African Journal of Mathematics and Computer Science Research Vol. 5(4), pp. 78-89, 15 February, 2012
Available online at http://www.academicjournals.org/AJMCSR
DOI: 10.5897/AJMCSR11.159
ISSN 2006-9731©2012 Academic Journals

Full Length Research Paper

Utilization of mixture of $\bar{x}$, $\bar{x}_1$ and $\bar{x}_2$ in imputation for missing data in post-stratification

D. Shukla[1], D. S. Thakur[2] and N. S. Thakur[3]

[1]Department of Mathematics and Statistics, Dr. H. S. Gour Central University, Sagar, (M.P.), India.
[2]School of Excellence, Sagar (M.P.), India.
[3]Centre for Mathematical Sciences (CMS), Banasthali University, Rajasthan, India.

Accepted 10 January, 2012

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 aforementioned 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.

Key words: Imputation, non-response, post-stratification, simple random sampling without replacement (SRSWOR), respondents (R).

INTRODUCTION

In sampling theory, the problem of mean estimation of a population is considered by many authors like Srivastava and Jhajj (1980, 1981), Sahoo (1984, 1986), Singh (1986), Singh et al. (1987), Singh and Singh (1991), Singh et al. (1994), Sahoo et al. (1995), Sahoo and Sahoo (2001), and 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). 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

*Corresponding author. E-mail: nst_stats@yahoo.co.in.

the auxiliary source of data.

Symbols and setup

Let $U = (U_1, U_2, \ldots\ldots\ldots, U_N)$ 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. Some symbols are as follows:

R-group, Respondents group or group of those who responses in survey;
NR-group, Non-respondents group or group of those who denied to response during 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:

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

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

ASSUMPTIONS

Consider following in light of Figure 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 $\bar{X}_1$ and $\bar{X}$ 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 $\bar{Y}$, $\bar{Y}_1$ and $\bar{Y}_2$.

PROPOSED CLASS OF ESTIMATION STRATEGY

To estimate population mean $\bar{Y}$, in setup of Figure 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 $\bar{Y}$. Shukla et al. (1991), Singh and Shukla (1993) and 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 $(\bar{x}_2^*)_4$ is adopted as:

$$(\bar{x}_2^*)_4 = \left[\frac{N\bar{X} - n\{f\bar{X}_1 + (1-f)\bar{X}_2\}}{N-n}\right] \qquad (1)$$

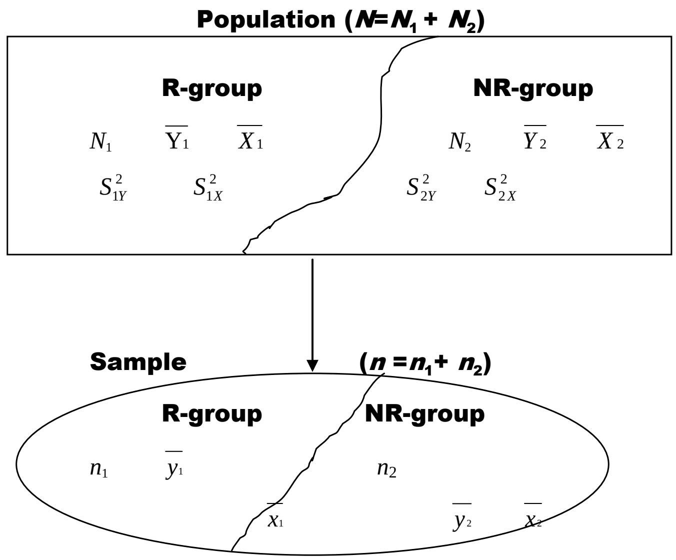

Figure 1. Setup to estimate population mean $\overline{Y}$.

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 $\bar{x}^{(2)}$ for $\bar{x}$ as describe as follows:

And

$$\bar{x}^{(4)} = \left[\frac{N_1\bar{x}_1 + N_2\left(\bar{x}_2^*\right)_4}{N_1+N_2}\right] \tag{2}$$

The proposed class of imputed factor-type estimator is:

$$[(\bar{y}_{FT})_D]_k = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left[\frac{(A+C)\overline{X} + fB\bar{x}^{(4)}}{(A+fB)\overline{X} + C\bar{x}^{(4)}}\right] \tag{3}$$

Where $0 < k < \infty$ and k is a constant and

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

LARGE SAMPLE APPROXIMATION

Consider the following for large n:

$$\left.\begin{aligned}\bar{y}_1 &= \overline{Y}_1(1+e_1)\\ \bar{y}_2 &= \overline{Y}_2(1+e_2)\\ \bar{x}_1 &= \overline{X}_1(1+e_3)\\ \bar{x}_2 &= \overline{X}_2(1+e_4)\end{aligned}\right\} \tag{4}$$

where, e_1, e_2, e_3 and e_4 are very small numbers $e_i > 0$ (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 (Cochran, 2005; Hansen et al., 1993; Sukhatme et al., 1984; Singh and Choudhary, 1986; Murthy, 1976), we get

$$\left.\begin{aligned}\mathrm{E}(e_1) &= \mathrm{E}\left[\mathrm{E}(e_1)\ |n_1\right] = 0\\ \mathrm{E}(e_2) &= \mathrm{E}\left[\mathrm{E}(e_2)\ |n_2\right] = 0\\ \mathrm{E}(e_3) &= \mathrm{E}\left[\mathrm{E}(e_3)\ |n_1\right] = 0\\ \mathrm{E}(e_4) &= \mathrm{E}\left[\mathrm{E}(e_4)\ |n_2\right] = 0\end{aligned}\right\} \tag{5}$$

