

Estimation of Population Mean in Two-Phase Stratified Sampling

Pitambar Das*

Abstract

In this paper, we have considered the problem related to estimation of population mean under stratified random sampling in presence of non-response and measurement error. Two different chain exponential to regression type estimators are formed separately for different structures of two-phase stratified sampling schemes. To reduce nuisance effect of random non responses in practical surveys, we have developed effective imputation techniques. The properties of the estimates are studied. The performances of the proposed strategy have been demonstrated through numerical evidences carried over the data set of natural population and population generated through simulation studies. We have also examined the performances of proposed strategy for different values of random non-response probabilities. Suitable recommendations have also been put forward to the survey statisticians for their application in real life.

Copyright © 2023 International Journals of Multidisciplinary Research Academy. All rights reserved.

Keywords:

Stratified Random Sampling;
Two-Phase Sampling;
Non-Response;
Imputation;
Mean Square Error.

Author correspondence:

Department of Mathematics, Netaji Nagar Day College,
170/436, N.S.C. Bose Road, Regent Estate, Kolkata-700092
Email: pitambardas.in@gmail.com

Mathematics Subject Classification: 62D05

1. Introduction

In sample survey, it is well known that to use of auxiliary variables to estimate unknown population variable in estimation stage in various sampling design. A number of sampling strategies utilize the advance information about an auxiliary variable. When such information is lacking, it is sometimes relatively cheap to take a large preliminary sample in which auxiliary variable alone is measured. The aim of this sample is to obtain a good estimate of the population mean of the auxiliary variable. This technique is known as double sampling or two-phase sampling. This paper mainly focuses on population mean estimators using different structures of two-phase stratified random sampling. Stratified sampling is one of the most widely used sampling techniques as it increases the precision of the estimate of the survey variable when units of the population are from different segment of population. For example, in socio-economic surveys, people may live in hospital, hostel, residential houses and jail etc. Thus, the problem of estimation is different for these different segments of the population.

It is evident from practical surveys that, there are two types of non sampling errors: response error and non-response error. Response error or measurement error occurs when the reported value differs from the true value due to some over reporting, under reporting, memory failure by respondents etc. For example, in surveys regarding household consumption/expenditure where the respondents are asked to report their catalogue, there is a great likelihood that the respondents may fail to recall precisely how much they spent on

* Department of Mathematics, Netaji Nagar Day College, Kolkata-700092, India.

various items over the interval. Many researchers have studied measurement errors like Shalabh [7], Manisha and Singh [4, 5], Singh and Karpe [11], Shukla, Pathak and Thakur [13] etc. Similarly, non-response error occurs when the researcher fails to collect information on one or more than one unit of survey or when surveying human populations as people hesitate to respond in surveys. Statisticians have identified for some time that failure to account for the stochastic nature of incompleteness can spoil inference. Rubin [6] advocated three concepts: missing at random (MAR), observed at random (OAR) and parameter distribution (PD). Rubin defined “the data are MAR if the probability of the observed missingness pattern, given the observed and unobserved data, does not depend on the value of the unobserved data”. Heitzan and Basu [3] have distinguished the meaning of missing at random (MAR) and missing completely at random (MCAR) in a very nice way. It is important to reduce the nuisance effect of non-response in practical surveys. Imputation technique is used to substitute values for missing data for reducing impact of non-responses. Recent contribution on this line could be considered due to Singh and Horn [9], Singh and Deo [8] Arnab and Singh [1] etc.

Some time, the researcher faces situations where some measurement error and non-response occur at the same time while collecting information. Motivated with the above discussion, we have suggested two exponential type estimators for population mean favourable for two different structure of two-phase stratified random sampling in presence of random non-responses (MAR) and measurement errors. We have formulated regression type imputation technique to reduce nuisance effect of non-response in practical surveys. The superiority of the suggested technique over the conventional ones have been justified through numerical evidences carried over the data set of natural population as well as population generated through simulation studies. The outcomes of the proposed methodology have been recommended for application by survey statisticians.

