_{FT3}$ is better than t_6, if $D_6 > 0 \Rightarrow \mathrm{N} > n$ which is always true.

It thus emerges that the proposed estimators $\bar{y}_{FT1}$ and $\bar{y}_{FT3}$ are better than the [12] imputation method and the estimators $\bar{y}_{FT2}$ and t_6 are equally efficient. It is thus established that $\bar{y}_{FT3}$ is the best estimator among the other existing and proposed estimators as discussed above.

4. Almost Unbiased Imputation Methods

In terms of expressions (3), (4) and (5), the bias of $\bar{y}_{FTj}; j = 1,2,3$ could be made zero up to the first order of approximation. This provides the following three equations:

$$-P\bar{Y}M_2\left(\theta_3 C_x^2 - \rho C_y C_x\right) = 0 \tag{16}$$

$$-P\bar{Y}M_3\left(\theta_3 C_x^2 - \rho C_y C_x\right) = 0 \tag{17}$$

$$-P\bar{Y}M_1\left(\theta_3 C_x^2 - \rho C_y C_x\right) = 0 \tag{18}$$

Using data from [5], in the above equations, we obtain that either

$$P = 0 \tag{19}$$

or

$$AV + fB(V - \alpha) + C(V - 1 + \alpha) = 0 \tag{20}$$

Equation (19) provides choice for either $\alpha = \alpha' = 4$ or $\alpha = \alpha_1' = 1.915, \ \alpha = \alpha_2' = 3, \alpha = \alpha_3' = 0.014$ where the proposed estimators are almost unbiased. Considering (20) shows polynomial equation in α, values $\alpha = \alpha_1'' = 1.9902$ while the rest of the roots appear to be imaginary which render the imputed estimator almost unbiased up to the first order of approximation.

5. Empirical Study

To compare the effect of missing data mechanism under the proposed and the existing imputation strategies the following three populations are considered:

Population A: (Source: [14]). *X* represents number of students and *Y* the number of teachers.

Population B: (Source: [15]) comprises of annual production data *X* (‘000 times) and corresponding area *Y* (‘000 ha.) over the time period of 1950-51 to 2015-16 (N=65).

Population C: (Source: [16]). Data collected by a market research company consists of N =2376 points of sale for which the sale area *Y* (in square meters) and the number of employees *X* are surveyed.

Using the whole data set, the following statistics are obtained.

Pop.	N	$\bar{Y}$	$\bar{X}$	C_Y	C_X	ρ	V
A	923 (Moderate)	436.43	11440.50	1.72	1.86	0.95	0.88
B	65 (Small)	3299.41	196536.60	0.31	0.47	0.99	0.65
C	2376 (Large)	1701.95	40.62	1.29	2.35	0.90	0.49

A preliminary sample of n units is selected, from the given populations of N units such that $n = f\,N$, where f denotes the finite sampling fraction being 5% to 25%. Further, r responding units are subsampled from the sample of n units and then selected response rate ranging between 5% to 95% are considered.

Percentage Relative Efficiency (P.R.E.) of the proposed imputation methods with respect to the mean, ratio, compromised, [4] and [12] imputation methods, using equation (21), are reported in Tables 3-5 for the three data sets respectively.

$$P.R.E.\left(t_k, \bar{y}_{FTj}\right) = \frac{MSE(t_k)}{MSE\left(\bar{y}_{FTj}\right)} \times 100\% = P.R.E.(k, FTj), \left(j = 1,2,3; k = 1,2,\ldots,6\right), \quad (21)$$

Table 3. P.R.E. of the Proposed Imputation Methods w.r.to the known Estimators for 5%-25% Sampled for Population A.

n	*r*	*f*	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
46	2	0.05	$\bar{y}_{FT1}$	104.54	15.46	13.87	94.23	13.86
			$\bar{y}_{FT2}$	753.88	111.52	100.04	679.55	100.00
			$\bar{y}_{FT3}$	1120.14	165.70	148.64	100.90	148.58
	9		$\bar{y}_{FT1}$	121.19	33.57	32.00	80.67	32.00
			$\bar{y}_{FT2}$	378.68	104.91	100.00	252.08	100.00
			$\bar{y}_{FT3}$	1120.14	310.33	295.86	545.66	295.81
	16		$\bar{y}_{FT1}$	144.56	59.00	57.47	75.38	57.46
			$\bar{y}_{FT2}$	251.57	102.67	100.01	131.17	100.00
			$\bar{y}_{FT3}$	1120.14	457.16	445.31	584.08	445.27
	23		$\bar{y}_{FT1}$	179.77	97.30	95.82	84.59	95.82
			$\bar{y}_{FT2}$	187.62	101.54	100.01	88.28	100.00
			$\bar{y}_{FT3}$	1120.14	606.25	597.06	527.06	597.03
	30		$\bar{y}_{FT1}$	238.84	161.55	160.16	122.97	160.16
			$\bar{y}_{FT2}$	149.13	100.87	100.00	76.78	100.00
			$\bar{y}_{FT3}$	1120.14	757.63	751.16	576.73	751.13
	37		$\bar{y}_{FT1}$	358.45	291.65	290.46	235.31	290.45
			$\bar{y}_{FT2}$	123.41	100.41	100.00	81.02	100.00
			$\bar{y}_{FT3}$	1120.14	911.38	907.65	735.31	907.64
	44		$\bar{y}_{FT1}$	729.54	695.29	694.67	654.62	694.67
			$\bar{y}_{FT2}$	105.02	100.09	100.00	94.23	100.00
			$\bar{y}_{FT3}$	1120.14	1067.54	1066.60	1005.09	1066.60
92	5	0.10	$\bar{y}_{FT1}$	104.30	15.03	13.61	93.99	13.61
			$\bar{y}_{FT2}$	766.44	111.74	100.04	690.68	100.00
			$\bar{y}_{FT3}$	1120.14	163.80	146.24	100.9	145.64
	18		$\bar{y}_{FT1}$	120.10	32.39	30.82	79.54	30.82
			$\bar{y}_{FT2}$	389.73	105.11	100.02	258.11	100.00
			$\bar{y}_{FT3}$	1120.14	300.40	283.98	540.33	287.11
	32		$\bar{y}_{FT1}$	142.33	56.57	55.03	72.98	55.41
			$\bar{y}_{FT2}$	258.66	102.80	100.01	132.64	100.00

