LINEAR COMBINATION BASED IMPUTATION METHOD FOR MISSING DATA IN SAMPLE

D. Shukla
Deptt. of Mathematics and Statistics, Dr. H. S. Gour Central University, Sagar , M.P. – 470003, INDIA
Narendra Singh Thakur
Centre for Mathematical Sciences (CMS), Banasthali University, Rajasthan-304022, INDIA
Dharmendra Singh Thakur
School of Excellence, Sagar, M.P.- 470002, INDIA
Sharad Pathak
Deptt. of Mathematics and Statistics, Dr. H. S. Gour Central University, Sagar , M.P. – 470003, INDIA

ABSTRACT
To estimate the population mean using auxiliary variable there are many estimators available in literature like- ratio, product, regression, dual-to-ratio estimator and so on. Suppose that all the information of the main variable is present in the sample but only a part of data of the auxiliary variable is available. Then, in this case none of the above estimators could be used. This paper presents an imputation based factor-type class of estimation strategy for population mean in presence of missing values of auxiliary variables. The non-sampled part of the population is used as an imputation technique in the proposed class. Some properties of estimators are discussed and numerical study is performed with efficiency comparison to the non-imputed estimator. An optimum sub-class is recommended.

Keywords: Imputation, Non-response, Post-stratification, Simple Random Sampling without Replacement (SRSWOR), Respondents (R).

1.0 INTRODUCTION:
In sampling theory, the problem of mean estimation of a population is considered by many authors like Srivastava and Jhajj (1980, 81), Sahoo (1984, 1986), Singh (1986), Singh, Upadhayaya and Iachan (1987), Singh and Singh (1991), Singh et al. (1994), Sahoo et al. (1995), Sahoo and Sahoo (2001), Singh and Singh (2001), Sometimes, in survey situations a small part of sample remains non-responded (or incomplete) due to many practical reasons. Techniques and estimation procedures are needed to develop for this purpose. The imputation is a well defined methodology by virtue of which this kind of problem could be partially solved. Ahmed et al. (2006), Rao and Sitter (1995), Rubin (1976) and Singh and Horn (2000) have given applications of various imputation procedures. Hinde and Chambers (1990) studied the non-response imputation with multiple source of non-response. The problem of non-response in sample surveys immensely looked into by Hansen and Hurwitz (1946), Grover and Couper (1998), Jackway and Boyce (1987), Khare (1987), Khot (1994), Lessler and Kalsbeek (1992).

When the “response” and “non-response” part of the sample is assumed into two groups, it is closed to call upon as post-stratification. Estimation problem in sample survey, in the setup of post-stratification, under non-response situation is studied due to Shukla and Dubey (2001, 2004, and 2006). Some other useful contributions to this area are by Holt and Smith (1979), Jagers et al. (1985), Jagers (1986), Smith (1991), Agrawal and Panda (1993), Shukla and Trivedi (1999, 2001, 2006), Wywial (2001), Shukla et al. (2002, 2006), Shukla and Thakur (2008), Shukla et al. (2009a), Shukla et al. (2009b). When a sample is full of response over main variable but some of auxiliary values are missing, it is hard to utilize the usual estimators. Traditionally, it is essential to estimate those missing observations first by some specific estimation techniques. One can think of utilizing the non-sampled part of the population in order to get estimates of missing observations in the sample. These estimates could be imputed into actual estimation procedures used for the population mean. The content of this research work takes into account the similar aspect for non-responding values of the sample assuming post-stratified setup and utilizing the auxiliary source of data.

1.1 SYMBOLS AND SETUP:

Let $U = (U_1, U_2, \ldots\ldots, U_N)$ be a finite population of N units with Y as a main variable and X the auxiliary variable. The population has two types of individuals like N_1 as number of "respondents (R)" and N_2 "non-respondents (NR)", $(N = N_1+N_2)$. Their population proportions are expressed like $W_1 = N_1/N$ and $W_2 = N_2/N$. Quantities W_1 and W_2 could be guessed by past data or by experience of the investigator. Further, let $\bar{Y}$ and $\bar{X}$ be the population means of Y and X respectively. In what follows following are symbols:

R-group	:	Respondents group or group of those who responses in survey.
$\bar{Y}_1$	:	Population mean of R-group of Y.
$\bar{Y}_2$	:	Population mean of NR-group of Y.
$\bar{X}_1$	:	Population mean of R-group of X.
$\bar{X}_2$	:	Population mean of NR-group of X.
S_{1Y}^2	:	Population mean square of R-group of Y.
S_{2Y}^2	:	Population mean square of NR-group of Y.
S_{1X}^2	:	Population mean square of R-group of X.
S_{2X}^2	:	Population mean square of NR-group of X.
C_{1Y}	:	Coefficient of Variation of Y in R-group.
C_{2Y}	:	Coefficient of Variation of Y in NR-group.
C_{1X}	:	Coefficient of Variation of X in R-group.
C_{2X}	:	Coefficient of Variation of X in NR-group.
ρ	:	Correlation Coefficient in population between X and Y.
n	:	Sample size from population of size N by SRSWOR.
n_1	:	Post-stratified sample size coming from R-group.
n_2	:	Post-stratified sample size from NR-group.
$\bar{y}_1$	:	Sample mean of Y based on n_1 observations of R-group.
$\bar{y}_2$	:	Sample mean of Y based on n_2 observations of NR-group.
$\bar{x}_1$	:	Sample mean of X based on n_1 observations of R-group.
$\bar{x}_2$	:	Sample mean of X based on n_2 observations of NR-group.
ρ_1	:	Correlation Coefficient of population of R-group.
ρ_2	:	Correlation Coefficient of population of NR-group.

Further, consider few more symbolic representations:

$$D_1 = E\left(\frac{1}{n_1}\right) = \left[\frac{1}{nW_1} + \frac{(N-n)(1-W_1)}{(N-1)n^2W_1^2}\right]; \quad D_2 = E\left(\frac{1}{n_2}\right) = \left[\frac{1}{nW_2} + \frac{(N-n)(1-W_2)}{(N-1)n^2W_2^2}\right]$$

$$\bar{Y} = \frac{N_1\bar{Y}_1 + N_2\bar{Y}_2}{N}; \quad \bar{X} = \frac{N_1\bar{X}_1 + N_2\bar{X}_2}{N}$$

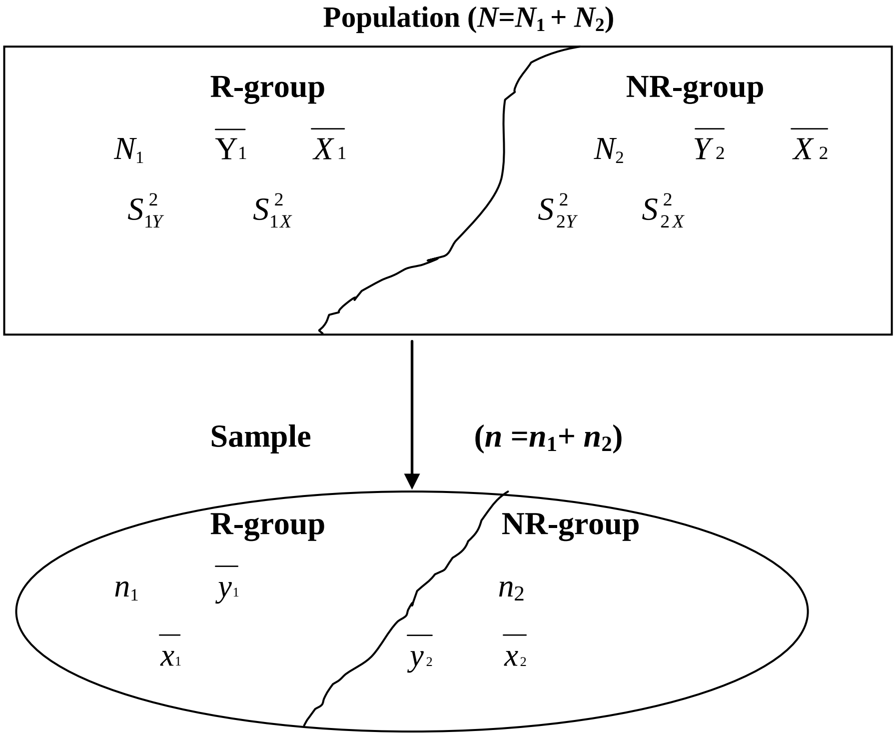

Fig : 1.1

2.0 ASSUMPTIONS :

Consider following in light of figure 1.1 before formulating an imputation based estimation procedure:

1. The sample of size n is drawn by SRSWOR and post-stratified into two groups of size n_1 and n_2 (n_1 + n_2 = n) according to R and NR group respectively
2. The information about Y variable in sample is completely available.
3. The sample means of both groups $\bar{y}_1$ and $\bar{y}_2$ are known such that

   $$\bar{y} = \frac{n_1\bar{y}_1 + n_2\bar{y}_2}{n}$$ which is sample mean on n units.
4. The population means $\overline{X}_1$ and $\overline{X}$ are known.
5. The population size N and sample size n are known. Also, N_1 and N_2 are known by past data, past experience or by guess of the investigator ($N_1 + N_2 = N$).
6. The sample mean of auxiliary information $\bar{x}_1$ is only known for R-Group, but information about $\bar{x}_2$ of NR-group is missing. Therefore

   $$\bar{x} = \frac{n_1\bar{x}_1 + n_2\bar{x}_2}{n}$$ could not be obtained due to absence of $\bar{x}_2$.
7. Other population parameters are assumed known, in either exact or in ratio from except the $\overline{Y}$, $\overline{Y}_1$ and $\overline{Y}_2$.