2. Research Method

2.1. Sampling Scheme

We consider a finite population $U = \{1, 2, 3, \dots, N\}$ of N identifiable units divided into K homogeneous strata with the h^{th} stratum $\{h = 1, 2, \dots, K\}$ having N_h units such that $\sum_{h=1}^K N_h = N$. Let y and (x, z) be the study variable and two auxiliary variables respectively taking values y_{hi} and (x_{hi}, z_{hi}) respectively, for the unit $i = 1, 2, \dots, N_h$ of the h th stratum. We have defined the population variables as:

$\bar{Y} = \sum_{h=1}^K W_h \bar{y}_h$, $\bar{X} = \sum_{h=1}^K W_h \bar{x}_h$ and $\bar{Z} = \sum_{h=1}^K W_h \bar{z}_h$ be the population mean of the variables y , x and z respectively where

$\bar{y}_h = \sum_{i=1}^{N_h} \frac{y_{hi}}{N_h}$, $\bar{x}_h = \sum_{i=1}^{N_h} \frac{x_{hi}}{N_h}$, $\bar{z}_h = \sum_{i=1}^{N_h} \frac{z_{hi}}{N_h}$ be the corresponding stratum and $W_h = \frac{N_h}{N}$ is the stratum

weight. To estimate the population mean \bar{Y} we have taken two-phase stratified random sampling. The first phase sample S_{nh} of size n_h is drawn at random without replacement from the each stratum containing N_h units $\{h=1, 2, \dots, K\}$. The second phase sample is selected by the following two different procedures.

Case 1: Select a second phase sample S_{mh} of size m_h using SRSWOR from each first phase sample of size $n_h \{h = 1, 2, \dots, K\}$.

Case 2: Select a second phase sample S_{mh} of size m_h using SRSWOR from remaining units of the h^{th} stratum $\{h = 1, 2, \dots, K\}$ the population.

We are considering the practical situation having (i) non-response situation only and (ii) joint presence of non-response and measurement error.

The occurrences of random non-response situation considered for the study variable y and noted the responses of the auxiliary variable x in the following ways

(i) It is to be noted from the Case 1 of sampling structure that the second phase sample S_{mh} is drawn from the first phase sample $S_{nh} \{h=1, 2, \dots, K\}$. In first phase sample, we consider that the units are responding for the variable x and second phase sample is drawn from it. Hence, it units of the sample S_{mh} give complete response for the variable x .

(ii) It is to be noted from the Case 2 of sampling structure that the second phase sample S_{mh} is drawn independently of the first phase sample $S_{nh} \{h=1, 2, \dots, K\}$. Therefore, we consider that units of first phase sample give complete response while non-response occurs on the variable x in second phase sample.

We have considered complete response for the variable z at all the units of the population.

2.2. Non-response Probability Model

If random non-response situations occur at the second phase sample S_{mh} for the stratum $h \{h=1, 2, \dots, K\}$ of size m_h and $r_h \{r_h = 0, 1, 2, \dots, (m_h - 2)\}$ denotes the number of sampling units on which information could

not be collected due to random non-response (non responding set denoted by Ar^c), then the observations of the respective variables on which random non-response occur can be taken from the remaining $(m_h -$

r_h) responding units of the second phase sample (responding set denoted by Ar). It is assumed that r_h is less than $(m_h - 1)$, that is, $0 \leq r_h \leq (m_h - 2)$. We also assume that if p_h denotes the probability of a non-response among the $(m_h - 2)$ possible values of non-response, then r_h has the following discrete distribution

$$P(r_h) = \frac{(m_h - r_h)}{m_h q_h + 2p_h} {}^{m_h - 2}C_r p_h^{r_h} q_h^{m_h - 2 - r_h}$$

where $r_h = 0, 1, 2, \dots, (m_h - 2)$ and $h=1, 2, \dots, K$.

where $q_h = 1 - p_h$ and ${}^{m_h - 2}C_{r_h}$ denote the total number of ways of obtaining r non-responses out of the $(m_h - 2)$ total possible non-responses, for instance, see Singh and Joarder [10].

It is to be noted, the probability model, defined in equation (1), is free from actual data values; hence, can be considered as a model suitable for MAR situation.

We have defined following variables based on the responding part of the sample as

$$\bar{x}_{mh}^* = \frac{1}{(m_h - r_h)} \sum_{i=1}^{m_h - r_h} x_{hi}, \bar{y}_{mh}^* = \frac{1}{m_h - r_h} \sum_{i=1}^{m_h - r_h} y_{hi} : \text{Sample means of the respective variables based on}$$

the responding part of the second phase sample S_{mh} for stratum $h \{h=1, 2, \dots, K\}$.