n	r	f	Estimators	$PRE(1, FTj)$	$PRE(2, FTj) = PRE(4, FTj)$	$PRE(3, FTj)$	$PRE(5, FTj)$	$PRE(6, FTj)$
			$\bar{y}_{FT3}$	1120.14	446.34	430.33	573.37	433.29
			$\bar{y}_{FT1}$	175.90	93.09	91.61	80.33	91.60
	46		$\bar{y}_{FT2}$	192.04	101.63	100.01	87.69	100.01
			$\bar{y}_{FT3}$	1120.14	600.02	570.56	581.10	579.33
			$\bar{y}_{FT1}$	232.51	154.66	153.27	115.81	153.26
	60		$\bar{y}_{FT2}$	151.72	100.92	100.01	75.57	100.01
			$\bar{y}_{FT3}$	1120.14	748.33	742.70	595.68	738.02
			$\bar{y}_{FT1}$	348.20	280.50	279.29	223.39	279.28
	74		$\bar{y}_{FT2}$	124.68	100.44	100.01	79.99	100.01
			$\bar{y}_{FT3}$	1120.14	900.66	892.91	721.22	898.45
			$\bar{y}_{FT1}$	716.49	681.08	680.45	639.05	680.45
	87		$\bar{y}_{FT2}$	105.30	100.10	100.01	93.92	100.01
			$\bar{y}_{FT3}$	1120.14	1064.64	1066.66	999.09	1066.65
			$\bar{y}_{FT1}$	104.07	14.95	13.36	93.75	13.35
	7		$\bar{y}_{FT2}$	779.48	111.97	100.04	702.24	100.00
			$\bar{y}_{FT3}$	1120.15	160.91	143.77	109.16	143.71
			$\bar{y}_{FT1}$	119.01	31.20	29.63	78.41	29.63
	28		$\bar{y}_{FT2}$	401.68	105.32	100.02	264.64	100.00
			$\bar{y}_{FT3}$	1120.15	293.70	278.92	564.36	420.44
			$\bar{y}_{FT1}$	140.08	54.12	52.58	70.57	52.58
	48		$\bar{y}_{FT2}$	266.44	102.94	100.01	134.24	100.00
			$\bar{y}_{FT3}$	1120.15	432.76	420.48	564.36	420.44
			$\bar{y}_{FT1}$	172.00	88.84	87.35	76.03	87.35
138	69	0.15	$\bar{y}_{FT2}$	196.92	101.71	100.01	87.04	100.01
			$\bar{y}_{FT3}$	1120.15	578.57	568.90	595.12	568.86
			$\bar{y}_{FT1}$	226.08	147.67	146.26	108.53	146.26
	90		$\bar{y}_{FT2}$	154.58	100.97	100.01	74.21	100.01
			$\bar{y}_{FT3}$	1120.15	731.63	724.69	600.74	724.66
			$\bar{y}_{FT1}$	337.66	269.03	267.81	211.15	267.80
	110		$\bar{y}_{FT2}$	126.09	100.47	100.01	78.85	100.01
			$\bar{y}_{FT3}$	1120.15	892.48	888.41	700.47	888.40
			$\bar{y}_{FT1}$	702.52	665.89	665.23	622.40	665.23
	131		$\bar{y}_{FT2}$	105.61	100.10	100.01	93.57	100.01
			$\bar{y}_{FT3}$	1120.15	1061.74	1060.70	992.41	1060.70
			$\bar{y}_{FT1}$	103.83	14.69	13.10	93.52	13.09
	9		$\bar{y}_{FT2}$	793.04	112.21	100.04	714.26	100.00
			$\bar{y}_{FT3}$	1120.15	158.50	141.31	108.88	141.25
184		0.20	$\bar{y}_{FT1}$	117.92	30.01	28.44	77.27	28.44
	37		$\bar{y}_{FT2}$	414.67	105.55	100.02	271.72	100.00
			$\bar{y}_{FT3}$	1120.15	285.12	270.19	734.00	270.14
	64		$\bar{y}_{FT1}$	137.82	51.66	50.12	68.15	50.11
			$\bar{y}_{FT2}$	275.02	103.09	100.02	136.00	100.00

