Kwaghe International
Journal of Sciences and Technology

ISSN : 0000-0000
Index: Harvard, Boston, Sydney University, Dimensions, Lens, ResearchGet Scilit, Semantic, Google Scholar, Base etc

https://doi.org/10.58578/KIJST.v1i2.4146

Regression-Type Imputation Scheme under Two-Stage with Unequal Chance of Random Non-Response at First Stage

I. Abubakar[1], A. Yahaya[2], J. Garba[3], Y. Aliyu[4]
[1]Kaduna State University Kaduna, Nigeria
[2,3,4]Ahmadu Bello University, Zaria, Kaduna, Nigeria
ibrahim.abubakar@kasu.edu.ng

Article Info:

Submitted:	Revised:	Accepted:	Published:
Oct 18, 2024	Nov 1, 2024	Nov 14, 2024	Nov 19, 2024

Abstract

This research focuses on the estimation of population mean in two-stage cluster sampling, where the firs-stage cluster units face unequal probabilities of random non-response. To address this, regression-type imputation schemes and estimators are developed, incorporating measurement error parameters for both the study and auxiliary variables. Analytical derivations and simulations demonstrate the efficiency of the proposed estimators. As shown in table 1, the proposed estimators which utilized the second auxiliary variable parameter outperform the usual mean per unit estimator and the Maji *et al.* (2018) estimator in Case A, both with and without measurement error. Similarly, in table 2, it can be observed that the suggested estimator (t_{14}^{*}), is more efficient, that the usual mean per unit estimator without auxiliary information and the Maji *et al.* (2018) estimator, while Maji *et al.* (2018) estimator performed better than other estimators in the same scenario when measurement error is absence for all non-response probability selections. In another scenario, when measurement is presence, the proposed estimators are more efficient for all non-response probabilities. These results confirm the practicality and robustness of the proposed methods for estimating finite population means in the presence of non-response and measurement error.

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

Introduction

In field surveys, obtaining a comprehensive list of every member of the research population is frequently impossible, making the practice of selecting a simple random sample more challenging. Cluster sampling is a viable alternative in these situations because it is often less expensive and does not necessitate an exhaustive population list [1].

Consider a study effort that tries to analyze patient health outcomes in a large city. In the lack of a comprehensive registry, it is difficult to sample a single patient. However, you may typically obtain a list of hospitals in the vicinity. In this situation, a random sample of hospitals (clusters) may be picked, and a survey may be undertaken with a sample of patients from each selected hospital. This method, known as two-stage cluster sampling, begins with the selection of clusters, also known as first-stage units (FSUs), and then selects a subset of elements, also known as second-stage units (SSUs), inside those clusters.

When compared to simple random sample methods, two-stage cluster sampling can improve precision. This method samples components within the clusters once they have been chosen. SSUs are individual components of FSUs, such as patients, households, or students, whereas FSUs are typically larger institutions, such as schools, hospitals, or geographic areas.

Survey sampling often presupposes that the observed variables are error-free and that all relevant data from the research population is collected consistently. However, measurement error and non-response are two difficulties that frequently cause this assumption to fail.

Non-response happens when some study participants do not supply the necessary information due to factors such as rejection, absence, or lack of interest. This difficulty hampers the data collection, calculation, and estimating operations. [2] were pioneers in handling non-response in the estimation of finite population means. [3] proposed notions like Missing at Random (MAR) and Observed at Random (OAR), stating that data are

MAR when the probability of missing data is independent of the unobserved data values. [4] distinguished between fully missing at random (MCAR) and MAR.

To deal with missing data, different imputation schemes have been proposed, which include replacing missing values with specified estimates to allow standard data analysis [5-14]. For example, regression imputation replaces missing values with a linear function derived from other variables. However, without a comprehensive list of population units, simple random sampling is impossible, rendering typical imputation approaches and estimators ineffective. Thus, this investigation employs two-stage cluster sampling.

Another important issue in survey research is measurement error, which is the difference between observed and true underlying values. For example, in a household income survey, respondents may record their incomes incorrectly due to recollection bias or intentional misreporting, resulting in erroneous results. If not appropriately addressed, these measurement mistakes can have a major impact on the validity of survey results. Many researchers have investigated measurement error, usually treating it distinct from non-response [15-24]. However, in practical survey circumstances, both non-response and measurement mistakes occur at the same time. The purpose of this work is to overcome these concerns through the use of two-stage cluster sampling.