3.0 PROPOSED CLASS OF ESTIMATION STRATEGY :

To estimate population mean $\overline{Y}$, in setup of fig. 1.1, a problem to face is of missing observations related to $\bar{x}_2$, therefore, usual ratio, product and regression estimators are not applicable. Singh and Shukla (1987) have proposed a factor type estimator for estimating population mean $\overline{Y}$. Shukla et al. (1991), Singh and Shukla (1993), Shukla (2002) have also discussed properties of factor-type estimators applicable for estimating population mean. But all these cannot be useful due to unknown information $\bar{x}_2$. In order to solve this, an imputation $\left(\bar{x}_2^*\right)_5$ is adopt as define,

$$\left(\bar{x}_2^*\right)_5 = \left[\frac{N\bar{X} - n\left\{f\bar{X}_1 + (1-f)\bar{x}_2^*\right\}}{N-n}\right] \qquad ...(3.1)$$

where, $\bar{x}_2^* = \left[\frac{N\bar{X} - n\bar{x}_1}{N-n}\right]$.

The logic for this imputation is to utilize the non-sampled part of the population of X for obtaining an estimate of missing $\bar{x}_2$ and generate $\bar{x}^{(5)}$ for $\bar{x}$ as describe below :

$$\bar{x}^{(5)} = \left[\frac{N_1\bar{x}_1 + N_2\left(\bar{x}_2^*\right)_5}{N_1 + N_2}\right] \qquad ...(3.2)$$

The proposed imputation based class of factor-type estimator is

$$[(\bar{y}_{FT})_E]_k = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left[\frac{(A+C)\bar{X} + fB\bar{x}^{(5)}}{(A+fB)\bar{X} + C\bar{x}^{(5)}}\right] \qquad ...(3.3)$$

where, $0 < k < \infty$ and k is a constant and

$A = (k-1)(k-2)$; $B = (k-1)(k-4)$; $C = (k-2)(k-3)(k-4)$; $f = n/N$.

4.0 LARGE SAMPLE APPROXIMATION :

Consider the following for large n:

$$\left.\begin{aligned} \bar{y}_1 &= \bar{Y}_1(1+e_1) \\ \bar{y}_2 &= \bar{Y}_2(1+e_2) \\ \bar{x}_1 &= \bar{X}_1(1+e_3) \\ \bar{x}_2 &= \bar{X}_2(1+e_4) \end{aligned}\right\} \qquad ...(4.1)$$

where, e_1, e_2, e_3 and e_4 are very small numbers and $|e_i| < 1 \quad (i = 1,2,3,4)$.

Using the basic concept of SRSWOR and the concept of post-stratification of the sample n into n_1 and n_2 [see Cochran (2005), Hansen et al. (1993), Sukhatme et al. (1984), Singh and Choudhary (1986), Murthy (1976)], we get

$$\left.\begin{aligned} E(e_1) &= E\left[E(e_1)\ |n_1\right] = 0 \\ E(e_2) &= E\left[E(e_2)\ |n_2\right] = 0 \\ E(e_3) &= E\left[E(e_3)\ |n_1\right] = 0 \\ E(e_4) &= E\left[E(e_4)\ |n_2\right] = 0 \end{aligned}\right\} \qquad ...(4.2)$$

Assume the independence of R-group and NR-group representation in the sample, the following expression could be obtained:

$$\begin{aligned} E\left[e_1^2\right] &= E\left[E\left(e_1^2\right)\ |n_1\right] \\ &= E\left[\left\{\left(\frac{1}{n_1} - \frac{1}{N}\right)C_{1Y}^2\right\}\Big| n_1\right] \\ &= \left[\left\{E\left(\frac{1}{n_1}\right) - \frac{1}{N}\right\}C_{1Y}^2\right] \\ &= \left[\left(D_1 - \frac{1}{N}\right)C_{1Y}^2\right] \end{aligned} \qquad ...(4.3)$$

$$\mathrm{E}\left[e_2^2\right]=\mathrm{E}\left[E\left(e_2^2\right)|n_2\right]$$

$$=\mathrm{E}\left[\left\{\left(\frac{1}{n_2}-\frac{1}{N}\right)C_{2Y}^2\right\}\Big|n_2\right]$$

$$=\left[\left(D_2-\frac{1}{N}\right)C_{2Y}^2\right] \qquad ...(4.4)$$

$$\mathrm{E}\left[e_3^2\right]=\mathrm{E}\left[\mathrm{E}\left(e_3^2\right)|n_1\right]$$

$$=\left[\left(D_1-\frac{1}{N}\right)C_{1X}^2\right] \qquad ...(4.5)$$

and $$\mathrm{E}\left[e_4^2\right]=\left[\left(D_2-\frac{1}{N}\right)C_{2X}^2\right] \qquad ...(4.5.1)$$

$$\mathrm{E}\left[e_1e_3\right]=\mathrm{E}\left[\mathrm{E}\left(e_1e_3\right)|n_1\right]$$

$$=\mathrm{E}\left[\left\{\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right\}\Big|n_1\right]$$

$$=\left[\left(D_1-\frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right] \qquad ...(4.6)$$

$$\mathrm{E}\left[e_1e_2\right]=\mathrm{E}\left[\mathrm{E}\left(e_1e_2\right)|n_1,n_2\right]=0 \qquad ...(4.7)$$

$$\mathrm{E}\left[e_1e_4\right]=0 \qquad ...(4.7.1)$$

$$\mathrm{E}\left[e_2e_3\right]=\mathrm{E}\left[\mathrm{E}\left(e_2e_3\right)|n_1,n_2\right]=0 \qquad ...(4.8)$$

$$\mathrm{E}\left[e_2e_4\right]=\left(D_2-\frac{1}{N}\right)\rho_2C_{2Y}C_{2X} \qquad ...(4.8.1)$$

$$\mathrm{E}\left[e_3e_4\right]=0 \qquad ...(4.8.2)$$

The expression (4.7), (4.7.1), (4.8) and (4.8.2) are true under the assumption of independent representation of R-group and NR-group units in the sample. This is introduced to simplify mathematical expressions.

THEOREM 4.1: The estimator $[(\bar{y}_{FT})_E]_k$ could be expressed under large sample approximation in following form :

$$[(\bar{y}_{FT})_E]_k=\delta_4\bar{Y}\,[1+s_1W_1e_1+s_2W_2e_2][1+(\alpha_4-\beta_4)e_3-(\alpha_4-\beta_4)\beta_4e_3^2+(\alpha_4-\beta_4)\,\beta_4^2\,e_3^3-\ldots.]$$

PROOF : Rewrite $(\bar{x}_2^*)_5$ as in (3.1):

$$(\bar{x}_2^*)_5=\frac{N\bar{X}-n\{f\bar{X}_1+(1-f)\bar{x}_2^*\}}{N-n}$$

where $$\bar{x}_2^*=\frac{N\bar{X}-n\bar{x}_1}{N-n}$$

$$(\bar{x}_2^*)_5=\frac{1}{N-n}\left[N\bar{X}-n\left\{f\bar{X}_1+(1-f)\left(\frac{N\bar{X}-n\bar{x}_1}{N-n}\right)\right\}\right]$$

$$=\left[N(N-n)\bar{X}-n(N-n)f\bar{X}_1-n(1-f)\left\{N\bar{X}-n\bar{x}_1\right\}\right]/(N-n)^2$$

$$=\left[\left\{N(N-n)-Nn(1-f)f\right\}\bar{X}-n(N-n)f\,\bar{X}_1+n^2(1-f)\bar{X}_1+n^2(1-f)\bar{X}_1e_3\right](N-n)^{-2}$$

