Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/100113

Logarithmic Estimation Strategy in Cluster Sampling in Presence of Random-Non Response and Measurement Error

Arnab Bandyapadhya[1], *Arindam Gupta*[2] *and Chowdhury Sadeka Nazrin*[2,∗]

[1]Department of Mathematics, Asansol Engineering College, India
[2]Department of Statistics, The University of Burdwan, India

Received: 15 Jun. 2019, Revised: 23 Dec. 2019, Accepted: 27 Feb. 2020
Published online: 1 Mar. 2021

Abstract: Our current study is about estimating the finite population mean under cluster sampling in existence of random non-response and measurement error state. Using material on a secondary variable, a logarithmic kind of estimator has been planned. We proposed this type of estimator as we have seen that in this field we have many types of estimators except logarithmic. Active imputation techniques have been proposed for a transaction with the random non-response and measurement error states. The possessions of the recommended estimation plans have been deliberate for diverse belongings of random non-response and measurement error states in real life studies. The Advantage of the recommended procedure over the usual sample mean estimator of population mean has been well-known over experiential studies approved through the data sets of usual population and artificially generated population. Appropriate recommendations to the study statistician are made.

Keywords: Cluster sampling, logarithmic estimation, random non-response, measurement error, imputation, auxiliary variable, study variable, mean square error, efficiency

1 Introduction

We have many types of sampling procedures in our sampling theory. Among them, the most cost effective, quick and easy is cluster sampling. We can use cluster sampling when it is inexpensively justified; when reduced costs can be used to overcome fatalities in accuracy. This is most likely to take place in the following circumstances. Building a complete list of population constituents is hard, expensive or impossible. In cluster sampling process, the population has been divided into discrete groups. Those groups are known as clusters. Then a simple random sample of clusters is selected from the population. Now we can conduct our analysis on data from the sampled clusters. Since cluster sampling is very cost effective and time saver. For this reason in literature, we have found many works on this sampling procedure. We name few of them in our paper.

In this article, we use cluster sampling technique in the existence of non response and measurement error. In literature, many works have been done in the presence of non response. Also, we found work in the existence of measurement error. As mentioned before, we use cluster sampling in the presence of non response and measurement error. Almost no work has been done in the present of these two problems together.

In our study, we use both of these two problems. As we think of an example where both non response and measurement error occur, almost no work has been done which considers all these problems together. For this reason, we want to see what would happen if in the case when non response and measurement error occur together, we use rank there. Also regression, exponential analysis done in this type of cases. Logarithmic estimation technique was never used in this type of case before.

In registered surveys, frequently it can be realized that a whole register of all the components of the study in the population is not accessible; it is pointed out that taking a simple random sample is not practicable. As for an example, in a socioeconomic study, a register of family units is not normally accessible where a register of inhabited houses, each

∗ Corresponding author e-mail: chowdhurynazrinsadeka@gmail.com

obliging a number of family units should be accessible by municipal and added suitable establishments. In these cases, it can be sensible to draw a simple random sample of houses and study, entirely, the family units fitting the sample family units. This technique is well-known as cluster sampling. Cluster sampling is a procedure where the whole population is alienated into groups or clusters, and a random sample of the chosen clusters are involved in the sample. Occasionally, in cluster sampling, it is not probable to take a sample of final units of attention, since the frame of these kind of units is not obtainable. Nevertheless, a register of some properly described superior units or primary stage units (psus) can be obtainable from which a sample (ssus) can be selected. In its place of entirely enumerating altogether, the ssus belonging to the particular psus as in case of cluster sampling, one can choose a sample of ssus from the register of all ssus fitting the particular psu. The sampling is consequently approved in two stages. This type of selection type, sampling technique, is therefore, named two stage sampling and henceforth, the names, primary stage units or first stage units (fsus) and second stage units (ssus). Obviously, a sampler has an extensive range of excellent sampling strategies to use in a two stage sampling procedure. As for an example in a socioeconomic study, villages can be considered as fsus and family unit as ssus for rural areas, where survey blocks can be considered as fsus and inhabited houses as ssus for a city. The major gain from two stage sampling is that it is more supple than one stage sampling; it reduces to one stage sampling. It is known that nonresponse is an inescapable phenomenon when studying practical populations. In studies covering human populations in maximum cases, information is not brought from the whole units in the survey at the first effort, even after certain call-backs. This is mainly true in mail surveys in which questionnaires are mailed to the tested respondents who are entreated to send back their returns by a deadline. As many respondents do not reply, the available sample of returns is incomplete. Thus, the observations may be missing for some of the time stages. Such type non-response (missingness) can have diverse patterns and causes. Rubin (1976) [8] addressed three concepts: missing at random (MAR), observed at random (OAR), and parameter distribution (PD). Rubin (1976) [8] 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. Sande (1979) [5] recommended imputation methods which make incomplete data sets mechanically complete. Imputation may be approved with the support of an auxiliary variable. As for an example, Lee et al. (1994) [3] has been using the evidence on an obtainable auxiliary variable for the imputation purpose. It has been seen that no significant effort has been taken to address the problems of random non-response situation (MAR) in cluster sampling. Statisticians have long identified that failure to justification for the stochastic nature of incompleteness can impair the real decision. An estimate gained from such incomplete data may be ambiguous, particularly when the respondents vary from the non-respondents, and the estimate founded on such explanations may lead to biased estimate. It is well-known that random non-response situation may be more destructive for large-scale household studies as this error could rise with the rise in sample size. The difference between a measured value of a quantity and its true value is called measurement error. An error is not a fault in statistics. Variability is a basic portion of the result of measurements and of the measurement technique. The term, response error, occasionally known as observational error and some other kinds of non-sampling errors; in survey-type situations, these errors can be faults in the collection of data, including both the incorrect recording of a response and the correct recording of a respondents incorrect response. These errors can be random or systematic. Random errors are produced by unintentional faults by respondents, interviewers or coders. Systematic error can arise if there is a systematic response of the respondents to the technique used to frame the survey question. Thus, the particular formulation of a survey question is vital, since it affects the level of measurement error. This encourages us to mature appropriate estimation procedures under diverse constructions of cluster sampling and address the problems of presence of non-response and measurement error situations in real life surveys.