Since, all the units related to the auxiliary variable z in sample S_{mh} are responding, therefore we consider

$$\bar{z}_{mh} = \frac{1}{m_h} \sum_{i=1}^{m_h} x_{hi} \text{ as sample mean of } z \text{ in second phase sample. Similarly, when units of the sample}$$

$S_{mh} \{h=1, 2, \dots, K\}$ give complete response for the variable x then we have respective sample mean

$$\bar{x}_{mh} = \frac{1}{m_h} \sum_{i=1}^{m_h} x_{hi}.$$

$$s_{x_{mh}}^{*2} = \frac{1}{(m_h - r_h - 1)} \sum_{i=1}^{m_h - r_h} (x_{hi} - \bar{x}_{mh}^*)^2 : \text{Sample variance of the variable } x \text{ based on the responding part of}$$

the second phase sample S .

$$s_{y_{mh}}^{*2} : \text{Sample variance of the study variable } y \text{ based on the responding part of the second phase sample } S.$$

2.3. Effect due to Measurement Error

Let (x^*, y^*, z^*) be the observed values and (X^*, Y^*, Z^*) be the true values on three characteristics (x, y, z) respectively. Let the measurement errors be (u, v, w) and are defined as follows:

$$u = y^* - Y^* \\ v = x^* - X^*$$

and

$$w = z^* - Z^*$$

The following notations are further used:

$$S_{yh} = \sqrt{\frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2}, S_{xh} = \sqrt{\frac{1}{N_h - 1} \sum_{i=1}^{N_h} (x_{hi} - \bar{X}_h)^2}, S_{zh} = \sqrt{\frac{1}{N_h - 1} \sum_{i=1}^{N_h} (z_{hi} - \bar{Z}_h)^2} : \text{population standard deviations for the variables } y, x \text{ and } z \text{ for the } h^{\text{th}} \text{ stratum } \{h = 1, 2, \dots, K\}.$$

$$C_{yh} = \frac{S_{yh}}{\bar{Y}_h}, C_{xh} = \frac{S_{xh}}{\bar{X}_h}, C_{zh} = \frac{S_{zh}}{\bar{Z}_h} : \text{coefficient of variations for the variables } y, x \text{ and } z \text{ for the } h^{\text{th}} \text{ stratum } \{h = 1, 2, \dots, K\}.$$

ρ_{yxh}, ρ_{yzh} and ρ_{xzh} : correlation coefficients between (y, x) , (y, z) , and (x, z) respectively in the h^{th} stratum.

$$\bar{x}_{nh} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi}, \bar{z}_{nh} = \frac{1}{n_h} \sum_{i=1}^{n_h} z_{hi}, \text{ sample mean for the respective variables for the } h^{\text{th}} \text{ stratum based on first phase sample } S_{nh}.$$

2.4. Imputation Technique

Since, information on the auxiliary variable z is readily available for the sample S_{mh} and motivated by the regression methods of imputation for estimating population mean as was suggested by Diana and Perri [2], we propose the following imputation method based on responding and non-responding units of the sample S_{mh} ($h=1, 2, \dots, K$) to estimate the population means \bar{Y}_h as

$$y_{m_{h,i}} = \begin{cases} \frac{m_h \bar{y}_{mh}^*}{(m_h - r_h)} + b_h (\bar{Z}_h - z_{hi}) & \text{if } i \in Ar \\ b_h (\bar{Z}_h - z_{hi}) & \text{if } i \in Ar^c \end{cases} \quad (1)$$

where b_h ($h=1, 2, \dots, K$) is a real scalar to be chosen suitably.

Under the above method of imputation, the estimator \hat{y}_{mh} for estimating \bar{Y}_h ($h=1, 2, \dots, K$) can be derived as

$$\hat{y}_{mh} = \frac{1}{m_h} \sum_{i \in S_{mh}} y_{m_{h,i}} = \frac{1}{m_h} \left[\sum_{i \in Ar} y_{m_{h,i}} + \sum_{i \in Ar^c} y_{m_{h,i}} \right]. \quad (2)$$

After simplification the estimator \hat{y}_{mh} takes the form of well-known regression estimator which is given as

$$\hat{y}_{mh} = \bar{y}_{mh}^* + b_h (\bar{Z}_h - \bar{z}_{mh}); (h=1, 2, \dots, K). \quad (3)$$

Similarly, we suggest regression type imputation technique based on responding and non-responding units of the sample S_{mh} ($h=1, 2, \dots, K$) to estimate the population means \bar{X}_h as

$$x_{m_{h,i}} = \begin{cases} \frac{m_h \bar{x}_{mh}^*}{(m_h - r_h)} + b'_h (\bar{Z}_h - z_{hi}) & \text{if } i \in Ar \\ b'_h (\bar{Z}_h - z_{hi}) & \text{if } i \in Ar^c \end{cases}$$

where b'_h ($h=1, 2, \dots, K$) is the real scalar.

After simplifying as above, we have \hat{x}_{mh} for estimating \bar{X}_h ($h=1, 2, \dots, K$) as

$$\hat{x}_{mh} = \bar{x}_{mh}^* + b'_h (\bar{Z}_h - \bar{z}_{mh}); (h=1, 2, \dots, K).$$

2.5. Formulation of Proposed Estimation Strategy

Motivated with the earlier work, we have developed the following class of chain exponential regression type estimators for population mean \bar{Y} based on the two different cases of two-phase stratified random sampling discussed in section 2 as presented below.