Here,

(i) $N(N-n) - Nn(1-f) = N(N-n) - Nn\left(1-\frac{n}{N}\right) = (N-n)(N-n)$

(ii) $n(N-n)f + n^2(1-f) = \frac{n^2}{N}(N-n) + n^2\left(1-\frac{n}{N}\right) = \frac{2n^2(N-n)}{N} = 2nf(N-n)$

(iii) $n^2(1-f) = nf(N-n)$

Therefore, $$(\overline{x_2^*})_5 = \frac{(N-n)\overline{X} - 2nf\,\overline{X_1} + nf\,\overline{X_1}e_3}{N-n} \quad ...(4.9)$$

We have, from (3.2)

$$\bar{x}^{(5)} = \frac{N_1\overline{x_1} + N_2(\overline{x_2^*})_5}{N}$$

by putting the value of $(\overline{x_2^*})_5$ from (4.9) and solving

$$\bar{x}^{(5)} = \left[\frac{(N-n)N_1\overline{X}_1(1+e_3) + N_2[(N-n)\overline{X} - 2nf\,\overline{X}_1 + nf\,\overline{X_1}e_3]}{N(N-n)}\right]$$

$$= W_1\overline{X_1}(1+e_3) + \left[\frac{N_2}{N(N-n)}\left\{(N-n)\overline{X} - 2nf\,\overline{X_1} + nf\,\overline{X_1}e_3\right\}\right]$$

$$= W_1\overline{X}_1 + W_1\overline{X}_1e_3 + [W_2\overline{X} - 2pf^2\,\overline{X_1} + pf^2\,\overline{X_1}e_3]$$

$$= W_2\overline{X} + (W_1 - 2pf^2)\overline{X_1} + (W_1 + pf^2)\overline{X_1}e_3$$

$$= \overline{X}\left[\{W_2 + (W_1 - 2pf^2)r_1\} + (W_1 + pf^2)r_1e_3\right]$$

$$\bar{x}^{(5)} = \overline{X}[t + ue_3] \quad ...(4.10)$$

where, $t = W_2 + (W_1 - 2pf^2)r_1$; $u = (W_1 + pf^2)r_1$; $p = \frac{N_2}{N-n}$.

Now, the estimator $[(\bar{y}_{FT})_E]_k$ under approximation and using (4.10) will be

$$[(\bar{y}_{FT})_E]_k = \left(\frac{N_1\overline{y_1} + N_2\overline{y_2}}{N}\right)\left[\frac{(A+C)\overline{X} + fB\bar{x}^{(5)}}{(A+fB)\overline{X} + C\bar{x}^{(5)}}\right]$$

$$= \left[\frac{N_1\overline{Y}_1(1+e_1) + N_2\overline{Y}_2(1+e_2)}{N}\right]\left[\frac{(A+C)\overline{X} + fB(t+ue_3)\overline{X}}{(A+fB)\overline{X} + C(t+ue_3)\overline{X}}\right]$$

$$= \overline{Y}[1 + s_1W_1e_1 + s_2W_2e_2]\left[\frac{(A+fBt+C) + fBue_3}{(A+fB+Ct) + Cue_3}\right]$$

$$= \overline{Y}[1 + s_1W_1e_1 + s_2W_2e_2]\left[\frac{\mu_1 + \mu_2e_3}{\mu_3 + \mu_4e_3}\right]$$

$$= \delta_4\overline{Y}[1 + s_1W_1e_1 + s_2W_2e_2](1 + \alpha_4e_3)(1 + \beta_4e_3)^{-1}$$

where $\mu_1 = A + fBt + C$; $\mu_2 = fBu$; $\mu_3 = A + fB + Ct$; $\mu_4 = Cu$;

$$r_1 = \frac{\overline{X_1}}{\overline{\overline{X}}};\ r_2 = \frac{\overline{X_2}}{\overline{\overline{X}}};\ s_1 = \frac{\overline{Y_1}}{\overline{\overline{Y}}};\ s_2 = \frac{\overline{Y_2}}{\overline{\overline{Y}}};\ \alpha_4 = \frac{\mu_2}{\mu_1};\ \beta_4 = \frac{\mu_4}{\mu_3};\ \delta_4 = \frac{\mu_1}{\mu_3};$$

We can further express the above into following:

$[(\bar{y}_{FT})_E]_k = \delta_4 \bar{Y} [1 + s_1 W_1 e_1 + s_2 W_2 e_2](1 + \alpha_4 e_3)(1 - \beta_4 e_3 + \beta_4^2 e_3^2 - \ldots\ldots)$

$= \delta_4 \bar{Y} [1 + s_1 W_1 e_1 + s_2 W_2 e_2][(1 - \beta_4 e_3 + \beta_4^2 e_3^2 - \ldots) + (\alpha_4 e_3 - \alpha_4 \beta_4 e_3^2 + \alpha_4 \beta_4^2 e_3^3 \ldots\ldots)]$

$[(\bar{y}_{FT})_E]_k = \delta_4 \bar{Y} [1 + s_1 W_1 e_1 + s_2 W_2 e_2] [1 + (\alpha_4 - \beta_4) e_3 - (\alpha_4 - \beta_4) \beta_4 e_3^2 + \ldots]$

5.0 BIAS AND MEAN SQUARED ERROR :

Define E(.) for expectation, B(.) for bias and M(.) for mean squared error, then the first order of approximations could be established; for $i, j = 1, 2, 3, \ldots$ as

$$\left.\begin{aligned} E\left[e_1^i e_2^j\right] &= 0 \quad \text{when } i+j > 2 \\ E\left[e_1^i e_3^j\right] &= 0 \quad \text{when } i+j > 2 \\ E\left[e_2^i e_3^j\right] &= 0 \quad \text{when } i+j > 2 \end{aligned}\right\} \quad \ldots.(5.1)$$

THEOREM 5.1 : The $[(\bar{y}_{FT})_E]_k$ is a biased estimator of $\bar{Y}$ with the amount of bias to the first order of approximation:

$$B[(\bar{y}_{FT})_E]_k = \bar{Y}\left[(\delta_4 - 1) - \delta_4 C_{1X}(\alpha_4 - \beta_4)\left(D_1 - \frac{1}{N}\right)\{\beta_4 C_{1X} - s_1 W_1 \rho_1 C_{1Y}\}\right]$$

PROOF : $B[(\bar{y}_{FT})_E]_k = E[\{(\bar{y}_{FT})_E\}_k - \bar{Y}]$

Using theorem 4.1 and taking expectations

$$E[(\bar{y}_{FT})_E]_k = \delta_4 \bar{Y} E[1 + (\alpha_4 - \beta_4)e_3 - (\alpha_4 - \beta_4)\beta_4 e_3^2 + s_1 W_1 e_1\{1 + (\alpha_4 - \beta_4)e_3 - (\alpha_4 - \beta_4)\beta_4 e_3^2\}$$
$$+ s_2 W_2 e_2 \{1 + (\alpha_4 - \beta_4)e_3 - (\alpha_4 - \beta_4)\beta_4 e_3^2\}]$$

$$= \delta_4 \bar{Y} [1 - (\alpha_4 - \beta_4)\beta_4 E(e_3^2) + s_1 W_1 (\alpha_4 - \beta_4) E(e_1 e_3) + s_2 W_2 (\alpha_4 - \beta_4) E(e_2 e_3)]$$

$$= \delta_4 \bar{Y}\left[1 - (\alpha_4 - \beta_4)\beta_4\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + (\alpha_4 - \beta_4)s_1 W_1\left(D_1 - \frac{1}{N}\right)\rho_1 C_{1Y} C_{1X}\right]$$

$$= \delta_4 \bar{Y}\left[1 - (\alpha_4 - \beta_4)\left(D_1 - \frac{1}{N}\right)C_{1X}\{\beta_4 C_{1X} - s_1 W_1 \rho_1 C_{1Y}\}\right]$$

Therefore,

$$B[(\bar{y}_{FT})_E]_k = E[\{(\bar{y}_{FT})_E\}_k - \bar{Y}]$$
$$= \bar{Y}\left[(\delta_4 - 1) - \delta_4 C_{1X}(\alpha_4 - \beta_4)\left(D_1 - \frac{1}{N}\right)\{\beta_4 C_{1X} - s_1 W_1 \rho_1 C_{1Y}\}\right]$$

THEOREM 5.2 : The mean squared error of $[(\bar{y}_{FT})_E]_k$ is

$M[(\bar{y}_{FT})_E]_k =$

$$\bar{Y}^2\left[(\delta_4 - 1)^2 + \left(D_1 - \frac{1}{N}\right)\{I_1 s_1^2 C_{1Y}^2 + I_2 C_{1X}^2 + 2 I_3 s_1 \rho_1 C_{1Y} C_{1X}\} + \left(D_2 - \frac{1}{N}\right)\delta_4^2 s_2^2 W_2^2 C_{2Y}^2\right]$$

where $I_1 = \delta_4^2 W_1^2$; $I_2 = \delta_4(\alpha_4 - \beta_4)\{\delta_4(\alpha_4 - \beta_4) - 2(\delta_4 - 1)\beta_4\}$;

$I_3 = W_1 \delta_4 (2\delta_4 - 1)(\alpha_4 - \beta_4)$.

