www.ijmer.com Vol.2, Issue.5, Sep-Oct. 2012 pp-3561-3571 ISSN: 2249-6645

Mean Estimation with Imputation in Two- Phase Sampling

Narendra Singh Thakur, Kalpana Yadav, Sharad Pathak*
Center for Mathematical Sciences (CMS), *Banasthali University,Banasthali, Rajasthan*
**Department of Mathematics and Statistics, Dr. H. S. Gour Central Univesity, Sagar* (M.P.)

Abstract: *Missing data is a problem encountered in almost every data collection activity but particularly in sample survey. The missing data naturally occurs in sample surveys when some, not all sampling units refuse or unable to participate in the survey or when data for specific items on a questionnaire completed for an otherwise cooperating unit are missing. Imputation is a methodology, which uses available data as a tool for the replacement of missing observations. Imputation methods used to fill the non responses and lead, under definite conditions, to suitable inference. This article suggests some imputation methods and discusses the properties of their mean estimators. Numerical study is performed over two populations using the expressions of bias and m.s.e and efficiency compared with existing estimators.*

I. Introduction

In literature, several imputation techniques are described, some of them are better over others. Rubin (1976) addressed three concepts: OAR (observed at random), MAR (missing at random), and PD (parametric distribution). He defined that if the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the value of the unobserved data, then data are MAR. The observed data are observed at random (OAR) if for each possible value of the missing data and the parameter ϕ the conditional probability of the observed pattern of missing data given the missing data and the observed data, is the same for all possible values of the observed data. Heitzen and Basu (1996) have distinguished the meaning of MAR and MCAR in a very nice way. In what follows MCAR (missing completely at random) is used.

Little and Rubin (1987) define three different classes of missingness. They defined the key terms used in discussing missingness in the literature. Data missing on *Y* are observed at random (OAR) if missingness on *Y* is not a function of *X*. Phrased another way, if *X* determines missingness on *Y*, the data are not OAR. Data missing on *Y* are missing at random (MAR) if missingness on *Y* is not a function of *Y*. Phrased another way, if *Y* determines missingness on *Y*, the data are not MAR. Data are Missing Completely at Random (MCAR) if missingness on *Y* is unrelated to *X* or *Y*. In other words MCAR=OAR + MAR. If the data are MCAR or at least MAR, then the missing data mechanism is considered "ignorable." Otherwise, the missing data mechanism is considered "non-ignorable."

There are different ways and means to control non-response. One way of dealing with the problem of non-response is to make more efforts to collect information by taking a sub-sample of units not responding at the first attempt. Another way of dealing with the problem of non-response is to estimate the probability of responding informants of their being at home at a specified point of time and weighting results with the inverse of this probability. A technique to deal with the problem of non-response was developed by Hansen and Hurwitz (1946). They assumed that the population is divided into two classes, a *response class* who respond in the first attempt and a *non-response class* who did not.

A questionnaire contains many questions that we call items. When item non-response occurs, substantial information about the non-respondent is usually available from other items on the questionnaire. Many imputation methods in literature use selection of these items as auxiliary variable in assigning values to the i^{th} non-respondent for item *y*. Rao and Sitter (1995), Singh and Horn (2000), Ahmed et al. (2006) and Shukla and Thakur (2008) have given applications of various imputation procedures.

Let the variable *Y* is of main interest and *X* be an auxiliary variable correlated with *Y* and the population mean $\bar{X}$ of auxiliary variable is unknown. A large preliminary simple random sample (without replacement) S' of n' units is drawn from the population $\Omega = (1, 2,..., N)$ to estimate $\bar{X}$ and a secondary sample *S* of size n ($n < n'$) drawn as a sub-sample of the sample S' to estimate the population mean of main variable. Let the sample *S* contains n_1 responding units and $n_2 = (n - n_1)$ non-responding units. Using the concept of post-stratification, sample may be divided into two groups: responding (R_1) and non-responding (R_2).

The sample may be considered as stratified into two classes namely a *response class* and *non-response class*, then the procedure is known as *post-stratification*. Sukhatme (1984) advocates that post-stratification procedure is as precise as the stratified sampling under proportional allocation if the sample size is large enough. Estimation problem in sample surveys, in the setup of post-stratification, under non-response situation is studied due to Shukla and Dubey (2004 and 2008). Shukla et al. (2009) have also given the concept of utilization of $\bar{X}_2$ (population mean of non-response group of *X*) in imputation for missing observations of auxiliary information due to non-response.

Now it may be consider the population has two types of individuals like N_1 as number of respondents (R_1) and N_2 non-respondents (R_2), Thus the total *N* units of the population will comprise N_1 and N_2, respectively, such that $N = N_1+N_2$. The population proportions of units in the R_1 and R_2 groups are expressed as $W_1 = N_1/N$ and $W_2 = N_2/N$ such that $W_1+W_2=1$. Further, let $\bar{Y}$ and $\bar{X}$ be the population means of *Y* and *X* respectively. For every unit $i \in R_1$, the value y_i is observed available. However, for the units $i \in R_2$, the y_i's are missing and imputed values are to be derived. The i^{th} value x_i of auxiliary variate is used as a source of imputation for missing data when $i \in R_2$. This is to assume that for sample *S*, the data $x_s = \{x_i : i \in S\}$ are known. The following notations are used in this paper:

$\bar{x}_n, \bar{y}_n$: the sample mean of X and Y respectively in *S*; $\bar{x}_1, \bar{y}_1$: the sample mean of X and Y respectively in R_1;

S_X^2, S_Y^2: the population mean squares of X and Y respectively; C_X, C_Y: the coefficient of variation of X and Y respectively; ρ: Correlation Coefficient in population between X and Y respectively.