n	r	f	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT3}$	1120.15	419.87	407.36	553.93	407.32
			$\bar{y}_{FT1}$	168.06	84.56	83.07	71.69	83.06
	92		$\bar{y}_{FT2}$	202.35	101.81	100.01	86.31	100.01
			$\bar{y}_{FT3}$	1120.15	563.58	553.64	477.82	553.60
			$\bar{y}_{FT1}$	219.56	140.57	139.16	101.15	139.16
	120		$\bar{y}_{FT2}$	157.79	101.02	100.01	72.69	100.01
			$\bar{y}_{FT3}$	1120.15	717.17	709.97	586.06	709.94
			$\bar{y}_{FT1}$	326.83	257.25	256.01	198.57	256.01
	147		$\bar{y}_{FT2}$	127.67	100.49	100.01	77.57	100.01
			$\bar{y}_{FT3}$	1120.15	881.68	877.42	680.56	877.41
			$\bar{y}_{FT1}$	687.54	649.60	648.92	604.56	648.92
	175		$\bar{y}_{FT2}$	105.96	100.11	100.01	93.17	100.01
			$\bar{y}_{FT3}$	1120.15	1058.33	1057.23	984.95	1057.22
			$\bar{y}_{FT1}$	103.59	14.43	12.84	93.28	12.83
	12		$\bar{y}_{FT2}$	807.15	112.46	100.04	726.77	100.00
			$\bar{y}_{FT3}$	1120.15	156.07	138.84	108.60	138.78
			$\bar{y}_{FT1}$	116.82	28.82	27.25	76.13	27.24
	46		$\bar{y}_{FT2}$	428.83	105.80	100.02	279.45	100.00
			$\bar{y}_{FT3}$	1120.15	276.36	261.27	529.95	261.22
			$\bar{y}_{FT1}$	135.54	49.19	47.64	65.72	47.64
	81		$\bar{y}_{FT2}$	284.54	103.26	100.02	137.96	100.00
			$\bar{y}_{FT3}$	1120.15	406.49	393.74	543.11	393.69
			$\bar{y}_{FT1}$	164.09	80.24	78.74	67.32	78.74
230	115	0.25	$\bar{y}_{FT2}$	208.42	101.92	100.01	85.50	100.01
			$\bar{y}_{FT3}$	1120.15	547.74	537.51	559.53	537.48
			$\bar{y}_{FT1}$	212.94	133.37	131.95	93.67	131.95
	150		$\bar{y}_{FT2}$	161.39	101.09	100.01	70.99	100.01
			$\bar{y}_{FT3}$	1120.15	701.59	694.11	592.71	694.09
			$\bar{y}_{FT1}$	315.70	245.14	243.88	185.64	243.88
	184		$\bar{y}_{FT2}$	129.46	100.52	100.01	76.12	100.01
			$\bar{y}_{FT3}$	1120.15	869.81	865.33	658.67	865.32
			$\bar{y}_{FT1}$	671.46	632.10	631.40	585.39	631.40
	219		$\bar{y}_{FT2}$	106.35	100.12	100.01	92.72	100.01
			$\bar{y}_{FT3}$	1120.15	1054.50	1053.32	976.56	1053.32

Table 4. P.R.E. of the Proposed Imputation Methods w.r.to the known Estimators for 5%-25% Sampled for Population B.

n	r	F	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT1}$	104.90	34.71	6.91	90.46	6.91
3	0	0.05	$\bar{y}_{FT2}$	1517.01	502.00	100.00	1308.16	100.00
			$\bar{y}_{FT3}$	5204.25	1722.16	343.06	4487.76	343.06