PROOF: $M[(\bar{y}_{FT})_E]_k = E[\{(\bar{y}_{FT})_E\}_k - \bar{Y}]^2$

Using theorem 4.1, we can express

$$M[(\bar{y}_{FT})_E]_k = E[\delta_4\bar{Y}\{1 + s_1W_1e_1 + s_2W_2e_2\}\{1 + (\alpha_4 - \beta_4)e_3 - (\alpha_4 - \beta_4)\beta_4 e_3^2 + (\alpha_3 - \beta_4)\beta_4 e_3^3 \ldots\} - \bar{Y}]^2$$

Using large sample approximations of (5.1) we could express

$$= \bar{Y}^2 E[(\delta_4 - 1) + \delta_4\{(\alpha_4 - \beta_4)e_3 - (\alpha_4 - \beta_4)\beta_4 e_3^2 + (s_1W_1e_1 + s_2W_2e_2) + (\alpha_4 - \beta_4)(s_1W_1e_1 + s_2W_2e_2)e_3\}]^2$$

$$= \bar{Y}^2 E[(\delta_4 - 1)^2 + \delta_4^2\{(\alpha_4 - \beta_4)^2 e_3^2 + s_1^2W_1^2e_1^2 + s_2^2W_2^2e_2^2 + 2s_1s_2W_1W_2e_1e_2 + 2(\alpha_4 - \beta_4)(s_1W_1e_1e_3 + s_2W_2e_2e_3)\} + 2\delta_4(\delta_4 - 1)\{(\alpha_4 - \beta_4)e_3 - (\alpha_4 - \beta_4)\beta_4 e_3^2 + (s_1W_1e_1 + s_2W_2e_2) + (\alpha_4 - \beta_4)(s_1W_1e_1e_3 + s_2W_2e_2e_3)\}]$$

Using (4.2), (4.7) and (4.8) we rewrite,

$$= \bar{Y}^2[(\delta_4 - 1)^2 + \delta_4^2\{(\alpha_4 - \beta_4)^2 E(e_3^2) + s_1^2W_1^2E(e_1^2) + s_2^2W_2^2E(e_2^2) + 2(\alpha_4 - \beta_4)s_1W_1E(e_1e_3)\} + 2\delta_4(\delta_4 - 1)\{-(\alpha_4 - \beta_4)\beta_4 E(e_3^2) + (\alpha_4 - \beta_4)s_1W_1E(e_1e_3)\}]$$

$$= \bar{Y}^2\left[(\delta_4 - 1)^2 + \delta_4^2\left\{(\alpha_4 - \beta_4)^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + s_1^2W_1^2\left(D_1 - \frac{1}{N}\right)C_{1Y}^2 + s_2^2W_2^2\left(D_2 - \frac{1}{N}\right)C_{2Y}^2 + 2(\alpha_4 - \beta_4)s_1W_1\left(D_1 - \frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right\} + 2\delta_4(\delta_4 - 1)\left\{-(\alpha_4 - \beta_4)\beta_4\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + (\alpha_4 - \beta_4)s_1W_1\left(D_1 - \frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right\}\right]$$

$$= \bar{Y}^2\left[(\delta_4 - 1)^2 + \left(D_1 - \frac{1}{N}\right)\{\delta_4^2s_1^2W_1^2C_{1Y}^2 + \delta_4(\alpha_4 - \beta_4)\{\delta_4(\alpha_4 - \beta_4) - 2(\delta_4 - 1)\beta_4\}C_{1X}^2 + 2\delta_4(2\delta_4 - 1)(\alpha_4 - \beta_4)s_1W_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right)\delta_4^2s_2^2W_2^2C_{2Y}^2\right]$$

$$= \bar{Y}^2\left[(\delta_4 - 1)^2 + \left(D_1 - \frac{1}{N}\right)\{I_1s_1^2C_{1Y}^2 + I_2C_{1X}^2 + 2I_3s_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right)\delta_4^2s_2^2W_2^2C_{2Y}^2\right]$$

6.0 SOME SPECIAL CASES :

The term A, B and C are functions of k. In particular, there are some special cases:

CASE I : When $k = 1$

$A = 0$; $B = 0$; $C = -6$; $\mu_1 = -6$; $\mu_2 = 0$; $\mu_3 = -6t$; $\mu_4 = -6u$; $\alpha_4 = 0$; $\beta_4 = \frac{u}{t}$; $\delta_4 = \frac{1}{t}$;

$I_1 = \frac{W_1^2}{t^2}$; $I_2 = \frac{u^2(3-2t)}{t^4}$; $I_3 = \frac{uW_1(t-2)}{t^3}$;

The estimator $[(\bar{y}_{FT})_E]_k$ along with bias and m.s.e. under case I is:

$$[(\bar{y}_{FT})_E]_{k=1} = \left[\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right]\left[\frac{\bar{X}}{\bar{x}^{(5)}}\right] \quad ...(6.1)$$

$$B\,[(\bar{y}_{FT})_E]_{k=1} = \bar{Y}\, t^{-3}\,[(1-t)t^2 + \left(D_1 - \frac{1}{N}\right) uC_{1X}\{uC_{1X} - ts_1W_1\rho_1C_{1Y}\}] \quad ...(6.2)$$

$$M\,[(\bar{y}_{FT})_E]_{k=1} = \bar{Y}^2\, t^{-4}[(1-t)^2t^2 + \left(D_1 - \frac{1}{N}\right)\{t^2W_1^2s_1^2C_{1Y}^2 + u^2(3-2t)C_{1X}^2 + 2t(t-2)uW_1s_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right) t^2W_2^2\, s_2^2\, C_{2Y}^2] \quad ...(6.3)$$

CASE II: When $k = 2$

$A = 0;\ B = -2;\ C = 0;\ \mu_1 = -2ft;\ \mu_2 = -2fu;\ \mu_3 = -2f;\ \mu_4 = 0;$

$\alpha_4 = ut^{-1};\ \beta_4 = 0;\ \delta_4 = t;\ I_1 = W_1^2t^2;\ I_2 = u^2;\ I_3 = uW_1(2t-1)$

$$[(\bar{y}_{FT})_E]_{k=2} = \left[\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right]\left[\frac{\bar{x}^{(5)}}{\bar{X}}\right] \quad ...(6.4)$$

$$B\,[(\bar{y}_{FT})_E]_{k=2} = \bar{Y}\left[(t-1) + \left(D_1 - \frac{1}{N}\right)us_1W_1\rho_1C_{1Y}C_{1X}\right] \quad ...(6.5)$$

$$M\,[(\bar{y}_{FT})_E]_{k=2} = \bar{Y}^2\,[(t-1)^2 + \left(D_1 - \frac{1}{N}\right)\{t^2W_1^2\, s_1^2\, C_{1Y}^2 + u^2C_{1X}^2 + 2W_1(2t-1)\,u\, s_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right) t^2W_2^2\, s_2^2\, C_{2Y}^2] \quad ...(6.6)$$

CASE III: When $k = 3$

$A = 2;\ B = -2;\ C = 0;\ \mu_1 = 2(1-ft);\ \mu_2 = -2fu;\ \mu_3 = 2(1-f);\ \mu_4 = 0;$

$\alpha_4 = \dfrac{-fu}{1-ft};\ \beta_4 = 0;\ \delta_4 = \dfrac{1-ft}{1-f};\ I_1 = \dfrac{(1-ft)^2W_1^2}{(1-f)^2};\ I_2 = \dfrac{u^2f^2}{(1-f)^2};\ I_3 = \dfrac{fuW_1\{2ft - f - 1\}}{(1-f)^2};$

$$[(\bar{y}_{FT})_E]_{k=3} = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left[\frac{\bar{X} + f\bar{x}^{(5)}}{(1-f)\bar{X}}\right] \quad ...(6.7)$$

$$B\,[(\bar{y}_{FT})_E]_{k=3} = \bar{Y}f(1-f)^{-1}\left[(1-t) - \left(D_1 - \frac{1}{N}\right)uW_1s_1\rho_1C_{1Y}C_{1X}\right] \quad ...(6.8)$$

$$M\,[(\bar{y}_{FT})_E]_{k=3} = \bar{Y}^2(1-f)^{-2}\,[f^2(1-t)^2 + \left(D_1 - \frac{1}{N}\right)\{(1-ft)^2\, s_1^2W_1^2C_{1Y}^2 + u^2f^2C_{1X}^2 + 2(2ft - f - 1)\}fuW_1s_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right)(1-ft)^2W_2^2\, s_2^2\, C_{2Y}^2] \quad ...(6.9)$$

