REGRESSION-TYPE IMPUTATION SCHEME UNDER SUBSAMPLING WITH EQUAL CHANCE OF RANDOM NON-RESPONSE AT FIRST STAGE.

Abubakar Ibrahim[1], Yahaya Abubakar [2], Garba Jamilu[3] and Aliyu Yakubu[4]

•

[1]Department of Mathematical Sciences, Kaduna State University Kaduna, Nigeria
[2, 3, 4]Department of Statistics, Ahmadu Bello University, Zaria, Kaduna, Nigeria
[1]ibrahim.abubakar@kasu.edu.ng. [2]ensiliyu2@yahoo.co.uk, [3]gyjamilu@gmail.com,
[4]yakubualiyu@abu.edu.ng

Abstract

The study addresses the challenges of estimating the population mean in two-stage cluster sampling, where there is an equal chance of random non-response at the first-stage unit. The researchers propose some regression-type imputation schemes and regression-type estimators that incorporate measurement error parameters for both the study and supplementary variables. The properties of the proposed estimators were derived and numerically compared using a simulated sample population. The proposed estimators outperformed the existing estimators consider in the study. The researchers conclude that their proposed methodology can be practically applied, using the actual responses of the respondents and including the measurement error parameters to estimate the finite population mean.

Keywords: First Stage Unit,Regression-type imputation scheme,Regression-type estimators,Random Non-Response

I. Introduction

In field surveys frequently indicate that it is not always possible to obtain a complete list of everyone who is a member of the research population, indicating that selecting a simple random sample is difficult. Cluster sampling can be used to collect data in this scenario because it is usually less expensive and does not require a list of all observations in the population [1].
Clusters are produced in cluster sampling by dividing the survey area into smaller sub-areas. Then using simple random sampling, some of these areas are chosen, and all elements of the chosen clusters are counted.

Assuming we are interested in the academic performance of all 400 level students in a given city. Because there is no sampling frame for such units, obtaining a complete list of everyone in the research population is extremely difficult. However, a list of university each student attends should be available. In these cases it is recommended to select a simple random sample of 400 level students from each university. The technique used in this study is two-stage cluster sampling. In cluster sampling, better precision is achieved by first selecting a cluster and then enumerating a specific number of elements from each cluster. Two-stage cluster sampling refers to the process of first picking clusters, which are the sampling units in the first stage, and then selecting a predetermined number of elements from each selected cluster, which are the sampling units in the second stage. The clusters that constitute the sampling units in the first stage are referred to as First

Stage Units (FSUs) or Primary Stage Units. The elements inside these clusters that form the sampling units in the second stage are referred to as Second Stage Units (SSUs).The key advantage of this two-stage cluster sampling approach is that it can provide better precision in the estimates compared to simpler random sampling methods.

In a sample survey, it is usually assumed that all information is obtained from the study population's unit and that the observed variables are obtained without error. Such an assumption is not always met because researchers face the issues of non-response and measurement error. Most human population surveys face the problem of non-response, where some units of the study population fail to provide the requested information for various reasons, such as refusal, absence, lack of interest, or adherence to ethical standards. This non-response causes issues during data collection, calculation, and estimation. The problem of non-response in estimating the finite population mean was first address by [2]. The typical approach is to return to the field and collect the missing values through a call-back method, but this requires additional resources like time, people, and money. Three concepts related to non-response: Missing at Random (MAR), Observed at Random (OAR), and Parameterized Distribution (PD), were discussed by [3]. According to [3], data are MAR when the probability of missing data does not depend on the value of the unobserved data. Missing completely at random (MCAR) and missing at random (MAR) was distinguished by [4]. Various imputation schemes have been used over time to address the problem of estimating unknown parameters in the presence of missing values in sample surveys. Imputation involves filling in missing values with specific substitutes so that standard data analysis methods can be applied. The regression imputation method is used to replace missing value with a linear function. The approach also predicts missing values using regression models using the other variables, and the fitted values are entered into the model. It is assumed that the value of one variable varies linearly with other variables. Several researchers, including [5-14], and many others have proposed imputation methods to handle missing data. However, drawing simple random samples is impractical without a full list of every unit in the population. As a result, the imputation schemes and their estimators suggested by the previous literature are not applicable when the complete list of all population units is not available (as is the case in non-response). Cluster sampling is a common sampling method when there is no complete list of all population units in a survey. Hence the adoption of two-stage cluster sampling method used in this study.

In addition to non-response, survey researchers face the issue of measurement error (ME). Several survey researchers work under the assumption that the information they acquire from respondents is correct, and some of the estimator attributes (biases and mean square errors) are derived from this assumption. The assumption that the observed data accurately represents the true values is not always correct, as researchers often face the problem of measurement error. Measurement error refers to the discrepancy between the observed values gathered from respondents and the real, underlying values. In other words, the observed data may not perfectly reflect the true information, and there is an element of error or inaccuracy introduced during the data collection process. This measurement error can be problematic and needs to be accounted for in the analysis, as relying solely on the observed data may lead to biased or inaccurate results. Let’s assume we want to collect information on the cumulative grade point average (CGPA) of some students; some may report a CGPA that is lower or higher than their real CGPA. As a result, the observed value remains erroneous, because the students did not provide their real CGPA. When the measurement error is insignificant, the inferences drawn on the observed value may be correct; nevertheless, when the measurement is not insignificantly small, the inferences taken on the observed value may have some unanticipated and unpleasant implications.

Several researchers have examined the issue of measurement error separately in their work, including [15-24].

Typically, non-response and ME are investigated separately using known supplementary variables. In reality, survey sampling results to simultaneous measurement errors and non-response. However, in this research, we will study both non-response and measurement errors.

II. Methods

2. Construction of sample structure

Suppose U is a finite population be divided into N FSU represented by $\left(U_1, U_2, \cdots U_N\right)$ in such a way that the quantity of SSU in every first stage unit is M. Assume y_{ij}, x_{1ij}, and x_{2ij} be the actual values for the character under study Y, first supplementary variable X_1 and second supplementary variable X_2, respectively. Also, assume that $y_{ij(e)}$, $x_{1ij(e)}$, $x_{2ij(e)}$ be the observed values for Y, X_1, X_2 on the j^{th} second stage units $\left(j=1,2,\cdots M\right)$ in the i^{th} first stage units $\left(i=1,2,\cdots N\right)$. Let U_{ij}, V_{ij} and W_{ij} be the ME parameters associated with the study variable, first supplementary variable, and second supplementary variable respectively. The MEs associated with these variables are thus defined as:

The ME associated with the character under study be

$$U_{ij} = y_{ij} - y_{ij(e)},\ U_{ij} \sim N\left(0, \sigma_U^2\right) \tag{1}$$

The ME associated with the first supplementary variable be

$$V_{ij} = x_{1ij} - x_{1ij(e)},\ V_{ij} \sim N\left(0, \sigma_V^2\right) \tag{2}$$

The ME associated with the second supplementary variable be

$$W_{ij} = x_{2ij} - x_{2ij(e)},\ W_{ij} \sim N\left(0, \sigma_W^2\right) \tag{3}$$

However, in this research work, we take into account a scenario in which the information on the first supplementary variable X_1 is not known at the level of first stage unit. Hence, information on the first supplementary variable x_1 can be gathered using the following strategy:

STRATEGY: At the level of first stage units, information on the first supplementary variable x_1 is gathered, and a sample of the first stage unit is chosen using the SRSWOR procedure.

Moreover, the above-mentioned strategy will be discussed under clusters with equal chance of random non-response which is detailed below.