n	*r*	*F*	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT1}$	123.19	53.25	25.55	72.51	25.55
	1		$\bar{y}_{FT2}$	482.05	208.39	100.00	283.73	100.00
			$\bar{y}_{FT3}$	5204.25	2249.75	1079.60	3063.17	1079.60
			$\bar{y}_{FT1}$	149.68	80.11	52.56	73.43	52.56
	1		$\bar{y}_{FT2}$	284.79	152.42	100.00	139.71	100.00
			$\bar{y}_{FT3}$	5204.25	2785.40	1827.39	2553.10	1827.39
			$\bar{y}_{FT1}$	191.50	122.51	95.18	100.44	95.18
	2		$\bar{y}_{FT2}$	201.19	128.71	100.00	105.53	100.00
			$\bar{y}_{FT3}$	5204.25	3329.28	2586.69	2729.63	2586.69
			$\bar{y}_{FT1}$	267.36	199.41	172.50	172.50	172.50
	2		$\bar{y}_{FT2}$	154.99	115.60	100.00	100.00	100.00
			$\bar{y}_{FT3}$	5204.25	3881.60	3357.75	3357.78	3357.75
			$\bar{y}_{FT1}$	447.34	381.86	355.93	360.61	355.93
	3		$\bar{y}_{FT2}$	125.68	107.29	100.00	101.31	100.00
			$\bar{y}_{FT3}$	5204.25	4442.55	4140.87	4195.23	4140.87
			$\bar{y}_{FT1}$	1414.43	1362.27	1341.61	1356.75	1341.61
	3		$\bar{y}_{FT2}$	105.43	101.54	100.00	101.13	100.00
			$\bar{y}_{FT3}$	5204.25	5012.33	4936.31	4992.03	4936.31
			$\bar{y}_{FT1}$	104.64	34.45	6.65	90.20	6.65
	0		$\bar{y}_{FT2}$	1573.02	517.89	100.00	1355.91	100.00
			$\bar{y}_{FT3}$	5204.25	1713.41	330.85	4485.96	330.85
			$\bar{y}_{FT1}$	121.97	52.02	24.32	71.28	24.32
	1		$\bar{y}_{FT2}$	501.61	213.93	100.00	293.14	100.00
			$\bar{y}_{FT3}$	5204.25	2219.60	1037.52	3041.32	1037.52
			$\bar{y}_{FT1}$	147.09	77.49	49.92	70.80	49.92
	2		$\bar{y}_{FT2}$	294.67	155.23	100.00	141.83	100.00
			$\bar{y}_{FT3}$	5204.25	2741.53	1766.16	2505.02	1766.16
			$\bar{y}_{FT1}$	186.77	117.71	90.36	95.62	90.36
7	3	0.10	$\bar{y}_{FT2}$	206.70	130.27	100.00	105.83	100.00
			$\bar{y}_{FT3}$	5204.25	3279.94	2517.80	2664.51	2517.80
			$\bar{y}_{FT1}$	258.83	190.76	163.80	163.80	163.80
	4		$\bar{y}_{FT2}$	158.01	116.46	100.00	100.00	100.00
			$\bar{y}_{FT3}$	5204.25	3835.62	3293.57	3293.60	3293.57
			$\bar{y}_{FT1}$	430.24	364.53	338.51	343.20	338.51
	5		$\bar{y}_{FT2}$	127.10	107.69	100.00	101.39	100.00
			$\bar{y}_{FT3}$	5204.25	4409.43	4094.63	4151.36	4094.63
			$\bar{y}_{FT1}$	1362.36	1309.48	1288.54	1303.89	1288.54
	6		$\bar{y}_{FT2}$	105.73	101.63	100.00	101.19	100.00
			$\bar{y}_{FT3}$	5204.25	5002.25	4922.25	4980.89	4922.25
			$\bar{y}_{FT1}$	104.38	34.19	6.39	89.94	6.39
10	0	0.15	$\bar{y}_{FT2}$	1633.63	535.08	100.00	1407.59	100.00
			$\bar{y}_{FT3}$	5204.25	1704.62	318.57	4484.15	318.57
	2		$\bar{y}_{FT1}$	120.76	50.79	23.08	70.05	23.08

n	r	F	Estimators	$PRE(1,FTj)$	$PRE(2,FTj)$ $=PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT2}$	523.27	220.08	100.00	303.55	100.00
			$\bar{y}_{FT3}$	5204.25	2188.83	994.56	3019.03	994.56
			$\bar{y}_{FT1}$	144.50	74.86	47.27	68.17	47.27
			$\bar{y}_{FT2}$	305.66	158.34	100.00	144.20	100.00
	3		$\bar{y}_{FT3}$	5204.25	2696.05	1702.65	2455.17	1702.65
			$\bar{y}_{FT1}$	182.03	112.90	85.52	90.79	85.52
	5		$\bar{y}_{FT2}$	212.84	132.01	100.00	106.16	100.00
			$\bar{y}_{FT3}$	5204.25	3227.93	2445.20	2595.87	2445.20
			$\bar{y}_{FT1}$	250.27	182.08	155.07	155.08	155.07
	6		$\bar{y}_{FT2}$	161.38	117.41	100.00	100.00	100.00
			$\bar{y}_{FT3}$	5204.25	3786.34	3224.76	3224.79	3224.76
			$\bar{y}_{FT1}$	413.02	347.07	320.95	325.66	320.95
	8		$\bar{y}_{FT2}$	128.68	108.14	100.00	101.47	100.00
			$\bar{y}_{FT3}$	5204.25	4373.30	4044.20	4103.50	4044.20
			$\bar{y}_{FT1}$	1308.84	1255.22	1233.99	1249.55	1233.99
	9		$\bar{y}_{FT2}$	106.07	101.72	100.00	101.26	100.00
			$\bar{y}_{FT3}$	5204.25	4991.06	4906.63	4968.52	4906.63
			$\bar{y}_{FT1}$	104.13	33.93	6.13	89.68	6.13
	1		$\bar{y}_{FT2}$	1699.45	553.76	100.00	1463.70	100.00
			$\bar{y}_{FT3}$	5204.25	1695.78	306.23	4482.33	306.23
			$\bar{y}_{FT1}$	119.54	49.56	21.84	68.82	21.84
	3		$\bar{y}_{FT2}$	547.41	226.93	100.00	315.16	100.00
			$\bar{y}_{FT3}$	5204.25	2157.42	950.71	2996.26	950.71
			$\bar{y}_{FT1}$	141.90	72.23	44.63	65.53	44.63
	5		$\bar{y}_{FT2}$	317.96	161.83	100.00	146.84	100.00
			$\bar{y}_{FT3}$	5204.25	2648.85	1636.76	2403.44	1636.76
			$\bar{y}_{FT1}$	177.27	108.08	80.68	85.96	80.68
13	7	0.20	$\bar{y}_{FT2}$	219.72	133.96	100.00	106.54	100.00
			$\bar{y}_{FT3}$	5204.25	3173.03	2368.56	2523.41	2368.56
			$\bar{y}_{FT1}$	241.67	173.37	146.32	146.32	146.32
	8		$\bar{y}_{FT2}$	165.17	118.49	100.00	100.00	100.00
			$\bar{y}_{FT3}$	5204.25	3733.37	3150.82	3150.85	3150.82
			$\bar{y}_{FT1}$	395.67	329.49	303.27	308.00	303.27
	10		$\bar{y}_{FT2}$	130.47	108.64	100.00	101.56	100.00
			$\bar{y}_{FT3}$	5204.25	4333.73	3988.96	4051.09	3988.96
			$\bar{y}_{FT1}$	1253.80	1199.43	1177.90	1193.68	1177.90
	12		$\bar{y}_{FT2}$	106.44	101.83	100.00	101.34	100.00
			$\bar{y}_{FT3}$	5204.25	4978.56	4889.18	4954.69	4889.18
			$\bar{y}_{FT1}$	103.87	33.67	5.86	89.42	5.86
	1		$\bar{y}_{FT2}$	1771.16	574.10	100.00	1524.85	100.00
16		0.25	$\bar{y}_{FT3}$	5204.25	1686.90	293.83	4480.50	293.83
			$\bar{y}_{FT1}$	118.32	48.32	20.60	67.59	20.60
	3		$\bar{y}_{FT2}$	574.46	234.60	100.00	328.17	100.00