CASE IV: When $k = 4$;

$A = 6;\ B = 0;\ C = 0;\ \mu_1 = 6;\ \mu_2 = 0;\ \mu_3 = 6;\ \mu_4 = 0;\ \alpha_4 = 0;\ \beta_4 = 0;\ \delta_4 = 1;\ I_1 = W_1^2;\ I_2 = 0;\ I_3 = 0;$

$$[(\bar{y}_{FT})_E]_{k=4} = \left[\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right] \quad ...(6.10)$$

$$B\,[(\bar{y}_{FT})_E]_{k=4} = 0 \quad ...(6.11)$$

$$V\,[(\bar{y}_{FT})_E]_{k=4} = \left(D_1 - \frac{1}{N}\right).W_1^2\bar{Y}_1^2C_{1Y}^2 + \left(D_2 - \frac{1}{N}\right).W_2^2\bar{Y}_2^2C_{2Y}^2 \quad ...(6.12)$$

7.0 ESTIMATOR WITHOUT IMPUTATION :

Throughout the discussion, the assumption is unknown value of $\bar{x}_2$. This is imputed by the term $(\bar{x}_2^*)_5$, to provide the generation of $\bar{x}^{(5)}$. [See eq.(3.1) and (3.2)]. Suppose the $\bar{x}_2$ is known, then there is no need of imputation and the proposed (3.2) and (3.3) reduces into :

$$\bar{x}^{(*)} = \left(\frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N}\right) \quad ...(7.1)$$

$$[(\bar{y}_{FT})_w]_k = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left(\frac{(A+C)\bar{X} + fB\bar{x}^{(*)}}{(A+fB)\bar{X} + C\bar{x}^{(*)}}\right) \quad ...(7.2)$$

where, k is a constant , $0 < k < \infty$ and
$A = (k-1)(k-2)$; $B = (k-1)(k-4)$; $C = (k-2)(k-3)(k-4)$; $f = n/N$.

THEOREM 7.1: The estimator $[(\bar{y}_{FT})_w]_k$ is biased for $\bar{Y}$ with the amount of bias

$$B[(\bar{y}_{FT})_w]_k = (\mu_1' - \mu_2')\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\{s_1\rho_1 C_{1Y} - \mu_2' r_1 C_{1X}\} + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\{s_2\rho_2 C_{2Y} - \mu_2' r_2 C_{2X}\}\right]$$

where,

$$\mu_1' = fB/(A+fB+C); \quad \mu_2' = C/(A+fB+C)$$

PROOF: The estimator $[(\bar{y}_{FT})_w]_k$ could be approximate like :

$$[(\bar{y}_{FT})_w]_k = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left(\frac{(A+C)\bar{X} + fB\bar{x}^{(*)}}{(A+fB)\bar{X} + C\bar{x}^{(*)}}\right)$$

$$= \left[\frac{N_1\bar{Y}_1(1+e_1) + N_2\bar{Y}_2(1+e_2)}{N}\right]\left[\frac{N(A+C)\bar{X} + fB\{N_1\bar{X}_1(1+e_3) + N_2\bar{X}_2(1+e_4)\}}{N(A+fB)\bar{X} + C\{N_1\bar{X}_1(1+e_3) + N_2\bar{X}_2(1+e_4)\}}\right]$$

$$= [\bar{Y} + W_1\bar{Y}_1e_1 + W_2\bar{Y}_2e_2]\left[\frac{(A+fB+C) + fB(W_1r_1e_3 + W_2r_2e_4)}{(A+fB+C) + C(W_1r_1e_3 + W_2r_2e_4)}\right]$$

$$= [\bar{Y} + W_1\bar{Y}_1e_1 + W_2\bar{Y}_2e_2][1 + \mu_1'(W_1r_1e_3 + W_2r_2e_4)]\,[1 + \mu_2'(W_1r_1e_3 + W_2r_2e_4)^{-1}]$$

Expanding above using binominal expansion, and ignoring $(e_i^k e_j^l)$ terms for $(k + l) > 2$, $(k, l = 0, 1, 2 ....)$, $(i, j = 1, 2, 3, 4)$; the estimator results into

$$[(\bar{y}_{FT})_w]_k = \bar{Y} + \bar{Y}(\Delta_1 - \Delta_2) + W_1\bar{Y}_1e_1(1 + \Delta_1 - \Delta_2) + W_2\bar{Y}_2e_2(1 + \Delta_1 - \Delta_2) \quad ...(7.3)$$

where. $\Delta_1 = (\mu_1' - \mu_2')\,(W_1r_1e_3 + W_2r_2e_4)$; $\Delta_2 = \mu_2'(\mu_1' - \mu_2')\,(W_1r_1e_3 + W_2r_2e_4)^2$

and $W_1 r_1 + W_2 r_2 = 1$ holds.

Further, one can derive up to first order of approximation, the following

(i) $E[\Delta_1] = 0$

(ii) $E[\Delta_1^2] = (\mu_1' - \mu_2')^2\left[W_1^2r_1^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + W_2^2r_2^2\left(D_2 - \frac{1}{N}\right)C_{2X}^2\right]$

(iii) $\mathrm{E}\left[\Delta_2\right] = \mu_2'\left(\mu_1' - \mu_2'\right)\left[W_1^2 r_1^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + W_2^2 r_2^2\left(D_2 - \frac{1}{N}\right)C_{2X}^2\right]$

(iv) $\mathrm{E}\left[e_1\,\Delta_1\right] = \left(\mu_1' - \mu_2'\right) W_1\, r_1\left(D_1 - \frac{1}{N}\right)\rho_1 C_{1Y} C_{1X}$

(v) $\mathrm{E}\left[e_2\,\Delta_1\right] = \left(\mu_1' - \mu_2'\right) W_2\, r_2\left(D_2 - \frac{1}{N}\right)\rho_2 C_{2Y} C_{2X}$

(vi) $\mathrm{E}\left[e_1\,\Delta_2\right] = 0 \quad \left[\text{under } o\left(n^{-1}\right)\right]$

(vii) $\mathrm{E}\left[e_2\,\Delta_2\right] = 0 \quad \left[\text{under } o\left(n^{-1}\right)\right]$

The bias of estimator without imputation is

$$\begin{aligned}
\mathrm{B}\left[(\bar{y}_{FT})_w\right]_k &= \mathrm{E}\left[\left\{(\bar{y}_{FT})_w\right\}_k - \bar{Y}\right] \\
&= \mathrm{E}\left[\bar{Y}(\Delta_1 - \Delta_2) + W_1\bar{Y}_1 e_1(1+\Delta_1-\Delta_2) + W_2\bar{Y}_2 e_2(1+\Delta_1-\Delta_2)\right] \\
&= \left[W_1\bar{Y}_1\mathrm{E}(e_1\Delta_1) + W_2\bar{Y}_2\mathrm{E}(e_2\Delta_1) - \bar{Y}\mathrm{E}(\Delta_2)\right] \\
&= \left(\mu_1' - \mu_2'\right)\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\left\{\bar{Y}_1\rho_1 C_{1Y} - \bar{Y}\mu_2' r_1 C_{1X}\right\}\right. \\
&\qquad \left. + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\left\{\bar{Y}_2\rho_2 C_{2Y} - \bar{Y}\mu_2' r_2 C_{2X}\right\}\right] \\
&= \bar{Y}\left(\mu_1' - \mu_2'\right)\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\left\{s_1\rho_1 C_{1Y} - \mu_2' r_1 C_{1X}\right\}\right. \\
&\qquad \left. + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\left\{s_2\rho_2 C_{2Y} - \mu_2' r_2 C_{2X}\right\}\right]
\end{aligned}$$