2.1. Clusters with equal chance of random non-response

2.1.1. Strategy: When supplementary information is gathered at Level of First Stage Unit

We take into account a situation where the population mean $\bar{X}_{1..}$ of the first supplementary variable x_1 is unknown at the level of first stage unit, so we used a two-phase or double sampling strategy to furnish the estimate. However, information on the second supplementary variable x_2 is known for every unit of the population. In order to estimate the population mean of Y, a first phase sample $S_{n'}\left(S_{n'} \subset U\right)$ of size n' first stage unit is taken out of N FSU from the population using SRSWOR method followed by a second phase sample S_1 of size n FSU $\left(n < n'\right)$ taken based on the subsequent two cases by using the SRSWOR technique to observe the character under study Y.

Case A: To create a S_1, a subsample of $S_{n'}\left(S_1 \subset S_{n'}\right)$ is taken.

Case B: In this case, S_1 is drawn independently of $S_{n'}$.

Furthermore, in order to estimate the population mean of Y, a second stage sample S_2 is obtained by selecting a portion of m second stage unit from M second stage unit for every one of the n chosen first stage unit in S_1 utilizing SRSWOR scheme.

At the second stage, it is assumed that the study variable y and the first supplementary variable x_1 have random failure to respond, but the sampled unit has full response for the second supplementary variable x_2. For such random non-response conditions, we consider the following probability model shown in section (2.1.1.1).

2.1.1.1. Probability of Non-Response Model

Since, we assume the occurrence of random non-response conditions on the study variable y and the first supplementary variable x_1 from the second stage sample; therefore, we are going to investigate random non-response conditions from the second stage sample S_2. Let $r\{r=0,1,\cdots,(m-2)\}$ represents the number of second stage sampling units that did not respond. Accordingly, we write A_r and A_r^c to represent the collection of respondent unit and non-respondent unit, respectively. The observations of the corresponding variables in which random non-response occurs could be obtained from the rest of the $(m-r)$ unit of each of the n first stage unit of the second stage sample (SSS).

We further suppose that if p represents the probability of random failure to respond among $(m-2)$ possible failure to respond cases, and then r follows the probability distribution shown in equation (4):

$$P(r)=\frac{m-r}{mq+2p}\,{}^{(m-2)}C_r p^r q^{m-2-r} \;;\; r=0,1,\cdots(m-2) \tag{4}$$

For example, see the work of [25-27], where $q=1-p$ and ${}^{(m-2)}C_r$ represent the overall possible methods to provide r failure to respond from $(m-2)$ total non-response.

Henceforth, the following notations will be used:

$\bar{Y}_{..}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}Y_{ij}$, Population average of the study variable y.

$\bar{X}_{1..}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}X_{1ij}$, Population average of the first supplementary variable x_1.

$\bar{X}_{2..}=\frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M}X_{2ij}$, Population average of the second supplementary variable x_2.

$\bar{y}^*_{i(e)}=\frac{1}{m}\sum_{j=1}^{m}y_{ij(e)}$, Sample average of the character under study on i^{th} FSU in S_2.

$\bar{y}^*_{i(m-r)(e)}=\frac{1}{m-r}\sum_{j=1}^{m-r}y_{ij(e)}$, Sample mean of y based on the respondent region of i^{th} FSU in S_2.

$\bar{x}^*_{1i(e)}=\frac{1}{m}\sum_{j=1}^{m}x_{1ij(e)}$, Sample average of the first supplementary variable on i^{th} FSU in S_2.

$\bar{x}^*_{1i(m-r)(e)}=\frac{1}{m-r}\sum_{j=1}^{m-r}x_{1ij(e)}$, Sample mean of x_1 based on the respondent region of i^{th} FSU in S_2.

$\bar{x}^*_{2i(e)}=\frac{1}{m}\sum_{j=1}^{m}x_{2ij(e)}$, Sample average of x_2 on i^{th} FSU in S_2.

$\bar{y}^{**}_{n(m-r)(e)}=\frac{1}{n}\sum_{i=1}^{n}\bar{y}^*_{i(e)}$, Sample average of the n FSU of the character under study.

$\bar{x}^{**}_{1n(m-r)(e)} = \frac{1}{n}\sum_{i=1}^{n}\bar{x}^{*}_{1i(e)}$, Sample average of the n FSU of the first variable.

$\bar{x}^{**}_{2nm(e)} = \frac{1}{n}\sum_{i=1}^{n}\bar{x}^{*}_{2i(e)}$, Sample average of the n FSU of the second supplementary variable.

$C_{X_2} = \frac{S_{X_2}}{\bar{X}_{2..}}$, Coefficient of variation of the second supplementary variable.

S_{X_2}, Standard deviation of the second supplementary variable.

$B_1(X_2)$, Population coefficient of skewness of the second supplementary variable.

$B_2(X_2)$, Population coefficient of skewness of the second supplementary variable.

2.1.1.2 Proposed Imputation Schemes and Estimators

We assumed that the second supplementary variable was generally accessible all through the population U. Inspired by the imputation schemes given by [28], we propose the following regression-type imputation schemes based on responding and non-responding units of the second stage sample S_2 to estimate population parameter under study $\bar{Y}_{..}$ as:

$$y_{ij(e)} = \begin{cases} y_{ij(e)}, & if\ j \in A_r \\ \dfrac{\bar{y}_{i(m-r)(e)} + b_{yx_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right), & if\ j \in A_r^c \end{cases} ;\ (i = 1,\ 2, \ldots, n) \tag{5}$$

where A_{x_2} and B_{x_2} are available functions of supplementary variable like coefficient of skewness, kurtosis, variation, standard deviation, $b_{yx_2(e)} = \sum_{j \in A_r} y_{ij(e)} \Big/ \sum_{j \in A_r} x_{2ij(e)}$, $b_{x_1x_2(e)} = \sum_{j \in A_r} x_{1ij(e)} \Big/ \sum_{j \in A_r} x_{2ij(e)}$.

Remark 1: Note that $A_{x_2} \neq B_{x_2}$ and $A_{x_2} \neq 0$

Under this approach, we derived the sample means of y on the i^{th} first stage units in S_2 denoted by $\bar{y}^{*}_{i(e)}$ as:

$$\bar{y}^{*}_{i(e)} = \frac{1}{m}\sum_{j=1}^{m} y_{ij(e)} = \frac{1}{m}\left[\sum_{j \in A_r} y_{ij(e)} + \sum_{j \in A_r^c} y_{ij(e)}\right] \tag{6}$$

$$\bar{y}^{*}_{i(e)} = \frac{1}{m}\sum_{j=1}^{m} y_{ij(e)} = \left(1 - \frac{r}{m}\right)\bar{y}_{i(m-r)(e)} + \frac{r}{m}\left(\frac{\bar{y}_{i(m-r)(e)} + b_{yx_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right)\right) \tag{7}$$

In S_2, the mean of n first stage unit of y is now:

$$\bar{y}^{**}_{n(m-r)(e)} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}^{*}_{i(e)} = \left(1 - \frac{r}{m}\right)\bar{y}_{n(m-r)(e)} + \frac{r}{m}\left(\frac{\bar{y}_{n(m-r)(e)} + b_{yx_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2mn(e)}\right)}{A_{x_2}\bar{x}_{2mn(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right)\right) \tag{8}$$

Likewise, for each unit in the second stage

$$x_{1ij(e)} = \begin{cases} x_{1ij(e)}, & j \in A_r \\ \dfrac{\bar{x}_{1i(m-r)(e)} + b_{x_1x_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right), & j \in A_r^c \end{cases} ;\ (i = 1, 2, \ldots, n) \tag{9}$$

