Journal of Reliability and Statistical Studies; ISSN (Print): 0974-8024, (Online):2229-5666
Vol. 5, Issue 2 (2012): 21-31

IMPUTATION USING REGRESSION ESTIMATORS FOR ESTIMATING POPULATION MEAN IN TWO-PHASE SAMPLING

Narendra Singh Thakur, Kalpana Yadav[1] and Sharad Pathak[2]
[1] Centre for Mathematical Sciences (CMS), Banasthali University, Rajasthan, India, Pin-304022
2 Department of Mathematics and Statistics, Dr. H. S. Gour Central University, Sagar (M.P.), India, Pin-470003.
E Mail:[2]sharadpathakstats@yahoo.com.

(Received May 10, 2012)

Abstract

This paper presents the estimation of mean in presence of missing data under two-phase sampling design using regression estimators as a tool for imputation while the size of responding (R_1) and non-responding (R_2) group is considered as a random variable. The bias and mean squared error of suggested estimators are derived in the form of population parameters using the concept of large sample approximation. Numerical study is performed over two populations by using the expressions of bias and mean squared error and efficiency compared with existing estimators.

Key Words: Estimation, Missing data, Regression estimators, Bias, Mean squared error (MSE), Two-phase sampling, SRSWOR, Large sample approximations.

1. Introduction

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 are used 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 for mean and variance estimation. Shukla et al. (2009a) have 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. Shukla et al. (2011) have discussed on the linear combination based imputation method for missing data for auxiliary information in the sample. Shukla et al. (2012) have given the concept of use of the mixture of $\bar{X}$, $\bar{X}_1$, $\bar{X}_2$ in imputation for missing observations of auxiliary information due to non-response. Thakur et al. (2011) have presented the estimation of mean in presence of missing data under two-phase sampling scheme. Thakur et al. (2012) and have given some imputation methods in double sampling scheme for estimation of population mean and Shukla et al. (2009) proposed the estimation of mean with imputation of missing data using factor- type estimator in two-phase sampling. Shukla et al. (2012) have discussed on a transformed estimator for estimation of population mean with missing data in sample-surveys. Shukla et al. (2012a) an estimator for mean estimation in presence of measurement error. Shukla et al. (2012b) have discussed on estimation of population mean using two auxiliary sources in sample surveys. Thakur et al. (2012a) advocated on mean estimation with imputation in two-phase sampling. Shukla et al. (2012) discussed on estimation of mean using improved ratio-cum-product type estimator with imputation for missing data. Some other useful and interesting contributions are due to Bhushan et al. (2008), Banerjie and Tiwari (2011), Singh et al. (2012).

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,* and then the procedure is known as *post-stratification*. Sukhatme et al. (1984) advocated that post-stratification technique is as precise as the stratified sampling technique under proportional allocation if the sample size is large enough.

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 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 chapter:

$\bar{x}_n, \bar{y}_n$: 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:

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

$$Q=\frac{nn_1^2(N-n')(N-1)}{n'N(N-n)(n-n_1)-2nn_1^2(N-n')(N-1)}.$$

2. Large sample approximation

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:

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

Also, $$E(e_1^2)=E\left[\left(\frac{\bar{y}_1-\bar{Y}}{\bar{Y}}\right)^2\middle|n_1\right]=\left(E\left(\frac{1}{n_1}\right)-\frac{1}{n^{'}}\right)C_y^2=\left(L-\frac{1}{n'}\right)C_y^2$$

Similarly, $$E(e_2^2)=\left(L-\frac{1}{n'}\right)C_X^2;\ E(e_3^2)=\left(\frac{1}{n}-\frac{1}{n'}\right)C_X^2;\ E(e'^2_3)=\left(\frac{1}{n'}-\frac{1}{N}\right)C_X^2;$$

$$E(e_1e_2)=E(e_1e_2/n_1)=E\left[\left(\frac{(\bar{y}_1-\bar{Y})(\bar{x}_1-\bar{X})}{\bar{Y}\,\bar{X}}\right)\middle|n_1\right]$$

$$=\left(E\left(\frac{1}{n_1}\right)-\frac{1}{n^{'}}\right)\rho C_yC_X=\left(L-\frac{1}{n'}\right)\rho C_yC_X$$