THEOREM 7.2 : The mean squared error of the estimator $\left[(\bar{y}_{FT})_w\right]_k$ is

$$\begin{aligned}
\mathrm{M}\left[(\bar{y}_{FT})_w\right]_k = \bar{Y}^2\left[\left(D_1 - \frac{1}{N}\right)W_1^2\left\{T_1^2 C_{1Y}^2 + T_2^2 C_{1X}^2 + 2T_1T_2\,\rho_1 C_{1Y} C_{1X}\right\}\right. \\
\left. + \left(D_2 - \frac{1}{N}\right)W_2^2\left\{S_1^2 C_{2Y}^2 + S_2^2 C_{2X}^2 + 2S_1S_2\rho_2 C_{2Y} C_{2X}\right\}\right]
\end{aligned}$$

where $T_1 = s_1$; $T_2 = \left(\mu_1' - \mu_2'\right)r_1$; $S_1 = s_2$; $S_2 = \left(\mu_1' - \mu_2'\right)r_2$;

PROOF :

$$\begin{aligned}
\mathrm{M}\left[(\bar{y}_{FT})_w\right]_k &= \mathrm{E}\left[\left\{(y_{FT})_w\right\}_k - \bar{Y}\right]^2 \\
&= \mathrm{E}\left[\bar{Y}(\Delta_1 - \Delta_2) + W_1\bar{Y}_1 e_1(1+\Delta_1-\Delta_2) + W_2\bar{Y}_2 e_2(1+\Delta_1-\Delta_2)\right]^2 \\
&= \bar{Y}^2\,\mathrm{E}\left(\Delta_1^2\right) + W_1^2\bar{Y}_1^2\,\mathrm{E}\left(e_1^2\right) + W_2^2\bar{Y}_2^2\mathrm{E}\left(e_2^2\right) + 2W_1\,\bar{Y}\bar{Y} \\
&\qquad + 2W_2\,\bar{Y}\bar{Y}_2\,\mathrm{E}\left(e_2\Delta_1\right) + 2W_1\,W_2\bar{Y}_1\bar{Y}_2\,\mathrm{E}\left(e_1 e_2\right) \\
&= \bar{Y}^2\left[\left(\mu_1' - \mu_2'\right)^2\left\{W_1^2 r_1^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + W_2^2 r_2^2\left(D_2 - \frac{1}{N}\right)C_{2X}^2\right\}\right. \\
&\qquad + \left\{W_1^2 s_1^2\left(D_1 - \frac{1}{N}\right)C_{1Y}^2 + W_2^2 s_2^2\left(D_2 - \frac{1}{N}\right)C_{2Y}^2\right\}
\end{aligned}$$

$$+2W_1 s_1\left(\mu_1' - \mu_2'\right)\left\{W_1 r_1\left(D_1 - \frac{1}{N}\right)\rho_1 C_{1Y} C_{1X}\right\}$$

$$\left.+2W_2 s_2\left(\mu_1' - \mu_2'\right)\left\{W_2 r_2\left(D_2 - \frac{1}{N}\right)\rho_2 C_{2Y} C_{2X}\right\}\right]$$

$$= \bar{Y}^2\left[\left(D_1 - \frac{1}{N}\right)W_1^2\left\{T_1^2 C_{1Y}^2 + T_2^2 C_{1X}^2 + 2T_1 T_2\,\rho_1 C_{1Y} C_{1X}\right\}\right.$$

$$\left.+\left(D_2 - \frac{1}{N}\right)W_2^2\left\{S_1^2 C_{2Y}^2 + S_2^2 C_{2X}^2 + 2S_1 S_2 \rho_2 C_{2Y} C_{2X}\right\}\right]$$

REMARK 7.1 :

At k = 1, k = 2, k = 3 and k = 4, there are some special cases of non-imputed estimators with the respective bias and mean squared error as laid down below :

CASE I : When k = 1

$A = 0$; $B = 0$; $C = -6$; $\mu_1' = 0$; $\mu_2' = 1$; $T_1 = s_1$; $T_2 = -r_1$; $S_1 = s_1$; $S_2 = -r_2$;

$$\left[(\bar{y}_{FT})_w\right]_{k=1} = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left(\frac{\bar{X}}{\bar{x}^{(*)}}\right)$$

$$\mathrm{B}\left[(\bar{y}_{FT})_w\right]_{k=1} = -\bar{Y}\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\left\{s_1\rho_1 C_{1Y} - r_1 C_{1X}\right\} + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\left\{s_2\rho_2 C_{2Y} - r_2 C_{2X}\right\}\right]$$

$$\mathrm{M}\left[(\bar{y}_{FT})_w\right]_{k=1} = \bar{Y}^2\left[\left(D_1 - \frac{1}{N}\right)W_1^2\left\{s_1^2 C_{1Y}^2 + r_1^2 C_{1X}^2 - 2s_1 r_1\,\rho_1 C_{1Y} C_{1X}\right\}\right.$$

$$\left.+\left(D_2 - \frac{1}{N}\right)W_2^2\left\{s_2^2 C_{2Y}^2 + r_2^2 C_{2X}^2 - 2s_2 r_2\,\rho_2 C_{2Y} C_{2X}\right\}\right]$$

CASE II : When k = 2

$A = 0$; $B = -2$; $C = 0$; $\mu_1' = 1$; $\mu_2' = 0$; $T_1 = s_1$; $T_2 = r_1$; $S_1 = s_2$; $S_2 = r_2$

$$\left[(\bar{y}_{FT})_w\right]_{k=2} = \left(\frac{N_1\bar{y}_2 + N_2 y_2}{N}\right)\left(\frac{\bar{x}^{(*)}}{\bar{X}}\right)$$

$$\mathrm{B}\left[(\bar{y}_{FT})_w\right]_{k=2} = \bar{Y}\left[\left(D_1 - \frac{1}{N}\right)W_1^2 s_1 r_1 \rho_1 C_{1X} C_{1Y} + \left(D_2 - \frac{1}{N}\right)W_2^2 s_2 r_2 \rho_2 C_{2X} C_{2Y}\right]$$

$$\mathrm{M}\left[(\bar{y}_{FT})_w\right]_{k=2} = \bar{Y}^2\left[\left(D_1 - \frac{1}{N}\right)W_1^2\left\{s_1^2 C_{1Y}^2 + r_1^2 C_{1X}^2 + 2s_1 r_1\,\rho_1 C_{1Y} C_{1X}\right\}\right.$$

$$\left.+\left(D_2 - \frac{1}{N}\right)W_2^2\left\{s_2^2 C_{2Y}^2 + r_2^2 C_{2X}^2 + 2s_2 r_2\,\rho_2 C_{2Y} C_{2X}\right\}\right]$$

CASE III : When k = 3

$A = 2$; $B = -2$; $C = 0$; $\mu_1' = -f(1-f)^{-1}$; $\mu_2' = 0$;

$T_1 = s_1$; $T_2 = r_1 f(1-f)^{-1}$; $S_1 = s_2$; $S_2 = r_2 f(1-f)^{-1}$.

$$\left[(\bar{y}_{FT})_w\right]_{k=3} = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left(\frac{\bar{X} - \bar{x}^{(*)}}{(1-f)\bar{X}}\right)$$

$$B[(\bar{y}_{FT})_w]_{k=3} = -\bar{Y}f(1-f)^{-1}\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 s_1 \rho_1 C_{1Y} C_{1X} + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 s_2 \rho_2 C_{2Y} C_{2X}\right]$$

$$M[(\bar{y}_{FT})_w]_{k=3} = \bar{Y}^2\left[\left(D_1 - \frac{1}{N}\right)W_1^2\{s_1^2 C_{1Y}^2 + (1-f)^{-2} f^2 r_1^2 C_{1X}^2 - 2(1-f)^{-1} f s_1 r_1 \rho_1 C_{1Y} C_{1X}\} + \left\{\left(D_2 - \frac{1}{N}\right)W_2^2 s_2^2 C_{2Y}^2 + (1-f)^{-2} f^2 r_2^2 C_{2X}^2 - 2(1-f)^{-1} f s_2 r_2 \rho_2 C_{2Y} C_{2X}\right\}\right]$$

Case IV: When $k = 4$,

$A = 6$; $B = 0$; $C = 0$; $\mu_1' = 0$; $\mu_2' = 0$; $T_1 = s_1$; $T_2 = 0$; $S_1 = s_2$; $S_2 = 0$;

$$[(\bar{y}_{FT})_w]_{k=4} = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)$$

$$B[(\bar{y}_{FT})_w]_{k=4} = 0$$

$$V[(\bar{y}_{FT})_w]_{k=4} = \left[\bar{Y}^2\left(D_1 - \frac{1}{N}\right)W_1^2 s_1^2 C_{1Y}^2 + \left(D_2 - \frac{1}{N}\right)W_2^2 s_2^2 C_{2Y}^2\right]$$

8.0 NUMERICAL ILLUSTRATION :

Consider two populations I and II given in appendix A and B. Both the populations and divided into two parts as R-group and NR-group having size N_1 and N_2 respectively ($N = N_1 + N_2$). The population parameters are displayed below :