Under this approach, we derived the sample means of x_{1i} on the i^{th} first stage units in S_2 denoted by $\bar{x}^*_{1i(e)}$ as:

$$\bar{x}^*_{1i(e)} = \frac{1}{m}\sum_{j=1}^{m} x_{1ij(e)} = \frac{1}{m}\left[\sum_{j\in A_r} x_{1ij(e)} + \sum_{j\in A_r^c} x_{1ij(e)}\right] \tag{10}$$

$$\bar{x}^*_{1i(e)} = \frac{1}{m}\sum_{j=1}^{m} x_{1ij(e)} = \left(1 - \frac{r}{m}\right)\bar{x}_{1i(m-r)(e)} + \frac{r}{m}\left(\frac{\bar{x}_{1i(m-r)(e)} + b_{x_1x_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right)\right) \tag{11}$$

In S_2, the mean of n first stage unit of $x_{1(e)}$ is now:

$$\bar{x}^{**}_{1n(m-r)(e)} = \frac{1}{n}\sum_{i=1}^{n} \bar{x}^*_{1i(e)} = \left(1 - \frac{r}{m}\right)\bar{x}_{1n(m-r)(e)} + \frac{r}{m}\left(\frac{\bar{x}_{1n(m-r)(e)} + b_{x_1x_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2mn(e)}\right)}{A_{x_2}\bar{x}_{2mn(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right)\right) \tag{12}$$

Hence the proposed estimator denoted by τ^* under the above proposed imputation scheme is obtained as:

$$\tau^* = \bar{y}^{**}_{n(m-r)(e)} + b^*_{(e)}\left(\bar{x}_{1n'M(e)} - \bar{x}^{**}_{1n(m-r)(e)}\right) \tag{13}$$

Where $b^*_{(e)}$ is a suitable constant chosen to minimize the mean square error of the proposed estimator τ^*.

2.1.1.3 Properties of the Proposed Estimators

Since τ^* is regression-type estimator, it is biased for $\bar{Y}_{..}$ the bias and mean square of τ^* up to the first order of approximations are derived under large sample approximations (ignoring f.p.c) using the following assumptions:

$\bar{y}_{n(m-r)(e)} = \bar{Y}_{..}\left(1 + \Delta_{0(e)}\right)$, $\bar{x}_{1n(m-r)(e)} = \bar{X}_{1..}\left(1 + \Delta_{1(e)}\right)$, $\bar{x}_{1n'M(e)} = \bar{X}_{1..}\left(1 + \Delta_{2(e)}\right)$, $\bar{x}_{2nm(e)} = \bar{X}_{2..}\left(1 + \Delta_{3(e)}\right)$

Such that $E\left(\Delta_i\right) = 0$ and $\left|\Delta_i\right| < 1$ for all $i = 0,1,2,3$.

Express $\left(\tau^*\right)$ in terms of errors $\Delta_{0(e)}$, $\Delta_{1(e)}$, $\Delta_{2(e)}$ and $\Delta_{3(e)}$.

$$\begin{aligned} \hat{\tau}_1 = \bar{Y}_{..}&\left[1 + \Delta_{0(e)} - \frac{r}{m}\left(\vartheta_X + 1\right)\Delta_{3(e)} + \frac{r}{m}\left(\vartheta_X^2 + \vartheta_X + 1\right)\Delta^2_{3(e)} - \left(\vartheta_X + 1\right)\Delta_{0(e)}\Delta_{3(e)}\right] + \\ b^*_{(e)}&\left[\left(1 + \Delta_{2(e)}\right)\bar{X}_{1..} - \bar{X}_{1..}\left(1 + \Delta_{1(e)} - \frac{r}{m}\left(\vartheta_X + 1\right)\Delta_{3(e)} + \frac{r}{m}\left(\vartheta_X^2 + \vartheta_X + 1\right)\Delta^2_{3(e)} - \frac{r}{m}\left(\vartheta_X + 1\right)\Delta_{1(e)}\Delta_{3(e)}\right)\right] \end{aligned} \tag{14}$$

$$\begin{aligned}\hat{\tau}_1-\bar{Y}_{\cdot\cdot}=\bar{Y}_{\cdot\cdot}\left[\Delta_{0(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{3(e)}+\frac{r}{m}\left(\vartheta_X^2+\vartheta_X+1\right)\Delta_{3(e)}^2-\left(\vartheta_X+1\right)\Delta_{0(e)}\Delta_{3(e)}\right]+\\ b_{(e)}^{*}\bar{X}_{1\cdot\cdot}\left[\Delta_{2(e)}-\Delta_{1(e)}+\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{3(e)}-\frac{r}{m}\left(\vartheta_X^2+\vartheta_X+1\right)\Delta_{3(e)}^2+\frac{r}{m}\left(\vartheta_X+1\right)\Delta_{1(e)}\Delta_{3(e)}\right]\end{aligned} \quad (15)$$

Where, $\vartheta_X=\left(\dfrac{A_{x_2}\bar{X}_{2\cdot\cdot}}{A_{x_2}\bar{X}_{2\cdot\cdot}+B_{x_2}}\right)$

We have separately derived the bias and mean square error of the estimator τ^* for Cases A and B of the two-phase sampling structure defined in Section 2, and these are presented below.

Case A: To create a S_1, a subsample of $S_{n'}\left(S_1\subset S_{n'}\right)$ is taken.

To obtain the expressions for the bias and mean square error in this situation, we will consider the following expected values of the sample statistics.

$$E\left(\Delta_{0(e)}^2\right)=\varphi_{n,N}\frac{S_y^{*2}+S_u^{*2}}{\bar{Y}_{\cdot\cdot}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_y^2+\bar{S}_u^2}{\bar{Y}_{\cdot\cdot}^2},\ E\left(\Delta_{1(e)}^2\right)=\varphi_{n,N}\frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1\cdot\cdot}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{x_1}^2+\bar{S}_v^2}{\bar{X}_{1\cdot\cdot}^2},$$

$$E\left(\Delta_{2(e)}^2\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1\cdot\cdot}^2},\ E\left(\Delta_{3(e)}^2\right)=\varphi_{n,N}\frac{S_{x_2}^{*2}+S_w^{*2}}{\bar{X}_{2\cdot\cdot}^2}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_2}^2+\bar{S}_w^2}{\bar{X}_{2\cdot\cdot}^2},$$

$$E\left(\Delta_{0(e)}\Delta_{1(e)}\right)=\varphi_{n,N}\frac{S_{yx_1}^{*}+S_{uv}^{*}}{\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{yx_1}+\bar{S}_{uv}}{\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}},\ E\left(\Delta_{0(e)}\Delta_{2(e)}\right)=\varphi_{n',N}\frac{S_{yx_1}^{*}+S_{uv}^{*}}{\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}},$$

$$E\left(\Delta_{0(e)}\Delta_{3(e)}\right)=\varphi_{n,N}\frac{S_{yx_2}^{*}+S_{uw}^{*}}{\bar{Y}_{\cdot\cdot}\bar{X}_{2\cdot\cdot}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{yx_2}+\bar{S}_{uw}}{\bar{Y}_{\cdot\cdot}\bar{X}_{2\cdot\cdot}},\ E\left(\Delta_{1(e)}\Delta_{2(e)}\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1\cdot\cdot}^2}=E\left(\Delta_{2(e)}^2\right),$$

$$E\left(\Delta_{1(e)}\Delta_{3(e)}\right)=\varphi_{n,N}\frac{S_{x_1x_2}^{*}+S_{vw}^{*}}{\bar{X}_{1\cdot\cdot}\bar{X}_{2\cdot\cdot}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_1x_2}+\bar{S}_{vw}}{\bar{X}_{1\cdot\cdot}\bar{X}_{2\cdot\cdot}},\ E\left(\Delta_{2(e)}\Delta_{3(e)}\right)=\varphi_{n',N}\frac{S_{x_1x_2}^{*}+S_{vw}^{*}}{\bar{X}_{1\cdot\cdot}\bar{X}_{2\cdot\cdot}}$$