2 Construction of Sampling Scheme

Let us take a finite population, U, to be divided into N first stage units (fsu) which is represented by U_1, U_2,......, U_N so that the quantity of second stage units (ssu) in each fsu be M. Let y_{ij}, x_{ij}, z_{ij} be the observed values and Y_{ij}, X_{ij} and Z_{ij} be the true values of the study character, first auxiliary variable and second auxiliary variable respectively on the j^{th} ssu (j = 1, 2, . . . , M) in the i^{th} fsu (i = 1, 2. . . N). Let us consider the condition when the information on auxiliary variable x is not obtainable, and therefore, it is collected by the following two different ways: i. On the auxiliary variable, x information is gathered at fsu level, and then a sample of fsus is chosen by SRSWOR (Simple Random Sampling Without Replacement) procedure.

ii. On the variable, x information is collected at ssu level, and after that, further sample of ssus is taken with SRSWOR scheme to observe study variable y. After that, we assume the occurrence of random non-response condition in the following ways. Here, it is assumed that y_{ij} and x_{ij}, are observed instead of true values Y_{ij} and X_{ij} respectively. The measurement errors are defined as

$u_{ij} = y_{ij}\text{-}Y_{ij}$

$v_{ij} = x_{ij}\text{-}X_{ij}$

u_{ij} and v_{ij} are random in nature with a mean of zero and differebnt variances σ_u^2 and σ_v^2, respectively. It is assumed that $\acute{u}_i$s and $\acute{v}_i$s are uncorrelated. It is also assumed that $\acute{u}_i$s and $\acute{v}_i$s are uncorrelated with $\acute{Y}_i$s and $\acute{X}_i$s, respectively. Let (μ_Y, μ_X) and (σ_Y^2, σ_X^2) be the mean and variance of (Y, X), i.e, study and auxiliary variables. ρ is the correlation coefficient between X and Y. Let $\bar{y} = \frac{1}{n}\sum y_i$, $\bar{x} = \frac{1}{n}\sum x_i$ be the unbiased estimators of population means μ_Y and μ_X, respectively.
$s_x^2 = \frac{1}{n-1}\sum(x_{ij}-\bar{x})^2$
and
$s_y^2 = \frac{1}{n-1}\sum(y_{ij}-\bar{y})^2$
are not unbiased estimators of population variances σ_X^2 and σ_Y^2. In presence of measurement errors, the expected value of s_y^2 and s_x^2 is given by E(s_y^2) = σ_Y^2 + σ_u^2 and E(s_x^2) = σ_X^2 + σ_v^2, respectively. Let error variances σ_u^2 and σ_v^2 be priors, then unbiased estimators of population variance under measurement errors are
$\hat{\sigma}_Y^2 = s_y^2 - \sigma_u^2 > 0$
$\hat{\sigma}_X^2 = s_x^2 - \sigma_v^2 > 0$

2.1 Non Response probability model

let us suppose that random non-response condition occurs on the 2^{nd} phase sample, S_2, of n fsu, every one containing m ssu. Then $r\{r = 0,1,......,(m-2)\}$ is the number of sampling established units on which information could not be collected due to random non-response (R^c set), and the observations of the respective variables on which random non-response occur can be taken from the remaining $(m-r)$ units of each of n fsus of the 2^{nd} stage sample (R set). If p is the probability of a non-response, then r has the following discrete distribution:
$P(r) = \frac{m-r}{mq+2p}\,{}^{(m-2)}C_r p^r q^{m-2-r};\ r = 0,1,......,(m-2)$
where q = 1-p and ${}^{(m-2)}C_r$ denote the total number of ways of obtaining r non-responses out of $(m-2)$ total possible non-responses, respectively, for instance, see Singh and Joarder (1998), Singh et al.(2000) and Singh et al. (2012).

Hence, onwards, we use the following notations:
$\overline{Y}_{..} = \frac{1}{NM}\sum\sum y_{ij}$, is the population mean of study variable y.
$\overline{X}_{..} = \frac{1}{NM}\sum\sum x_{ij}$, is the population mean of first auxiliary variable x.
$\overline{Z}_{..} = \frac{1}{NM}\sum\sum z_{ij}$, is the population mean of second auxiliary variable z.
$\bar{y}_i = \frac{1}{m}\sum y_{ij}$, is the sample mean of i^{th} fsu in S_2.
$\bar{y}_{i(m-r)} = \frac{1}{m-r}\sum y_{ij}$, is the sample mean of y based on the respondents region of i^{th} fsu in S_2.
$\bar{x}_i = \frac{1}{m}\sum x_{ij}$, is the sample mean of x on i^{th} fsu in S_2.
$\bar{x}_{i(m-r)} = \frac{1}{m-r}\sum x_{ij}$, is the sample mean of x based on the respondents region of i^{th} fsu in S_2.
$\bar{z}_i = \frac{1}{m}\sum z_{ij}$, is the sample mean of z on i^{th} fsu in S_2.
$\overline{X}_{nM} = \frac{1}{n/M}\sum\sum x_{ij}$, $\bar{z}_{nm} = \frac{1}{n}\sum z_i$
$\bar{y}_{nm} = \frac{1}{n}\sum y_i$, $\bar{x}_{nm} = \frac{1}{n}\sum x_i$
$\bar{y}_1 = \frac{1}{m}\sum y_{1j} = \bar{y}_{1(m-r_1)}exp(\frac{\overline{Z}_{..}-\bar{z}_{nm}}{\overline{Z}_{..}+\bar{z}_{nm}})$ is the sample mean of y on 1st fsu in S_2
$\bar{y}_2 = \frac{2}{m}\sum y_{2j} = \bar{y}_{2(m-r_2)}exp(\frac{\overline{Z}_{..}-\bar{z}_{nm}}{\overline{Z}_{..}+\bar{z}_{nm}})$ is the sample mean of y on 2nd fsu in S_2

$\bar{x}_1 = \frac{1}{m}\sum x_{1j} = \bar{x}_{1(m-r_1)}exp(\frac{\overline{Z}_{..}-\bar{z}_{nm}}{\overline{Z}_{..}+\bar{z}_{nm}})$ is the sample mean of x on the 1st fsu in S_2.