Further, consider few more symbolic representations:

$$L = E\left(\frac{1}{n_1}\right) = \left[\frac{1}{nW_1} + \frac{(N-n)(1-W_1)}{(N-1)n^2W_1^2}\right], \quad M = \frac{(N-n)(n-n_1)n'N}{nn_1^2(N-1)(N-n')}, \quad Q = \frac{nn_1^2(N-n')(N-1)}{n'N(N-n)(n-n_1) - 2nn_1^2(N-n')(N-1)}.$$

Under this setup as describe above in case of simple random sampling without replacement and assuming $\bar{X}$ is known, some well known imputation methods are given below:

1.1. Mean Method of Imputation:

For y_i define $y_{\bullet i}$ as $$y_{\bullet i} = \begin{cases} y_i & if \quad i \in R_1 \\ \bar{y}_r & if \quad i \in R_2 \end{cases} \qquad \ldots(1.1)$$

Using above, the imputation-based estimator of population mean $\bar{Y}$ is: $\bar{y}_m = \frac{1}{r}\sum_{i \in R} y_i = \bar{y}_r$ …(1.2)

Lemma 1.1: The bias and mean squared error is given by: **(i)** $B(\bar{y}_m) = 0$ …(1.3)

(ii) $V(\bar{y}_m) \approx \left(\frac{1}{r} - \frac{1}{N}\right)S_y^2$ …(1.4)

1.2. Ratio Method of Imputation:

For y_i and x_i, define $y_{\bullet i}$ as: $$y_{\bullet i} = \begin{cases} y_i & if \quad i \in R_1 \\ \hat{b}x_i & if \quad i \in R_2 \end{cases} \quad \text{where } \hat{b} = \sum_{i \in R} y_i \Big/ \sum_{i \in R} x_i \qquad \ldots(1.5)$$

Under this, the imputation-based estimator is: $\bar{y}_S = \frac{1}{n}\sum_{i \in S} y_{\bullet i} = \bar{y}_r\left(\frac{\bar{x}_n}{\bar{x}_r}\right) = \bar{y}_{RAT}$ …(1.6)

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

Lemma 1.2: The bias and mean squared error of $\bar{y}_{RAT}$ is given by: **(i)** $B(\bar{y}_{RAT}) = \bar{Y}\left(\frac{1}{r} - \frac{1}{n}\right)\left(C_x^2 - \rho C_y C_x\right)$ …(1.7)

(ii) $M(\bar{y}_{RAT}) \approx \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2 + \left(\frac{1}{r} - \frac{1}{n}\right)\left[S_y^2 + R_1^2 S_x^2 - 2R_1 S_{xy}\right]$ where $R_1 = \frac{\bar{Y}}{\bar{X}}$ …(1.8)

1.3. Compromised Method of Imputation:

Singh and Horn (2000) proposed Compromised imputation procedure as given below:

$$y_{\bullet i} = \begin{cases} (\alpha n/r)y_i + (1-\alpha)\hat{b}x_i & if \quad i \in R_1 \\ (1-\alpha)\hat{b}x_i & if \quad i \in R_2 \end{cases} \qquad \ldots(1.9)$$

where α is a suitably chosen constant, such that the resultant variance of the estimator is optimum. The imputation-based estimator, for this case, Estimator of population mean is $\bar{y}_{COMP} = \left[\alpha\bar{y}_r + (1-\alpha)\bar{y}_r\frac{\bar{x}_n}{\bar{x}_r}\right]$ …(1.10)

Lemma 1.3: The bias, mean squared error and minimum mean squared error at $\alpha = 1 - \rho\frac{C_Y}{C_X}$ of $\bar{y}_{COMP}$ is given by

(i) $B(\bar{y}_{COMP}) = \bar{Y}(1-\alpha)\left(\frac{1}{r} - \frac{1}{n}\right)\left(C_X^2 - \rho C_Y C_X\right)$ …(1.11)

(ii) $M\left(\bar{y}_{COMP}\right)\approx\left\{\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2+\left(\frac{1}{r}-\frac{1}{n}\right)\left[S_y^2+R_1^2S_z^2-2R_1S_{xy}\right]\right\}-\left(\frac{1}{r}-\frac{1}{n}\right)\alpha^2\bar{Y}^2C_x^2$ …(1.12)

(iii) $M\left(\bar{y}_{COM}\right)_{\min}=\left[\left(\frac{1}{r}-\frac{1}{N}\right)-\left(\frac{1}{r}-\frac{1}{n}\right)\rho^2\right]S_Y^2$ …(1.13)

1.4. Ahmed's Methods:

(A) $y_{7i}=\begin{cases} y_i & \text{if } i\in R_1 \\ \bar{y}_r+\frac{nk_1}{(n-r)}\left(\bar{X}-\bar{x}\right)+k_2\left(x_i-\bar{x}_r\right) & \text{if } i\in R_2 \end{cases}$ …(1.14)

Under this method, the point estimator of $\bar{Y}$ is: $t_7=\bar{y}_r+k_1\left(\bar{X}-\bar{x}\right)+k_2\left(\bar{x}-\bar{x}_r\right)$ …(1.15)

Lemma 1.4: The bias, variance and minimum variance at $k_1=k_2=\frac{S_{XY}}{S_X^2}$ of t_7 is given by: **(i)** $B[t_7]=0$ …(1.16)

(ii) $V(t_7)=\left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2-2S_{XY}\left[k_1\left(\frac{1}{n}-\frac{1}{N}\right)+k_2\left(\frac{1}{r}-\frac{1}{n}\right)\right]+S_X^2\left[k_1^2\left(\frac{1}{n}-\frac{1}{N}\right)+k_2^2\left(\frac{1}{r}-\frac{1}{n}\right)\right]$ …(1.17)

(iii) $V(t_7)_{\min}=\left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2\left(1-\rho^2\right)$ …(1.18)

(B) $y_{8i}=\begin{cases} y_i & \text{if } i\in R_1 \\ \left[\frac{\bar{y}_r\left(x_i+\frac{r}{n-r}\bar{x}_r\right)}{\theta_1\bar{x}_r+(1-\theta_1)\bar{x}}-\frac{r}{n-r}\bar{y}_r\right] & \text{if } i\in R_2 \end{cases}$ …(1.19)

under this setup, the point estimator of $\bar{Y}$ is : $t_8=\frac{\bar{y}_r\bar{x}}{\theta_1\bar{x}_r+(1-\theta_1)\bar{x}}$ …(1.20)

Lemma 1.5: The bias, mean squared error and minimum mean squared error at $\theta_1=\rho\frac{C_Y}{C_X}$ of t_8 is given by

(i) $B(t_8)\approx\left(\frac{1}{r}-\frac{1}{n}\right)\bar{Y}\left(\theta_1^2C_X^2-\theta_1\rho C_YC_X\right)$ …(1.21)

(ii) $M(t_8)\approx\bar{Y}^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)C_Y^2+\theta_1^2\left(\frac{1}{r}-\frac{1}{n}\right)C_X^2-2\theta_1\left(\frac{1}{r}-\frac{1}{n}\right)\rho C_YC_X\right]$ …(1.22)