For simplicity we let

$$\left.\begin{aligned}&\zeta_{0(e)}=E\left(\Delta_{0(e)}^2\right),\zeta_{1(e)}=E\left(\Delta_{1(e)}^2\right),\zeta_{2(e)}=E\left(\Delta_{2(e)}^2\right)=E\left(\Delta_{1(e)}\Delta_{2(e)}\right),\\&\zeta_{3(e)}=E\left(\Delta_{3(e)}^2\right),\zeta_{4(e)}=E\left(\Delta_{0(e)}\Delta_{1(e)}\right),\zeta_{5(e)}=E\left(\Delta_{0(e)}\Delta_{2(e)}\right),\\&\zeta_{6(e)}=E\left(\Delta_{0(e)}\Delta_{3(e)}\right),\zeta_{7(e)}=E\left(\Delta_{1(e)}\Delta_{3(e)}\right),\zeta_{8(e)}=E\left(\Delta_{2(e)}\Delta_{3(e)}\right)\end{aligned}\right\} \quad (16)$$

The following notations and expectation will be use under this strategy, when measurement error is not taking into account.

$$E\left(\Delta_0^2\right)=\varphi_{n,N}\frac{S_y^{*2}}{\bar{Y}_{\cdot\cdot}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_y^2}{\bar{Y}_{\cdot\cdot}^2},\ E\left(\Delta_1^2\right)=\varphi_{n,N}\frac{S_{x_1}^{*2}}{\bar{X}_{1\cdot\cdot}^2}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{x_1}^2}{\bar{X}_{1\cdot\cdot}^2},\ E\left(\Delta_2^2\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}}{\bar{X}_{1\cdot\cdot}^2},$$

$$E\left(\Delta_3^2\right)=\varphi_{n,N}\frac{S_{x_2}^{*2}}{\bar{X}_{2\cdot\cdot}^2}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_2}^2}{\bar{X}_{2\cdot\cdot}^2},\ E\left(\Delta_0\Delta_1\right)=\varphi_{n,N}\frac{S_{yx_1}^{*}}{\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}}+\frac{1}{n}\varphi_{m,r}\frac{\bar{S}_{yx_1}}{\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}},\ E\left(\Delta_0\Delta_2\right)=\varphi_{n',N}\frac{S_{yx_1}^{*}}{\bar{Y}_{\cdot\cdot}\bar{X}_{1\cdot\cdot}},$$

$$E\left(\Delta_0\Delta_3\right)=\varphi_{n,N}\frac{S_{yx_2}^{*}}{\bar{Y}_{\cdot\cdot}\bar{X}_{2\cdot\cdot}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{yx_2}}{\bar{Y}_{\cdot\cdot}\bar{X}_{2\cdot\cdot}},\ E\left(\Delta_1\Delta_2\right)=\varphi_{n',N}\frac{S_{x_1}^{*2}}{\bar{X}_{1\cdot\cdot}^2}=E\left(\Delta_2^2\right),$$

$$E\left(\Delta_1\Delta_3\right)=\varphi_{n,N}\frac{S_{x_1x_2}^{*}}{\bar{X}_{1\cdot\cdot}\bar{X}_{2\cdot\cdot}}+\frac{1}{n}\varphi_{m,M}\frac{\bar{S}_{x_1x_2}}{\bar{X}_{1\cdot\cdot}\bar{X}_{2\cdot\cdot}},\ E\left(\Delta_2\Delta_3\right)=\varphi_{n',N}\frac{S_{x_1x_2}^{*}}{\bar{X}_{1\cdot\cdot}\bar{X}_{2\cdot\cdot}}$$

Similarly, for simplicity we let

$$\left.\begin{aligned}&\zeta_0=E\left(\Delta_0^2\right),\ \zeta_1=E\left(\Delta_1^2\right),\zeta_2=E\left(\Delta_2^2\right)=E\left(\Delta_1\Delta_2\right),\zeta_3=E\left(\Delta_3^2\right),\ \zeta_4=E\left(\Delta_0\Delta_1\right),\\&\zeta_5=E\left(\Delta_0\Delta_2\right),\zeta_6=E\left(\Delta_0\Delta_3\right),\ \ \zeta_7=E\left(\Delta_1\Delta_3\right),\ \zeta_8=E\left(\Delta_2\Delta_3\right)\end{aligned}\right\} \quad (17)$$

Where

$$\bar{Y}_{..}=\frac{1}{N}\sum_{i=1}^{N}\bar{Y}_{i.},\ \bar{Y}_{i.}=\frac{1}{M}\sum_{j=1}^{M}y_{ij},\ \bar{X}_{1..}=\frac{1}{N}\sum_{i=1}^{N}\bar{X}_{1i.},\ \bar{X}_{1i.}=\frac{1}{M}\sum_{j=1}^{M}x_{1ij},\ \bar{X}_{2..}=\frac{1}{N}\sum_{i=1}^{N}\bar{X}_{2i.},\ \bar{X}_{2i.}=\frac{1}{M}\sum_{j=1}^{M}x_{2ij},$$

$$S_{y}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)^{2},\ \bar{S}_{y}^{2}=\frac{1}{N}\sum_{i=1}^{N}S_{y_i}^{2},\ S_{y_{i.}}^{2}=\frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij}-\bar{Y}_{i.}\right)^{2}$$

$$S_{x_1}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{1i.}-\bar{X}_{1..}\right)^{2},\ \bar{S}_{x_1}^{2}=\frac{1}{N}\sum_{i=1}^{N}S_{x_{1i}}^{2},\ S_{x_{1i}}^{2}=\frac{1}{M-1}\sum_{j=1}^{M}\left(x_{1ij}-\bar{X}_{1i.}\right)^{2}$$

$$S_{x_2}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{2i.}-\bar{X}_{2..}\right)^{2},\ \bar{S}_{x_2}^{2}=\frac{1}{N}\sum_{i=1}^{N}S_{x_{2i}}^{2},\ S_{x_{2i}}^{2}=\frac{1}{M-1}\sum_{j=1}^{M}\left(x_{2ij}-\bar{X}_{2i.}\right)^{2}$$

$$S_{yx_1}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)\left(\bar{X}_{1i.}-\bar{X}_{1..}\right),\ \bar{S}_{yx_1}=\frac{1}{N}\sum_{i=1}^{N}S_{yx_{1i}},\ S_{yx_{1i}}=\frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij}-\bar{Y}_{i.}\right)\left(x_{1ij}-\bar{X}_{1i.}\right)$$

$$S_{yx_2}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{Y}_{i.}-\bar{Y}_{..}\right)\left(\bar{X}_{2i.}-\bar{X}_{2..}\right),\ \bar{S}_{yx_2}=\frac{1}{N}\sum_{i=1}^{N}S_{yx_{2i}},\ S_{yx_{2i}}=\frac{1}{M-1}\sum_{j=1}^{M}\left(y_{ij}-\bar{Y}_{i.}\right)\left(x_{2ij}-\bar{X}_{2i.}\right)$$

$$S_{x_1x_2}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{X}_{1i.}-\bar{X}_{1..}\right)\left(\bar{X}_{2i.}-\bar{X}_{2..}\right),\ \bar{S}_{x_1x_2}=\frac{1}{N}\sum_{i=1}^{N}S_{x_1x_{2i}},\ S_{x_1x_{2i}}=\frac{1}{M-1}\sum_{j=1}^{M}\left(x_{1ij}-\bar{X}_{1i.}\right)\left(x_{2ij}-\bar{X}_{2i.}\right)$$