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

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

3. 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) $$y_{V4i}=\begin{cases} y_i & if \quad i\in R_1 \\ \left[\bar{y}_1+a(x_i-\bar{x}_1)\right] & if \quad i\in R_2 \end{cases} \qquad \ldots(3.1)$$

where a is a constant, such that the variance of the estimator is minimum.
Under this, the point estimator of $\bar{Y}$ is

$$T_{V4}=\bar{y}_1+a(\bar{x}_n-\bar{x}_1) \qquad \ldots(3.2)$$

$$\textbf{(2)}\quad y_{V5i} = \begin{cases} y_i & if \quad i \in R_1 \\ \\ \bar{y}_1 + \dfrac{b}{(1-W_1)}\left(\bar{x}^{'} - \bar{x}_n\right) & if \quad i \in R_2 \end{cases} \qquad \ldots(3.3)$$

where b is a constant, such that the variance of the estimator is minimum.
Under this, the point estimator of $\bar{Y}$ is

$$T_{V5} = \bar{y}_1 + b\left(\bar{x}^{'} - \bar{x}_n\right) \qquad \ldots(3.4)$$

$$\textbf{(3)}\quad y_{V6i} = \begin{cases} y_i & if \quad i \in R_1 \\ \\ \bar{y}_1 + \dfrac{c}{(1-W_1)}\left(\bar{x}^{'} - \bar{x}_1\right) & if \quad i \in R_2 \end{cases} \qquad \ldots(3.5)$$

where c is a constant, such that the variance of the estimator is minimum.
Under this, the point estimator of $\bar{Y}$ is

$$T_{V6} = \bar{y}_1 + c\left(\bar{x}^{'} - \bar{x}_1\right) \qquad \ldots(3.6)$$

4. Bias and mean squared error (MSE) of proposed methods

Let $B(.)$ and $M(.)$ denote the bias and mean squared error (MSE) 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_{V4} in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$T_{V4} = \bar{Y}(1+e_1) + a\bar{X}(e_3 - e_2) \qquad \ldots(4.1)$$

Proof: $T_{V4} = \bar{y}_1 + a\left(\bar{x} - \bar{x}_1\right)$

$$= \bar{Y}(1+e_1) + a\bar{X}(e_3 - e_2)$$

(2) T_{V4} is an unbiased estimator, i.e. $B[T_{V4}] = 0$ …(4.2)

Proof: $B(T_{V4}) = E\left[T_{V4} - \bar{Y}\right] = \bar{Y} - \bar{Y} = 0$

(3) The variance of T_{V4} upto first order of approximation could be written as

$$V(T_{V4}) = \left(L - \frac{1}{n'}\right)S_Y^2 + \left(L - \frac{1}{n}\right)\left(a^2 S_X^2 - 2a\rho S_Y S_X\right) \qquad \ldots(4.3)$$

Proof: $V(T_{V4}) = E\left[T_{V4} - \bar{Y}\right]^2 = E\left[\bar{Y}e_1 + a\bar{X}(e_3 - e_2)\right]^2$

$$= E\left[\bar{Y}^2 e_1^2 + a^2\bar{X}^2(e_3 - e_2)^2 + 2a\bar{Y}\bar{X}(e_3 - e_2)e_1\right]$$

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

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

(4) The minimum variance of the T_{V4} is

$$\left[V\left(T_{V4}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{1}{n}\right)\rho^2\right]S_Y^2,\ \text{when}\ \ a=\rho\frac{S_Y}{S_X} \qquad \ldots(4.4)$$

Proof: By differentiating (4.3) with respect to a and equate to zero

$$\frac{d}{da}\left[V\left(T_{V4}\right)\right]=0\ \Rightarrow a=\rho\frac{S_Y}{S_X}$$