(iii) $M(t_8)_{\min}\approx\left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2-\left(\frac{1}{r}-\frac{1}{n}\right)\frac{S_{XY}^2}{S_X^2}$ …(1.23)

(C) $y_{9i}=\begin{cases} y_i & \text{if } i\in R_1 \\ \frac{1}{(n-r)}\left[\frac{n\bar{y}_r\bar{X}}{\theta_2\bar{x}+(1-\theta_2)\bar{X}}-r\bar{y}_r\right] & \text{if } i\in R_2 \end{cases}$ …(1.24)

Under this method, the point estimator of $\bar{Y}$ is : $t_9=\frac{\bar{y}_r\bar{X}}{\theta_2\bar{x}+(1-\theta_2)\bar{X}}$ …(1.25)

Lemma 1.6: The bias, mean squared error and minimum mean squared error at $\theta_2=\rho\frac{C_Y}{C_X}$ of t_8 is given by

(i) $B(t_9)\approx\left(\frac{1}{n}-\frac{1}{N}\right)\bar{Y}\left(\theta_2^2C_X^2-\theta_2\rho C_YC_X\right)$ …(1.26)

(ii) $M(t_9) \approx \bar{Y}^2\left[\left(\frac{1}{r}-\frac{1}{N}\right)C_Y^2+\theta_2^2\left(\frac{1}{n}-\frac{1}{N}\right)C_X^2-2\theta_2\left(\frac{1}{n}-\frac{1}{N}\right)\rho C_Y C_X\right]$ …(1.27)

(iii) $M(t_9)_{\min} \approx \left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2-\left(\frac{1}{n}-\frac{1}{N}\right)\frac{S_{XY}^2}{S_X^2}$ …(1.28)

(d) $$y_{10i}=\begin{cases} y_i & \text{if } i\in R_1 \\ \frac{1}{(n-r)}\left[\frac{n\,\bar{y}_r\bar{X}}{\theta_3\bar{x}+(1-\theta_3)\bar{X}}-r\,\bar{y}_r\right] & \text{if } i\in R_2\end{cases}$$ …(1.29)

under this, the point estimator of population mean $\bar{Y}$ is: $t_{10}=\frac{\bar{y}_r\bar{X}}{\theta_3\bar{x}_r+(1-\theta_3)\bar{X}}$ …(1.30)

Lemma 1.7: The bias, mean squared error and minimum mean squared error at $\theta_3=\rho\frac{C_Y}{C_X}$ of t_{10} is given by

(i) $B(t_{10})\approx\left(\frac{1}{r}-\frac{1}{N}\right)\bar{Y}\left(\theta_3^2C_X^2-\theta_3\rho C_Y C_X\right)$ …(1.31)

(ii) $M(t_{10})\approx\bar{Y}^2\left(\frac{1}{r}-\frac{1}{N}\right)\left[C_Y^2+\theta_3^2C_X^2-2\theta_3\rho C_Y C_X\right]$ …(1.32)

(iii) $M(t_{10})_{\min}=\left(\frac{1}{r}-\frac{1}{N}\right)S_Y^2(1-\rho^2)$ …(1.33)

II. Large Sample Approximations:

Let $\bar{y}_1=\bar{Y}(1+e_1)$; $\bar{x}_1=\bar{X}(1+e_2)$; $\bar{x}_n=\bar{X}(1+e_3)$ and $\bar{x}'=\bar{X}(1+e_3')$, which implies the results $e_1=\frac{\bar{y}_1}{\bar{Y}}-1$; $e_2=\frac{\bar{x}_1}{\bar{X}}-1$; $e_3=\frac{\bar{x}_n}{\bar{X}}-1$ and $e_3'=\frac{\bar{x}'}{\bar{X}}-1$. Now by using the concept of two-phase sampling and the mechanism of MCAR, for given n_1, n and n' [see Rao and Sitter (1995)] we have:

$$E(e_1)=E\left[E(e_1)|n_1\right]=E\left[\left(\frac{\bar{y}_1-\bar{Y}}{\bar{Y}}\right)\Big|n_1\right]=\frac{\bar{Y}-\bar{Y}}{\bar{Y}}=0;\ \text{Similarly, } E(e_2)=E(e_3)=E(e_3')=0;$$

$$E(e_1^2)=E\left[\left(\frac{\bar{y}_1-\bar{Y}}{\bar{Y}}\right)^2\Big|n_1\right]=\left(E\left(\frac{1}{n_1}\right)-\frac{1}{n'}\right)C_y^2=\left(L-\frac{1}{n'}\right)C_y^2;\ E(e_2^2)=\left(L-\frac{1}{n'}\right)C_X^2\ \ E(e_3^2)=\left(\frac{1}{n}-\frac{1}{n'}\right)C_X^2;$$

$$E(e'^2_3)=\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2;\ E(e_1e_2)=E(e_1e_2/n_1)=E\left[\left(\frac{(\bar{y}_1-\bar{Y})(\bar{x}_1-\bar{X})}{\bar{Y}\bar{X}}\right)\Big|n_1\right]=\left(E\left(\frac{1}{n_1}\right)-\frac{1}{n'}\right)\rho C_yC_X$$

$$=\left(L-\frac{1}{n'}\right)\rho C_yC_X\ \ E(e_1e_3)=\left(\frac{1}{n}-\frac{1}{n'}\right)\rho C_yC_X;\ E(e_1e_3')=\left(\frac{1}{n'}-\frac{1}{N}\right)\rho C_yC_X;\ E(e_2e_3)=\left(\frac{1}{n}-\frac{1}{n'}\right)C_X^2;$$

$$E(e_2e_3')=\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2;\ E(e_3e_3')=\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2$$

III. Proposed Different Imputation Methods

Let y_{vji} denotes the i^{th} available observation for the j^{th} imputation. We suggest the following imputation methods:

(1) $$y_{V7i}=\begin{cases} y_i & \text{if } i\in R_1 \\ \bar{y}_1+\frac{1}{(1-W_1)}\left[h(\bar{x}'-\bar{x}_n)+(1-W_1)k(x_i-\bar{x}_1)\right] & \text{if } i\in R_2\end{cases}$$ …(3.1)

where h and k are suitably chosen constants, such that the variance the resultant estimator is minimum. Under this

$$T_{V7}=\bar{y}_1+h(\bar{x}'-\bar{x}_n)+k(\bar{x}_n-\bar{x}_1)$$ …(3.2)