Construction of sample structure

Assume Ω is a finite population that is divided into N FSU represented by $(v_1, v_2, \dots, v_N)$ such that the number of SSU in each first stage unit is M. Assume that y_{ij}, x_{1ij} and x_{2ij} are the actual values for the character Y, first supplemental variable X_1, and second supplementary variable X_2, respectively. Assume $y_{ij(e)}$, $x_{1ij(e)}$ and $x_{2ij(e)}$ are the observed values for Y, X_1, X_2 on the j^{th} second stage units $(j = 1,\ 2, \dots, M)$ in the i^{th} first stage units $(i = 1,\ 2, \dots, N)$. Let U_{ij}, V_{ij} and W_{ij} denote the measurement error (ME) parameters associated with the study variable, first supplemental variable, and second supplementary variable, respectively. Measurement errors related with these variables are thus defined as:

The ME related to the character under study be

$$U_{ij} = y_{ij} - y_{ij(e)}, \qquad U_{ij} \sim N(0,\ \sigma_U^2) \quad (2.1)$$

The ME related to the first supplementary variable be

$$V_{ij} = x_{1ij} - x_{1ij(e)}, \qquad V_{ij} \sim N(0,\ \sigma_V^2) \qquad (2.2)$$

The ME related to the second supplementary variable be

$$W_{ij} = x_{2ij} - x_{2ij(e)}, \qquad W_{ij} \sim N(0,\ \sigma_W^2) \qquad (2.3)$$

However, in this study, we consider a scenario in which the first supplementary variable $X_1$is unknown at the first stage unit level. As a result, information on the first supplementary variable x_1 can be obtained using the following strategy:

Strategy:

At the first stage unit level, information on the first supplementary variable x_1 is acquired, and a first stage unit sample is selected using the SRSWOR technique. Furthermore, the above-mentioned technique will be addressed in clusters with an equal possibility of random non-response, as shown below.

Clusters with unequal chance of random non-response

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

We consider a scenario in which the population mean $\bar{X}_{1..}$ of the first supplemental variable x_1 is unknown at the first stage unit level, so we employed a two-phase or double sampling technique to get the estimate. The second supplemental variable,x_2, is, nevertheless, known for each unit of the population. To estimate the population mean of Y, a first phase sample $S_{n'}(S_{n'} \subset \Omega)$ of size n' first stage unit is taken from the population Ω using the SRSWOR method, followed by a second phase sample S_1 of size n FSU $(n \leq n')$ taken based on the following two cases using the SRSWOR technique to observe the character under study.

Case A, a subsample of $S_{n'}$ $(S_1 \subset S_{n'})$ is used to create S_1.

Case B, S_1 is drawn independently of $S_{n'}$.

To estimate the population mean of Y , a second stage sample S_2 is obtained by selecting a portion of m second stage units from M second stage units for each of the n chosen first stage units in S_1 using the SRSWOR method.

In the second stage, it is assumed that the study variable y and the first supplemental variable x_1 have random non-response, while the sampled unit responds fully to the second supplementary variablex_2. For such random non-response conditions, we consider the probability model described in section (2.1.1.1).

Probability of Non-Response Model

We assume that random non-response conditions on the study variable y and the first supplementary variable x_1 occur in the second stage sample S_2 of n first stage units, each with m second stage units, and that the random non-response varies among the various first stage units chosen in the second stage sampleS_2. Let $r_i\{r_i = 0, 1, \dots, (m-2); i = 1, 2, \dots, n\}$ be the number of non-responding second stage units in each first stage unit of the second stage sample. As a result, we write A and A^c to denote the collection of respondent units and non-respondent units, respectively. The observations of the associated variables in which random non-response occurs could be collected from the remainder of the $(m - r_i)$ unit of each of the n first stage units of the second stage sample.

We further assume that if p_i represents the chance of random non-response among $(m - 2)$ possible non-response occurrences, and then r_i follows the probability distribution described in equation (2.4).

$$P(r_i) = \frac{m-r_i}{mq_i+2p_i}\binom{m-2}{r_i}p_i^{r_i}q_i^{m-r_i-2}; \text{ for } r_i = 0, 1, \dots, (m-2); i = 1, 2, \dots, n \quad (2.4)$$

For example, consider the work of [25-27], where $q_i = 1 - p_i$ and $\binom{m-2}{r_i}$ denote the total number of ways to provide r_i non-response from $(m - 2)$ total non-responses.