TABLE 8.1 : PARAMETERS OF POPULATION - I (IN APPENDIX A)

	Entire Population	For R-group	For NR-group
Size	$N = 180$	$N_1 = 100$	$N_2 = 80$
Mean *Y*	$\bar{Y} = 159.03$	$\bar{Y}_1 = 173.60$	$\bar{Y}_2 = 140.81$
Mean *X*	$\bar{X} = 113.22$	$\bar{X}_1 = 128.45$	$\bar{X}_2 = 94.19$
Mean Square *Y*	$S_Y^2 = 2205.18$	$S_{1Y}^2 = 2532.36$	$S_{2Y}^2 = 1219.90$
Mean Square *X*	$S_X^2 = 1972.61$	$S_{1X}^2 = 2300.86$	$S_{2X}^2 = 924.17$
Coefficient of Variation of *Y*	$C_Y = 0.295$	$C_{1Y} = 0.2899$	$C_{2Y} = 0.248$
Coefficient of Variation of *X*	$C_X = 0.392$	$C_{1X} = 0.373$	$C_{2X} = 0.323$
Correlation coefficient	$\rho_{XY} = 0.897$	$\rho_{1XY} = 0.857$	$\rho_{2XY} = 0.956$

TABLE 8.2: PARAMETERS OF POPULATION - II (IN APPENDIX B)

	Entire Population	For R-group	For NR-group
Size	$N = 150$	$N_1 = 90$	$N_2 = 60$
Mean *Y*	$\bar{Y} = 63.77$	$\bar{Y}_1 = 66.33$	$\bar{Y}_2 = 59.92$
Mean *X*	$\bar{X} = 29.2$	$\bar{X}_1 = 30.72$	$\bar{X}_2 = 26.92$
Mean Square *Y*	$S_Y^2 = 299.87$	$S_{1Y}^2 = 349.33$	$S_{2Y}^2 = 206.35$
Mean Square *X*	$S_X^2 = 110.43$	$S_{1X}^2 = 112.67$	$S_{2X}^2 = 100.08$
Coefficient of Variation of *Y*	$C_Y = 0.272$	$C_{1Y} = 0.282$	$C_{2Y} = 0.2397$
Coefficient of Variation of *X*	$C_X = 0.3599$	$C_{1X} = 0.345$	$C_{2X} = 0.3716$
Correlation coefficient	$\rho_{XY} = 0.8093$	$\rho_{1XY} = 0.8051$	$\rho_{2XY} = 0.8084$

Let samples of size $n = 40$ and $n = 30$ are drawn from population I and II respectively by SRSWOR and post-stratified into R and NR-groups. The sample values are in Table 6.8.3 and 6.8.4.

TABLE 8.3: SAMPLE VALUES FOR POPULATION - I

	Entire Sample	R-group	NR-group
Size	$n = 40$	$n_1 = 28$	$n_2 = 12$
Fraction	$f = 0.22$	-	-

TABLE 8.4 : SAMPLE VALUES FOR POPULATION - II

	Entire Sample	R-group	NR-group
Size	$n = 30$	$n_1 = 20$	$n_2 = 10$
Fraction	$f = 0.2$	-	-

Table 8.5 : Bias and M.S.E. Comparisons of $[(\bar{y}_{FT})_E]_k$

	Population I		Population II	
Estimator	**Bias**	**M.S.E.**	**Bias**	**M.S.E.**
$[(\bar{y}_{FT})_E]_{k=1}$	-1.7802	18.4419	0.7708	7.6442
$[(\bar{y}_{FT})_E]_{k=2}$	2.0992	222.1439	-0.5739	48.9692
$[(\bar{y}_{FT})_E]_{k=3}$	-0.5913	9.1005	-0.3507	6.3859
$[(\bar{y}_{FT})_E]_{k=4}$	0	43.6500	0	9.2675

TABLE 8.6 : BIAS AND M.S.E. COMPARISON OF $[(\bar{y}_{FT})_w]_k$

	Population I		Population II	
Estimator	**Bias**	**M.S.E.**	**Bias**	**M.S.E.**
$[(\bar{y}_{FT})_w]_{k=1}$	0.1433	12.9589	0.1095	6.0552
$[(\bar{y}_{FT})_w]_{k=2}$	0.3141	216.3024	0.1599	46.838
$[(\bar{y}_{FT})_w]_{k=3}$	–0.096	4.327	–0.031	5.2423
$[(\bar{y}_{FT})_w]_{k=4}$	0	43.65	0	9.2662

9.0 DISCUSSION :

The idea of utilizing a composition of $\bar{X}$, $\bar{X}_1$ and non-sampled part $\bar{x}_2^*$ is taken into consideration as an imputation methodology in equation (3.3). The class of imputed type estimator is proposed which has several members as special cases. Bias and mean squared error of the class is obtained via theorem 5.1 and 5.2. The class of non-imputed estimator is also derived in equation (7.2) and compared with the imputed one. Expressions for special cases are also derived. The computation over two populations is performed for bias and mean squared error. Over first population M. S. E. for non-imputed are very close to the imputed estimator. The similar pattern is found in population II also. At k = 3, the m. s. e. of population two is lowest. At k = 2, the class of estimator bears highest m.s.e. For all cases k = 1, 2, 3 the estimators are biased but in small amount. The choice k = 3 seems with lowest bias and therefore, is recommended for a good selection.

10.0 CONCLUSIONS :

The proposed imputation technique is useful and effective for obtaining population mean $\bar{Y}$ using factor-type estimation strategy. The choice k = 3 performs best in terms of lowest bias and lowest mean squared error. The imputation based mean squared error are little higher than non-imputed but very close in performance. Therefore, the composition of $\bar{X}$, $\bar{X}_1$ and non-sampled part of population plays an important role in driving imputation methodology for missing observation $\bar{x}_2$.

REFERENCES

1. **Agrawal, M.C. and Panda, K.B.(1993):** *An efficient estimator in post-stratification,* Metron, 51, 3-4, 179-187.
2. **Ahmed, M.S., Al-titi, Omar, Al-Rawi, Ziad and Abu-dayyeh, Walid(2006):** *Estimation of a population mean using different imputation method,* Statistics in Transition, 7(6), 1247-1264.
3. **Cochran, W.G.(2005):** *Sampling Techniques*, Third Edition, John Willey & Sons, New Delhi.
4. **Grover and Couper(1998):** *Non-response in household surveys,* John Wiley and Sons, New York.
5. **Hansen, M.H. and Hurwitz, W.N. (1946):** *The problem of non-response in sample surveys,* Jour. Amer. Stat. Asso., 41, 517-529.
6. **Hansen, M.H., Hurwitz, W.N. and Madow, W.G.(1993):** *Sample Survey Methods and Theory*, John Wiley and Sons, New York.
7. **Hinde, R.L. and Chambers, R.L. (1990):** *Non-response imputation with multiple source of non-response,* Jour. Off. Statistics, 7, 169-179.
8. **Holt., D. and Smith, T.M.F.(1979):** *Post-stratification,* J. R. Stat. Soc., Ser. A. 143, 33-46.
9. **Jackway, P.T. and Boyce, R.A. (1987):** *Response including techniques of mail surveys,* Aus. Jour. of Stat., 29, 255-263.
10. **Jagers, P. (1986):** *Post-stratification against bias in sampling,* International Statistical Review, 54, 159-167.
11. **Jagers, P., Oden, A. and Trulsoon, L. (1985):** *Post-stratification and ratio estimation,* Int. Stat. Rev., 53, 221-238.
12. **Khare, B.B.(1987):** *Allocation in stratified sampling in presence of non-response,* Metron , 45(I/II), 213-221.
13. **Khot, P.S.(1994):** *A note on handling non-response in sample surveys,* Jour. Amer. Stat. Assoc., 89, 693-696.
14. **Lessler and Kalsbeek(1992):** *Non-response error in surveys,* John Wiley and Sons, New York.
15. **Murthy, M.N. (1976):** *Sampling Theory and Methods,* Statistical Publishing Society, Calcutta
16. **Rao, J.N.K. and Sitter, R.R.(1995):** *Variance estimation under two phase sampling with application to imputation for missing data,* Biometrika, 82, 453-460.
17. **Rubin, D.B.(1976):** *Inference and missing data,* Biometrika, 63, 581-593.
18. **Sahoo, L.N.(1984):** *A note on estimation of the population mean using two auxiliary variables,* Ali. Jour. Stat., 3-4, 63-66.
19. **Sahoo, L.N.(1986):** *On a class of unbiased estimators using multi auxiliary information,* Jour. Ind. Soc. Ag. Stat, 38-3, 379-382.
20. **Sahoo, L.N. and Sahoo, R. K. (2001):** *Predictive estimation of finite population mean in two-phase sampling using two auxiliary variables,* Ind. Soc. Ag. Stat., 54 (4), 258-264.
21. **Sahoo, L.N., Sahoo J. and Mohanty, S.(1995):** *New predictive ratio estimator,* Jour. Ind. Soc. Ag. Stat., 47(3), 240-242.
22. **Shrivastava, S.K. and Jhajj, H. S. (1980):** *A class of estimators using auxiliary information for estimation finite population variance,* Sankhya, C, 42, 87-96.
23. **Shrivastava, S.K. and Jhajj, H. S. (1981):** *A class of the population mean in a survey sampling using auxiliary information,* Biometrika, 68, 341-343.
24. **Shukla, D. (2002):** *F-T estimator under two phase sampling*, Metron, 59, 1-2, 253-263.
25. **Shukla, D. and Dubey, J. (2001):** *Estimation in mail surveys under PSNR sampling scheme*, Ind. Soc. Ag. Stat., 54, 3, 288-302.