(2) $$y_{V8i} = \begin{cases} y_i & if \quad i \in R_1 \\ \dfrac{\bar{y}_1}{(1-W_1)}\left[\dfrac{\left(x_i(1-W_1)+W_1\bar{x}_1\right)}{\theta\bar{x}_1+(1-\theta)\bar{x}_n} - W_1\right] & if \quad i \in R_2 \end{cases} \quad \text{…(3.3)}$$

where θ is suitably chosen constant, such that the variance the resultant estimator is minimum. Under this

$$T_{V8} = \frac{\bar{y}_1\bar{x}_n}{\theta\bar{x}_1+(1-\theta)\bar{x}_n} \quad \text{…(3.4)}$$

(3) $$y_{V9i} = \begin{cases} y_i & if \quad i \in R_1 \\ \dfrac{\bar{y}_1}{(1-W_1)}\left[\dfrac{\bar{x}_1}{\varphi\bar{x}_n+(1-\varphi)\bar{x}'} - W_1\right] & if \quad i \in R_2 \end{cases} \quad \text{…(3.5)}$$

where φ is suitably chosen constant, such that the variance the resultant estimator is minimum.

Under this, the point estimator of population mean $\bar{Y}$ is

$$T_{V9} = \frac{\bar{y}_1\bar{x}'}{\varphi\bar{x}_n+(1-\varphi)\bar{x}'} \quad \text{…(3.6)}$$

(4) $$y_{V10i} = \begin{cases} y_i & if \quad i \in R_1 \\ \dfrac{\bar{y}_1}{(1-W_1)}\left[\dfrac{\bar{x}'}{\psi\bar{x}_r+(1-\psi)\bar{x}'} - W_1\right] & if \quad i \in R_2 \end{cases} \quad \text{…(3.7)}$$

where ψ is suitably chosen constant, such that the variance the resultant estimator is minimum. Under this, the point estimator of population mean $\bar{Y}$ is

$$T_{V10} = \frac{\bar{y}_1\bar{x}'}{\psi\bar{x}_r+(1-\psi)\bar{x}'} \quad \text{…(3.8)}$$

IV. Bias and M.S.E of Proposed Methods:

Let $B(.)$ and $M(.)$ denote the bias and mean squared error (*M.S.E.*) of an estimator under a given sampling design. The properties of estimators are derived in the following theorems respectively.

Theorem 4.1:

(1) The estimator T_{V7} in terms of e_1, e_2, e_3 and e_3' is:

$$T_{V7} = \bar{Y}(1+e_1) + h\bar{X}(e_3' - e_3) + k\bar{X}(e_3 - e_2) \quad \text{…(4.1)}$$

Proof: $T_{V7} = \bar{y}_1 + h(\bar{x}' - \bar{x}_n) + k(\bar{x}_n - \bar{x}_1) = \bar{Y}(1+e_1) + h\bar{X}(e_3' - e_3) + k\bar{X}(e_3 - e_2)$

(2) The estimator T_{V7} is unbiased i.e. $B[T_{V7}] = 0$ …(4.2)

Proof: $B(T_{V7}) = E[T_{V7} - \bar{Y}] = \bar{Y} - \bar{Y} = 0$

(3) The variance of T_{V7} is

$$V(T_{V7}) = \left(L - \frac{1}{n'}\right)S_Y^2 + \left(\frac{1}{n} - \frac{2}{n'} + \frac{1}{N}\right)\left(h^2S_X^2 - 2h\rho S_Y S_X\right) + \left(L - \frac{1}{n}\right)\left(k^2S_X^2 - 2k\rho S_Y S_X\right) \quad \text{…(4.3)}$$

Proof: $V(T_{V7}) = E[T_{V7} - \bar{Y}]^2 = E[\bar{Y}e_1 + h\bar{X}(e_3' - e_3) + k\bar{X}(e_3 - e_2)]^2$

$$= E\Big[\bar{Y}^2e_1^2 + h^2\bar{X}^2(e_3' - e_3)^2 + k^2\bar{X}^2(e_3 - e_2)^2 + 2h\bar{Y}\bar{X}(e_3' - e_3)e_1 + 2hk\bar{X}^2(e_3' - e_3)(e_3 - e_2) + 2k\bar{Y}\bar{X}(e_3 - e_2)e_1\Big]$$

$$= E\Big[\bar{Y}^2e_1^2 + h^2\bar{X}^2(e_3'^2 + e_3^2 - 2e_3e_3') + k^2\bar{X}^2(e_3^2 + e_2^2 - 2e_2e_3) + 2h\bar{Y}\bar{X}(e_1e_3' - e_1e_3) + 2hk\,\bar{X}^2(e_3e_3' - e_3^2 - e_2e_3' + e_2e_3) + 2k\bar{Y}\bar{X}(e_1e_3 - e_1e_2)\Big]$$

$$=\left(L-\frac{1}{n'}\right)S_Y^2+\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(h^2S_X^2-2h\rho S_YS_X\right)+\left(L-\frac{1}{n}\right)\left(k^2S_X^2-2k\rho S_YS_X\right)$$

(4) The minimum variance of the T_{V7} is

$$\left[V\left(T_{V7}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_Y^2 \qquad \text{…(4.4)}$$

Proof: By differentiating (4.3) with respect to h and k then equate to zero

$$\frac{d}{dh}\left[V\left(T_{V7}\right)\right]=0 \Rightarrow h=\rho\frac{S_Y}{S_X}$$

and $$\frac{d}{dk}\left[V\left(T_{V7}\right)\right]=0 \Rightarrow k=\rho\frac{S_Y}{S_X}$$

After replacing value of h and k in (4.3), we obtained

$$\left[V\left(T_{V7}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_Y^2$$

Theorem 4.2:

(5) The estimator T_{V8} in terms of e_1, e_2, e_3 and e_3' is :

$$T_{V8}=\bar{Y}\left[1+e_1+\theta\left(e_3-e_2-e_1e_2+e_1e_3+(1-2\theta)e_2e_3+\theta e_2^2-(1-\theta)e_3^2\right)\right] \qquad \text{…(4.5)}$$