$\bar{x}_2 = \frac{1}{m}\sum x_{2j} = \bar{x}_{2(m-r_2)}exp(\frac{\overline{Z}_{..}-\bar{z}_{nm}}{\overline{Z}_{..}+\bar{z}_{nm}})$ is the sample mean of x on the 2nd fsu in S_2.

2.1.1 Construction of the Proposed Estimator:

In this topic, we can find exponential type estimator in several articles. Here, we propose a logarithemic type estimator, which is absolutely new in this context. If we take log in any calculation then that becomes easily solvable. We use logarithemic so that our calculation becomes easy.
In this paper, we have total N clusters in the first stage unit (fsu) and n clusters in the second stage unit (ssu). Where at every stage, each cluster size is M. Then we propose an estimator of $\overline{Y}_{..}$ as

t = $\overline{y}^*_{n(m-r)} + \log[1 + \alpha(\frac{\overline{x}^{/}_{nM} - \overline{x}^*_{n(m-r)}}{\overline{x}^{/}_{nM} + \overline{x}^*_{n(m-r)}})]$, where α is a real constant.

Where

$\overline{y}^*_{nm} = \overline{y}_{nm} exp(\frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}})$

$\overline{x}^*_{nm} = \overline{x}_{\acute{n}m} exp(\frac{\overline{z}_{..} - \overline{z}_{nm}}{\overline{z}_{..} + \overline{z}_{nm}})$

$\overline{x}_{\acute{n}M} = \frac{1}{\acute{n}M} \sum\sum x_{ij}$

Imputation for T_1:

The information on the second auxiliary variable, z, is readily available for all of population U. Therefore, we are trying to develop a new imputation technique which is a logarithimic from. We suggest the following logarithimic kind imputation method based on responding and non-responding units of the second stage sample, S_2, to estimate the population parameter under study $\overline{Y}_{..}$ as

$y_{ij} = y_{ij}$ if j $\in$ R [(m-r) unit]

= $\overline{y}_{i(m-r)}$ j $\in R^c$ (r unit)

Under the above mentioned imputation technique, we can derive the sample mean of y on the i^{th} fsu in S_2 as

$\overline{y}_{i.} = \frac{1}{m} \sum_{j=1}^{m} y_{ij}$

= $\frac{1}{m}[\sum_{j \in R} y_{ij} + \sum_{j \in R^c} y_{ij}]$

= $\frac{1}{m}[(m-r)\overline{y}_{i(m-r)} + r\overline{y}_{i(m-r)}]$

$\overline{y}_{i.} = \overline{y}_{i(m-r)}$

Now, the mean of n fsu of y in S_2 is

$\overline{y}^*_{n(m-r)} = \frac{1}{n} \sum_{i=1}^{n} y_{i.}$

= $\frac{1}{n} \sum_{i=1}^{n} y_{i(m-r)}$

= $\overline{y}_{n(m-r)}$

Similarly, for each second stage ssu

$x_{i.} = x_i \in R$

= $\overline{x}_r \in$ NR (Non response size (m-r))

Under the above mentioned imputation technique, the sample mean of x on the i^{th} fsu in S_2 can be derived as

$\overline{x}_i = \frac{1}{r} \sum_{i=1}^{r} x_i$

$\overline{x} = \frac{1}{m} \sum_{i=1}^{m} x_i$ = $\frac{1}{m}[\sum_{i=1}^{r} x_i + \sum_{i=1}^{m-r} \overline{x}_r$ = $\frac{1}{m}[(m-r)\overline{x}_r + r\overline{x}_r]$

$\overline{x}_{i.} = \overline{x}_{i(m-r)}$

Now, the mean of n fsu of x in S_2 is

$\overline{x}^* = \frac{1}{n} \sum_{i=1}^{n} x_i$

= $\frac{1}{n} \sum_{i=1}^{n} x_{i(m-r)}$

= $\overline{x}_{n(m-r)}$

2.1.2 Properties of the Proposed Estimator

Here, T_1 is a logarithimic kind of estimator. The bias and mean square error of the proposed estimators to the first order of approximations are derived under large sample approximations (ignoring f.p.c), using the following assumptions:

To get the mean square error (MSE) of the estimator, we have the following expected expressions of the sample statistics by using large sample approximation using the below mentioned transformations.

$\overline{y}_{nm} = \overline{Y}_{..}(1 + e_0)$

$\overline{x}_{nm} = \overline{X}_{..}(1 + e_1)$

$\overline{x}_{n/m} = \overline{X}_{..}(1 + e_2)$

$\overline{z}_{n/m} = \overline{z}_{..}(1 + e_3)$

We have obtained the following parametric values.

$$E(e_0^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_0^{*2}}{\overline{Y}_{..}^2} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_0^2}{\overline{Y}_{..}^2}$$
$$E(e_1^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}_{..}^2} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}_{..}^2}$$
$$E(e_2^2) = (\frac{1}{n/} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}_{..}^2}$$
$$E(e_3^2) = (\frac{1}{\acute{n}} - \frac{1}{N})\frac{S_2^{*2}}{\overline{Z}^2}$$
$$E(e_0e_1) = (\frac{1}{n} - \frac{1}{N})\frac{S_{01}^{*}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}$$
$$E(e_2e_3) = (\frac{1}{n/ - \frac{1}{N}})\frac{S_{12}^{*}}{\overline{X}_{..}\overline{z}_{..}}$$
$$E(e_0e_2) = (\frac{1}{n/} - \frac{1}{N})\frac{S_{01}^{*}}{\overline{Y}_{..}\overline{X}_{..}}$$
$$E(e_0e_3) = (\frac{1}{n/} - \frac{1}{N})\frac{S_{02}^{*}}{\overline{Y}_{..}\overline{z}_{..}}$$
$$E(e_1e_2) = (\frac{1}{n/} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}_{..}^2}$$
$$E(e_1e_3) = (\frac{1}{n/} - \frac{1}{N})\frac{S_{12}^{*}}{\overline{X}_{..}\overline{z}_{..}}$$

After applying the above transformation, we have obtained the following form of the proposed estimator.