n	r	F	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT3}$	5204.25	2125.35	905.93	2973.02	905.93
			$\bar{y}_{FT1}$	139.30	69.59	41.98	62.90	41.98
	6		$\bar{y}_{FT2}$	331.83	165.77	100.00	149.82	100.00
			$\bar{y}_{FT3}$	5204.25	2599.84	1568.35	2349.72	1568.35
			$\bar{y}_{FT1}$	172.51	103.26	75.83	81.11	75.83
	8		$\bar{y}_{FT2}$	227.50	136.17	100.00	106.96	100.00
			$\bar{y}_{FT3}$	5204.25	3115.00	2287.54	2446.81	2287.54
			$\bar{y}_{FT1}$	233.05	164.63	137.53	137.53	137.53
	11		$\bar{y}_{FT2}$	169.46	119.70	100.00	100.00	100.00
			$\bar{y}_{FT3}$	5204.25	3676.29	3071.13	3071.16	3071.13
			$\bar{y}_{FT1}$	378.20	311.77	285.47	290.21	285.47
	13		$\bar{y}_{FT2}$	132.48	109.22	100.00	101.66	100.00
			$\bar{y}_{FT3}$	5204.25	4290.21	3928.19	3993.43	3928.19
			$\bar{y}_{FT1}$	1197.19	1142.04	1120.20	1136.21	1120.20
	15		$\bar{y}_{FT2}$	106.87	101.95	100.00	101.43	100.00
			$\bar{y}_{FT3}$	5204.25	4964.50	4869.55	4939.15	4869.55

Table 5. P.R.E. of the Proposed Imputation Methods w.r.to the known Estimators for 5%-25% Sampled for Population C.

n	r	f	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT1}$	104.01	107.94	23.77	88.59	23.77
	6		$\bar{y}_{FT2}$	437.50	454.02	100.00	372.64	100.00
			$\bar{y}_{FT3}$	526.32	546.19	120.30	448.29	120.30
			$\bar{y}_{FT1}$	118.41	122.20	40.90	68.36	40.90
	24		$\bar{y}_{FT2}$	289.47	298.75	100.00	167.12	100.00
			$\bar{y}_{FT3}$	526.32	543.17	181.82	303.85	181.82
			$\bar{y}_{FT1}$	137.76	141.38	63.94	70.21	63.94
	42		$\bar{y}_{FT2}$	215.46	221.11	100.00	109.81	100.00
			$\bar{y}_{FT3}$	526.32	540.12	244.27	268.25	244.27
			$\bar{y}_{FT1}$	165.18	168.54	96.57	96.58	96.57
119	59	0.05	$\bar{y}_{FT2}$	171.05	174.53	100.00	100.01	100.00
			$\bar{y}_{FT3}$	526.32	537.01	307.69	307.72	307.69
			$\bar{y}_{FT1}$	207.03	210.00	146.36	152.41	146.36
	77		$\bar{y}_{FT2}$	141.45	143.48	100.00	104.13	100.00
			$\bar{y}_{FT3}$	526.32	533.86	372.09	387.47	372.09
			$\bar{y}_{FT1}$	278.75	281.05	231.71	249.75	231.71
	95		$\bar{y}_{FT2}$	120.30	121.29	100.00	107.79	100.00
			$\bar{y}_{FT3}$	526.32	530.66	437.50	471.57	437.50
			$\bar{y}_{FT1}$	430.07	430.97	411.78	427.36	411.78
	113		$\bar{y}_{FT2}$	104.44	104.66	100.00	103.78	100.00
			$\bar{y}_{FT3}$	526.32	527.41	503.94	523.00	503.94
	12		$\bar{y}_{FT1}$	103.80	107.73	23.53	88.38	23.53