Proof:

$$T_{V8}=\frac{\bar{y}_1\bar{x}}{\theta\bar{x}_1+(1-\theta)\bar{x}_n}=\frac{\bar{Y}\bar{X}(1+e_1)(1+e_3)}{\theta\bar{X}(1+e_2)+(1-\theta)\bar{X}(1+e_3)}$$

$$=\bar{Y}(1+e_1)(1+e_3)(1+e_3+\theta e_2-\theta e_3)^{-1}$$

$$=\bar{Y}(1+e_1)(1+e_3)(1+\theta e_2+(1-\theta)e_3)^{-1}$$

$$=\bar{Y}(1+e_1)(1+e_3)\left(1-\theta e_2-(1-\theta)e_3+\{\theta e_2+(1-\theta)e_3\}^2-.....\right)$$

$$=\bar{Y}\left[1+e_1+\theta\left(e_3-e_2-e_1e_2+e_1e_3+(1-2\theta)e_2e_3+\theta e_2^2-(1-\theta)e_3^2\right)\right]$$

(6) The bias of the estimator T_{V8} is

$$B(T_{V8})=\bar{Y}\left(L-\frac{1}{n}\right)\left(\theta^2C_X^2-\theta\rho C_YC_X\right) \qquad \text{…(4.6)}$$

Proof: $$B(T_{V8})=E\left[T_{V8}-\bar{Y}\right]=\bar{Y}E\left[1+e_1+\theta\left(e_3-e_2-e_1e_2+e_1e_3+(1-2\theta)e_2e_3+\theta e_2^2-(1-\theta)e_3^2\right)-1\right]$$

$$=\bar{Y}\left(L-\frac{1}{n}\right)\left(\theta^2C_X^2-\theta\rho C_YC_X\right)$$

(7) Mean squared error of T_{V8} is

$$M(T_{V8})=\bar{Y}^2\left[\left(L-\frac{1}{n'}\right)C_Y^2+\left(L-\frac{1}{n}\right)\left(\theta^2C_X^2-2\theta\rho C_YC_X\right)\right] \qquad \text{…(4.7)}$$

Proof: $$M(T_{V8})=E\left[T_{V8}-\bar{Y}\right]^2=\bar{Y}^2E\left[1+e_1+\theta\left(e_3-e_2-e_1e_2+e_1e_3+(1-2\theta)e_2e_3+\theta e_2^2-(1-\theta)e_3^2\right)-1\right]^2$$

$$=\bar{Y}^2E\left[e_1^2+\theta^2\left(e_3^2+e_2^2-2e_2e_3\right)+2\theta\left(e_1e_3-e_1e_2\right)\right]$$

$$=\bar{Y}^2\left[\left(L-\frac{1}{n'}\right)C_Y^2+\left(L-\frac{1}{n}\right)\left(\theta^2C_X^2-2\theta\rho C_YC_X\right)\right]$$

(8) The minimum m.s.e. of T_{V8} is

$$\left[M\left(T_{V8}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{1}{n}\right)\rho^2\right]S_Y^2 \text{ when } \theta=\rho\frac{C_Y}{C_X} \qquad \text{…(4.8)}$$

Proof: By differentiating (4.7) with respect to θ then equate to zero

$$\frac{d}{d\theta}\left[M\left(T_{V8}\right)\right]=0 \Rightarrow \theta=\rho\frac{C_Y}{C_X}$$

After replacing value of θ in (4.7), we obtained

$$\left[M\left(T_{V8}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{1}{n}\right)\rho^{2}\right]S_{Y}^{2}$$

Theorem 4.3:

(9) The estimator T_{V9} in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$T_{V9}=\overline{Y}\left[1+e_1+\varphi\left(e_3^{'}-e_3+e_1e_3^{'}-e_1e_3\right)-\left(1+2\varphi\right)e_3e_3^{'}+\varphi e_3^{2}+\left(1+\varphi\right)e_3^{'2}\right] \qquad ...(4.9)$$

Proof: $$T_{V9}=\frac{\overline{y}_1\overline{x}^{'}}{\varphi\overline{x}_n+\left(1-\varphi\right)\overline{x}^{'}}=\frac{\overline{Y}\,\overline{X}\left(1+e_1\right)\left(1+e_3^{'}\right)}{\varphi\overline{X}\left(1+e_3\right)+\left(1-\varphi\right)\overline{X}\left(1+e_3^{'}\right)}=\overline{Y}\left(1+e_1\right)\left(1+e_3^{'}\right)\left(1+e_3^{'}+\varphi e_3-\varphi e_3^{'}\right)^{-1}$$

$$=\overline{Y}\left(1+e_1\right)\left(1+e_3^{'}\right)\left(1+\varphi e_3+\left(1-\varphi\right)e_3^{'}\right)^{-1}=\overline{Y}\left(1+e_1\right)\left(1+e_3^{'}\right)\left(1-\varphi e_3-\left(1-\varphi\right)e_3^{'}+\left\{\varphi e_3+\left(1-\varphi\right)e_3^{'}\right\}^{2}-.....\right)^{-1}$$

$$=\overline{Y}\left[1+e_1+\varphi\left(e_3^{'}-e_3+e_1e_3^{'}-e_1e_3\right)-\left(1+2\varphi\right)e_3e_3^{'}+\varphi e_3^{2}+\left(1+\varphi\right)e_3^{'2}\right]$$

(10) The bias of the estimator T_{V9} is

$$B\left(T_{V9}\right)=\overline{Y}\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(\varphi_2^{2}C_X^{2}-\varphi_2\rho C_YC_X\right) \qquad \ldots(4.10)$$

Proof: $$B\left(T_{V9}\right)=E\left[T_{V9}-\overline{Y}\right]$$

$$=\overline{Y}E\left[1+e_1+\varphi\left(e_3^{'}-e_3+e_1e_3^{'}-e_1e_3\right)-\left(1+2\varphi\right)e_3e_3^{'}+\varphi e_3^{2}+\left(1+\varphi\right)e_3^{'2}-1\right]$$

$$=\overline{Y}\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(\varphi_2^{2}C_X^{2}-\varphi_2\rho C_YC_X\right)$$

(11) Mean squared error of T_{V9} is

$$M\left(T_{V9}\right)=\overline{Y}^{2}\left[\left(L-\frac{1}{n'}\right)C_Y^{2}+\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(\varphi_2^{2}C_X^{2}-2\varphi_2\rho C_YC_X\right)\right] \qquad \ldots(4.11)$$