$$\bar{U}_{..}=\frac{1}{N}\sum_{i=1}^{N}\bar{U}_{i.},\bar{U}_{i.}=\frac{1}{M}\sum_{j=1}^{M}u_{ij},\bar{V}_{..}=\frac{1}{N}\sum_{i=1}^{N}\bar{V}_{i.},\bar{V}_{i.}=\frac{1}{M}\sum_{j=1}^{M}v_{ij},\bar{W}_{..}=\frac{1}{N}\sum_{i=1}^{N}\bar{W}_{i.},\bar{W}_{i.}=\frac{1}{M}\sum_{j=1}^{M}w_{ij}$$

$$S_{u}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{U}_{i.}-\bar{U}_{..}\right)^{2},\ \bar{S}_{u}^{2}=\frac{1}{N}\sum_{i=1}^{N}S_{u_i}^{2},\ S_{u_{i.}}^{2}=\frac{1}{M-1}\sum_{j=1}^{M}\left(u_{ij}-\bar{U}_{i.}\right)^{2}$$

$$S_{v}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{V}_{i.}-\bar{V}_{..}\right)^{2},\ \bar{S}_{v}^{2}=\frac{1}{N}\sum_{i=1}^{N}S_{v_i}^{2},\ S_{v_i}^{2}=\frac{1}{M-1}\sum_{j=1}^{M}\left(v_{ij}-\bar{V}_{i.}\right)^{2}$$

$$S_{w}^{*2}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{W}_{i.}-\bar{W}_{..}\right)^{2},\ \bar{S}_{w}^{2}=\frac{1}{N}\sum_{i=1}^{N}S_{w_i}^{2},\ S_{w_i}^{2}=\frac{1}{M-1}\sum_{j=1}^{M}\left(w_{ij}-\bar{W}_{i.}\right)^{2}$$

$$S_{uv}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{U}_{i.}-\bar{U}_{..}\right)\left(\bar{V}_{i.}-\bar{V}_{..}\right),\ \bar{S}_{uv}=\frac{1}{N}\sum_{i=1}^{N}S_{uv_i},\ S_{uv_i}=\frac{1}{M-1}\sum_{j=1}^{M}\left(u_{ij}-\bar{U}_{i.}\right)\left(v_{ij}-\bar{V}_{i.}\right)$$

$$S_{uw}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{U}_{i.}-\bar{U}_{..}\right)\left(\bar{W}_{i.}-\bar{W}_{..}\right),\ \bar{S}_{uw}=\frac{1}{N}\sum_{i=1}^{N}S_{uw_i},\ S_{uw_i}=\frac{1}{M-1}\sum_{j=1}^{M}\left(u_{ij}-\bar{U}_{i.}\right)\left(w_{ij}-\bar{W}_{i.}\right)$$

$$S_{vw}^{*}=\frac{1}{N-1}\sum_{i=1}^{N}\left(\bar{V}_{i.}-\bar{V}_{..}\right)\left(\bar{W}_{i.}-\bar{W}_{..}\right),\ \bar{S}_{vw}=\frac{1}{N}\sum_{i=1}^{N}S_{vw_i},\ S_{vw_i}=\frac{1}{M-1}\sum_{j=1}^{M}\left(v_{ij}-\bar{V}_{i.}\right)\left(w_{ij}-\bar{W}_{i.}\right)$$

$$S_{b}^{2}=\frac{1}{N}\sum_{i=1}^{N}\left(\bar{y}_{i}-\bar{Y}_{..}\right)^{2},\ S_{w}^{2}=\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{1}{M-1}\sum_{j=i}^{M}\left(\bar{y}_{ij}-\bar{y}_{i}\right)^{2}\right\},\ S_{b(U)}^{2}=\frac{1}{N}\sum_{i=1}^{N}\left(\bar{U}_{i}-\bar{U}\right)^{2},\ S_{w(U)}^{2}=\frac{1}{N}\left\{\sum_{i=1}^{N}\frac{1}{M-1}\sum_{j=i}^{M}\left(U_{ij}-\bar{U}_{i}\right)^{2}\right\}$$

$$\varphi_{n,N}=\left(\frac{1}{n}-\frac{1}{N}\right),\ \varphi_{n',N}=\left(\frac{1}{n'}-\frac{1}{N}\right),\varphi_{m,M}=\left(\frac{1}{m}-\frac{1}{M}\right),\varphi_{m,r}=\left(\frac{1}{mq+2p}-\frac{1}{M}\right)$$

Taking expectation on both sides of (15) and applying the results of (16) we obtain the bias of τ^* as:

$$B\left(\hat{\tau}_1\right)_I=\frac{r}{m}\left[\left(\vartheta_X^2+\vartheta_X+1\right)\left(\bar{Y}_{..}-b^*_{(e)}\bar{X}_{1..}\right)\zeta_{3(e)}-\left(\vartheta_X+1\right)\left(\bar{Y}_{..}-b^*_{(e)}\bar{X}_{1..}\right)\left(\zeta_{6(e)}-\zeta_{7(e)}\right)\right] \tag{18}$$

The mean square error (MSE) of τ^* is obtained by taking expectation and square on both sides of (15) and applying the results of (16)

$$\begin{aligned}MSE\left(\tau^*\right)_I=&\bar{Y}_{..}^2\left[\zeta_{0(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{6(e)}\right]-\\&2b^*_{(e)}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}+\zeta_{4(e)}-\zeta_{5(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}-\zeta_{8(e)}\right)\right]+\\&b^{*2}_{(e)}\bar{X}_{1..}^2\left[\zeta_{1(e)}-\zeta_{2(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{7(e)}-\zeta_{8(e)}\right)\right]\end{aligned} \tag{19}$$

To obtain expression for that minimize $MSE\left(\tau^*\right)_I$, differentiate (19) partially with respect to $b^*_{(e)}$ and equate the result to zero.

$$b^*_{(e)opt}=\frac{\bar{Y}_{..}\left(\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}+\zeta_{4(e)}-\zeta_{5(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}-\zeta_{8(e)}\right)\right)}{\bar{X}_{1..}\left(\zeta_{1(e)}-\zeta_{2(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{7(e)}-\zeta_{8(e)}\right)\right)} \tag{20}$$

Substituting the value of $b^*_{opt(e)}$ in (19), gives the minimum value of $MSE\left(\tau^*\right)_I$ as:

$$\begin{aligned}MSE_{\min}\left(\tau^*\right)_I=&\bar{Y}_{..}^2\left[\zeta_{0(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{6(e)}\right]-\\&2b^*_{opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}+\zeta_{4(e)}-\zeta_{5(e)}-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}-\zeta_{8(e)}\right)\right]+\\&b^{*2}_{opt(e)}\bar{X}_{1..}^2\left[\zeta_{1(e)}-\zeta_{2(e)}+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)}-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{7(e)}-\zeta_{8(e)}\right)\right]\end{aligned} \tag{21}$$

The mean square error without measurement error is given by:

$$\begin{aligned}MSE_{\min}\left(\tau\right)_I=&\bar{Y}_{..}^2\left[\zeta_0+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3-2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_6\right]-\\&2b^*_{opt}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3+\zeta_4-\zeta_5-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_6+\zeta_7-\zeta_8\right)\right]+\\&b^{*2}_{opt}\bar{X}_{1..}^2\left[\zeta_1-\zeta_2+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_7-\zeta_8\right)\right]\end{aligned} \tag{22}$$

where,

$$b^*_{opt}=\frac{\bar{Y}_{..}\left(\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3+\zeta_4-\zeta_5-\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_6+\zeta_7-\zeta_8\right)\right)}{\bar{X}_{1..}\left(\zeta_1-\zeta_2+\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3-2\frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_7-\zeta_8\right)\right)} \tag{23}$$

Case B: In this case, S_1 is drawn independently of $S_{n'}$.