After replacing value of a in (4.3), we obtained

$$\left[V\left(T_{V4}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(L-\frac{1}{n}\right)\rho^2\right]S_Y^2$$

Theorem 4.2

(5) The estimator T_{V5} in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$T_{V5}=\overline{Y}\left(1+e_1\right)+b\,\overline{X}\left(e_3^{'}-e_3\right) \qquad \ldots(4.5)$$

Proof: $T_{V5}=\overline{y}_1+b\left(\overline{x}^{'}-\overline{x}_n\right)=\overline{Y}\left(1+e_1\right)+b\,\overline{X}\left(e_3^{'}-e_3\right)$

(6) The estimator T_{V5} is unbiased, i.e. $B\left[T_{V5}\right]=0$ $\qquad \ldots(4.6)$

Proof: $B\left(T_{V5}\right)=E\ \left[T_{V5}-\overline{Y}\right]=\overline{Y}-\overline{Y}=0$

(7) The variance of T_{V5} is

$$V\left(T_{V5}\right)=\left(L-\frac{1}{n'}\right)S_Y^2+\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(b^2S_X^2-2b\,\rho S_YS_X\right) \qquad \ldots(4.7)$$

Proof:
$$V\left(T_{V5}\right)=E\left[T_{V5}-\overline{Y}\right]^2=E\left[\overline{Y}e_1+b\overline{X}\left(e_3^{'}-e_3\right)\right]^2$$
$$=E\left[\overline{Y}^2e_1^2+b^2\overline{X}^2\left(e_3^{'}-e_3\right)^2+2b\overline{Y}\overline{X}\left(e_3^{'}-e_3\right)e_1\right]$$
$$=E\left[\overline{Y}^2e_1^2+b^2\overline{X}^2\left(e_3^{'2}+e_3^2-2e_3e_3^{'}\right)+2b\overline{Y}\overline{X}\left(e_1e_3^{'}-e_1e_3\right)\right]$$
$$=\left[\left(L-\frac{1}{n'}\right)S_Y^2+\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\left(b^2S_X^2-2b\,\rho S_YS_X\right)\right]$$

(8) The minimum variance of the T_{V5} is

$$\left[V\left(T_{V5}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_Y^2 \qquad \ldots(4.8)$$

Proof: By differentiating (4.7) with respect to b and equate to zero

$$\frac{d}{db}\left[V\left(T_{V5}\right)\right]=0\ \Rightarrow b=\rho\frac{S_Y}{S_X}$$

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

$$\left[V\left(T_{V5}\right)\right]_{Min}=\left[\left(L-\frac{1}{n'}\right)-\left(\frac{1}{n}-\frac{2}{n'}+\frac{1}{N}\right)\rho^2\right]S_Y^2$$

Theorem 4.3

(9) The estimator T_{V6} in terms of e_1, e_2, e_3 and $e_3^{'}$ is :

$$T_{V6} = \overline{Y}(1+e_1) + c\overline{X}\left(e_3^{'} - e_2\right) \qquad \ldots(4.9)$$

Proof: $T_{V6} = \overline{y}_1 + c\left(\overline{x}^{'} - \overline{x}_1\right) = \overline{Y}(1+e_1) + c\,\overline{X}\left(e_3^{'} - e_2\right)$

(10) The bias estimator T_{V6} is $B[T_{V6}] = 0$ …(4.10)

Proof: $B(T_{V6}) = E\left[T_{V6} - \overline{Y}\right] = \overline{Y} - \overline{Y} = 0$

(11) The variance of T_{V6} is

$$[V(T_{V6})] = \left[\left(L - \frac{1}{n'}\right)S_Y^2 + \left(L - \frac{2}{n'} + \frac{1}{N}\right)\left(c^2 S_X^2 - 2c\rho S_Y S_X\right)\right] \qquad \ldots(4.11)$$