Estimator based on Case 1 of two-phase stratified random sampling has been suggested as

$$T_{p1} = \sum_{h=1}^K W_h \hat{y}_{mh} \exp\left(\frac{\bar{X}_{nh} - \bar{X}_{mh}}{\bar{X}_{nh} + \bar{X}_{mh}}\right) \quad (4)$$

Similarly, the estimator based on Case 2 of two-phase stratified random sampling has been constructed as

$$T_{p2} = \sum_{h=1}^K W_h \bar{y}_{mh}^* \exp\left(\frac{\bar{X}_{nh} - \hat{X}_{mh}}{\bar{X}_{nh} + \hat{X}_{mh}}\right) \quad (5)$$

2.6. Mean Square Error of Proposed Estimator T_{p1} and T_{p2}

Since, T_{p1} and T_{p2} are regression and chain exponential to regression type estimators. The mean square error

$M(\cdot)$ up to the first order of approximations are derived using large sample approximations given below:

$$\bar{y}_{mh}^* = \bar{Y}_h (1 + e_0), \bar{x}_{mh} = \bar{X}_h (1 + e_1), \bar{x}_{nh} = \bar{X}_h (1 + e_3), \bar{z}_{mh}^* = \bar{Z}_h (1 + e_2), \bar{x}_{mh}^* = \bar{X}_h (1 + e_4).$$

Using above transformations the estimators T_{p1} and T_{p2} may be represented by discarding higher order terms of e 's as

$$T_{p1} = \sum_{h=1}^K W_h \left(\bar{Y}_h (1 + e_0) - b_h \bar{Z}_h e_2 + \bar{Y}_h \frac{(e_3 - e_1)}{2} \right) \quad (6)$$

and

$$T_{p2} = \sum_{h=1}^K W_h \bar{Y}_h \left[1 + \frac{1}{2} (e_3 - e_4 + \frac{b'_h \bar{Z}_h}{\bar{X}_h} e_2) + e_0 \right] \quad (7)$$

Mean Square Errors (MSEs) of the estimators T_{p1} and T_{p2} are obtained in (i) non-response situation only and (ii) joint presence of non-response and measurement error up to first order of sample size as

(i) MSEs of T_{p1} and T_{p2} under non-response situation only

$$MSE(T_{p1}) = E(T_{p1} - \bar{Y})^2 = \sum_{h=1}^k W_h^2 \bar{Y}_h^2 [A_1 + b_1^2 \bar{Z}_h^2 B_1 - 2b_1 \bar{Y}_h \bar{Z}_h C_1] \quad (8)$$

and

$$\text{MSE}(T_{p2}) = E(T_{p2} - \bar{Y})^2 = \sum_{h=1}^k \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \left[A_2 + \frac{b_2^2 \bar{Z}^2}{4\bar{X}^2} B_2 + \frac{b_2 \bar{Z}}{\bar{X}} C_2 \right] \quad (9)$$

where

$$\begin{aligned} A_1 &= \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} \right) \frac{S_{y_h}^2}{\bar{Y}_h^2} + \frac{1}{4} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \left[\rho_{y_x h} C_{y_h} C_{x_h} - \frac{1}{4} C_{x_h}^2 \right] \\ B_1 &= \sum_{h=1}^k \bar{Y}_h^2 W_h^2 \left(\frac{1}{m_h} - \frac{1}{N_h} \right) C_{z_h}^2 \\ C_1 &= \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \rho_{y_z h} C_{y_h} C_{z_h} + \frac{1}{2} \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{n_h} \right) W_h^2 \bar{Y}_h^2 \rho_{x_z h} C_{x_h} C_{z_h} \\ A_2 &= \sum_{h=1}^k \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 C_{y_h}^2 + \frac{1}{4\bar{X}_h^2} \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} + \frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 C_{x_h}^2 \\ &\quad - \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \rho_{y_x h} C_{y_h} C_{x_h} \\ B_2 &= \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 C_{z_h}^2 \\ C_2 &= \left\{ \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \rho_{y_z h} C_{y_h} C_{z_h} - \frac{1}{2} \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \rho_{x_z h} C_{x_h} C_{z_h} \right\} \end{aligned}$$