3 Bias and MSE of the Proposed Estimation Strategy

The bias and mean square of T_2 to the first order of approximation are derived under large sample approximations (ignoring f.p.c), using the following transformations:
$\overline{y}_{nm} = \overline{Y}_{..}(1+e_4)$, $\overline{x}_{n(\acute{m}-r)} = \overline{X}_{..}(1+e_5)$, $\overline{z}_{n\acute{m}} = \overline{Z}_{..}(1+e_6)$, $\overline{x}_{nm} = \overline{X}_{..}(1+e_7)$
Such that $E(e_i) = 0$ and $|e_i| < 1$, $\forall i = 4,5,6,7$.
Now, for obtaining the expressions for the bias and mean square error, we have pondered the given below expectation.

$$E(e_4^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_0^{*2}}{\overline{Y}_{..}^2} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_0^{*2}}{\overline{Y}_{..}^2},\ E(e_5^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}_{..}^2} + \frac{1}{n}(\frac{1}{\acute{m}q+2p} - \frac{1}{M})\frac{\overline{S}_1^{*2}}{\overline{X}_{..}^2}$$
$$E(e_6^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_2^{*2}}{\overline{Z}_{..}^2} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_2^{*2}}{\overline{Z}_{..}^2},\ E(e_7^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}_{..}^2} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_{01}^{*2}}{\overline{X}_{..}^2}$$
$$E(e_4e_5) = (\frac{1}{n} - \frac{1}{N})\frac{S_{01}^{*}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2p} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}$$
$$E(e_4e_6) = (\frac{1}{n} - \frac{1}{N})\frac{S_{02}^{*}}{\overline{Y}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_{02}}{\overline{Y}_{..}\overline{Z}_{..}}$$
$$E(e_4e_7) = (\frac{1}{n} - \frac{1}{N})\frac{S_{01}^{*}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2p} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}$$
$$E(e_5e_6) = (\frac{1}{n} - \frac{1}{N})\frac{S_{12}^{*}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}}$$

$$E(e_5e_7) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}_{..}^2} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}^2}$$
$$E(e_6e_7) = (\frac{1}{n} - \frac{1}{N})\frac{S_{12}^{*}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}}$$

Where f = $\frac{1}{n} - \frac{1}{N}$, $f_{\acute{m}} = \frac{1}{\acute{m}} - \frac{1}{M}$, $f_m = \frac{1}{m} - \frac{1}{M}$ and $f_{\acute{m}r} = \frac{1}{\acute{m}q+2p} - \frac{1}{M}$.
Now, taking expectation from both expectation on both side, we have our bias as
$B(T_2) = E(T_2 - \overline{Y}_{..})$

So, estimator T_2 takes the form given below.
$T_2 = \overline{Y}_{..}(1+e_4) + \beta\overline{X}[e_5 - \frac{1}{2}e_6 - e_7 + \frac{3}{8}e_6^2 - \frac{1}{2}e_5e_6]$
Taking expectation from the equation, we get the following expression of bias and mean square errors of the proposed estimators to first order of sample size as

$$M(t_1) = E(t_1 - \overline{Y}_{..})^2$$

Where $M(t_P) = \overline{Y}^2 A + \alpha^2 B + 2\overline{Y}_{..} C\alpha$
Now, the value of the coefficients becomes

$$A = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_0^{*2}}{\overline{Y}_{..}^2} + \frac{1}{n}\left(\frac{1}{mq+2P} - \frac{1}{M}\right)\frac{\overline{S}_0^2}{\overline{Y}_{..}^2}$$

$$B = \left(\frac{1}{4}\right)\left[\left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_2^{*2}}{\overline{z}_{..}^2} + \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_1^{*2}}{\overline{X}_{..}^2} + \frac{1}{n}\left(\frac{1}{m} - \frac{1}{M}\right)\frac{\overline{S}_1^2}{\overline{X}_{..}^2}\right]$$

$$C = \left(\frac{1}{n}\right)\left(\frac{1}{mq+2P} - \frac{1}{M}\right)\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}} - \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_{02}^{*}}{\overline{Y}_{..}\overline{z}_{..}}$$

Our minimum mean square error of the suggested estimators becomes,
$M(t_P) = \overline{Y}^2 A + \alpha^2 B + 2\overline{Y}_{..} C\alpha$
It may be noted that MSE of the proposed estimators depends on the value of α. Therefore, it is required to obtain the minimum values of the proposed class of the estimators. Thus, differentiating both sides of the equations, and with respect to α, we have obtained the optimum value of α as
$\frac{dM(t_P)}{d\alpha} = 0$

After solving this, we get α as
$\alpha = -\frac{\overline{Y}C}{\beta}$

Now after substituting the optimum values of α in the equations (MSEs), we have the minimum mean square error of the proposed class of estimators as

$$M(t_P) = \overline{Y}^2 (A - \frac{C^2}{\beta})$$

In the above expression, we only have random non-response, but we are missing measurement errors.

After adding the measurement error, we get the expression as
$M(t_P) = \overline{Y}^2 (A - \frac{C^2}{B}) + \overline{Y}^2 (A^{/} - \frac{C^{/2}}{B})$

$$A^{/} = (\frac{1}{n} - \frac{1}{N})\frac{S_U^{*2}}{\overline{Y}^2} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_U^2}{\overline{Y}^2}$$

$$C^{/} = 0$$

$$B^{/} = \frac{1}{2}\left\{(\frac{1}{n^{/}} - \frac{1}{N})\frac{S_V^{*}}{\overline{X}^2}\right\} + \frac{1}{8}\left\{\left((\frac{1}{n^{/}} - \frac{1}{N})\right)\frac{S_w^{*2}}{\overline{z}^2}\right\} - \left\{(\frac{1}{n^{/}} - \frac{1}{N})\frac{S_X^{*2}}{\overline{X}^2}\right\} + \frac{1}{2}\left\{(\frac{1}{n} - \frac{1}{N})\frac{S_V^{*2}}{\overline{X}^2}\right\} + \frac{1}{n}\left\{\left(\frac{1}{mq+2P} - \frac{1}{M}\right)\frac{\overline{S}_V^2}{\overline{X}^2}\right\}$$

Now,
$A^{//} = A + A^{/}$
$B^{//} = B + B^{/}$

Non response and measurement error, both are present, then the final model becomes
$M(t_p)_{opt} = \overline{Y}^{//2}\left(A^{//} - \frac{C^2}{B^{//}}\right)$.