Proof: $V(T_{V6}) = E\left[T_{V6} - \overline{Y}\right]^2 = \left[\overline{Y}(1+e_1) + c\overline{X}\left(e_3^{'} - e_2\right) - Y\right]^2$

$$= E\left[\overline{Y}^2 e_1^2 + c\overline{X}^2\left(e_3^{'} - e_2\right)^2 + 2c\overline{Y}\overline{X}\left(e_3^{'} - e_2\right)e_1\right]$$

$$= E\left[\overline{Y}^2 e_1^2 + c\overline{X}^2\left(e_3^{'2} + e_2^2 - 2e_2 e_3^{'}\right) + 2c\overline{Y}\overline{X}\left(e_1 e_3^{'} - e_1 e_2\right)\right]$$

$$= \left[\overline{Y}^2\left(L - \frac{1}{n'}\right)C_Y^2 + b_3^2\overline{X}^2\left(\left(\frac{1}{n'} - \frac{1}{N}\right)C_X^2 + \left(L - \frac{1}{n'}\right)C_X^2 - 2\left(\frac{1}{n'} - \frac{1}{N}\right)C_X^2\right)\right.$$

$$\left. + 2b_3\overline{Y}\overline{X}\left(\left(\frac{1}{n'} - \frac{1}{N}\right)\rho C_Y C_X - \left(L - \frac{1}{n'}\right)\rho\, C_Y C_X\right)\right]$$

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

(12) The minimum variance of the T_{V6} is

$$[V(T_{V6})]_{Min} = \left[\left(L - \frac{1}{n'}\right) - \left(L - \frac{2}{n'} + \frac{1}{N}\right)\rho^2\right] S_Y^2 \qquad \ldots(4.12)$$

Proof: By differentiating (4.11) with respect to c and equate to zero

$$\frac{d}{dc}[V(T_{V6})] = 0 \Rightarrow c = \rho\frac{S_Y}{S_X}$$

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

$$[V(T_{V6})]_{Min} = \left[\left(L - \frac{1}{n'}\right) - \left(L - \frac{2}{n'} + \frac{1}{N}\right)\rho^2\right] S_Y^2$$

5. Comparison

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

(1) $D_4 = \min[M(t_4)] - \min[M(T_{V4})]$

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

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

(T_{V4}) is better than t_4,

if $D_4 > 0$

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

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

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

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

and $(N-n)(n-n_1)n'N<nn_1^2(N-1)(N-n')$

(2) $D_5=\min\left[M(t_5)\right]-\min\left[M(T_{V5})\right]$

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

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

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

(T_{V5}) is better than t_5, if $D_5 > 0$

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

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

$$\Rightarrow \rho<\pm\sqrt{\frac{1}{2}-M} \qquad \Rightarrow -\sqrt{\frac{1}{2}-M}<\rho<+\sqrt{\frac{1}{2}-M}$$

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

(3) $D_6=\min\left[M(t_6)\right]-\min\left[M(T_{V6})\right]$

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

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

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

(T_{V6}) is better than t_6, if $D_6 > 0 \Rightarrow \rho^2 < \dfrac{\left[\frac{1}{n_1}-\frac{1}{N}-L+\frac{1}{n'}\right]}{\left[\frac{1}{n_1}-\frac{2}{N}-L+\frac{2}{n'}\right]}$

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

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

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

6. Numerical Illustrations

We considered 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:

Population	N	$\bar{Y}$	$\bar{X}$	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

Table 6.1 Parameters of Populations A and B

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 MSE of suggested estimators (using the expressions of bias and MSE of Section 5) and Ahmed et al. (2006) methods are given in table 6.2 and 6.3 for population **A** and **B** respectively.