(ii) MSEs of T_{p1} and T_{p2} under joint presence of non-response and measurement errors situation

$$\text{MSE}(T_{p1}) = E(T_{p1} - \bar{Y})^2 = \sum_{h=1}^k W_h^2 \bar{Y}_h^2 \left[\hat{A}_1 + b_1^2 \bar{Z}^2 \hat{B}_1 - 2b_1 \bar{Y} \bar{Z} C_1 \right] \quad (10)$$

and

$$\text{MSE}(T_{p2}) = E(T_{p2} - \bar{Y})^2 = \sum_{h=1}^k W_h^2 \bar{Y}_h^2 \left[\hat{A}_2 + \frac{b_2^2 \bar{Z}^2}{4\bar{X}^2} \hat{B}_2 + \frac{b_2 \bar{Z}}{\bar{X}} C_2 \right] \quad (11)$$

where $\hat{A}_1 = A_1 + A'_1$; $\hat{B}_1 = B_1 + B'_1$

$\hat{A}_2 = A_2 + A'_2$; $\hat{B}_2 = B_2 + B'_2$;

$$A'_1 = \sum_{h=1}^k \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} \right) \frac{S_U^2}{\bar{Y}_h^2} - \frac{1}{4} \left(\frac{1}{n_h} - \frac{1}{N_h} \right) \frac{S_V^2}{\bar{X}_h^2}$$

$$B'_1 = \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{N_h} \right) \frac{S_W^2}{\bar{Z}_h^2} ; \quad C'_1 = 0$$

$$A'_2 = \sum_{h=1}^k \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \frac{S_U^2}{\bar{Y}_h^2} + \frac{1}{4\bar{X}_h^2} \left(\frac{1}{m_h q_h + 2p_h} - \frac{1}{N_h} + \frac{1}{n_h} - \frac{1}{N_h} \right) W_h^2 \bar{Y}_h^2 \frac{S_V^2}{\bar{X}_h^2}$$

$$B'_2 = \sum_{h=1}^k \left(\frac{1}{m_h} - \frac{1}{N_h} \right) \frac{S_W^2}{\bar{Z}_h^2} ; \quad C'_2 = 0$$

Remark: It may be noted that the following expectations

$E(e_1 e_2) = E(e_2 e_3) = E(e_4 e_3) = E(e_0 e_3) = 0$ as in Case 2 of sampling scheme, we have drawn second phase sample independent of first phase sample.

2.7. Minimum Mean Square Error of Proposed Estimators T_{p1} and T_{p2}

It may be noted from equation (8) and (9) that the expressions for $M(T_{p1})$ and $M(T_{p2})$ depends on the values of b_h and b'_h ($h = 1, 2, \dots, K$) which are real constants. Therefore, we need to find the optimum values of b_h and b'_h which can minimize the MSEs of the estimators T_{p1} and T_{p2} respectively.

The optimum values of b_1 and b_2 are found as

$$b_h = \frac{\bar{Y}_h C_1}{\bar{Z}_h B_1} \quad \text{and} \quad b'_h = -\frac{2C_2 \bar{X}_h}{\bar{Z}_h B_2} ; h = (1, 2, \dots, K) \quad (12)$$

Substituting the optimum values of b_h and b'_h in from equation (12) into the equations (8) and (9), we have the minimum MSE of the estimators T_{p1} and T_{p2} in presence of non-response situation as

$$\text{Min. } M(T_{p1}) = \sum_{h=1}^k W_h^2 \bar{Y}_h^2 \left(A_1 - \frac{C_1^2}{B_1} \right) \quad (13)$$

$$\text{and Min. } M(T_{p2}) = \sum_{h=1}^k W_h^2 \bar{Y}_h^2 \left(A_2 - \frac{C_2^2}{B_2} \right) \quad (14)$$

Similarly, in joint presence of non-response and measurement error the expressions of Minimum MSE for the estimators T_{p1} and T_{p2} may be found as

$$\text{Min. } M(T_{p1}) = \sum_{h=1}^k W_h^2 \bar{y}_h^2 \left(\hat{A}_1 - \frac{\hat{C}_1^2}{\hat{B}_1} \right) \quad (15)$$

and

$$\text{Min. } M(T_{p2}) = \sum_{h=1}^k W_h^2 \bar{y}_h^2 \left(\hat{A}_2 - \frac{\hat{C}_2^2}{\hat{B}_2} \right) \quad (16)$$

3. Results and Analysis

3.1. Performance of the Proposed Strategy

It is important to investigate the performances of the proposed strategy. To compare the efficiency of the proposed estimator T_{p1} and T_{p2} , we have considered the natural sample mean estimator \bar{y}_{mh} (i.e., mean of the study variable y based on second phase sample S_{mh}) of \bar{Y}_h ($h=1, 2, \dots, K$) in absence of any non-response and measurement error.