Proof: $$M\left(T_{V9}\right)=E\left[T_{V9}-\overline{Y}\right]^{2}$$

$$=\overline{Y}^{2}E\left[1+e_1+\varphi\left(e_3^{'}-e_3+e_1e_3^{'}-e_1e_3\right)-\left(1+2\varphi\right)e_3e_3^{'}+\varphi e_3^{2}+\left(1+\varphi\right)e_3^{'2}-1\right]^{2}$$

$$=\overline{Y}^{2}E\left[e_1^{2}+\varphi_2^{2}\left(e_3^{'2}+e_3^{2}-2e_3e_3^{'}\right)+2\varphi_2\left(e_1e_3^{'}-e_1e_3\right)\right]$$

$$=\overline{Y}^{2}\left[\left(L-\frac{1}{n'}\right)C_Y^{2}+\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(\varphi_2^{2}C_X^{2}-2\varphi_2\rho C_YC_X\right)\right]$$

(12) The minimum m.s.e. of T_{V9} is

$$\left[M\left(T_{V9}\right)\right]_{Min}=\left[L-\left(\frac{1}{n'}\right)-\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\rho^{2}\right]S_Y^{2}\ \text{ when }\ \varphi=\rho\frac{C_Y}{C_X} \qquad \ldots(4.12)$$

Proof: By differentiating (4.11) with respect to φ then equate to zero

$$\frac{d}{d\varphi}\left[M\left(T_{V9}\right)\right]=0\ \Rightarrow\varphi=\rho\frac{C_Y}{C_X}$$

After replacing value of φ in (4.11), we obtained

$$\left[M\left(T_{V9}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\rho^{2}\right]S_Y^{2}$$

Theorem 4.4:

(13) The estimator T_{V10} in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$T_{V10}=\overline{Y}\left[1+e_1+\psi\left(e_3^{'}-e_2+e_1e_3^{'}-e_1e_2+\psi e_2^{2}+\psi e_3^{'2}-e_3^{'2}-e_2e_3^{'}\right)\right] \qquad \ldots(4.13)$$

Proof: $$T_{V10}=\frac{\overline{y}_1\overline{x}^{'}}{\psi\,\overline{x}_1+\left(1-\psi\right)\overline{x}^{'}}=\frac{\overline{Y}\,\overline{X}\left(1+e_1\right)\left(1+e_3^{'}\right)}{\psi\overline{X}\left(1+e_2\right)+\left(1-\psi\right)\overline{X}\left(1+e_3^{'}\right)}=\overline{Y}\left(1+e_1\right)\left(1+e_3^{'}\right)\left(1+e_3^{'}+\psi e_2-\psi e_3^{'}\right)^{-1}$$

$$=\bar{Y}(1+e_1)(1+e_3^{'})(1+\psi e_2+(1-\psi)e_3^{'})^{-1}=\bar{Y}(1+e_1)(1+e_3^{'})(1-\psi e_2-(1-\psi)e_3^{'}+\{\psi e_2+(1-\psi)e_3^{'}\}^2-\ldots)$$
$$=\bar{Y}\left[1+e_1+\psi\left(e_3^{'}-e_2+e_1e_3^{'}-e_1e_2+\psi e_2^2+\theta_3 e_3^{'2}-e_3^{'2}-e_2e_3^{'}\right)\right]$$

(14) The bias of the estimator T_{V10} is

$$B(T_{V10})=\bar{Y}\left(\psi^2\left(L-\frac{1}{N}\right)C_X^2-2\psi\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2-\psi\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho C_Y C_X\right) \quad \ldots(4.14)$$

Proof: $B(T_{V10})=E\left[T_{V10}-\bar{Y}\right]=\bar{Y}E\left[1+e_1+\psi\left(e_3^{'}-e_2+e_1e_3^{'}-e_1e_2+\psi e_2^2+\psi e_3^{'2}-e_3^{'2}-e_2e_3^{'}\right)-1\right]$

$$=\bar{Y}\left(\psi^2\left(L-\frac{1}{N}\right)C_X^2-2\psi\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2-\psi\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho C_Y C_X\right)$$

(15) Mean squared error of T_{V10} is

$$M(T_{V10})=\bar{Y}^2\left[\left(L-\frac{1}{n'}\right)C_Y^2+\left(L-\frac{2}{n'}+\frac{1}{N}\right)\left(\psi_3^2C_X^2-2\psi\rho C_Y C_X\right)\right] \quad \ldots(4.15)$$

Proof: $M(T_{V10})=E\left[T_{V10}-\bar{Y}\right]^2$

$$=\bar{Y}^2E\left[1+e_1+\psi_3\left(e_3^{'}-e_2+e_1e_3^{'}-e_1e_2+\psi e_2^2+\psi e_3^{'2}-e_3^{'2}-e_2e_3^{'}\right)-1\right]^2$$
$$=\bar{Y}^2E\left[e_1^2+\psi^2\left(e_3^{'2}+e_2^2-2e_2e_3^{'}\right)+2\psi\left(e_1e_3^{'}-e_1e_2\right)\right]$$
$$=\bar{Y}^2\left[\left(L-\frac{1}{n'}\right)C_Y^2+\left(L-\frac{2}{n'}+\frac{1}{N}\right)\left(\psi_3^2C_X^2-2\psi\rho C_Y C_X\right)\right]$$

(16) The minimum m.s.e. of T_{V10} is

$$\left[M(T_{V10})\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_Y^2 \text{ when } \psi=\rho\frac{C_Y}{C_X} \quad \ldots(4.16)$$

Proof: By differentiating (4.15) with respect to ψ then equate to zero

$$\frac{d}{d\psi}\left[M(T_{V10})\right]=0 \Rightarrow \psi=\rho\frac{C_Y}{C_X}$$

After replacing value of ψ in (4.15), we obtained

$$\left[M(T_{V10})\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_Y^2$$

V. Comparisons

In this section we derived the conditions under which the suggested estimators are superior to the Ahmed et al. (2006).