The following notations will now 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 variablex_1.

$\bar{X}_{2..} = \frac{1}{NM}\sum_{i=1}^{N}\sum_{j=1}^{M} X_{2ij}$, Population average of the second supplementary variablex_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 inS_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 inS_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 inS_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 inS_2.

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

Proposed Imputation Schemes and Estimators

We assumed that the second supplementary variable was accessible throughout the population Ω. Inspired by the imputation schemes presented by [28], we suggest the following regression-type imputation strategies based on responding and non-responding units of the second stage sample S_2 to estimate population parameter $\bar{Y}_{..}$ as:

$$y_{ij(e)} = \begin{cases} y_{ij(e)}\,, & if\ j\epsilon A \\ \dfrac{\bar{y}_{i(m-r_i)(e)} + b_{yx_2(e)}\left(\bar{X}_{2..} - \bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)} + B_{x_2}}\left(A_{x_2}\bar{X}_{2..} + B_{x_2}\right), & if\ j\epsilon A^c \end{cases}$$

$$(i = 1, 2, \dots, n) \qquad (2.5)$$

Where $A_{x_2}$and $B_{x_2}$are available functions of supplementary variable like coefficient of skewness, kurtosis, variation, standard deviation, $b_{yx_2(e)} = \sum_{j\epsilon A} y_{ij(e)} / \sum_{j\epsilon A} x_{2ij(e)}$, $b_{x_1x_2(e)} = \sum_{j\epsilon A} x_{1ij(e)} / \sum_{j\epsilon A} x_{2ij(e)}$, for; $i = 1, 2, \dots, n$.

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:

$$\bar{y}^{*}_{i(e)} = \frac{1}{m}\sum_{j=1}^{m} y_{ij(e)} = \frac{1}{m}\left[\sum_{j\in A_{r_i}} y_{ij(e)} + \sum_{j\in A^{c}_{r_i}} y_{ij(e)}\right] \qquad (2.6)$$

$$\bar{y}^*_{i(e)} = \frac{1}{m}\sum_{j=1}^{m} y_{ij(e)} = \left(1-\frac{r_i}{m}\right)\bar{y}_{i(m-r_i)(e)} + \frac{r_i}{m}\left(\frac{\bar{y}_{i(m-r_i)(e)}+b_{yx_2(e)}\left(\bar{X}_{2..}-\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right)\right) \quad (2.7)$$

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

$$\bar{y}^{**}_{n(m-r_i)(e)} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}^*_{i(e)} = \left(1-\frac{r_i}{m}\right)\bar{y}_{n(m-r_i)(e)} + \frac{r_i}{m}\left(\frac{\bar{y}_{n(m-r_i)(e)}+b_{yx_2(e)}\left(\bar{X}_{2..}-\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right)\right) \quad (2.8)$$

Likewise, for each unit in the second stage

$$x_{1ij(e)} = \begin{cases} x_{1ij(e)}, & if\ j\epsilon A \\ \frac{\bar{x}_{1i(m-r_i)(e)}+b_{x_1x_2(e)}\left(\bar{X}_{2..}-\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right), & if\ j\epsilon A^c \end{cases} \quad (i = 1, 2, \dots, n) \quad (2.9)$$

Under this approach, we derived the sample means of x_{1i} on the i^{th} first stage units in S_2denoted by $\bar{x}^*_{1i(e)}$as:

$$\bar{x}^*_{1i(e)} = \frac{1}{m}\sum_{j=1}^{m} x_{1ij(e)} = \frac{1}{m}\left[\sum_{j\epsilon A} x_{1ij(e)} + \sum_{j\epsilon A^c} x_{1ij(e)}\right] \quad (2.10)$$

$$\bar{x}^*_{1i(e)} = \frac{1}{m}\sum_{j=1}^{m} x_{1ij(e)} = \left(1-\frac{r_i}{m}\right)\bar{x}_{1i(m-r_i)(e)} + \frac{r_i}{m}\left(\frac{\bar{x}_{1i(m-r_i)(e)}+b_{x_1x_2(e)}\left(\bar{X}_{2..}-\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right)\right) \quad (2.11)$$