Imputation technique for T_2:

$y_{ij} = y_{ij} + log[1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}}]$, $j \in R$:
$= \overline{y}_{i(m-r)} + log[1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}}]$, $j \in R^c$
$\overline{y}_{i.} = \frac{1}{m}\sum_{j=1}^{m} y_{ij}\ \overline{y}_{i.} = \overline{y}_{i(m-r)} + log[1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}}]$
$\overline{y}^{**}_{n(m-r)} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_{i.} = \overline{y}_{n(m-r)} + log[1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}}]$
Similarly, for each second stage ssu
$x_{ij} = x_{ij} + log[1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}}]$
$\overline{x}^{**}_{n(m-r)} = \overline{x}_{n(m-r)} + log[1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}}]$

Proposed estimator:

We get
$x_{ij} = x_{ij} + log(1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}})$

$\overline{x}^{*}_{n(m-r)} = \overline{x}_{n(m-r)} + log(1 + \frac{\overline{Z}_{..} - \overline{z}_{nm}}{\overline{Z}_{..} + \overline{z}_{nm}})$

Our estimator under imputation becomes

$t_2 = \overline{y}^{*}_{n(m-r)} + log(1 + \alpha_2(\frac{\overline{x}_{\acute{n}m} - \overline{x}^{*}_{n(m-r)}}{\overline{x}_{\acute{n}m} + \overline{x}^{*}_{n(m-r)}}))$

Mean square error of the proposed estimator:

The mean square error of our proposed estimator to first order of approximation are obtained under large sample approximations (ignoring f.p.c), by using the following transformation:

$\overline{y}_{nm} = \overline{Y}_{..}(1+e_4), \overline{x}^{/} = \overline{X}_{..}(1+e_5), \overline{z}_{n\acute{m}} = \overline{Z}_{..}(1+e_6)$

$\overline{x}_{nm} = \overline{X}_{..}(1+e_7), \overline{x}\acute{n}m = \overline{X}_{..}(1+e_2), \overline{y}_{n(m-r)} = \overline{Y}_{..}(1+e_0)$

s.t. $E(e_i) = 0$, and $|e_i| < 1, \forall i = 4, 5, 6, 7$

So, our estimator, T_2, takes the following form:

So, our MSE becomes

MSE = $E[T_2 - \overline{Y}_{..}]^2$

$= E[\overline{Y}_{..}e_4 + \beta\overline{X}(e_5 - \frac{1}{2}e_6 - e_7)]^2$

$\overline{x}^{/}_{n(m-r)} = \overline{x}_{n(m-r)} + log[1 + \frac{\overline{z}_{..} - \overline{z}_{nm}}{\overline{z}_{..} + \overline{z}_{nm}}]$

$= \overline{X}_{..}(1+e_5) + log[1 + \frac{\overline{z}_{..} - \overline{z}(1+e_3)}{\overline{z}_{..} + \overline{z}(1+e_3)}]$

$= \overline{X}_{..}(1+e_5) + log[1 + \frac{-\overline{z}_{..}e_3}{2\overline{z}_{..}(1+e_{3/2})}]$

$= \overline{X}_{..}(1+e_5) + log[1 - \frac{1}{2}e_3(1 + \frac{e_3}{2})]$

$= \overline{X}_{..}(1+e_5) + log[1 - \frac{1}{2}e_3(1 - \frac{e_3}{2} + \frac{e_3^2}{2})]$

$= \overline{X}_{..}(1+e_5) + [log(1 - \frac{e_3}{2} - \frac{e_3^2}{4})]$

So now,

$\overline{x}_{n(m-r)} = [\overline{X}_{..}(1+e_5) + \frac{e_3}{2}]$

Similarly,

$\overline{y}_{n(m-r)} = [\overline{Y}_{..}(1+e_0) + \frac{e_3}{2}]$

So, our estimator becomes

$T_2 = \overline{y}^{*}_{n(m-r)} + log[1 + \alpha_2(\frac{\overline{x}_{\acute{n}m} - \overline{x}^{*}_{n(m-r)}}{\overline{x}_{\acute{n}m} - \overline{x}^{*}_{n(m-r)}})]$

$= \overline{Y}_{..}(1+e_0) + log[1 + \alpha_2(\frac{\overline{X}_{..}(1+e_2) - \overline{X}_{..}(1+e_5) + \frac{e_3}{2})}{\overline{X}_{..}(1+e_2) + \overline{X}_{..}(1+e_5) - \frac{e_3}{2})})]$

$= \overline{Y}_{..}(1+e_0) + log[1 + \alpha_2\frac{\overline{X}_{..}(e_2 - e_5) + \frac{e_3}{2}}{\overline{X}_{..}(2 - e_2 + e_5) - \frac{e_3}{2}}]$

$= \overline{Y}_{..}(1+e_0) + [\alpha_2(\frac{e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}}}{1 + \frac{e_2+e_5}{2} - \frac{e_5}{4\overline{X}_{..}}})]$

$= \overline{Y}_{..}(1+e_0) + [\alpha_2(1 + \frac{e_2+e_5}{2} - \frac{e_5}{4\overline{X}_{..}})^{-1}(e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}})]$

$= \overline{Y}_{..}(1+e_0) + \alpha_2(1 - \frac{e_2+e_5}{2} + \frac{e_5}{4\overline{X}_{..}})(e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}})$

$= \overline{Y}_{..}(1+e_0) + \alpha_2(e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}})$

$M(T_P) = E[T_P - \overline{Y}_{..}]^2$

$= E[\overline{Y}_{..}e_0 + \alpha_2(e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}})]^2$

Now, after taking expectation from both sides, we get our parametric values as follows:

$E(e_0^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_0^{*2}}{\overline{Y}^2_{..}} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_0^2}{\overline{Y}^2_{..}}$

$E(e_1^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}^2_{..}} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}^2_{..}}$

$E(e_2^2) = (\frac{1}{n^{/}} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}^2_{..}}$