n	*r*	*f*	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
238		0.10	$\bar{y}_{FT2}$	441.24	457.94	100.00	375.66	100.00
			$\bar{y}_{FT3}$	526.32	546.24	119.28	448.09	119.28
			$\bar{y}_{FT1}$	117.48	121.28	39.80	67.31	39.80
	48		$\bar{y}_{FT2}$	295.18	304.73	100.00	169.14	100.00
			$\bar{y}_{FT3}$	526.32	543.35	178.30	301.58	178.30
			$\bar{y}_{FT1}$	135.94	139.57	61.77	68.08	61.77
	83		$\bar{y}_{FT2}$	220.07	225.94	100.00	110.21	100.00
			$\bar{y}_{FT3}$	526.32	540.37	239.16	263.57	239.16
			$\bar{y}_{FT1}$	162.25	165.64	93.08	93.09	93.08
	119		$\bar{y}_{FT2}$	174.31	177.95	100.00	100.01	100.00
			$\bar{y}_{FT3}$	526.32	537.30	301.94	301.97	301.94
			$\bar{y}_{FT1}$	202.75	205.76	141.28	147.40	141.28
	154		$\bar{y}_{FT2}$	143.51	145.64	100.00	104.34	100.00
			$\bar{y}_{FT3}$	526.32	534.13	366.73	382.64	366.73
			$\bar{y}_{FT1}$	273.16	275.51	225.06	243.51	225.06
	190		$\bar{y}_{FT2}$	121.37	122.42	100.00	108.20	100.00
			$\bar{y}_{FT3}$	526.32	530.85	433.64	469.19	433.64
			$\bar{y}_{FT1}$	425.98	426.92	406.92	423.16	406.92
	226		$\bar{y}_{FT2}$	104.68	104.91	100.00	103.99	100.00
			$\bar{y}_{FT3}$	526.32	527.47	502.76	522.83	502.76
			$\bar{y}_{FT1}$	103.59	107.52	23.28	88.16	23.28
	18		$\bar{y}_{FT2}$	445.07	461.95	100.00	378.75	100.00
			$\bar{y}_{FT3}$	526.32	546.29	118.26	447.90	118.26
			$\bar{y}_{FT1}$	116.54	120.35	38.69	66.27	38.69
	71		$\bar{y}_{FT2}$	301.24	311.09	100.00	171.28	100.00
			$\bar{y}_{FT3}$	526.32	543.52	174.72	299.26	174.72
			$\bar{y}_{FT1}$	134.11	137.75	59.59	65.92	59.59
	125		$\bar{y}_{FT2}$	225.06	231.18	100.00	110.63	100.00
			$\bar{y}_{FT3}$	526.32	540.63	233.86	258.71	233.86
			$\bar{y}_{FT1}$	159.28	162.69	89.54	89.55	89.54
356	178	0.15	$\bar{y}_{FT2}$	177.88	181.70	100.00	100.01	100.00
			$\bar{y}_{FT3}$	526.32	537.59	295.87	295.91	295.87
			$\bar{y}_{FT1}$	198.36	201.41	136.05	142.26	136.05
	232		$\bar{y}_{FT2}$	145.80	148.04	100.00	104.57	100.00
			$\bar{y}_{FT3}$	526.32	534.41	360.99	377.46	360.99
			$\bar{y}_{FT1}$	267.31	269.72	218.10	236.98	218.10
	285		$\bar{y}_{FT2}$	122.56	123.67	100.00	108.65	100.00
			$\bar{y}_{FT3}$	526.32	531.06	429.43	466.59	429.43
			$\bar{y}_{FT1}$	421.53	422.51	401.62	418.58	401.62
	339		$\bar{y}_{FT2}$	104.96	105.20	100.00	104.22	100.00
			$\bar{y}_{FT3}$	526.32	527.53	501.46	522.63	501.46
475	24	0.20	$\bar{y}_{FT1}$	103.38	107.32	23.03	87.94	23.03
			$\bar{y}_{FT2}$	448.98	466.06	100.00	381.91	100.00

n	r	f	Estimators	$PRE(1,FTj)$	$PRE(2,FTj) = PRE(4,FTj)$	$PRE(3,FTj)$	$PRE(5,FTj)$	$PRE(6,FTj)$
			$\bar{y}_{FT3}$	526.32	546.34	117.22	447.70	117.22
			$\bar{y}_{FT1}$	115.61	119.43	37.57	65.21	37.57
	95		$\bar{y}_{FT2}$	307.69	317.86	100.00	173.57	100.00
			$\bar{y}_{FT3}$	526.32	543.70	171.05	296.89	171.05
			$\bar{y}_{FT1}$	132.25	135.92	57.38	63.74	57.38
	166		$\bar{y}_{FT2}$	230.48	236.87	100.00	111.09	100.00
			$\bar{y}_{FT3}$	526.32	540.90	228.35	253.68	228.35
			$\bar{y}_{FT1}$	156.25	159.69	85.94	85.95	85.94
	238		$\bar{y}_{FT2}$	181.82	185.82	100.00	100.01	100.00
			$\bar{y}_{FT3}$	526.32	537.91	289.47	289.51	289.47
			$\bar{y}_{FT1}$	193.85	196.94	130.68	136.98	130.68
	309		$\bar{y}_{FT2}$	148.34	150.70	100.00	104.82	100.00
			$\bar{y}_{FT3}$	526.32	534.71	354.81	371.90	354.81
			$\bar{y}_{FT1}$	261.19	263.66	210.82	230.14	210.82
	380		$\bar{y}_{FT2}$	123.89	125.06	100.00	109.17	100.00
			$\bar{y}_{FT3}$	526.32	531.28	424.81	463.75	424.81
			$\bar{y}_{FT1}$	416.67	417.69	395.83	413.58	395.83
	451		$\bar{y}_{FT2}$	105.26	105.52	100.00	104.48	100.00
			$\bar{y}_{FT3}$	526.32	527.60	500.00	522.42	500.00
			$\bar{y}_{FT1}$	102.84	106.76	22.70	87.44	22.70
	30		$\bar{y}_{FT2}$	332.32	344.99	73.36	282.56	73.36
			$\bar{y}_{FT3}$	365.87	379.83	80.77	311.09	80.77
			$\bar{y}_{FT1}$	112.96	116.73	35.91	63.20	35.91
	119		$\bar{y}_{FT2}$	257.69	266.30	81.92	144.18	81.92
			$\bar{y}_{FT3}$	365.87	378.09	116.31	204.71	116.31
			$\bar{y}_{FT1}$	126.43	130.00	53.48	59.68	53.48
	208		$\bar{y}_{FT2}$	207.31	213.17	87.70	97.86	87.70
			$\bar{y}_{FT3}$	365.87	376.21	154.77	172.71	154.77
			$\bar{y}_{FT1}$	145.23	148.52	78.01	78.02	78.01
594	297	0.25	$\bar{y}_{FT2}$	171.01	174.89	91.86	91.87	91.86
			$\bar{y}_{FT3}$	365.87	374.16	196.53	196.55	196.53
			$\bar{y}_{FT1}$	173.31	176.18	114.64	120.49	114.64
	386		$\bar{y}_{FT2}$	143.61	145.99	95.00	99.85	95.00
			$\bar{y}_{FT3}$	365.87	371.94	242.02	254.37	242.02
			$\bar{y}_{FT1}$	219.79	221.96	175.28	192.35	175.28
	475		$\bar{y}_{FT2}$	122.20	123.41	97.45	106.95	97.45
			$\bar{y}_{FT3}$	365.87	369.50	291.79	320.20	291.79
			$\bar{y}_{FT1}$	311.56	312.37	295.01	309.11	295.01
	564		$\bar{y}_{FT2}$	105.00	105.28	99.43	104.18	99.43
			$\bar{y}_{FT3}$	365.87	366.83	346.44	363.00	346.44