$E\left(\Delta_{0(e)}\Delta_{2(e)}\right)=E\left(\Delta_{1(e)}\Delta_{2(e)}\right)=E\left(\Delta_{2(e)}\Delta_{3(e)}\right)=0$, and other expectation are the same as stated in Case A.

Following the procedure used in Case A, we have obtained the minimum mean square error of τ^* as:

$$MSE_{\min}\left(\tau^*\right)_{II} = \bar{Y}_{..}^2\left[\zeta_{0(e)} + \frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)} - 2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{6(e)}\right] - 2b^*_{opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)} + \zeta_{4(e)} - \frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}\right)\right] + b^{*2}_{opt(e)}\bar{X}_{1..}^2\left[\zeta_{1(e)}+\zeta_{2(e)} + \frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)} - 2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{7(e)}\right] \tag{24}$$

Where,

$$b^*_{opt(e)} = \frac{\bar{Y}_{..}\left(\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)} + \zeta_{4(e)} - \frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_{6(e)}+\zeta_{7(e)}\right)\right)}{\bar{X}_{1..}\left(\zeta_{1(e)}+\zeta_{2(e)} + \frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_{3(e)} - 2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_{7(e)}\right)} \tag{25}$$

The mean square error without measurement error is given by:

$$MSE_{\min}\left(\tau\right)_{II} = \bar{Y}_{..}^2\left[\zeta_0 + \frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3 - 2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_6\right] - 2b^*_{opt}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3 + \zeta_4 - \frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_6+\zeta_7\right)\right] + b^{*2}_{opt}\bar{X}_{1..}^2\left[\zeta_1+\zeta_2 + \frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3 - 2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_7\right] \tag{26}$$

Where,

$$b^*_{opt} = \frac{\bar{Y}_{..}\left(\frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3 + \zeta_4 - \frac{r}{m}\left(\vartheta_X+1\right)\left(\zeta_6+\zeta_7\right)\right)}{\bar{X}_{1..}\left(\zeta_1+\zeta_2 + \frac{r^2}{m^2}\left(\vartheta_X+1\right)^2\zeta_3 - 2\frac{r}{m}\left(\vartheta_X+1\right)\zeta_7\right)} \tag{27}$$

2.1.1.4 Efficiency comparison

To evaluate the efficiency of the proposed estimators, we compare them with the usual mean per unit estimator without supplementary information and with the [29] estimators of the population mean in a two-stage cluster sampling scheme, using the strategy discussed in Section 2.1.1.
The mean per unit estimator τ_0^* and its variance in the presence of measurement error are given by:

$$\tau_0^* = \bar{y}_{nm(e)} \tag{28}$$

$$V\left(\tau_0^*\right) = \varphi_{n,N}\left(S_b^2 + S_{b(U)}^2\right) + \frac{1}{n}\varphi_{m,M}\left(S_w^2 + S_{w(U)}^2\right) \tag{29}$$

The mean per unit estimator τ_0 and its variance in the absence of measurement error are given by:

$$\tau_0 = \bar{y}_{nm} \tag{30}$$

$$V\left(\tau_0\right) = \varphi_{n,N}S_b^2 + \frac{1}{n}\varphi_{m,M}S_w^2 \tag{31}$$

The following estimators of population and their mean square error under case A and case B in the absence of measurement error were proposed by [29].

$$\tau_{MSB} = \bar{y}^*_{n(m-r)(e)} + B\left(\bar{x}_{1n'M(e)} - \bar{x}^*_{1n(m-r)(e)}\right) \tag{32}$$

The $MSE_{\min}\left(\tau_{MSB}\right)$, for both Case A and Case B are given by:

Case A

$$MSE_{\min}\left(\tau_{MSB}\right)_I = \bar{Y}_{..}^2\left(\zeta_0+\frac{1}{4}\zeta_3-\zeta_6\right)-2B_{opt}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6+\frac{1}{2}\zeta_8-\frac{1}{2}\zeta_7-\zeta_5+\zeta_4\right)+ B_{opt}^2\bar{X}_{1..}^2\left(\frac{1}{4}\zeta_3+\zeta_1-\zeta_2+\zeta_8-\zeta_7\right) \tag{33}$$

Where,

$$B_{opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6+\frac{1}{2}\zeta_8-\frac{1}{2}\zeta_7-\zeta_5+\zeta_4\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_3+\zeta_1-\zeta_2+\zeta_8-\zeta_7\right)} \tag{34}$$

Case B

$$MSE_{\min}\left(\tau_{MSB}\right)_{II} = \bar{Y}_{..}^2\left(\zeta_0+\frac{1}{4}\zeta_3-\zeta_6\right)-2B_{1opt}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6-\frac{1}{2}\zeta_7+\zeta_4\right)+ B_{1opt}^2\bar{X}_{1}^2\left(\frac{1}{4}\zeta_3+\zeta_1+\zeta_2-\zeta_7\right) \tag{35}$$

Where,

$$B_{opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_3-\frac{1}{2}\zeta_6-\frac{1}{2}\zeta_7+\zeta_4\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_3+\zeta_1+\zeta_2-\zeta_7\right)} \tag{36}$$

To compare it with our proposed estimator in the presence of measurement, we include the contribution of measurement error parameters in the [29] estimators.

The mean square error, accounting for measurement error, for both Case A and Case B, is as follows:

Case A

$$MSE_{\min}\left(\tau_{MSB}^*\right)_I = \bar{Y}_{..}^2\left(\zeta_{0(e)}+\frac{1}{4}\zeta_{3(e)}-\zeta_{6(e)}\right)-2B_{1opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}+\frac{1}{2}\zeta_{8(e)}-\frac{1}{2}\zeta_{7(e)}-\zeta_{5(e)}+\zeta_{4(e)}\right)+ B_{opt(e)}^2\bar{X}_{1..}^2\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}+\zeta_{8(e)}-\zeta_{7(e)}\right) \tag{37}$$

where,

$$B_{(e)opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}+\frac{1}{2}\zeta_{8(e)}-\frac{1}{2}\zeta_{7(e)}-\zeta_{5(e)}+\zeta_{4(e)}\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}+\zeta_{8(e)}-\zeta_{7(e)}\right)} \tag{38}$$

Case B

$$MSE_{\min}\left(\tau_{MSB}^*\right)_{II} = \bar{Y}_{..}^2\left(\zeta_{0(e)}+\frac{1}{4}\zeta_{3(e)}-\zeta_{6(e)}\right)-2B_{opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}-\frac{1}{2}\zeta_{7(e)}+\zeta_{4(e)}\right)+ B_{opt(e)}^2\bar{X}_{1}^2\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}-\zeta_{7(e)}\right) \tag{39}$$

Where,

$$B_{(e)opt}=\frac{\bar{Y}_{..}\left(\frac{1}{4}\zeta_{3(e)}-\frac{1}{2}\zeta_{6(e)}-\frac{1}{2}\zeta_{7(e)}+\zeta_{4(e)}\right)}{\bar{X}_{1..}\left(\frac{1}{4}\zeta_{3(e)}+\zeta_{1(e)}-\zeta_{2(e)}-\zeta_{7(e)}\right)} \tag{40}$$

To demonstrate the performance of our suggested estimators, we compared their percentage relative efficiency (PRE) to the traditional mean per unit estimator, which is based on the normal two-stage design technique without supplementary information, as well as [29] estimators. The empirical study was carried out employing simulated population data sets.

The PRE of an estimator τ^* relative to the natural mean per unit estimator τ_0^* is defined as:

$$PRE = \frac{\left(V\left(\tau_0^*\right)\right)}{MSE_{\min}\left(\tau^*\right)} \times 100 \qquad (41)$$

III. Results

3.1. 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 [6], [30], and [29], who used artificial population generation techniques.