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

$$\bar{x}^{**}_{1n(m-r_i)(e)} = \frac{1}{n}\sum_{i=1}^{n} \bar{x}^*_{1i(e)} = \left(1-\frac{r_i}{m}\right)\bar{x}_{1n(m-r_i)(e)} + \frac{r_i}{m}\left(\frac{\bar{x}_{1n(m-r_i)(e)}+b_{x_1x_2(e)}\left(\bar{X}_{2..}-\bar{x}_{2nm(e)}\right)}{A_{x_2}\bar{x}_{2nm(e)}+B_{x_2}}\left(A_{x_2}\bar{X}_{2..}+B_{x_2}\right)\right) \quad (2.12)$$

Hence the proposed estimator for $\bar{Y}_{..}$ denoted by t^*_p for $p = 1, 2, \dots, 17,$ under the above proposed imputation scheme is obtained as:

$$t^*_p = \bar{y}^{**}_{n(m-r_i)(e)} + \alpha^*_{1(e)}\left(\bar{x}_{1n'M(e)} - \bar{x}^{**}_{1n(m-r_i)(e)}\right) \quad (2.13)$$

where $\alpha^*_{1(e)}$ is a suitable constant chosen so that the proposed estimator t^*_p mean square error is as small as possible.

Properties of the Proposed Estimators t_p^*

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

$\bar{y}_{n(m-r_i)(e)} = \bar{Y}_{..}(1+\Delta^*_{0(e)}),\ \bar{x}_{1n(m-r_i)(e)} = \bar{X}_{1..}(1+\Delta^*_{1(e)}),\ \bar{x}_{1n'M(e)} = \bar{X}_{1..}(1+\Delta^*_{2(e)}),$

$\bar{x}_{2nm(e)} = \bar{X}_{2..}(1+\Delta^*_{3(e)})$

Such that $E(\Delta^*_i) = 0$ and $|\Delta^*_i| < 1$ for all $i = 0, 1, 2, 3.$

Express (t_p^*) in terms of errors$\Delta^*_{0(e)}, \Delta^*_{1(e)}, \Delta^*_{2(e)}, \Delta^*_{3(e)}$.

$$t_p^* = \bar{Y}_{..}\left[1+\Delta^*_{0(e)} - \frac{r_i}{m}(\delta_X+1)\Delta^*_{3(e)} + \frac{r_i}{m}(\delta_X^2+\delta_X+1)\Delta^*_{3(e)} - (\delta_X+1)\Delta^*_{0(e)}\Delta^*_{3(e)}\right] + \left[\alpha^*_{1(e)}(1+\Delta^*_{2(e)})\bar{X}_{1..} - \bar{X}_{1..}\left(1+\Delta^*_{1(e)} - \frac{r_i}{m}(\delta_X+1)\Delta^*_{3(e)} + \frac{r_i}{m}(\delta_X^2+\delta_X+1)\Delta^{*2}_{3(e)} - \frac{r_i}{m}(\delta_X+1)\Delta^*_{1(e)}\Delta^*_{3(e)}\right)\right] \quad (2.14)$$

$$t_p^* - \bar{Y}_{..} = \bar{Y}_{..}\left[\Delta^*_{0(e)} - \frac{r_i}{m}(\delta_X+1)\Delta^*_{3(e)} + \frac{r_i}{m}(\delta_X^2+\delta_X+1)\Delta^*_{3(e)} - (\delta_X+1)\Delta^*_{0(e)}\Delta^*_{3(e)}\right] + \alpha^*_{1(e)}\bar{X}_{1..}\left[\Delta^*_{2(e)} - \Delta^*_{1(e)} + \frac{r_i}{m}(\delta_X+1)\Delta^*_{3(e)} - \frac{r_i}{m}(\delta_X^2+\delta_X+1)\Delta^{*2}_{3(e)} + \frac{r_i}{m}(\delta_X+1)\Delta^*_{1(e)}\Delta^*_{3(e)}\right] \quad (2.15)$$

Where, $\delta_X = \frac{A_{x_2}\bar{X}_{2..}}{A_{x_2}\bar{X}_{2..}+B_{x_2}}$

The bias and mean square error of the estimator t_p^* for Cases A and B of the two-phase sample structure outlined in Section 2 have been calculated individually and are shown below.