(1) $D_7=\min\left[M(t_7)\right]-\min\left[M(T_{V7})\right]$

$$=\left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho^2\right]S_Y^2-\left[\left(L-\frac{1}{n^{'}}\right)-\left(L-\frac{2}{n^{'}}+\frac{1}{N}\right)\rho^2\right]S_y^2$$
$$=\left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(L-\frac{1}{n^{'}}\right)\right]S_Y^2+\left[-\left(\frac{1}{n_1}-\frac{1}{N}\right)+\left(L-\frac{2}{n^{'}}-\frac{1}{N}\right)\right]\rho^2S_y^2$$
$$=\left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]S_Y^2-\left[\frac{1}{n_1}-\frac{2}{N}-L+\frac{2}{n'}\right]\rho^2S_y^2$$

(T_{V7}) is better than t_7, if

$$D_7>0 \Rightarrow \rho^2<\frac{\left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]}{\left[\frac{1}{n_1}-\frac{2}{N}-L+\frac{2}{n'}\right]} \Rightarrow \rho^2<1+\frac{nn_1^2(N-n')(N-1)}{n'N(N-n)(n-n_1)-2nn_1^2(N-n')(N-1)}$$

$$\Rightarrow \rho < \pm\sqrt{1+Q} \qquad \Rightarrow -\sqrt{1+Q} < \rho < +\sqrt{1+Q}$$

where $Q > 1 \Rightarrow nn_1^2(N-n')(N-1) > n'N(N-n)(n-n_1) - 2nn_1^2(N-n')(N-1)$

(2) $D_8 = \min[M(t_8)] - \min[M(T_{V8})]$

$$= \left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(\frac{1}{n_1}-\frac{1}{n}\right)\rho^2\right]S_Y^2 - \left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{1}{n}\right)\rho^2\right]S_y^2 = \left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]S_Y^2 - \left[\frac{1}{n_1}-L\right]\rho^2 S_y^2$$

(T_{V8}) is better than t_8, if

$$D_8 > 0 \qquad \Rightarrow \left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]S_Y^2 - \left[\frac{1}{n_1}-L\right]\rho^2 S_y^2 > 0$$

$$\Rightarrow \left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]S_Y^2 - \left[\frac{1}{n_1}-L\right]\rho^2 S_y^2 > 0 \Rightarrow \rho^2 < \frac{\left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]}{\left[\frac{1}{n_1}-L\right]}$$

$$\Rightarrow \rho^2 < \left[\frac{(N-n)(n-n_1)n'N}{nn_1^2(N-1)(N-n')}\right]^{-1} - 1 \Rightarrow \rho < \pm\sqrt{\frac{1-M}{M}}$$

where $M < 1 \Rightarrow (N-n)(n-n_1)n'N < nn_1^2(N-1)(N-n')$

(3) $D_9 = \min[M(t_9)] - \min[M(T_{V9})]$

$$= \left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(\frac{1}{n}-\frac{1}{N}\right)\rho^2\right]S_Y^2 - \left[\left(L-\frac{1}{n'}\right)-\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_y^2$$

$$= \left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(L-\frac{1}{n'}\right)\right]S_Y^2 + \left[-\left(\frac{1}{n}-\frac{1}{N}\right)+\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\right]\rho^2 S_y^2$$

$$= \left[\frac{1}{n_1}-\frac{1}{N}-L-\frac{1}{n'}\right]S_Y^2 - 2\left[\frac{1}{n'}-\frac{1}{N}\right]\rho^2 S_y^2$$

(T_{V9}) is better than t_9, if

$$D_9 > 0 \Rightarrow \left[\frac{1}{n_1}-\frac{1}{N}-L-\frac{1}{n'}\right]S_Y^2 - 2\left[\frac{1}{n'}-\frac{1}{N}\right]\rho^2 S_y^2 > 0 \Rightarrow \rho^2 < \frac{1}{2}\frac{\left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]}{\left[\frac{1}{n'}-\frac{1}{N}\right]}$$

$$\Rightarrow \rho^2 < \frac{1}{2} - \frac{(N-n)(n-n_1)n'N}{nn_1^2(N-1)(N-n')} \qquad \Rightarrow \rho < \pm\sqrt{\frac{1}{2}-M} \qquad \Rightarrow -\sqrt{\frac{1}{2}-M} < \rho < +\sqrt{\frac{1}{2}-M}$$

where $M < \frac{1}{2} \Rightarrow 2(N-n)(n-n_1)n'N < nn_1^2(N-1)(N-n')$

(4) $D_{10} = \min[M(t_{10})] - \min[M(T_{V10})]$

$$= \left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(\frac{1}{n_1}-\frac{1}{N}\right)\rho^2\right]S_Y^2 - \left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_y^2$$

$$= \left[\left(\frac{1}{n_1}-\frac{1}{N}\right)-\left(L-\frac{1}{n'}\right)\right]S_Y^2 + \left[-\left(\frac{1}{n_1}-\frac{1}{N}\right)+\left(L-\frac{2}{n'}-\frac{1}{N}\right)\right]\rho^2 S_y^2$$

$$= \left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]S_Y^2 - \left[\frac{1}{n_1}-\frac{2}{N}-L+\frac{2}{n'}\right]\rho^2 S_y^2$$

(T_{V10}) is better than t_{10}, if $D_{10} > 0$

$$\Rightarrow \rho^2 < \frac{\left[\frac{1}{n_1} - \frac{1}{N} - L + \frac{1}{n'}\right]}{\left[\frac{1}{n_1} - \frac{2}{N} - L + \frac{2}{n'}\right]} \qquad \Rightarrow \rho^2 < 1 + \frac{nn_1^2(N-n')(N-1)}{n'N(N-n)(n-n_1) - 2nn_1^2(N-n')(N-1)}$$

$$\Rightarrow \rho < \pm\sqrt{1+Q} \qquad \Rightarrow -\sqrt{1+Q} < \rho < +\sqrt{1+Q}$$

where $Q > 1 \Rightarrow nn_1^2(N-n')(N-1) > n'N(N-n)(n-n_1) - 2nn_1^2(N-n')(N-1)$

VI. Numerical Illustrations

We consider two populations A and B, first one is the artificial population of size N = 200 [source Shukla et al. (2009a)] and another one is from Ahmed et al. (2006) with the following parameters:

Table 6.1 Parameters of Populations A and B

Population	N	$\overline{Y}$	$\overline{X}$	S_Y^2	S_X^2	ρ	C_X	C_Y
A	200	42.485	18.515	199.0598	48.5375	0.8652	0.3763	0.3321
B	8306	253.75	343.316	338006	862017	0.522231	2.70436	2.29116

Let n' = 60, n = 40, n_1 = 35 for population **A** and n' = 2000, n = 500, n_1 = 450 for population **B** respectively. Then the bias and M.S.E of suggested estimators (using the expressions of bias and m.s.e. of Section 5) and other existing estimators with Ahmed et al. (2006) methods are given in table 6.2 and 6.3 for population **A** and **B** respectively.