Thus, the natural sample mean estimator \hat{y} of \bar{Y} is found as

$$\bar{y}_{mh} = \sum_{h=1}^k W_h \bar{y}_{mh} \quad (17)$$

The expression of variance of \hat{y} is found as:

$$V(\hat{y}) = \sum_{h=1}^k W_h^2 \bar{Y}_h^2 \left(\frac{1}{m_h} - \frac{1}{N_h} \right) \frac{S_{y_h}^2}{\bar{Y}_h^2} \quad (18)$$

Thus, to have a tangible idea about the performance of the estimators T_{p1} and T_{p2} , we have computed Percent Relative Losses in efficiencies L_1 and L_2 of the estimators T_{p1} and T_{p2} with respect to \hat{y} have been derived as

We have

$$L_1 = \frac{V(\hat{y}) - M(T_{p1})}{M(T_{p1})} \times 100 \quad (19)$$

$$\text{and } L_2 = \frac{V(\hat{y}) - M(T_{p2})}{M(T_{p2})} \times 100 \quad (20)$$

3.2. Numerical Illustration through Population Generated Artificially by Simulation Studies

An important aspect of simulation is that one builds a simulation model to replicate the actual system. Simulation allows comparison of analytical techniques and helps in concluding whether a newly developed technique is better than the existing ones. Motivated by Singh and Deo [8] and Singh *et al.* [12] who have been adopted the artificial population generation techniques, we have generated five sets of independent random numbers of size N ($N = 100$) namely $x'_{1k}, y'_{1k}, x'_{2k}, y'_{2k}$ and z'_k ($k=1, 2, 3, \dots, N$) from a standard normal distribution with the help of R-software. By varying the correlation coefficients ρ_{yx} and ρ_{xz} , we have generated the following transformed variables of the population U with the values of $\sigma_y^2 = 50$ (variance of y), $\mu_y = 10$ (mean of y), $\sigma_x^2 = 50$ (variance of x), $\mu_x = 10$ (mean of x), $\sigma_z^2 = 40$ (variance of z) and $\mu_z = 20$ (mean of z) by the following transformations

$$y_{1k} = \mu_y + \sigma_y \left[\rho_{xy} x'_{1k} + \left(\sqrt{1 - \rho_{xy}^2} \right) y'_{1k} \right]$$

$$x_{1k} = \mu_x + \sigma_x x'_{1k}$$

$$z_k = \mu_z + \sigma_z \left[\rho_{xz} x'_{1k} + \left(\sqrt{1 - \rho_{xz}^2} \right) z'_k \right]$$

$$y_{2k} = y_{1k}$$

$$\text{and } x_{2k} = x_{1k}.$$

We have spitted total population of size 100 sequentially into 5 strata each of size 20 (i.e., $N=100, N_h=20, h=1, 2, \dots, 5$).

We have computed the losses in efficiencies L_1 and L_2 of the estimators T_{p1} and T_{p2} with respect to the sample mean estimator based on the above population for different values of the non-response probability p_h on h^{th} strata ($h=1, 2, \dots, 5$), correlation coefficients ρ_{yx} and ρ_{xz} and sample sizes (i.e., n_h and m_h) and the findings are displayed in Tables 1-4.

Table 1. Loss in Efficiency of the estimator T_{p_1} in presence of Non-response situation only

$\rho_{yx} = 0.7, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_1	p_h	L_1
0.1 (fixed for all strata)	-10.035	0.1 (fixed for all strata)	-7.452
0.15 (fixed for all strata)	-2.1531	0.15 (fixed for all strata)	-2.062
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-2.5136	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-1.981
$\rho_{yx} = 0.9, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_1	p_h	L_1
0.1 (fixed for all strata)	-14.163	0.1 (fixed for all strata)	-11.134
0.15 (fixed for all strata)	-6.442	0.15 (fixed for all strata)	-3.251
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-4.831	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-2.753

Table 2. Loss in Efficiency of the estimator T_{p_1} in joint presence of Non-response and Measurement Error

$\rho_{yx} = 0.7, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_1	p_h	L_1
0.1 (fixed for all strata)	-7.271	0.1 (fixed for all strata)	-5.142
0.15 (fixed for all strata)	-13.796	0.15 (fixed for all strata)	-8.517
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-10.224	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-7.248
$\rho_{yx} = 0.9, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_1	p_h	L_1
0.1 (fixed for all strata)	-11.427	0.1 (fixed for all strata)	-9.314
0.15 (fixed for all strata)	-14.358	0.15 (fixed for all strata)	-10.891
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-12.275	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-8.761

Table 3. Loss in Efficiency of the estimator T_{p_2} in presence of Non-response situation only