6. Discussion

Based on empirical investigation summarized in Table 3 we conclude that the proposed estimator $\bar{y}_{FT1}$ performs better than the prevalent contemporary estimators when the response rate is higher than 50% and continues to gain in efficiency as response rate increases. At 65% response rate, the relative efficiency gain is observed as 238.84%, 161.55%, 160.16%, 122.97% and 160.16% which grows to gain further 358.45%, 291.65%, 290.46%, 235.31% and 290.45% at 80% response rate with respect to the mean, ratio, compromised, [4] and [12] methods of imputation respectively. The proposed estimator $\bar{y}_{FT1}$ is therefore more suitable for imputing data in real survey situations where fractional response rate is 65% and higher, which isthe most common data collection scenario in the real sample surveys observed so far.

As response rate decreases the proposed estimator $\bar{y}_{FT2}$ performs better. For example, if the response rate is 5%, then a gain of 753.88%, 111.52%, 100.04%, 679.55% and 100% (i.e. equal) with respect to mean, ratio, compromised, [4] and [12] methods of imputation is expected by the proposed method $\bar{y}_{FT2}$ respectively. At response rate of 35%, the corresponding relative efficiency gains are 251.57%, 102.67%, 100.01% (i.e. equal), 131.17% and 100.00% (i.e. equal) respectively. Hence, the proposed estimator $\bar{y}_{FT2}$ is suited for imputing missing data in real survey situation where higher fractions of non-respondents occur. Relative Efficiency of the proposed estimator $\bar{y}_{FT3}$ increases as the response rate in the survey increases (similar to the behavior exhibited by the proposed estimator $\bar{y}_{FT1}$) except under the mean method of imputation where relative efficiency of the proposed estimator $\bar{y}_{FT3}$ is substantially high at 1120.14%. The median gain in efficiency for $\bar{y}_{FT3}$ in comparison to the prevalent ratio, compromised, [4] and [12] methods of imputation for the proposed estimator is observed to be 606.25%, 597.09%, 527.06% and 597.03% respectively at 50% response rate.

Similarly, based on empirical summary from Table 4, we conclude that for the proposed estimator $\bar{y}_{FT1}$ is more efficient than the other five considered estimators, when the response rate is higher than 50% and continues to gain in efficiency as response rate increases. At 65% response rate, the relative efficiency gain is observed as 267.36%, 199.40%, 172.50%, 172.50% and 172.50% which substantially grows to gain further 1414.43%, 1362.27%, 1341.61%, 1356.75% and 1341.61% at 95% response rate with respect to the mean, ratio, compromised, [4] and [12] methods of imputation respectively.

As response rate decreases the proposed estimator $\bar{y}_{FT2}$ performs better. For example, if the response rate is 5%, then a gain of 1517.01%, 502.00%, 100.00%, 1308.16% and 100% (i.e. equal) is achieved by the proposed method $\bar{y}_{FT2}$ with respect to mean, ratio, compromised, [4] and [12] methods of imputation respectively. At response rate of 35%, the corresponding relative efficiency gains are 284.79%, 152.42%, 100.00% (i.e. equal), 139.71% and 100.00% (i.e. equal) respectively. Hence the proposed estimator $\bar{y}_{FT2}$ is suited for imputing missing data in real survey situation wherehigher fractions of non-respondents occur.

Relative Efficiency of the proposed estimator $\bar{y}_{FT3}$ increases as the response rate in the survey increases (similar to the behavior exhibited by the proposed estimator $\bar{y}_{FT1}$) except under the mean method of imputation where relative efficiency of the proposed estimator $\bar{y}_{FT3}$ is substantially high at 5204.25%. The median gain in efficiency in comparison to the prevalent ratio, compromised, [4] and [12] methods of imputation for the proposed estimator is observed at 3329.28%, 2586.69%, 2729.63% and 2586.69% respectively at 50% response rate.