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

$$\mathrm{E}\left[e_1^2\right] = \mathrm{E}\left[\mathrm{E}\left(e_1^2\right)\ |n_1\right]$$

$$= \mathrm{E}\left[\left\{\left(\frac{1}{n_1} - \frac{1}{N}\right)C_{1Y}^2\right\}\middle| n_1\right]$$

$$= \left[\left\{\mathrm{E}\left(\frac{1}{n_1}\right) - \frac{1}{N}\right\}C_{1Y}^2\right]$$

$$= \left[\left(D_1 - \frac{1}{N}\right)C_{1Y}^2\right] \tag{6}$$

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

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

$$=\left[\left(D_2-\frac{1}{N}\right)C_{2Y}^2\right] \tag{7}$$

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

$$=\left[\left(D_1-\frac{1}{N}\right)C_{1X}^2\right] \tag{8}$$

and

$$\mathrm{E}\left[e_4^2\right]=\left[\left(D_2-\frac{1}{N}\right)C_{2X}^2\right] \tag{9}$$

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

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

$$=\left[\left(D_1-\frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right] \tag{10}$$

$$\mathrm{E}\left[e_1e_2\right]=\mathrm{E}\left[\mathrm{E}\left(e_1e_2\right)\,|n_1,n_2\right]\;=0 \tag{11}$$

$$\mathrm{E}\left[e_1e_4\right]=0 \tag{12}$$

$$\mathrm{E}\left[e_2e_3\right]=\mathrm{E}\left[\mathrm{E}\left(e_2e_3\right)\,|n_1,n_2\right]=0 \tag{13}$$

$$\mathrm{E}\left[e_2e_4\right]=\left(D_2-\frac{1}{N}\right)\rho_2C_{2Y}C_{2X} \tag{14}$$

$$\mathrm{E}\left[e_3e_4\right]=0 \tag{15}$$

The expressions (11), (12), (13) and (15) 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 1

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

$$[(\bar{y}_{FT})_D]_k = \delta_3\bar{Y}\,[1+s_1W_1e_1+s_2W_2e_2][1+(\alpha_3-\beta_3)e_3-(\alpha_3-\beta_3)\beta_3\,e_3^2+(\alpha_3-\beta_3)\,\beta_3^2\,e_3^3\ldots..$$

Proof

Rewrite $\bar{x}^{(4)}$ as in Equation 2:

$$\bar{x}^{(4)}\;=\left[\frac{N_1\bar{x}_1+N_2\left(\bar{x}_2^*\right)_4}{N_1+N_2}\right]$$

Where,

$$\left(\bar{x}_2^*\right)_4\;=\left[\frac{N\bar{X}-n\left\{f\bar{X}_1+(1-f)\bar{X}_2\right\}}{N-n}\right]$$

$$\Rightarrow\qquad \bar{x}^{(4)}=\frac{1}{N}\left[N_1\bar{X}_1(1+e_3)+N_2\left\{\frac{N\bar{X}-n\left\{f\bar{X}_1+(1-f)\bar{X}_2\right\}}{N-n}\right\}\right]$$

$$=\frac{N_1\bar{X}_1(1+e_3)+p\left[N\bar{X}-n\left(f\bar{X}_1+(1-f)\bar{X}_2\right)\right]}{N}$$

$$=\left[\frac{N_1\bar{X}_1+pN\bar{X}-pnf\bar{X}_1-pn(1-f)\bar{X}_2+N_1\bar{X}_1e_3}{N}\right]$$

$$=\left[\frac{pN\bar{X}+\left(N_1-pnf\right)\bar{X}_1-pn(1-f)\bar{X}_2+N_1\bar{X}_1e_3}{N}\right]$$

$$=\bar{X}\left[p+\left(W_1-pf^2\right)r_1-pf(1-f)r_2+W_1r_1e_3\right]$$

$$=\bar{X}\left[\mu+W_1r_1e_3\right] \tag{16}$$

Where,

$p=\dfrac{N_2}{N-n}$; $\qquad \mu = p + (W_1 - pf^2)r_1 - pf(1-f)\,r_2$.

Now, the estimator $[(\bar{y}_{FT})_D]_k$ under approximation and using (16) is

$$[(\bar{y}_{FT})_D]_k\;=\left(\frac{N_1\bar{y}_1+N_2\bar{y}_2}{N}\right)\left[\frac{(A+C)\bar{X}+fB\bar{x}^{(4)}}{(A+fB)\bar{X}+C\bar{x}^{(4)}}\right]$$

$$=\left[\frac{N_1\bar{Y}_1(1+e_1)+N_2\bar{Y}_2(1+e_2)}{N}\right]\left[\frac{(A+C)\bar{X}+fB(\mu+W_1r_1e_3)\bar{X}}{(A+fB)\bar{X}+C(\mu+W_1r_1e_3)\bar{X}}\right]$$

$$=\;\bar{Y}\left[1+s_1W_1e_1+s_2W_2e_2\right]\left[\frac{(A+fB\mu+C)+fBW_1r_1e_3}{(A+fB+C\mu)+CW_1r_1e_3}\right]$$

$$=\;\bar{Y}\left[1+s_1W_1e_1+s_2W_2e_2\right]\left[\frac{\xi_1+\xi_2e_3}{\xi_3+\xi_4e_3}\right]$$

$$=\;\delta_3\bar{Y}\left[1+s_1W_1e_1+s_2W_2e_2\right](1+\alpha_3e_3)(1+\beta_3e_3)^{-1}$$

Where,

$\xi_1 = A + fB\mu + C$; $\xi_2 = fBW_1r_1$; $\xi_3 = A + fB + C\mu$; $\xi_4 = CW_1r_1$;

$r_1 = \frac{\overline{X_1}}{\overline{X}}$; $r_2 = \frac{\overline{X_2}}{\overline{X}}$; $s_1 = \frac{\overline{Y_1}}{\overline{Y}}$; $s_2 = \frac{\overline{Y_2}}{\overline{Y}}$; $\alpha_3 = \frac{\xi_2}{\xi_1}$; $\beta_3 = \frac{\xi_4}{\xi_3}$; $\delta_3 = \frac{\xi_1}{\xi_3}$.