3.1.1. Simulation Results

This simulation exercise consists of the following steps:

1. Six independent variables (normally distributed) are simulated (a total of N times M)
2. The simulated data is then split into M distinct clusters each of size N . The variables Y , X_1 and X_2 are constructed following the relationship defined in the work of [29] only that here, the error component is added.
3. A random sample of m (or m' then m) clusters is selected out of the M total clusters. This is called the first sample units (fsu).
4. A random sample of n (or n' then n) units are sampled from each of m selected clusters. This is called the second sample units (ssu).
5. All the different estimators of the Mean Square Error are calculated based on the observed data and compared.
6. Steps 3 to 5 are repeated a hundred times for each specific case and the estimates of Mean Square Error are all saved in arrays after which the means are calculated and compared.

3.2. Numerical Illustration using Artificial Population

Population 1

$Y = \mu_y + \sigma_y \left(\rho_{x_1 y} \times pop\left[,2\right] + \sqrt{1-\rho_{x_1 y}^2} \times pop\left[,1\right]\right) + U$, $X_1 = \mu_{x_1} + \sigma_{x_1} \times pop\left[,2\right] + V$

$X_2 = \mu_{x_2} + \sigma_{x_2} \left(\rho_{x_1 x_2} \times pop\left[,2\right] + \sqrt{1-\rho_{x_1 x_2}^2}\, pop\left[,1\right]\right) + W$,

$U \sim N\left(0,3\right), V \sim N\left(0,8\right)$, $W \sim N\left(0,12\right)$, $\rho_{x_1 y} = 0.7$, $\rho_{x_1 x_2} = 0.6$, $\sigma_y^2 = 5$, $\sigma_{x_1}^2 = 12$, $\sigma_{x_2}^2 = 9$, $\mu_y = 20$, $\mu_{x_1} = 50$, $\mu_{x_2} = 40$, $N = 10$, $M = 10$, $n' = 9$, $n = 5$, $m = 7$.

Table 1: *Percentage Relative Efficiency (PRE) of Estimators in the Presence of Measurement Error, under Case A.*

Estimators	Auxiliary parameter	*P*=0.05 *q*=0.95	*P*=0.1 *q*=0.9	*P*=0.15 *q*=0.85	*P*=0.2 *q*=0.8
τ_0^*	Not applicable	100.00	100.00	100.00	100.00
τ_{MSB}^*	Not applicable	199.88	192.13	184.52	177.02
τ_1^*	$A_{x_2}=1, B_{x_2}=0$	201.78	194.51	186.67	178.33
τ_2^*	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	201.78	194.51	186.66	178.32
τ_3^*	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	201.77	194.5	186.67	178.34
τ_4^*	$A_{x_2}=1, B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33
τ_5^*	$A_{x_2}=1, B_{x_2}=S_{X_2}$	201.76	194.5	186.69	178.41
τ_6^*	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	201.79	194.52	186.68	178.33
τ_7^*	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33
τ_8^*	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	201.68	194.09	186.05	177.58
τ_9^*	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	201.78	194.51	186.66	178.32
τ_{10}^*	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33
τ_{11}^*	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	201.77	194.51	186.69	178.38
τ_{12}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	201.77	194.48	186.62	178.24
τ_{13}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	201.70	194.44	186.72	178.59
τ_{14}^*	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	201.60	194.40	186.88	179.10
τ_{15}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	201.78	194.51	186.67	178.33
τ_{16}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	201.78	194.51	186.67	178.33
τ_{17}^*	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	201.78	194.51	186.67	178.33

Table 2: *Percentage Relative Efficiency (PRE) of Estimators in the Presence of Measurement Error, under Case B.*

Estimators	Auxiliary parameter	P=0.05 q=0.95	P=0.1 q=0.9	P=0.15 q=0.85	P=0.2 q=0.8
τ_0^*	Not applicable	100.00	100.00	100.00	100.00
τ_{MSB}^*	Not applicable	214.13	206.07	198.13	190.28
τ_1^*	$A_{x_2}=1, B_{x_2}=0$	217.60	209.49	200.87	191.83
τ_2^*	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	217.60	209.49	200.87	191.83
τ_3^*	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	217.60	209.49	200.88	191.85
τ_4^*	$A_{x_2}=1, B_{x_2}=C_{X_2}$	217.60	209.49	200.87	191.83
τ_5^*	$A_{x_2}=1, B_{x_2}=S_{X_2}$	217.60	209.50	200.92	191.94
τ_6^*	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	217.63	209.53	200.91	191.85
τ_7^*	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	217.60	209.50	200.87	191.83
τ_8^*	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	217.64	209.42	200.49	191.14
τ_9^*	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	217.60	209.49	200.87	191.83
τ_{10}^*	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	217.60	209.49	200.87	191.83
τ_{11}^*	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	217.60	209.50	200.90	191.89
τ_{12}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	217.61	209.50	200.85	191.78
τ_{13}^*	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	217.56	209.50	201.03	192.21
τ_{14}^*	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	217.51	209.56	201.33	192.87
τ_{15}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	217.60	209.49	200.87	191.83
τ_{16}^*	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	217.60	209.49	200.87	191.83
τ_{17}^*	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	217.60	209.49	200.87	191.83

Table 3: *Percentage Relative Efficiency (PRE) of Estimators in the Absence of Measurement Error, under Case A.*

Estimators	Auxiliary parameter	*P*=0.05 *q*=0.95	*P*=0.1 *q=0.9*	*P*=0.15 *q*=0.85	*P*=0.2 *q*=0.8
τ_0	Not applicable	100.00	100.00	100	100
τ_{MSB}	Not applicable	226.48	217.5	208.73	200.14
τ_1	$A_{x_2}=1, B_{x_2}=0$	229.00	220.21	211.02	201.47
τ_2	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	229.00	220.21	211.02	201.47
τ_3	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	229.00	220.21	211.03	201.49
τ_4	$A_{x_2}=1, B_{x_2}=C_{X_2}$	229.00	220.21	211.02	201.47
τ_5	$A_{x_2}=1, B_{x_2}=S_{X_2}$	228.99	220.22	211.06	201.56
τ_6	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	229.01	220.23	211.04	201.48
τ_7	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	229.00	220.22	211.02	201.47
τ_8	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	228.90	219.73	210.29	200.6
τ_9	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	229.00	220.21	211.02	201.47
τ_{10}	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	229.00	220.21	211.02	201.47
τ_{11}	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	229.00	220.22	211.05	201.53
τ_{12}	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	228.98	220.18	210.96	201.39
τ_{13}	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	228.96	220.20	211.13	201.76
τ_{14}	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	228.93	220.27	211.41	202.38
τ_{15}	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	229.00	220.21	211.02	201.47
τ_{16}	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	229.00	220.21	211.02	201.47
τ_{17}	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	229.00	220.21	211.02	201.47

Table 4: *Percentage Relative Efficiency (PRE) of Estimators in the Absence of Measurement Error, under Case B.*