$E(e_3^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_2^{*2}}{\overline{z}^2_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_2^2}{\overline{z}^2_{..}}$

$E(e_0e_1) = (\frac{1}{n} - \frac{1}{N})\frac{S_{01}^{*}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{mq+2P} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}$

$E(e_0e_2) = (\frac{1}{n^{/}} - \frac{1}{N})\frac{S_{01}^{*}}{\overline{Y}_{..}\overline{X}_{..}}$

$E(e_0e_3) = (\frac{1}{n} - \frac{1}{N})\frac{S_{02}^{*}}{\overline{Y}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_{02}}{\overline{Y}_{..}\overline{z}_{..}}$

$E(e_1e_2) = (\frac{1}{\acute{n}} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}^2_{..}}$

$E(e_1e_3) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{12}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{z}_{..}}$
$E(e_2e_3) = (\frac{1}{\acute{n}} - \frac{1}{N})\frac{S^*_{12}}{\overline{X}_{..}\overline{z}_{..}}$
$E(e_4^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_0^{*2}}{\overline{Y}^2_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_0^2}{\overline{Y}^2_{..}}$, $E(e_5^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}^2_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}^2_{..}}$
$E(e_6^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_2^{*2}}{\overline{Z}^2_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_2^2}{\overline{Z}^2_{..}}$, $E(e_7^2) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}^2_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}^2_{..}}$
$E(e_4e_5) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{01}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}$
$E(e_4e_6) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{02}}{\overline{Y}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_{02}}{\overline{Y}_{..}\overline{Z}_{..}}$
$E(e_4e_7) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{01}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}$
$E(e_5e_6) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{12}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}}$
$E(e_5e_7) = (\frac{1}{n} - \frac{1}{N})\frac{S_1^{*2}}{\overline{X}^2_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}^2_{..}}$
$E(e_6e_7) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{12}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}}$
$E(e_3e_5) = (\frac{1}{n} - \frac{1}{N})\frac{S^*_{12}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}}$
$E(e_2e_5) = E(e_5^2)$, $E(e_0e_5) = E(e_4e_5)$

Where , $\overline{Y}_{..} = \frac{1}{N}\sum_{i=1}^{N}\overline{Y}_i, \overline{Y}_{i.} = \frac{1}{M}\sum_{j=1}^{M} y_{ij}$

$\overline{X}_{..} = \frac{1}{N}\sum_{i=1}^{N}\overline{X}_i, \overline{X}_{i.} = \frac{1}{M}\sum_{j=1}^{M} x_{ij}$

$\overline{Z}_{..} = \frac{1}{N}\sum_{i=1}^{N}\overline{Z}_i, \overline{Z}_{i.} = \frac{1}{M}\sum_{j=1}^{M} z_{ij}$

$S_0^{*2} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_{i.} - \overline{Y}_{..})^2, \overline{S}_0^2 = \frac{1}{N}\sum_{i=1}^{N} S_{0i}^2$

$S_{0i}^2 = \frac{1}{M-1}\sum_{j=1}^{M}(y_{ij} - \overline{Y}_{i.})^2$

$S_1^{*2} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{X}_{i.} - \overline{X}_{..})^2, \overline{S}_1^2 = \frac{1}{N}\sum_{i=1}^{N} S_{1i}^2$

$S_{1i}^2 = \frac{1}{M-1}\sum_{j=1}^{M}(x_{ij} - \overline{X}_{i.})^2$

$S_2^{*2} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Z}_{i.} - \overline{Z}_{..})^2, \overline{S}_2^2 = \frac{1}{N}\sum_{i=1}^{N} S_{2i}^2$

$S_{2i}^2 = \frac{1}{M-1}\sum_{j=1}^{M}(z_{ij} - \overline{Z}_{i.})^2$

$S^*_{01} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_{i.} - \overline{Y}_{..})(\overline{X}_{i.} - \overline{X}_{..})$

$\overline{S}_{01} = \frac{1}{N}\sum_{i=1}^{N} S_{01i}, S_{01i} = \frac{1}{M-1}\sum_{j=1}^{M}(y_{ij} - \overline{Y}_{i.})(x_{ij} - \overline{X}_{i.})$

$S^*_{02} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{Y}_{i.} - \overline{Y}_{..})(\overline{Z}_{i.} - \overline{Z}_{..})$

$\overline{S}_{02} = \frac{1}{N}\sum_{i=1}^{N} S_{02i}, S_{02i} = \frac{1}{M-1}\sum_{j=1}^{M}(y_{ij} - \overline{Y}_{i.})(z_{ij} - \overline{Z}_{i.})$

$S^*_{12} = \frac{1}{N-1}\sum_{i=1}^{N}(\overline{X}_{i.} - \overline{X}_{..})(\overline{Z}_{i.} - \overline{Z}_{..})$

$\overline{S}_{12} = \frac{1}{N}\sum_{i=1}^{N} S_{12i}, S_{12i} = \frac{1}{M-1}\sum_{j=1}^{M}(x_{ij} - \overline{X}_{i.})(z_{ij} - \overline{Z}_{i.})$

and f = $\frac{1}{n} - \frac{1}{N}, \acute{f} = \frac{1}{\acute{n}} - \frac{1}{N}, f_m = \frac{1}{m} - \frac{1}{M}, f_{mr} = \frac{1}{mq+2p} - \frac{1}{M}$

$f_1 = \frac{1}{n} - \frac{1}{\acute{n}}$

$M(T_P) = \overline{Y}^2\acute{A} + \alpha_2^2\acute{B} + 2\overline{Y}\alpha_2\acute{C}$

Now, differentiating partially both sides with respect to α_2, we get

$\frac{M(T_P)}{\delta\alpha_2} = 0$

$2\alpha_2\acute{B} + 2\overline{Y}_{..}\acute{C} = 0$

$\alpha_2 = -\frac{\acute{C}\overline{Y}_{..}}{B}$

$M(T_P) = \overline{Y}^2_{..}(\acute{A} - \frac{\acute{C}^2}{\acute{B}})$

Where

$\acute{A} = E(e_0^2)$

= $(\frac{1}{n} - \frac{1}{N})\frac{S_0^{*2}}{\overline{Y}^2_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_0^2}{\overline{Y}^2_{..}}$