We can further express the above into following:

$= \delta_3\overline{Y}\,[1 + s_1W_1e_1 + s_2W_2e_2]\,(1 + \alpha_3e_3)\,(1 - \beta_3e_3 + \beta_3^2\,e_3^2 - \beta_3^3\,e_3^3 \;.......)$

$= \delta_3\overline{Y}\,[1 + s_1W_1e_1 + s_2W_2e_2][1 + (\alpha_3 - \beta_3)e_3 - (\alpha_3 - \beta_3)\beta_3e_3^2 + (\alpha_3 - \beta_3)\,\beta_3^2\,e_3^3 \;....]$

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, ... as

$$\left.\begin{aligned} \mathrm{E}\left[e_1^ie_2^j\right] &= 0 \quad \text{when } i+j > 2 \\ \mathrm{E}\left[e_1^ie_3^j\right] &= 0 \quad \text{when } i+j > 2 \\ \mathrm{E}\left[e_2^ie_3^j\right] &= 0 \quad \text{when } i+j > 2 \end{aligned}\right\} \tag{17}$$

Theorem 2

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

$$\mathrm{B}\left[(\overline{y}_{FT})_D\right]_k = \overline{Y}\left[(\delta_3 - 1) - \delta_3C_{1X}(\alpha_3 - \beta_3)\left(D_1 - \frac{1}{N}\right)\left\{\beta_3\,C_{1X} - s_1W_1\,\rho_1C_{1Y}\right\}\right]$$

Proof

$$\mathrm{B}\left[(\overline{y}_{FT})_D\right]_k = \mathrm{E}\left[\left\{(\overline{y}_{FT})_D\right\}_k - \overline{Y}\right]$$

Using theorem 1 and taking expectations

$\mathrm{E}[(\overline{y}_{FT})_D]_k = \delta_3\overline{Y}\,\mathrm{E}[1 + (\alpha_3 - \beta_3)e_3 - (\alpha_3 - \beta_3)\,\beta_3\,e_3^2 + s_1W_1e_1\{1 + (\alpha_3 - \beta_3)\,e_3 - (\alpha_3 - \beta_3)\,\beta_3\,e_3^2\}$

$+ s_2W_2e_2\,\{1 + (\alpha_3 - \beta_3)\,e_3 - (\alpha_3 - \beta_3)\,\beta_3e_3^2\}]$

$= \delta_3\overline{Y}\,[1 - (\alpha_3 - \beta_3)\beta_3\,\mathrm{E}(e_3^2) + s_1W_1(\alpha_3 - \beta_3)\,\mathrm{E}(e_1e_3) + s_2W_2(\alpha_3 - \beta_3)\mathrm{E}(e_2e_3)]$

$$= \delta_3\overline{Y}\left[1 - (\alpha_3 - \beta_3)\,\beta_3\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + (\alpha_3 - \beta_3)s_1W_1\left(D_1 - \frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right]$$

$$= \delta_3\overline{Y}\left[1 - (\alpha_3 - \beta_3)\left(D_1 - \frac{1}{N}\right)C_{1X}\left\{\beta_3C_{1X} - s_1W_1\rho_1C_{1Y}\right\}\right]$$

Therefore,

$$\mathrm{B}\left[(\overline{y}_{FT})_D\right]_k = \mathrm{E}\left[\left\{(\overline{y}_{FT})_D\right\}_k - \overline{Y}\right]$$

$$= \overline{Y}\left[(\delta_3 - 1) - \delta_3C_{1X}(\alpha_3 - \beta_3)\left(D_1 - \frac{1}{N}\right)\left\{\beta_3C_{1X} - s_1W_1\rho_1C_{1Y}\right\}\right]$$

Theorem 3

The mean squared error of $[(\overline{y}_{FT})_D]_k$ is

$\mathrm{M}\,[(\overline{y}_{FT})_D]_k$

$$= \overline{Y}^2\left[(\delta_3 - 1)^2 + \left(D_1 - \frac{1}{N}\right)\left\{J_1s_1^2C_{1Y}^2 + J_2C_{1X}^2 + 2J_3s_1\rho_1C_{1Y}C_{1X}\right\} + \left(D_2 - \frac{1}{N}\right)\delta_3^2s_2^2W_2^2C_{2Y}^2\right]$$

Where

$$J_1 = \delta_3^2W_1^2,\; J_2 = \delta_3(\alpha_3 - \beta_3)\left\{\delta_3(\alpha_3 - \beta_3) - 2(\delta_3 - 1)\beta_3\right\};\; J_3 = W_1\delta_3(2\delta_3 - 1)(\alpha_3 - \beta_3)$$

Proof

$$\mathrm{M}\,[(\overline{y}_{FT})_D]_k = \mathrm{E}[\{(\overline{y}_{FT})_D\}_k - \overline{Y}\,]^2$$

Using Theorem 1, we can express

$\mathrm{M}[(\overline{y}_{FT})_D]_k = \mathrm{E}[\,\delta_3\overline{Y}\,\{1 + s_1W_1e_1 + s_2W_2e_2\}\{1 + (\alpha_3 - \beta_3)e_3 - (\alpha_3 - \beta_3)\beta_3e_3^2 + (\alpha_3 - \beta_3)\,\beta_3^2\,e_3^3$