Similarly, from Table 5, the proposed estimator $\bar{y}_{FT1}$ is evidenced to perform better than the prevalent contemporary estimators when the response rate is higher than 50% and the gain in efficiency continues as response rate increases. At 65% response rate, the relative efficiency gain is observed as 207.03%, 210.00%, 146.36%, 152.41% and 146.36% which grows to gain further 430.07%, 430.97%, 411.78%, 427.36% and 411.78% at 95% response rate with respect to the mean, ratio, compromised, [4] and [12] methods of imputation respectively. The proposed estimator $\bar{y}_{FT1}$ is therefore more suitable for imputing data in real survey situations where fractional response rate is 65% and higher, which isthe most common data collection scenario in the real sample surveys observed so far.

As response rate decreases the proposed estimator $\bar{y}_{FT2}$ performs better. For example, if the response rate is 5%, then a gain of 430.50%, 454.02%, 100.00%, 372.64% and 100% (i.e. equal) with respect to mean, ratio, compromised, [4] and [12] methods of imputation respectively. At response rate of 35%, the corresponding relative efficiency gains are 215.46%, 221.11%, 100.00% (i.e. equal), 109.81% and 100.00% (i.e. equal) respectively. Hence the proposed estimator $\bar{y}_{FT2}$ is suited for imputing missing data in real survey situation where higher fractions of non-respondents occur.

Relative Efficiency of the proposed estimator $\bar{y}_{FT3}$ increases as the response rate in the survey increases (similar to the behavior exhibited by the proposed estimator $\bar{y}_{FT1}$) except under the mean method of imputation where relative efficiency of the proposed estimator $\bar{y}_{FT3}$ is substantially high at 526.32%. The median gain in efficiency in comparison to the prevalent ratio, compromised, [4] and [12] methods of imputation for the proposed estimator is observed at 537.01%, 307.69% and 307.72% and 307.69 respectively at 50% response rate.

7. Conclusion

Thus, the present work provides threemore efficient alternative imputation strategies for situations involving higher fraction of non-respondents ($\bar{y}_{FT2}$) as well as for the situations which involve smaller data loss ($\bar{y}_{FT1}$ and $\bar{y}_{FT3}$) on the study characteristic in a bi-variate sample data.The new estimators are formulated by transforming Singh's Walsh-type estimator to Factor-Type estimators. The proposed estimator $\bar{y}_{FT3}$ shows highest improvement in terms of relative efficiency among all the three proposed estimators. All the three proposed alternative estimators are theoretically as well as empirically found to have higher PRE and therefore are regarded superior to the existing mean, ratio, compromised, [4] and [12] methods of imputation respectively. The present paper is therefore an important contribution for the practitioners in the area of missing data analysis as it offers improved estimators than the existing ones for imputing lost or missing data.

Acknowledgments

Part of the work of the first author is supported by R & D Grant from University of Delhi and DU-DST PURSE Grant.

References

1. J. R. Carpenter and M. G. Kenward, Multiple Imputation and Its Application (Wiley, (2013). http://dx.doi.org/10.1002/9781119942283
2. D. B. Rubin, Biometrika **63**, 581 (1976). http://dx.doi.org/10.1093/biomet/63.3.581
3. H. Toutenburg, and V. K. Srivastava, Metrika **48**, 177 (1998). http://dx.doi.org/10.1007/PL00003973
4. H. Toutenburg, V. K. Srivastava, and Shalabh, Statistical Papers **49**, 237 (2008). http://dx.doi.org/10.1007/s00362-006-0009-4
5. J. Shabbir and N. S. Khan, Commun. Statistics-Theory Meth. **42(22)**, 4127 (2013). http://dx.doi.org/10.1080/03610926.2011.643851
6. G. Kalton and D. Kaspryzk, Imputing for Missing Survey Responses – *Proc. of the Section on Survey Research Methods* (Am. Statistical Assoc., 1982) pp. 22-31.
7. S. Singh and S. Horn, Metrika **51**, 267 (2000). http://dx.doi.org/10.1007/s001840000054
8. S. Singh, and B. Deo, Statistical Papers **44**, 555 (2003). http://dx.doi.org/10.1007/BF02926010
9. M. S. Ahmed, O. Al-Titi, Z. Al-Rawi, and W. Abu-Dayyeh, Statistics Transition **7(6)**, 1247 (2006).
10. R. Pandey, N. S. Thakur, and K. Yadav, J. Ind. Statistical Assoc. **53(1,2)**, 89 (2015).
11. D. Shukla and N. S. Thakur, Statistics Transition **9(1)**, 33 (2008).
12. S. Singh, Statistics: J. Theor. Appl. Statistics **43(5)**, 499 (2009).
13. V. N. Reddy, Sankhya C **40**, 29 (1978).
14. N. Koyuncu and C. Kadilar, J. Statistical Plann. Inference **139**, 2552 (2009). http://dx.doi.org/10.1016/j.jspi.2008.11.009
15. Indian Institute of Sugarcane Research: www.iisr.nic.in
16. G. Diana and P. F. Perri, Commun. Statistics-Theory Meth. **39**, 3245 (2010). http://dx.doi.org/10.1080/03610920903009400