$\acute{B} = E[e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}}]^2$

= $\frac{S_1^{*2}}{\overline{X}^2_{..}}[(\frac{1}{\acute{n}} - \frac{1}{N}) - (\frac{1}{n} - \frac{1}{N})] + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_1^2}{\overline{X}^2_{..}} - \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{4\overline{X}_{..}}\frac{S_2^{*2}}{\overline{Z}^2_{..}}[(\frac{1}{n} - \frac{1}{N}) + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})]$

$\acute{C} = E[e_0(e_2 - e_5 + \frac{e_3}{2\overline{X}_{..}})]$

= $(\frac{1}{\acute{n}} - \frac{1}{N})\frac{S^*_{01}}{\overline{Y}_{..}\overline{X}_{..}} - [(\frac{1}{n} - \frac{1}{N})\frac{S^*_{01}}{\overline{Y}_{..}\overline{X}_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_{01}}{\overline{Y}_{..}\overline{X}_{..}}] + \frac{1}{2\overline{X}_{..}}[(\frac{1}{n} - \frac{1}{N})\frac{S^*_{02}}{\overline{Y}_{..}\overline{Z}_{..}} + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})\frac{\overline{S}_{02}}{\overline{Y}_{..}\overline{X}_{..}}]$

Now, after applying the measurement error, the parameters becomes

$A^{//} = (\frac{1}{n} - \frac{1}{N})\frac{S_U^{*2}}{\overline{Y}^2_{..}} + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_U^2}{\overline{Y}^2_{..}}$

$B^{//} = \frac{S_V^{*2}}{\overline{X}^2_{..}}[(\frac{1}{\acute{n}} - \frac{1}{N}) - (\frac{1}{n} - \frac{1}{N})] + \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_V^2}{\overline{X}^2_{..}} - \frac{1}{n}(\frac{1}{\acute{m}q+2P} - \frac{1}{M})\frac{\overline{S}_{12}}{\overline{X}_{..}\overline{Z}_{..}} + \frac{1}{4\overline{X}_{..}}\frac{S_W^{*2}}{\overline{Z}^2_{..}}[(\frac{1}{n} - \frac{1}{N}) + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})]$

$C^{//} = 0$

After measurement error, we have our final estimator as

$M(T_{P_2}) = \overline{Y}^2_{..}(A^* - \frac{C^{*2}}{B^*})$

Where

$A^* = A^{/} + A^{//}$

$B^* = B^{/} + B^{//}$

$C^* = C^{/} + C^{//}$

3.1 Comparison of Efficiency

Now, we are going to have a comparison of our suggested estimators in two-stage design with the natural mean per unit estimator $\overline{y}_{nm}$ (where no extra information is used) of population mean $\overline{Y}_{..}$ in two-stage sampling scheme to support the effectiveness of our suggested methodology. The mean per unit estimator $\overline{y}_{nm}$ is an unbiased estimator and its variance is given as

$V(\overline{y}_{nm}) = (\frac{1}{n} - \frac{1}{N})S_b^2 + \frac{1}{n}(\frac{1}{m} - \frac{1}{M})S_W^2$

Where S_b^2 = mean square of study character y between the cluster means

= $\frac{1}{N-1}\sum_{i=1}^{N}(\overline{y}_i - \overline{Y}_{..})^2$

Where S_W^2 = mean square of study character y within the cluster

= $\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{M-1}\sum_{j=1}^{M}(y_{ij} - \overline{y}_i)^2\right\}$.

To elucidate the performance of our proposed procedure, we have observed percent relative losses in accuracy of our proposed estimators with respect to the usual mean per unit estimator built on the usual two-stage design approach with no use of extra information. The empirical studies are approved through natural population and simulated population

Table 1: Artificial population study

Artificial Population Study					
Non Response Probability		Estimator: T_{P_1} (N=10, M=10, ń=7, n=5, m=7)		Estimator: T_{P_2} (N=10, M=10, n=5, ḿ=8, m=7)	
		In Absence of Measurement Error	In Presence of Measurement Error	In Absence of Measurement Error	In Presence of Measurement Error
p	q	L_1	L_1	L_2	L_2
0.05	0.95	-81.13	-97.53	-85.28	-92.24
0.1	0.9	-100.01	-110.21	-105.34	-110.12
0.15	0.85	-133.29	-130.35	-128.15	-139.98
0.2	0.8	-140.34	-155.65	-136.62	-147.80

data sets.
The percent relative losses in precision of suggested estimators T_{P_i} (i= 1, 2) with respect to natural mean per unit estimator $\bar{y}_{nm}$ are given as
$L_i = \frac{M(T_{P_i})-V(\bar{y}_{nm})}{M(T_{P_i})} \times 100$; (i = 1,2)
We have derived percent relative loss in efficiency of proposed estimator T_{P_i} (i = 1, 2) with respect to natural sample mean estimator $\bar{y}_{nm}$ for different choices of non-response prbabilities, as shown in tables 1 and 2. It may be noted that the measurement error we consider is given by
U = Y + N (0,1)
V = X + N (0,1)
W = Z + N(0,1)

3.2 Empirical Investigation

3.2.1 Study Using Artificially Generated Population

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 (2003) and Singh et al.(2001) who have 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,....., N) from a standard normal distribution by using R-software. We have varied correlation coefficients ρ_{yx} and ρ_{xz} and produced the following transformed variables of the population U with the values of n=70, m=50, u=20, $\sigma_y^2 = 50$, $\mu_y = 10$, $\sigma_x^2 = 100$, $\mu_x = 50$, $\sigma_z^2 = 50$ and $\mu_z = 20$ as

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

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

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

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

3.2.2 Numerical Illustration Using Natural Population

We have taken into account the following real population to show the efficacy of the proposed estimation strategies in Tables 2. The community hospitals of the United States (Table No- 174)
Y: Number of hospitals of a state in United States in 2009
X: Total number of patients admitted in all hospitals of a state in United States in 2009