$...\} - \overline{Y}\,]^2$

Using large sample approximations of (17) we could express

$= \overline{Y}^2\,\mathrm{E}\,[(\delta_3 - 1) + \delta_3\{(\alpha_3 - \beta_3)e_3 - (\alpha_3 - \beta_3)\beta_3e_3^2 + (s_1W_1e_1 + s_2W_2e_2) + (\alpha_3 - \beta_3)(s_1W_1e_1 + s_2W_2e_2)$

$e_3\}]^2$

$= \overline{Y}^2\,\mathrm{E}\,[(\delta_3 - 1)^2 + \delta_3^2\{(\alpha_3 - \beta_3)^2\,e_3^2 + s_1^2\,W_1^2\,e_1^2 + s_2^2\,W_2^2\,e_2^2 + 2s_1s_2W_1W_2e_1e_2$

$+ 2(\alpha_3 - \beta_3)(s_1W_1e_1\,e_3 + s_2W_2e_2e_3)\} + 2\delta_3(\delta_3 - 1)\,\{(\alpha_3 - \beta_3)e_3 - (\alpha_3 - \beta_3)\beta_3e_3^2$

$+ (s_1W_1e_1 + s_2W_2e_2) + (\alpha_3 - \beta_3)\,(s_1W_1e_1e_3 + s_2W_2e_2e_3)\}]$

Using (5), (11) and (12) we rewrite,

$= \overline{Y}^2\,[\,(\delta_3 - 1)^2 + \delta_3^2\{(\alpha_3 - \beta_3)^2\,\mathrm{E}(e_3^2) + s_1^2\,W_1^2\,\mathrm{E}(e_1^2) + s_2^2\,W_2^2\,\mathrm{E}(e_2^2) + 2(\alpha_3 - \beta_3)\,s_1W_1\,\mathrm{E}(e_1\,e_3)\}$

$+ 2\delta_3(\delta_3 - 1)\{-(\alpha_3 - \beta_3)\beta_3\,\mathrm{E}(e_3^2) + (\alpha_3 - \beta_3)s_1W_1\,\mathrm{E}(e_1e_3)\}]$

$$= \overline{Y}^2\left[(\delta_3 - 1)^2 + \delta_3^2\left\{(\alpha_3 - \beta_3)^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + s_1^2W_1^2\left(D_1 - \frac{1}{N}\right)C_{1Y}^2\right.\right.$$

$$\left. + s_2^2W_2^2\left(D_2 - \frac{1}{N}\right)C_{2Y}^2 + 2(\alpha_3 - \beta_3)s_1W_1\left(D_1 - \frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right\}$$

$$\left. + 2\delta_3(\delta_3 - 1)\left\{-(\alpha_3 - \beta_3)\beta_3\left(D_1 - \frac{1}{N}\right)C_{1X}^2\right\} + (\alpha_3 - \beta_3)s_1W_1\left(D_1 - \frac{1}{N}\right)\rho_1C_{1Y}C_{1X}\right]$$

$$= \overline{Y}^2\left[(\delta_3 - 1)^2 + \left(D_1 - \frac{1}{N}\right)\left\{\delta_3^2\,s_1^2W_1^2C_{1Y}^2 + \delta_3(\alpha_3 - \beta_3)\{\delta_3(\alpha_3 - \beta_3) - 2(\delta_3 - 1)\beta_3\}C_{1X}^2\right.\right.$$

$$\left.\left. + 2\delta_3(\delta_3 - 1)(\alpha_3 - \beta_3)s_1W_1\rho_1C_{1Y}C_{1X}\right\} + \left(D_2 - \frac{1}{N}\right)\delta_3^2s_2^2W_2^2C_{2Y}^2\right]$$

$$= \overline{Y}^2\left[(\delta_3 - 1)^2 + \left(D_1 - \frac{1}{N}\right)\left\{J_1s_1^2C_{1Y}^2 + J_2C_{1X}^2 + 2J_3s_1\rho_1C_{1Y}C_{1X}\right\} + \left(D_2 - \frac{1}{N}\right)\delta_3^2s_2^2W_2^2C_{2Y}^2\right]$$

SOME SPECIAL CASES

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

Case 1

When $k = 1$

$A = 0$; $B = 0$; $C = -6$; $\xi_1 = -6$; $\xi_2 = 0$; $\xi_3 = -6\mu$; $\xi_4 = -6r_1W_1$; $\alpha_3 = 0$; $\beta_3 = \frac{r_1W_1}{\mu}$; $\delta_3 = \mu^{-1}$;

$$J_1 = \frac{W_1^2}{\mu^2};\quad J_2 = \frac{r_1^2W_1^2(3-2\mu)}{\mu^4};\quad J_3 = \frac{r_1W_1^2(\mu-2)}{\mu^3};$$

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

$$[(\bar{y}_{FT})_D]_{k=1} = \left[\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right]\left[\frac{\bar{X}}{\bar{x}^{(4)}}\right] \tag{18}$$

$$B[(\bar{y}_{FT})_D]_{k=1} = \bar{Y}\mu^{-3}[(1-\mu)\mu^2 + \left(D_1 - \frac{1}{N}\right) r_1W_1^2C_{1X}\{r_1C_{1X} - \mu s_1\rho_1C_{1Y}\}] \tag{19}$$

$$M[(\bar{y}_{FT})_D]_{k=1} = \bar{Y}^2\mu^{-4}[(1-\mu)^2\mu^2 + W_1^2\left(D_1 - \frac{1}{N}\right)\{\mu^2 s_1^2C_{1Y}^2 + (3-2\mu)\, r_1^2C_{1X}^2 + 2(\mu-2)\,\mu\, r_1s_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right)\mu^2W_2^2s_2^2C_{2Y}^2] \tag{20}$$