$\rho_{yx} = 0.7, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_2	p_h	L_2
0.1 (fixed for all strata)	-9.273	0.1 (fixed for all strata)	-7.258
0.15 (fixed for all strata)	-11.090	0.15 (fixed for all strata)	-8.147
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-10.489	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-9.247
$\rho_{yx} = 0.9, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_2	p_h	L_2
0.1 (fixed for all strata)	-12.374	0.1 (fixed for all strata)	-10.318
0.15 (fixed for all strata)	-14.291	0.15 (fixed for all strata)	-12.873
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-11.269	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-9.716

Table 4. Loss in Efficiency of the estimator T_{p_2} in joint presence of Non-response and Measurement Error

$\rho_{yx} = 0.7, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_2	p_h	L_2
0.1 (fixed for all strata)	-22.840	0.1 (fixed for all strata)	-20.452
0.15 (fixed for all strata)	-23.321	0.15 (fixed for all strata)	-21.286
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-22.753	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-18.835
$\rho_{yx} = 0.9, \rho_{xz} = 0.5$			
$n_h = 12, m_h = 8$		$n_h = 10, m_h = 6$	
p_h	L_2	p_h	L_2
0.1 (fixed for all strata)	-24.346	0.1 (fixed for all strata)	-22.458
0.15 (fixed for all strata)	-25.287	0.15 (fixed for all strata)	-23.817
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-23.290	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-20.183

Numerical Illustration using known natural population

Natural data set: We have chosen natural population datasets on distribution of population by literates and literacy rates by sex in the different states of India. Statistical Abstract of India: (Census-2011) to elucidate the efficacious performance of our proposed estimator. The nature of the variables y , x and z and the values of the various parameters are given below.

y , x , z : percentage of literates (female, male) and the overall literacy rate in the year 2011 respectively.

We have formed 6 different (equal size and zone wise) out of 30(=N) states of India as shown in Table 5.

Table 5. Stratification of natural data set and related parameters

Strata	Constituent States	Statistical Parameters
		No of strata = 6 Units in each strata = 5 {i.e., $N_h = 5$; $h = \{1, 2, \dots, 6\}$ }. $\bar{Y}_1 = 19.38, \bar{X}_1 = 7.5, \bar{Z}_1 = 42$
Strata 1	Andhra Pradesh ,Karnataka ,Kerala ,Tamil Nadu , Madhya Pradesh	$\bar{Y}_2 = 19.78, \bar{X}_2 = 6.82, \bar{Z}_2 = 39.14$
Strata 2	Goa, Gujarat, Maharashtra, Punjab , Rajasthan	$\bar{Y}_3 = 22.4, \bar{X}_3 = 6.96, \bar{Z}_3 = 47.8$
Strata 3	West Bengal ,Jharkhand , Odisha , Bihar, Sikkim.	$\bar{Y}_4 = 20.42, \bar{X}_4 = 6.9, \bar{Z}_4 = 40.4$
Strata 4	Manipur , Meghalaya ,Arunachal Pradesh, Assam ,Tripura	$\bar{Y}_5 = 21.74, \bar{X}_5 = 6.98, \bar{Z}_5 = 51.6$
Strata 5	Haryana, Himachal Pradesh, Jammu & Kashmir Uttar Pradesh , Uttarakhand	$\bar{Y}_6 = 16.48, \bar{X}_6 = 5.72, \bar{Z}_6 = 30$
Strata 6	A.&N.Islands, Chandigarh, Delhi, Lakshadweep, Puducherry	

Note: It is found that mean of the respective variables i.e. x , y and z is found for the above data as $\bar{Y}_{..} = 20.03, \bar{X}_{..} = 6.75, \bar{Z}_{..} = 86.86667$.

We have computed the losses in efficiencies of the proposed estimators T_{p1} and T_{p2} with respect to the sample mean estimator based on this population for different values of the non-response probability p_h on h^{th} strata ($h = 1, 2, \dots, 6$) and the findings are displayed in Tables 6 and 7.

Table 6. Loss in Efficiency of the estimator T_{p1} in presence of Non-response situation and Measurement error

Presence of Non-response situation Only		Joint Presence of non-response and measurement Error	
$n_h = 4, m_h = 3$			
p_h	L_1	p_h	L_1
0.1 (fixed for all strata)	-70.1401	0.1 (fixed for all strata)	-53.2496
0.15 (fixed for all strata)	-58.0818	0.15 (fixed for all strata)	-42.8518
$p_1 = 0.05$ (first strata) $p_2 = 0.1$ (second strata) $p_3 = 0.15$ (third strata) $p_4 = 0.2$ (forth strata) $p_5 = 0.25$ (fifth strata)	-49.2818	$p_1 = 0.05$ (first strata) $p_2 = 0.1$ (second strata) $p_3 = 0.15$ (third strata) $p_4 = 0.2$ (forth strata) $p_5 = 0.25$ (fifth strata)	-48.3655