Table 6.2 Bias and MSE for Population A and B

Estimators	**Population A**		**Population B**	
	Bias	**MSE**	**Bias**	**MSE**
T_{V7}	0	2.338387	0	458.4694
T_{V8}	-0.000001	1.841686	0.000003	561.7505
T_{V9}	0.000001	2.882792	0.000001	478.9972
T_{V10}	-0.025350	2.338387	-0.347570	458.4694

Table 6.3 Bias and MSE for Population A and B for Ahmed et al. (2006)

Estimators	**Population A**		**Population B**	
	Bias	**MSE**	**Bias**	**MSE**
$\bar{y}_r$	0	4.692124	0	710.4302
$\bar{y}_{RAT}$	0.005080	4.908211	0.22994	768.7752
$\bar{y}_{COMP}$	0.003879	4.188044	0.050411	689.9429
t_7	0	1.179736	0	516.6780
t_8	-0.000001	4.159944	0.000003	689.9452
t_9	-0.000006	1.711916	0.000002	537.1631
t_{10}	-0.000008	1.179736	0.000003	516.6780

The sampling efficiency of suggested estimators over Ahmed et al. (2006) is defined as:

$$E_i = \frac{Opt[M(T_{Vi})]}{Opt[M(t_i)]}; \qquad i = 7,8,9,10 \qquad \ldots(6.1)$$

The efficiency for population **A** and population **B** are given in table 6.4

Table 6.4 Efficiency for Population A and B over Ahmed et al. (2006)

Efficiency	**Population A**	**Population B**
E_7	1.982128	0.887341
E_8	0.442719	0.814196
E_9	1.683957	0.891717
E_{10}	1.982128	0.887341

VII. Discussion

The idea of two-phase sampling is used while considering, the auxiliary population mean is unknown and numbers of available observations are considered as random variable. Some strategies are suggested in Section 3 and the estimator of population mean derived. Properties of derived estimators like bias and m.s.e are discussed in the Section 4. The optimum value of parameters of suggested estimators is obtained as well in same section. Ahmed et al. (2006) estimators are considered for comparison purpose and two populations **A** and **B** considered for numerical study first one from Shukla et al. (2009) and another one is Ahmed et al. (2006). The sampling efficiency of suggested estimator over Ahmed et al. (2006) is obtained and suggested strategy is found very close with Ahmed et al. (2006) when $\bar{X}$ is not known.

VIII. Conclusions

The proposed estimators are useful when some observations are missing in the sampling and population mean of auxiliary information is unknown. For population **A** proposed estimators T_{v8} are found to be more efficient than the existing estimators. For population **B** proposed estimators T_{v7}, T_{v8}, T_{v9} and T_{v10} are found to be more efficient than the existing estimators.

REFERENCES

[1] Ahmed, M.S., Al-Titi, O., Al-Rawi, Z. and Abu-Dayyeh, W. (2006): *Estimation of a population mean using different imputation methods*, Statistics in Transition, 7, 6, 1247-1264.
[2] Hansen, M. H. and Hurwitz, W. N. (1946): *The problem of non-response in Sample Surveys*. Journal of the American Statistical Association, 41, 236, 517-529.
[3] Heitjan, D. F. and Basu, S. (1996): *Distinguishing 'Missing at random' and 'Missing completely at random'*. The American Statistician, 50, 207-213.
[4] Kalton, G., Kasprzyk, D. and Santos, R. (1981): *Issues of non-response and imputation in the Survey of Income and Program Participation*. Current Topics in Survey Sampling, (D. Krewski, R. Platek and J.N.K. Rao, eds.), 455-480, Academic Press, New York.
[5] Little, R.J.A., and Rubin, D.B. (1987): *Statistical Analysis with Missing Data*, John Wiley and Sons, NY.
[6] Lee, H., Rancourt, E. and Sarndal, C. E. (1994): *Experiments with variance estimation from survey data with imputed values*. Journal of official Statistics, 10, 3, 231-243.
[7] Lee, H., Rancourt, E. and Sarndal,C. E. (1995): *Variance estimation in the presence of imputed data for the generalized estimation system.* Proc. of the American Statist. Assoc. (Social Survey Research Methods Section), 384-389.
[8] Rao, J. N. K. and Sitter, R. R. (1995): *Variance estimation under two-phase sampling with application to imputation for missing data*, Biometrica, 82, 453-460.
[9] Rubin, D. B. (1976): *Inference and missing data.* Biometrica, 63, 581-593.
[10] Sande, I. G. (1979): *A personal view of hot deck approach to automatic edit and imputation*. Journal Imputation Procedures. Survey Methodology, 5, 238-246.
[11] Shukla, D. (2002): *F-T estimator under two-phase sampling*, Metron, 59, 1-2, 253-263.
[12] Shukla, D. and Dubey, J. (2001): *Estimation in mail surveys under PSNR sampling scheme*. Ind. Soc. Ag. Stat., 54, 3, 288-302.
[13] Shukla, D. and Dubey, J. (2004): *On estimation under post-stratified two phase non-response sampling scheme in mail survey*. International Journal of Management and Systems, 20, 3, 269-278.
[14] Shukla, D. and Dubey, J. (2006): *On earmarked strata in post-stratification.* Statistics in Transition, 7, 5, 1067-1085.
[15] 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.
[16] Shukla, D., Thakur, N. S., Pathak, S. and Rajput D. S. (2009): *Estimation of mean with imputation of missing data using factor- type estimator in two-phase sampling*, Statistics in Transition, 10, 3, 397-414.
[17] Shukla, D., Thakur, N. S., Thakur, D. S. (2009a):*Utilization of non-response auxiliary population mean in imputation for missing observations*, Journal of Reliability and Statistical Studies, 2, 28-40.
[18] Shukla, D., Thakur, N. S., Thakur, D. S. and Pathak, S. (2011): *Linear combination based imputation method for missing data in sample*, International Journal of Modern Engineering Research, 1, 2, 580-596.
[19] 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.
[20] Singh, S. and Horn, S. (2000): *Compromised imputation in survey sampling*, Metrika, 51, 266-276.
[21] Thakur, N. S., Yadav, K. and Pathak, S. (2011): *Estimation of mean in presence of missing data under two-phase sampling scheme*, Journal of Reliability and Statistical Studies, 4, 2, 93-104.
[22] Thakur, N. S., Yadav, K. and Pathak, S. (2012): *Some imputation methods in double sampling scheme for estimation of population mean*, International Journal of Modern Engineering Research, 2, 1, pp-200-207.