Case 2

When $k = 2$

$A = 0$; $B = -2$; $C = 0$; $\xi_1 = -2f\mu$; $\xi_2 = -2fW_1r_1$; $\xi_3 = -2f$; $\xi_4 = 0$;

$\alpha_3 = r_1W_1\mu^{-1}$; $\beta_3 = 0$; $\delta_3 = \mu$; $J_1 = W_1^2\mu^2$; $J_2 = r_1^2W_1^2$; $J_3 = r_1W_1^2(2\mu-1)$;

$$[(\bar{y}_{FT})_D]_{k=2} = \left[\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right]\left[\frac{\bar{x}^{(4)}}{\bar{X}}\right] \tag{21}$$

$$B[(\bar{y}_{FT})_D]_{k=2} = \bar{Y}\left[(\mu-1) + \left(D_1 - \frac{1}{N}\right)W_1^2r_1s_1\rho_1C_{1Y}C_{1X}\right] \tag{22}$$

$$M[(\bar{y}_{FT})_D]_{k=2} = \bar{Y}^2[(\mu-1)^2 + W_1^2\left(D_1 - \frac{1}{N}\right)\{\mu^2s_1^2C_{1Y}^2 + r_1^2C_{1X}^2 + 2(2\mu-1)s_1r_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right)\mu^2W_2^2s_2^2C_{2Y}^2] \tag{23}$$

Case 3

When $k = 3$

$A = 2$; $B = -2$; $C = 0$; $\xi_1 = 2(1-f\mu)$; $\xi_2 = -2fW_1r_1$; $\xi_3 = 2(1-f)$; $\xi_4 = 0$; $\alpha_3 = \frac{-fW_1r_1}{1-f\mu}$

$\beta_3 = 0$; $\delta_3 = \frac{1-f\mu}{1-f}$; $J_1 = \frac{(1-f\mu)^2W_1^2}{(1-f)^2}$; $J_2 = \frac{f^2W_1^2r_1^2}{(1-f)^2}$; $J_3 = \frac{fW_1^2r_1\{2f\mu - f - 1\}}{(1-f)^2}$;

$$[(\bar{y}_{FT})_D]_{k=3} = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left[\frac{\bar{X} + f\bar{x}^{(4)}}{(1-f)\bar{X}}\right] \tag{24}$$

$$B[(\bar{y}_{FT})_D]_{k=3} = \bar{Y}f(1-f)^{-1}\left[(1-\mu) - \left(D_1 - \frac{1}{N}\right)W_1^2r_1s_1\rho_1C_{1Y}C_{1X}\right] \tag{25}$$

$$M[(\bar{y}_{FT})_D]_{k=3} = \bar{Y}^2(1-f)^{-2}[f^2(1-\mu)^2 + \left(D_1 - \frac{1}{N}\right)W_1^2\{(1-f\mu)^2s_1^2C_{1Y}^2 + f^2r_1^2C_{1X}^2 + 2(2f\mu - f - 1)fr_1s_1\rho_1C_{1Y}C_{1X}\} + \left(D_2 - \frac{1}{N}\right)(1-f\mu)^2W_2^2s_2^2C_{2Y}^2] \tag{26}$$

Case 4

When $k = 4$;

$A = 6$; $B = 0$; $C = 0$; $\xi_1 = 6$; $\xi_2 = 0$; $\xi_3 = 6$; $\xi_4 = 0$; $\alpha_3 = 0$; $\beta_3 = 0$; $\delta_3 = 1$; $J_1 = W_1^2$; $J_2 = 0$; $J_3 = 0$;

$$[(\bar{y}_{FT})_D]_{k=4} = \left[\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right] \tag{27}$$

$$B[(\bar{y}_{FT})_D]_{k=4} = 0 \tag{28}$$

$$V[(\bar{y}_{FT})_D]_{k=4} = \left(D_1 - \frac{1}{N}\right)W_1^2\bar{Y}_1^2C_{1Y}^2 + \left(D_2 - \frac{1}{N}\right)W_2^2\bar{Y}_2^2C_{2Y}^2 \tag{29}$$

ESTIMATOR WITHOUT IMPUTATION

Throughout the discussion, the assumption is unknown value of $\bar{x}_2$. This is imputed by the term $(\bar{x}_2^*)_4$, to provide the generation of $\bar{x}^{(4)}$ [Equations (1) and (2)]. Suppose the $\bar{x}_2$ is known, then there is no need of imputation and the proposed (2) and (3) reduces into:

$$\bar{x}^{(*)} = \left(\frac{N_1\bar{x}_1 + N_2\bar{x}_2}{N}\right) \tag{30}$$

$$\left[(\bar{y}_{FT})_w\right]_k = \left(\frac{N_1\bar{y}_1 + N_2\bar{y}_2}{N}\right)\left(\frac{(A+C)\bar{X} + fB\bar{x}^{(*)}}{(A+fB)\bar{X} + C\bar{x}^{(*)}}\right) \quad (31)$$

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 4

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

$$B\left[(\bar{y}_{FT})_w\right]_k = (\xi_1' - \xi_2')\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\left\{s_1\rho_1 C_{1Y} - \xi_2' r_1 C_{1X}\right\}\right.$$

$$\left. + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\left\{s_2\rho_2 C_{2Y} - \xi_2' r_2 C_{2X}\right\}\right]$$