Estimators	Auxiliary parameter	*P*=0.05 *q*=0.95	*P*=0.1 *q*=0.9	*P*=0.15 *q*=0.85	*P*=0.2 *q*=0.8
τ_0	Not applicable	100.00	100.00	100.00	100.00
τ_{MSB}	Not applicable	242.19	232.93	223.84	214.91
τ_1	$A_{x_2}=1, B_{x_2}=0$	246.09	236.61	226.72	216.51
τ_2	$A_{x_2}=1, B_{x_2}=B_1(X_1)$	246.09	236.61	226.72	216.51
τ_3	$A_{x_2}=1, B_{x_2}=B_2(X_2)$	246.08	236.62	226.74	216.54
τ_4	$A_{x_2}=1, B_{x_2}=C_{X_2}$	246.09	236.61	226.72	216.51
τ_5	$A_{x_2}=1, B_{x_2}=S_{X_2}$	246.08	236.63	226.78	216.63
τ_6	$A_{x_2}=B_1(X_2), B_{x_2}=B_2(X_2)$	246.11	236.65	226.77	216.54
τ_7	$A_{x_2}=B_1(X_2), B_{x_2}=C_{X_2}$	246.09	236.61	226.73	216.51
τ_8	$A_{x_2}=B_1(X_2), B_{x_2}=S_{X_2}$	246.77	237.29	227.05	216.41
τ_9	$A_{x_2}=B_2(X_2), B_{x_2}=B_1(X_2)$	246.09	236.61	226.72	216.51
τ_{10}	$A_{x_2}=B_2(X_2), B_{x_2}=C_{X_2}$	246.09	236.61	226.72	216.51
τ_{11}	$A_{x_2}=B_2(X_2), B_{x_2}=S_{X_2}$	246.09	236.63	226.76	216.57
τ_{12}	$A_{x_2}=C_{X_2}, B_{x_2}=B_1(X_2)$	246.10	236.62	226.71	216.47
τ_{13}	$A_{x_2}=C_{X_2}, B_{x_2}=B_2(X_2)$	246.07	236.66	226.93	216.94
τ_{14}	$A_{x_2}=C_{X_2}, B_{x_2}=S_{X_2}$	246.07	236.80	227.32	217.68
τ_{15}	$A_{x_2}=S_{X_2}, B_{x_2}=B_1(X_2)$	246.09	236.61	226.72	216.51
τ_{16}	$A_{x_2}=S_{X_2}, B_{x_2}=B_2(X_2)$	246.09	236.61	226.73	216.51
τ_{17}	$A_{x_2}=S_{X_2}, B_{x_2}=C_{X_2}$	246.09	236.61	226.72	216.51

IV. Discussion

From table 1, table 2, table 3 and table 4, it can be observed that our proposed estimator, which utilizes the second supplementary variable parameter, is more efficient, with higher percentage relative efficiencies (PREs) than the usual mean per unit estimator without supplementary information and the [29] estimator in both cases scenario for all choices of probabilities. Therefore, it can be concluded that our proposed methodology can be practically applied, utilizing the actual responses of the respondents and including the measurement error parameters in estimating the finite population mean.

References

[1] Sandarl, C. E., Swensson, B., and Wretman, J. (2003). Model assisted survey sampling. Springer Science & Business Media.

[2] Hansen, M. H. And Hurwitz, W. N. (1946).The problem of non-response in sample surveys.Journal of the American Statistical Association, 41(236):517-529.

[3] Rubin, D. B. (1976). Inference and missing data. Biometrika, 63 (3):581–92.

[4] Heitjan, D. F., and Basu, S. (1996). Distinguishing 'missing at random' and 'missing completely at random. American Statistician, 50 (3):207–13.

[5] Singh, S. and Horn, S. (2000). Compromised imputation in survey sampling. Metrika 51(3):267-276.

[6] Singh, S and Deo, B. (2003).Imputation by power transformation. Statistical Papers, 44(4):555-579.

[7] Kadilar, C. and Cingi, H. (2008). Estimators for the population mean in the case of missing data, Communications in Statistics- Theory and Methods, 37, 2226-2236.

[8] Singh, H. P and Karpe, N. (2009). On the estimation of ratio and product of two population means using supplementary information in presence. Statistica, 69(1): 27-47.

[9] Singh, A. K., Singh, P. and Singh, V. K. (2014a). Exponential-Type Compromised Imputation in Survey Sampling, Journal of Statistics & probability, 3(2):211-217.

[10] Gira, A. A. (2015). Estimation of population mean with a New Imputation Methods. Applied Mathematical Sciences, 9(34), 1663-1672.

[11] Singh, G. N., Maurya, S. Khetan, M. and Kadilar, C. (2016). Some imputation methods for missing data in sample surveys.Hacettepe Journal of Mathematics and Statistics, 45(6):1865-1880.

[12] Bhushan, S. and Pandey, A. P. (2016). Optimal imputation of missing data for estimation of population mean. Journal of Statistics and Management Systems, 19 (6), 755-769.

[13] Pandey, A. K., Singh, G. N., Bhattacharyya, D., Ali, A. Q., Al-Thubaiti, S., and Yakout, H. (2021). Some classes of logarithmic-type imputation techniques for handling missing data. Computational Intelligence and Neuroscience, pages11.

[14] Singh, G., Jaiswal, A. K., Singh, C., and Usman, M. (2022). An improved alternative method of imputation for missing data is survey sampling. Hacettepe Journal of Statistics Applications and Probability, 11(2):535-543.

[15] Cochran, W. G. (1968). Errors of Measurement in Statistics. Technometrics, 10(4):637-666.

[16] Fuller, W. A. (1995). Estimation in the presence of measurement error. International Statistical Review/Revue Internationale de Statistique, 121-141.

[17] Shalabh, S.(1997). Ratio method of estimation in the presence of measurement errors.Journal of the Indian Society of Agricultural Statistics, 52:150-155.

[18] Manisha and Singh, K. R. (2001). An estimation of population mean in the presence of measurement errors. Journal of the Indian Society of Agricultural Statistics, 54(1):13-18.

[19] Allen, J., Singh, H. P., and Smarandache, F. (2003). A family of estimators of population mean using multiauxiliary information in presence of measurement errors. International Journal of Social Economics, 30(7):837-848.

[20] Singh, S. (2009). A new method of imputation in survey sampling. Statistics, 43(5):499-511.

[21] Salas, C. and Gregoire, T. G. (2010).Statistical analysis of ratio estimators and their estimators of variances when the auxiliary variate is measured with error. European Journal of Forest Research, 129:847-861.

[22] Kumar, M., Rajesh, S., Singh, A. K., and Smarandache, F. (2011).Some ratio type estimators under measurement errors.WorldApplied Sciences Journal, 14(2):272.

[23] Shukla, D., Pathak, S., and Thakur, N. S. (2012).An estimator for mean estimation in presence of measurement error.A Journal of Statistics, 1(1):1-8.

[24] Singh, R., Malik, S., and Kohshnivesan, M. (2014b). An alternative estimator for estimating the finite population mean in presence of measurement errors with the view to financial modelling. Science Journal of Applied Mathematics and Statistics, 2:107-111.

[25] Singh, S., and A. H. Joarder. (1998). Estimation of finite population variance using random non-response in survey sampling. Metrika, 47 (1):241–9.

[26] Singh, S., A. H. Joarder, and D. S. Tracy. (2000). Regression type estimators of random non-response in survey sampling. Statistica, 60(1):39-44.

[27] Singh, H. P., R. Tailor, J. M. Kim, and S. Singh. (2012). Families of estimators of finite population variance using a random non-response in survey sampling. The Korean Journal of Applied Statistics, 25 (4):681–95.

[28] Audu, A., Ishaq, O. O., Abubakar, A., Akintola, K. A., Isah, U., Rashida, A., and Muhammed, S. (2021). Regression-type Imputation Class of Estimators using Auxiliary Attributes. Asian Research Journal of Mathematics, 17(5), 1-13.

[29] Maji, R., Singh, G. N. and Bandyopadhyay, A. (2018). Estimation of population mean in presence of random non response in two-stage cluster sampling, Communications in Statistics - Theory and Methods, 48(14):3586-3608.

[30] Singh, S., A. H. Joarder, and D. S. Tracy. 2001. Median estimation using double sampling. Australian & New Zealand Journal of Statistics 43 (1):33–46.