26. **Shukla, D. and Dubey, J.(2004):** *On estimation under post-stratified two phase non-response sampling scheme in mail surveys,* International Jour. of Management and Systems, 20, 3, 269-278.
27. **Shukla, D. and Dubey, J. (2006):** *On earmarked strata in post-stratification,* Statistics in Transition, 7, 5, 1067-1085.
28. **Shukla, D. and Thakur, N. S. (2008):** *Estimation of mean with imputation of missing data using factor type estimator*, Statistics in Transition, 9, 1, 33-48.
29. **Shukla, D., Thakur, N. S., Pathak, S. and Rajput, D. S. (2009a):** *Estimation of mean under imputation of missing data using factor-type estimator in two-phase sampling*, Statistics in Transition, 10, 3, 397-414.
30. **Shukla, D., Thakur, N. S. and Thakur, D. S. (2009b):** *Utilization of non-response auxiliary population mean in imputation for missing observations*, Journal of Reliability and Statistical Studies, 2, 1, 28-40.
31. **Shukla, D. and Trivedi, M.(1999):** *A new estimator for post-stratified sampling scheme,* Proceedings of NSBA-TA (Bayesian Analysis), Eds. Rajesh Singh, 206-218.
32. **Shukla, D. and Trivedi, M. (2001):** *mean estimation in deeply stratified population under post-stratification*. Ind. Soc. Ag. Stat., 54(2), 221-235.
33. **Shukla, D. and Trivedi, M.(2006):** *Examining stability of variability ratio with application in post-stratification*, International Jour. of Management and Systems, 22, 1, 59-70.
34. **Shukla, D., Bankey, A. and Trivedi, M.(2002):** *Estimation in post-stratification using prior information and grouping strategy*, Ind. Soc. Ag. Stat., 55, 2, 158-173.
35. **Shukla, D., Singh, V.K. and Singh, G.N.(1991):** *On the use of transformation in factor type estimator,* Metron, 49 (1-4), 359-361.
36. **Shukla, D., Trivedi, M. and Singh, G.N. (2006):** *Post-stratification two-way deeply stratified population*, Statistics in Transition, 7, 6, 1295-1310.
37. **Singh G.N. and Singh V.K.(2001):** *On the use of auxiliary information in successive sampling,* Jour. Ind. Soc. Ag. Stat., 54(1), 1-12.
38. **Singh, D. and Choudhary, F.S.(1986):** *Theory and Analysis of Sample Survey Design, Wiley Eastern Limited,* New Delhi.
39. **Singh, H.P.(1986):** *A generalized class of estimators of ratio product and mean using supplementary information on an auxiliary character in PPSWR scheme,* Guj. Stat. Rev., 13, 2, 1-30.
40. **Singh, H.P. and Upadhyaya, L.N. and Iachan, R.(1987):** *Unbiased product estimators,* Guj. Stat. Rev., 14, 2, 41-50.
41. **Singh, S. and Horn, S.(2000):** *Compromised imputation in survey sampling,* Metrika, 51, 266-276.
42. **Singh, V.K. and Shukla, D.(1987):** *One parameter family of factor type ratio estimator,* Metron, 45 (1-2), 273-283.
43. **Singh, V.K. and Shukla, D.(1993):** *An efficient one-parameter family of factor type estimator in sample survey,* Metron, 51, 1-2, 139-159.
44. **Singh, V.K. and Singh, G.N. (1991):** *Chain type estimator with two auxiliary variables under double sampling scheme,* Metron, 49, 279-289.
45. **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), 209-221.
46. **Smith, T.M.F.(1991):** *Post-stratification,* The Statisticians, 40, 323-330.
47. **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.
48. **Wywial, J.(2001):** *Stratification of population after sample selection,* Statistics in Transition, 5, 2, 327-348.

Appendix A

Population I (N= 180)

R-group: (N_1=100)

Y:	110	75	85	165	125	110	85	80	150	165	135	120	140	135	145
X:	80	40	55	130	85	50	35	40	110	115	95	60	70	85	115
Y:	200	135	120	165	150	160	165	145	215	150	145	150	150	195	190
X:	150	85	80	100	25	130	135	105	185	110	95	75	70	165	160
Y:	175	160	165	175	185	205	140	105	125	230	230	255	275	145	125
X:	145	110	135	145	155	175	80	75	65	170	170	190	205	105	85
Y:	110	110	120	230	220	280	275	220	145	155	170	195	170	185	195
X:	75	80	90	165	160	205	215	190	105	115	135	145	135	110	145
Y:	180	150	185	165	285	150	235	125	165	135	130	245	255	280	150
X:	135	110	135	115	125	205	100	195	85	115	75	190	205	210	105
Y:	205	180	150	205	220	240	260	185	150	155	115	115	220	215	230
X:	110	105	110	175	180	215	225	110	90	95	85	75	175	185	190
Y:	210	145	135	250	265	275	205	195	180	115					
X:	170	85	95	190	215	200	165	155	150	175					

NR-group: (N_2=80)

Y:	85	75	115	165	140	110	115	13.5	120	125	120	150	145	90	105
X:	55	40	65	115	90	55	60	65	70	75	80	120	105	45	65
Y:	110	90	155	130	120	95	100	125	140	155	160	145	90	90	95
X:	70	60	85	95	80	55	60	75	90	105	125	95	45	55	65
Y:	115	140	180	170	175	190	160	155	175	195	90	90	80	90	80
X:	75	105	120	115	125	135	110	115	135	145	45	55	50	60	50
Y:	105	125	110	120	130	145	160	170	180	`145	130	195	200	160	110
X:	65	75	70	80	85	105	110	115	130	95	65	135	130	115	55
Y:	155	190	150	180	200	160	155	170	195	200	150	165	155	180	200
X:	115	130	110	120	125	145	120	105	100	95	90	105	125	130	145
Y:	160	155	170	195	200										
X:	120	115	120	135	150										

Appendix B

Population II (N=150)

R-group (N_1=90)

Y:	90	75	70	85	95	55	65	80	65	50	45	55	60	60	95
X:	30	35	30	40	45	25	40	50	35	30	15	20	25	30	40
Y:	100	40	45	55	35	45	35	55	85	95	65	75	70	80	65
X:	50	10	25	25	10	15	10	25	35	55	35	40	30	45	40
Y:	90	95	80	85	55	60	75	85	80	65	35	40	95	100	55
X:	40	50	35	45	35	25	30	40	25	35	10	15	45	45	25
Y:	45	40	40	35	55	75	80	80	85	55	45	70	80	90	55
X:	15	15	20	10	30	25	30	40	35	20	25	30	40	45	30
Y:	65	60	75	75	85	95	90	90	45	40	45	55	60	65	60
X:	25	40	35	30	40	35	40	35	15	25	15	30	30	25	20
Y:	75	70	40	55	75	45	55	60	85	55	60	70	75	65	80
X:	25	20	35	30	45	10	30	25	40	15	25	30	35	30	45

NR-group (N_2=60)

Y:	40	90	95	70	60	65	85	55	45	60	65	60	55	55	45
X:	10	30	30	30	25	30	40	25	15	20	30	30	35	25	20
Y:	65	80	55	65	75	55	50	55	60	45	40	75	75	45	70
X:	35	45	30	30	40	15	15	20	30	15	10	40	45	10	30
Y:	65	70	55	35	35	50	55	35	55	60	30	35	45	55	65
X:	30	40	30	10	15	25	30	15	20	30	10	20	15	30	30
Y:	75	65	70	65	70	45	55	60	85	55	60	70	75	65	80
X:	30	35	40	25	45	10	30	25	40	15	25	30	35	30	45