Where,

$$\xi_1' = fB/(A+fB+C); \qquad \xi_2' = C/(A+fB+C).$$

Proof

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

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

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

$$= \left[\bar{Y} + W_1\bar{Y}_1 e_1 + W_2\bar{Y}_2 e_2\right]\left[\frac{(A+fB+C) + fB(W_1r_1e_3 + W_2r_2e_4)}{(A+fB+C) + C(W_1r_1e_3 + W_2r_2e_4)}\right]$$

$$= \left[\bar{Y} + W_1\bar{Y}_1 e_1 + W_2\bar{Y}_2 e_2\right]\left[1 + \xi_1'(W_1r_1e_3 + W_2r_2e_4)\right]\left[1 + \xi_2'(W_1r_1e_3 + W_2r_2e_4)\right]^{-1}$$

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

$$\left[(\bar{y}_{FT})_w\right]_k = \bar{Y} + \bar{Y}(\Delta_1 - \Delta_2) + W_1\bar{Y}_1 e_1(1 + \Delta_1 - \Delta_2) + W_2\bar{Y}_2 e_2(1 + \Delta_1 - \Delta_2) \quad (32)$$

Where,

$\Delta_1 = (\xi_1' - \xi_2')(W_1r_1e_3 + W_2r_2e_4)$; $\Delta_2 = \xi_2'(\xi_1' - \xi_2')(W_1r_1e_3 + W_2r_2e_4)^2$

And

$W_1 r_1 + W_2 r_2 = 1$ holds.

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

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

(ii) $E[\Delta_1^2] = (\xi_1' - \xi_2')^2\left[W_1^2 r_1^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + W_2^2 r_2^2\left(D_2 - \frac{1}{N}\right)C_{2X}^2\right]$

(iii) $E[\Delta_2] = \xi_2'(\xi_1' - \xi_2')\left[W_1^2 r_1^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + W_2^2 r_2^2\left(D_2 - \frac{1}{N}\right)C_{2X}^2\right]$

(iv) $E[e_1\Delta_1] = (\xi_1' - \xi_2')W_1 r_1\left(D_1 - \frac{1}{N}\right)\rho_1 C_{1X} C_{1Y}$

(v) $E[e_2\Delta_1] = (\xi_1' - \xi_2')W_2 r_2\left(D_2 - \frac{1}{N}\right)\rho_2 C_{2X} C_{2Y}$

(vi) $E[e_1\Delta_2] = 0$ [under $0(n^{-1})$]

(vii) $E[e_2\Delta_2] = 0$ [under $0(n^{-1})$]

The bias of estimator without imputation is

$$\mathrm{B}\left[(\bar{y}_{FT})_w\right]_k = \mathrm{E}\left[\left\{(\bar{y}_{FT})_w\right\}_k - \bar{Y}\right]$$

$$= E\left[\bar{Y}(\Delta_1 - \Delta_2) + W_1\bar{Y}_1 e_1(1 + \Delta_1 - \Delta_2) + W_2\bar{Y}_2 e_2(1 + \Delta_1 - \Delta_2)\right]$$

$$= \left[W_1\bar{Y}_1 E(e_1\Delta_1) + W_2\bar{Y}_2 E(e_2\Delta_1) - \bar{Y}E(\Delta_2)\right]$$

$$= (\xi_1' - \xi_2')\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\left\{\bar{Y}_1\rho_1 C_{1Y} - \bar{Y}\xi_1' r_1 C_{1X}\right\}\right.$$

$$\left. + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\left\{\bar{Y}_2\rho_2 C_{2Y} - \bar{Y}\xi_1' r_2 C_{2X}\right\}\right]$$

$$= \bar{Y}(\xi_1' - \xi_2')\left[\left(D_1 - \frac{1}{N}\right)W_1^2 r_1 C_{1X}\left\{s_1\rho_1 C_{1Y} - \xi_2' r_1 C_{2X}\right\}\right.$$

$$\left. + \left(D_2 - \frac{1}{N}\right)W_2^2 r_2 C_{2X}\left\{s_2\rho_2 C_{2Y} - \xi_2' r_2 C_{2X}\right\}\right]$$

Theorem 5

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

$$\mathrm{M}\left[(\bar{y}_{FT})_w\right]_k = \bar{Y}^2\left[\left(D_1 - \frac{1}{N}\right)W_1^2\left\{T_1^2 C_{1Y}^2 + T_2^2 C_{1X}^2 + 2T_1T_2\rho_1 C_{1Y} C_{1X}\right\}\right.$$

$$\left. + \left(D_2 - \frac{1}{N}\right)W_2^2\left\{S_1^2 C_{2Y}^2 + S_2^2 C_{2X}^2 + 2S_1S_2\rho_2 C_{2Y} C_{2X}\right\}\right]$$

Where,

$T_1 = s_1$; $T_2 = (\xi_1' - \xi_2')r_1$; $S_1 = s_2$; $S_2 = (\xi_1' - \xi_2')r_2$;