Table 7. Loss in Efficiency of the estimator T_{p_2} in presence of Non-response situation and Measurement error

Presence of non-response situation only		Joint Presence of non-response and measurement Error	
$n_h = 4, m_h = 3$			
p_h	L_2	p_h	L_2
0.1 (fixed for all strata)	-257.949	0.1 (fixed for all strata)	-175.566
0.15 (fixed for all strata)	-214.002	0.15 (fixed for all strata)	-146.854
$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-225.18	$p_1=0.05$ (first strata) $p_2= 0.1$ (second strata) $p_3= 0.15$ (third strata) $p_4= 0.2$ (forth strata) $p_5= 0.25$ (fifth strata)	-152.131

4. Conclusion

(i) From Table 1 – 7, it is found that losses in efficiencies L_1 and L_2 of the estimators T_{p_1} and T_{p_2} are all negative. Negative loss indicates gain in efficiencies. It may be noted that loss in efficiency is usual because we are comparing our proposed methodology which face non-response and measurement errors situation with respect to natural sample mean estimator under complete response situation and in absence of measurement error. However, the structure of the estimator proposed and imputation technique suggested in this present work found to be effective and may cope with the nuisance effect of non-response situation as well as produce precise estimates.

(ii) From Tables 1-7, it is also observed that for different choices of non-response probability our suggested methodology performs well.

(iii) It may be observed that natural population data set is heterogeneous as parametric values are found significantly different from strata to strata while population generated through simulation studies is homogeneous as parametric values are almost similar between strata to strata. From Tables 6 and 7, it is clear that our suggested estimator T_{p_1} and T_{p_2} produce efficient estimates in these practical situation.

(iv) From Tables 1-4, it may be seen that for increasing values of the correlation coefficients ρ_{yx} , loss in efficiencies L_1 and L_2 are decreasing and we are getting more precise estimates. This phenomenon justifies the effectiveness of suggested methodologies employed here. It also claims that our suggested strategy performs well if highly correlated auxiliary variable present.

Moreover, looking on the encouraging findings, we are happy to recommend the proposed strategy to the survey statisticians for their application in real life.

References

- [1] Arnab, R. and Singh, S., "On the estimation of variance from missing data," *Metron*, vol. LXIV2, pp. 161 - 171, 2006.
- [2] Diana, G. and Perri, P.F., "Estimation of finite population mean using multi auxiliary information," *Metron*, vol. LXV, pp. 99 - 112, 2007.
- [3] Heitjan, D.F. and Basu, S., "Distinguishing 'missing at random' and missing completely at Random'," *The American Statistician*, vol. 50, pp. 207- 213, 1966.
- [4] Manisha and Singh R.K., "An estimation of population mean in the presence of measurement errors," *Journal of Indian Society Agriculture Statistician*, vol. 54(1), pp. 13 - 18, 2001.
- [5] Manisha and Singh R.K., "Role of regression estimator involving measurement errors," *Brazilians Journal of Statistics*, vol. 16, pp. 39 – 46, 2002.
- [6] Rubin, R. B., "Inference and missing data," *Biometrika*, vol. 63(3), pp.581 - 592, 1976.
- [7] Shalabh "Ratio method of estimation in presence of measurement errors," *Journal of Indian Society of Agriculture Statistician*, vol. 50(2), pp. 150 - 155, 1997.
- [8] Singh, S., and Deo, B., "Imputation by power transformation," *Statistical Papers*, vol. 44 (4), pp. 555–579, 2003.
- [9] Singh, S., and Horn, S., "Compromised imputation in survey sampling," *Metrika*, vol. 51 (3), pp. 267–276, 2000.
- [10] Singh, S. and Joarder, A.H., "Estimation of finite population variance using random non response in Survey sampling," *Metrika*, vol. 47, pp. 241- 249, 1998.
- [11] Singh, H.P. and Karpe, N., "Ratio product estimator for population mean in presence of measurement errors," *Journal of Applied Statistical Science.*, vol. 16, pp. 49 – 64, 2008.
- [12] Singh, S., A. H. Joarder, and D. S. Tracy., "Median estimation using double sampling," *Australian & New Zealand Journal of Statistics*, vol. 43 (1), pp.33- 46, 2001.
- [13] Sukhla, D, Pathak, S. and Thakur, N.S., "An estimator for mean estimation in presence of measurement error, Research and Reviews," *A Journal of Statistics*, vol. 1(1), pp.1- 8, 2012.