Table 2: Natural population study

Natural Population Study					
Non Response Probability		Estimator: T_{P_1} (N=10, M=10, $\acute{n}$=7, n=5, m=3)		Estimator: T_{P_2} (N=10, M=5, n=6, $\acute{m}$=4, m=3)	
		In Absence of Measurement Error	In Presence of Measurement Error	In Absence of Measurement Error	In Presence of Measurement Error
p	q	L_1	L_1	L_2	L_2
0.05	0.95	-205.45	-223.12	-210.15	-276.75
0.1	0.9	-245.29	-270.26	-278.32	-300.18
0.15	0.85	-273.44	-288.48	-320.28	-345.24
0.2	0.8	-280.78	-304.25	-368.16	-380.45

Z: Total number of beds in all hospitals of a state in United States in 2009
To show the efficacy of our newly suggested methodology, we have taken into account the following clusters for the case when random non-response does not vary between the clusters:
Cluster 1 = Alabama, Alaska, Arizona Arkansas, California
Cluster 2 =Colorado, Connecticut, District of Colombia, Florida, Georgia
Cluster 3 = Hawaii, Idaho, Illinois, Indiana, Iowa
Cluster 4 = Kansas, Kentucky, Louisiana, Maine, Maryland
Cluster 5 = Massachusetts, Michigan, Minnesota, Mississippi, Missouri
Cluster 6 = Montana, Nebraska, Nevada, New Hampshire, New Jersey
Cluster 7 = New Mexico, New York, North Carolina, North Dakota, Ohio
Cluster 8 = Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina
Cluster 9 = South Dakota, Tennessee, Texas, Utah, Vermont
Cluster 10 = Virginia, Washington, West Virginia, Wisconsin, Wyoming
where N = 10, M = 5, $\overline{Y}_{..} = 100.02$, $\overline{X}_{..} = 708500.88$, $\overline{Z}_{..} = 16068.14$
Now, consider the case when random non-response diverges between the clusters, the following clusters are built to intricate the result:
Cluster 1 = [Alabama, Alaska, Arizona Arkansas, California, Colorado, Connecticut,District of Colombia, Florida, Georgia]
Cluster 2 = [Hawaii, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland]
Cluster 3 = [Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey]
Cluster 4 = [New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina]
Cluster 5 = [South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming]
where N = 5 M = 10, $\overline{Y}_{..} = 100.02$, $\overline{X}_{..} = 708500.88$ and $\overline{Z}_{..} = 16068.14$
Survey data are collected from the Statistical Abstract of the United States, 2012, published by the United States Census Bureau.

4 Perspective

1. We can say from Table 1 and 2 that in the presence or absence of measurement error, our loss is increasing. our proposed methodology is higher than the standard cluster sampling scheme as we notice that PRE of our estimator with respect to the standard mean per unit estimator is becoming higher in all cases and also for all choices of probabilities of non-response. Also, negative loss indicates that our proposed procedure has less variance than the standard two-stage sampling scheme. Therefore, the proposed procedure is more acceptable in comparison with the standard two stage cluster sampling, when no auxiliary information has been used. Hence, is suggested for survey specialists for their use in real life problems.

2. Table 1 and 2 demonstrate about the outcomes and trends for increasing non response rate. For increasing values of non-response probabilities, losses in efficiencies of our proposed estimators are also increasing. This phenomenon is highly looked-for as it indicates effectiveness of the proposed strategies. It also justifies that capability of our suggested methodologies in handling random non-response situations in practical surveys.
3. Measurement error is very important for our research. Measurement error is the difference between a measured quantity and its true value. It also includes random error. We have done this research to see what is the effect if there is a measurement error in the case of non response.
4. Lack of practical support of earlier work (Cluster sampling-random non-response and measurement error together; what happens when non-response and measurement error, together, take place (as written in introduction part also)) and we need to justify the effectiveness of our work and finding with respect to the above remark.

We can say from Table 1 and 2 that for artificially produced population datasets, our proposed procedure is greater than the standard cluster sampling structure as we notice that PRE of this estimator with respect to the standard mean per unit estimator is becoming higher in all cases and also for all selections of probabilities of non-response. Also, the negative loss shows that our proposed procedure has less variance than the standard two-stage sampling scheme.

Acknowledgement

Authors are thankful to Science and Engineering Research Board, Department of Science and Technology, Govt. of India for funding the research project (File No : EMR/2016/001320), which help us to carry out this research work.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] G.N. Singh and M. Khalid. An effective estimation strategy for population mean under random non-response situations in two-phase successive sampling. *Communications in Statistics-Theory and Methods*, **47(19)**, 4641-4660 (2018).
[2] G. N. Singh, M. Khalid and A. K. Sharma. Some efficient classes of estimators of population mean in two-phase successive sampling under random non response. *Communications in Statistics-Theory and Methods*, **46(24)**, 12194-12209 (2017).
[3] H. Lee, E. Rancourt and C. E. Sarndall. Experiments with variance estimation from survey data with imputed values. *Journal of Official Statistics*, **10 (3)**, 231-243 (1994).
[4] H. P. Singh and R. S. Solanki. Estimation of finite population mean using random non-response in survey sampling. *Pakistan Journal of Statistics and Operation Research*, **VII (1)**, 21-41 (2011).
[5] I. G. Sande: A personal view of hot–deck imputation procedures. *Survey Methodology, Statistics Canada*, **5**, 238–247 (1979).
[6] L. N. Upadhyaya and H. P. Singh. Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, **41 (5)**, 627-636 (1999).
[7] M. Azeem and M. Hanif. On estimation of population mean in the presence of measurement error and non-response. *Pak. J. Statist.*, **31(5)**, 657-670 (2015).
[8] R. B. Rubin. Inference and missing data. *Biometrika*, **63(3)**, 581–592 (1976).
[9] S. Kumar, S. Bhougal, N. S. Nataraja and M. Viswanathaiah. Estimation of population mean in the presence of non-response and measurement Error. *Revista Colombiana de Estadistica*, **38(1)**, 145-161 (2015).
[10] S. Mishra, D. K. Yadav and Dipika. An Efficient estimator for estimating population variance in presence of measurement errors. *International Journal of Mathematics and its Applications*, **4(2-D)**, 23-28 (2016).
[11] S. Singh and A. H. Joarder. Estimation of finite population variance using random non-response in survey sampling. *Metrika*, **47**, 241-249 (1998).