Proof

$$M\left[\left(\bar{y}_{FT}\right)_w\right]_k = E\left[\left\{\left(y_{FT}\right)_w\right\}_k - \bar{Y}\right]^2$$

$$\left[\left(\bar{y}_{FT}\right)_w\right]_k = E\left[\bar{Y}\left(\Delta_1 - \Delta_2\right) + W_1\bar{Y}_1 e_1\left(1 + \Delta_1 - \Delta_2\right) + W_2\bar{Y}_2 e_2\left(1 + \Delta_1 - \Delta_2\right)\right]^2$$

$$= \bar{Y}^2 E\left(\Delta_1^{\,2}\right) + W_1^2\bar{Y}_1^2 E\left(e_1^2\right) + W_2^2\bar{Y}_2^2 E\left(e_2^2\right) + 2W_1\,\bar{Y}\bar{Y}_1 E\left(e_1\Delta_1\right)$$

$$+ 2W_2\,\bar{Y}\bar{Y}_2 E\left(e_2\Delta_1\right) + 2W_1\,W_2\bar{Y}_1\bar{Y}_2 E\left(e_1 e_2\right)$$

$$= \bar{Y}^2\left[\left(\xi_1' - \xi_2'\right)^2\left\{W_1^2 r_1^2\left(D_1 - \frac{1}{N}\right)C_{1X}^2 + W_2^2 r_2^2\left(D_2 - \frac{1}{N}\right)C_{2X}^2\right\}\right.$$

$$+ \left\{W_1^2 s_1^2\left(D_1 - \frac{1}{N}\right)C_{1Y}^2 + W_2^2 s_2^2\left(D_2 - \frac{1}{N}\right)C_{2Y}^2\right\}$$

$$+ 2W_1 s_1\left(\xi_1' - \xi_2'\right)\left\{W_1 r_1\left(D_1 - \frac{1}{N}\right)\rho_1 C_{1Y} C_{1X}\right\}$$

$$\left. + 2W_2 s_2\left(\xi_1' - \xi_2'\right)\left\{W_2 r_2\left(D_2 - \frac{1}{N}\right)\rho_2 C_{2Y} C_{2X}\right\}\right]$$

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

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

Remark 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 with the following

Case 5

When k = 1

$A = 0$; $B = 0$; $C = -6$; $\xi_1' = 0$; $\xi_2' = 1$; $T_1 = s_1$;

$T_2 = -r_1$; $S_1 = s_1$; $S_2 = -r_2$;

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

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

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

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

Case 6

When k = 2

$A = 0$; $B = -2$; $C = 0$; $\xi_1' = 1$; $\xi_2' = 0$; $T_1 = s_1$; $T_2 = r_1$; $S_1 = s_2$; $S_2 = r_2$.

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

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

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

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

Case 7

When k = 3

$A = 2$; $B = -2$; $C = 0$; $\xi_1' = -f/(1-f)^{-1}$; $\xi_2' = 0$; $T_1 = s_1$; $T_2 = r_1 f\,(1-f)^{-1}$;
$S_1 = s_2$; $S_2 = r_2 f\,(1-f)^{-1}$;

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

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

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

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

Case 8

When k = 4, then

$A = 6$; $B = 0$; $C = 0$; $\xi_1' = 0$; $\xi_2' = 0$; $T_1 = s_1$; $T_2 = 0$; $S_1 = s_2$; $S_2 = 0$;

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

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

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

NUMERICAL ILLUSTRATION

Consider two populations I and II given in Appendix A and B. Both populations are 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 in Tables 1 to 6.

DISCUSSION

Using the imputation for $\bar{x}_2$ by the mixture of three

Table 1. Parameters of population - I (from Appendix A).

Parameter	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 2. Parameters of population - ii (from Appendix B).

Parameter	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 Tables 3 and 4.

Table 3. Sample values for population – i.

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

Table 4. Sample values for population – ii

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

population means is performed. The proposed class is in Equation (3) with bias in theorem 2 and mean squared error in theorem 3. The class contains some special imputed estimators for value k = 1, k = 2, k = 3 and k = 4. A non-imputed class of estimator is also developed which has bias and mean squared error derived in theorem 4 and 5. This class also has some special cases. The computation over two population is made whose description is given in appendix. The two random sample of size n = 40 and n = 30 are drawn from populations **I** and **II** respectively and post-stratified into two groups. The Tables 5 and 6 are presenting a numerical comparison between imputation and non-imputation class over the two populations in terms of their bias and m. s. e. The imputation technique (1) is effective because there is not much increase in the mean squared error due to

Table 5. Bias and M.S.E. comparisons of $\left[\left(\bar{y}_{FT}\right)_D\right]_k$.

Estimator	Population I		Population II	
	Bias	M.S.E.	Bias	M.S.E.
$[(\bar{y}_{FT})_D]_{k=1}$	-2.8775	18.6000	-1.1874	9.8618
$[(\bar{y}_{FT})_D]_{k=2}$	3.3127	232.4162	3.4360	41.4794
$[(\bar{y}_{FT})_D]_{k=3}$	-4.0370	8.4710	-0.6377	6.3540
$[(\bar{y}_{FT})_D]_{k=4}$	0	43.6500	0	9.2675

Table 6. Bias and MSE comparison of $\left[\left(\bar{y}_{FT}\right)_w\right]_k$.

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

imputation. The k = 3 seems to be a good choice. The k = 2 performs worst for both sots of data.

CONCLUSIONS

The technique of mixture of $\bar{X}$, $\bar{X}_1$, $\bar{X}_2$ performs well and the imputed factor-type estimator is very close to non-imputed in terms of mean squared error when k = 1, 2, and 3 holds. The choice k = 3 is better over the other two. Performance over population II is superior than I. it seems that factor type lass is able to replace the non-responded observation in a nice manner.

ACKNOWLEDGEMENT

Authors are thankful to the referee for his valuable suggestions.