Estimators	**Population A**		**Population B**	
	Bias	**MSE**	**Bias**	**MSE**
T_{V4}	0	1.841686	0	561.7505
T_{V5}	0	2.882792	0	478.9972
T_{V6}	0	2.338387	0	458.4694

Table 6.2 Bias and MSE for Population A and B

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_4	0	4.159944	0	689.9452
t_5	0	1.711916	0	537.1631
t_6	0	1.179736	0	516.6780

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

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

$$E_i = \frac{Opt\left[M\left(T_{Vi}\right)\right]}{Opt\left[M\left(t_i\right)\right]}; \qquad i = 4,5,6 \qquad \ldots (6.1)$$

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

Efficiency	Population A	Population B
E_4	0.442719	0.814195
E_5	1.683957	0.891717
E_6	1.982128	0.887340

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

7. Discussion and Conclusions:

The form regression type estimators are used as a source of imputation under the setup of two-phase sampling under the assumption that, the auxiliary population mean is unknown and the sizes of the respondent and non-respondent group 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 MSE are discussed in the Section 4. The suggested estimators are unbiased and the optimum value of parameters of suggested estimators is obtained as well in same section. Other existing estimators are considered for comparison purpose and two populations **A** and **B** used for numerical study first one from Shukla et al. (2009) and another one is from 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. 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_{V4} is found to be more efficient than the existing estimators and T_{V5} and T_{V6} results are also very close with Ahmed estimators. For population **B** proposed estimators T_{V4}, T_{V5}, T_{V6} are found to be more efficient than the existing estimators.

Acknowledgement

Authors are thankful to the Editorial Board of JRSS and referees for their valuable suggestions and for recommending the manuscript for publication.

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), p. 1247-1264.
2. Banerjie, J. and Tiwari, N (2011). Improved ratio type estimator using jackknife method of estimation, Journal of Reliability and Statistical Studies, 4(1), p. 53-63.
3. Bhushan, S., Pandey, A and Katara, S. (2008). A class of estimators in double sampling using two auxiliary variables, Journal of Reliability and Statistical Studies, 1(1), p. 67-73.
4. Rao, J. N. K. and Sitter, R. R. (1995). Variance estimation under two-phase sampling with application to imputation for missing data, Biometrica, 82, p. 453-460.
5. Shukla, D. and Thakur, N. S. (2008). Estimation of mean with imputation of missing data using factor-type estimator, Statistics in Transition, 9 (1), p. 33-48.
6. Shukla, D., Pathak, S. and Thakur, N. S. (2012). A transformed estimator for estimation of population mean with missing data in sample-surveys, Journal of Current Engineering Research, 2 (1).
7. Shukla, D., Pathak, S. and Thakur, N. S. (2012a). An estimator for mean estimation in presence of measurement error, Research & Reviews: A Journal of Statistics, 2 (1).
8. Shukla, D., Pathak, S. and Thakur, N. S. (2012b). Estimation of population mean using two auxiliary sources in sample surveys, Statistics in Transition-new series, 13 (1), p. 21-36.
9. Shukla, D., Thakur, N. S. and Thakur, D. S. (2012). Utilization of mixture of $\overline{X}$, $\overline{X}_1$ and $\overline{X}_2$ in imputation for missing data in post-stratification, African Journal of Mathematics and Computer Science Research, 5 (4), p. 78-89.
10. 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), p. 397-414.
11. Shukla, D., Thakur, N. S., Pathak, S. and Yadav, K. (2012). Estimation of mean using improved ratio-cum-product type estimator with imputation for missing data, International Journal of Mathematics and Computational Methods in Science & Technology, 2, 1.
12. 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(1), p. 28-40.

13. 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), p. 580-596.
14. Singh, S. and Horn, S. (2000): Compromised imputation in survey sampling, Metrika, 51, p. 266-276.
15. Singh, R., Malik, S., Chaudhary, M. K., Verma, H. K. and Adewara, A. A. (2012). A general family of ratio-type estimators in systematic sampling, Journal of Reliability and Statistical Studies, 5(1), p. 73-82.
16. 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.
17. 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), p. 93-104.
18. 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), p. 200-207.
19. Thakur, N. S., Yadav, K. and Pathak, S. (2012a). Mean estimation with imputation in two-phase sampling, International Journal of Modern Engineering Research, 2(5), p. 3561-3571.