REFERENCES

Agrawal MC, Panda KB (1993). An efficient estimator in post-stratification. Metron, 51(3-4): 179-187.
Ahmed MS, Al-titi O, Al-Rawi Z, Abu-dayyeh W (2006). Estimation of a population mean using different imputation method. Stat. Transit., 7(6): 1247-1264.
Cochran WG (2005). Sampling Techniques. Third Edition, John Willey & Sons, New Delhi.
Grover RM, Couper MP (1998). Non-response in household surveys. John Wiley and Sons, New York.
Hansen MH, Hurwitz WN (1946). The problem of non-response in sample surveys. J. Am. Stat. Assoc., 41: 517-529.
Hansen MH, Hurwitz WN, Madow WG (1993). Sample Survey Methods and Theory. John Wiley and Sons, New York.
Hinde RL, Chambers RL (1990). Non-response imputation with multiple source of non-response. J. Off. Stat., 7: 169-179.
Holt D, Smith TMF (1979). Post-stratification. J. R. Stat. Soc., Ser. A., 143: 33-46.
Jackway PT, Boyce RA (1987). Response including techniques of mail surveys. Aus. J. Stat., 29: 255-263.
Jagers P (1986). Post-stratification against bias in sampling. Int. Stat. Rev., 54: 159-167.
Jagers P, Oden A, Trulsoon L (1985). Post-stratification and ratio estimation. Int. Stat. Rev., 53: 221-238.
Khare BB (1987). Allocation in stratified sampling in presence of non-response. Metron, 45(I/II): 213-221.
Khot PS (1994). A note on handling non-response in sample surveys. J. Am. Stat. Assoc., 89: 693-696.
Lessler JT, Kalsbeek WD (1992). Non-response error in surveys. John Wiley and Sons, New York.
Murthy MN (1976). Sampling Theory and Methods. Statistical Publishing Society, Calcutta.
Rao JNK, Sitter RR (1995). Variance estimation under two phase sampling with application to imputation for missing data. Biometrika, 82: 453-460.
Rubin DB (1976). Inference and missing data. Biometrika, 63: 581-593.
Sahoo LN (1984). A note on estimation of the population mean using two auxiliary variables. Ali. J. Stat., pp. 3-4, 63-66.
Sahoo LN (1986). On a class of unbiased estimators using multi auxiliary information. J. Ind. Soc. Ag. Stat, 38(3): 379-382.
Sahoo LN, Sahoo RK (2001). Predictive estimation of finite population mean in two-phase sampling using two auxiliary variables. Ind. Soc. Ag. Stat., 54(4): 258-264.

Sahoo LN, Sahoo J, Mohanty S (1995). New predictive ratio estimator. J. Ind. Soc. Ag. Stat., 47(3): 240-242.

Shrivastava SK, Jhajj HS (1980). A class of estimators using auxiliary information for estimation finite population variance. Sankhya C, 42: 87-96.

Shrivastava SK, Jhajj HS (1981). A class of the population mean in a survey sampling using auxiliary information. Biometrika, 68: 341-343.

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

Shukla D, Dubey J (2001). Estimation in mail surveys under PSNR sampling scheme. Ind. Soc. Ag. Stat., 54(3): 288-302.

Shukla D, Dubey J (2004). On estimation under post-stratified two phase non-response sampling scheme in mail surveys. Int. J. Manage. Syst., 20(3): 269-278.

Shukla D, Dubey J (2006). On earmarked strata in post-stratification. Stat. Transit., 7(5): 1067-1085.

Shukla D, Trivedi M (1999). A new estimator for post-stratified sampling scheme. Proceedings of NSBA-TA (Bayesian Analysis), Eds. Rajesh Singh, pp. 206-218.

Shukla D, Trivedi M (2001). Mean estimation in deeply stratified population under post-stratification. Ind. Soc. Ag. Stat., 54(2): 221-235.

Shukla D, Trivedi M (2006). Examining stability of variability ratio with application in post-stratification. Int. J. Manage. Syst., 22(1): 59-70.

Shukla D, Bankey A, Trivedi M (2002). Estimation in post-stratification using prior information and grouping strategy. Ind. Soc. Ag. Stat., 55(2): 158-173.

Shukla D, Singh VK, Singh GN (1991). On the use of transformation in factor type estimator. Metron, 49(1-4): 359-361.

Shukla D, Trivedi M, Singh GN (2006). Post-stratification two-way deeply stratified population. Stat. Transit., 7(6): 1295-1310.

Singh GN, Singh VK (2001). On the use of auxiliary information in successive sampling. J. Ind. Soc. Ag. Stat., 54(1): 1-12.

Singh D, Choudhary FS (1986). Theory and Analysis of Sample Survey Design. Wiley Eastern Limited, New Delhi.

Singh HP (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.

Singh HP, Upadhyaya LN, Lachan R (1987). Unbiased product estimators. Guj. Stat. Rev., 14(2): 41-50.

Singh S, Horn S (2000). Compromised imputation in survey sampling. Metrika, 51: 266-276.

Singh VK, Shukla D (1987). One parameter family of factor type ratio estimator. Metron, 45(1-2): 273-283.

Singh VK, Shukla D (1993). An efficient one-parameter family of factor type estimator in sample survey. Metron, 51(1-2): 139-159.

Singh VK, Singh GN (1991). Chain type estimator with two auxiliary variables under double sampling scheme. Metron, 49: 279-289.

Singh VK, Singh GN, Shukla D (1994). A class of chain ratio estimator with two auxiliary variables under double sampling scheme. Sankhya, Ser. B., 46(2): 209-221.

Smith TMF (1991). Post-stratification. Statisticians, 40: 323-330.

Sukhatme PV, Sukhatme BV, Sukhatme S, Ashok C (1984). Sampling Theory of Surveys with Applications. Iowa State University Press, I.S.A.S. Publication, New Delhi.

Wywial J (2001). Stratification of population after sample selection. Stat. Transit., 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