Case A: A subsample of $S_{n'}$ $(S_1 \subset S_{n'})$ is used to constructS_1. In this case, we will use the expected values of the sample statistics to calculate the bias and mean square error.

$E(\Delta^{*2}_{0(e)}) = \phi_1 \frac{S_y^{*2}+S_u^{*2}}{\bar{Y}^2_{..}} + \frac{1}{Nn}\sum_{i=1}^N \phi_4 \frac{\bar{S}^2_{y_i}+\bar{S}^2_{u_i}}{\bar{Y}^2_{..}}$, $E(\Delta^{*2}_{1(e)}) = \phi_1 \frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}^2_{1..}} + \frac{1}{Nn}\sum_{i=1}^N \phi_4 \frac{\bar{S}^2_{x_{1i}}+\bar{S}^2_{v_i}}{\bar{X}^2_{1..}}$

$$E\left(\Delta_{2(e)}^{*2}\right) = \phi_2 \frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1..}^2}, E\left(\Delta_{3(e)}^{*2}\right) = \phi_1 \frac{S_{x_2}^{*2}+S_w^{*2}}{\bar{X}_{2..}^2} + \frac{1}{n}\phi_3 \frac{\bar{S}_{x_2}^2+\bar{S}_w^2}{\bar{X}_{2..}^2}$$

$$E\left(\Delta_{0(e)}^*\Delta_{1(e)}^*\right) = \phi_1 \frac{S_{yx_1}^*+S_{uv}^*}{\bar{Y}_{..}\bar{X}_{1..}} + \frac{1}{Nn}\sum_{i=1}^N \phi_4 \frac{\bar{S}_{yx_{1i}}+\bar{S}_{uv_i}}{\bar{Y}_{..}\bar{X}_{1..}}, E\left(\Delta_{0(e)}^*\Delta_{2(e)}^*\right) = \phi_2 \frac{S_{yx_1}^*+S_{uv}^*}{\bar{Y}_{..}\bar{X}_{1..}}$$

$$E\left(\Delta_{0(e)}^*\Delta_{3(e)}^*\right) = \phi_1 \frac{S_{yx_2}^*+S_{uw}^*}{\bar{Y}_{..}\bar{X}_{2..}} + \frac{1}{n}\phi_3 \frac{\bar{S}_{yx_2}+\bar{S}_{uw}}{\bar{Y}_{..}\bar{X}_{2..}}, E\left(\Delta_{1(e)}^*\Delta_{2(e)}^*\right) = \phi_2 \frac{S_{x_1}^{*2}+S_v^{*2}}{\bar{X}_{1..}^2} = E\left(\Delta_{2(e)}^{*2}\right),$$

$$E\left(\Delta_{1(e)}^*\Delta_{3(e)}^*\right) = \phi_1 \frac{S_{x_1x_2}^*+S_{vw}^*}{\bar{X}_{1..}\bar{X}_{2..}} + \frac{1}{n}\phi_3 \frac{\bar{S}_{x_1x_2}+\bar{S}_{vw}}{\bar{X}_{1..}\bar{X}_{2..}}, E\left(\Delta_{2(e)}^*\Delta_{3(e)}^*\right) = \phi_2 \frac{S_{x_1x_2}^*+S_{vw}^*}{\bar{X}_{1..}\bar{X}_{2..}}$$

For simplicity, we will let

$$\left.\begin{array}{c}\theta_{0(e)}^* = E\left(\Delta_{0(e)}^{*2}\right), \theta_{1(e)}^* = E\left(\Delta_{1(e)}^{*2}\right), \theta_{2(e)}^* = E\left(\Delta_{2(e)}^{*2}\right) = E\left(\Delta_{1(e)}^*\Delta_{2(e)}^*\right), \\ \theta_{3(e)}^* = E\left(\Delta_{3(e)}^{*2}\right), \theta_{4(e)}^* = E\left(\Delta_{0(e)}^*\Delta_{1(e)}^*\right), \theta_{5(e)}^* = E\left(\Delta_{0(e)}^*\Delta_{2(e)}^*\right), \\ \theta_{6(e)}^* = E\left(\Delta_{0(e)}^*\Delta_{3(e)}^*\right), \theta_{7(e)}^* = E\left(\Delta_{1(e)}^*\Delta_{3(e)}^*\right), \theta_{8(e)}^* = E\left(\Delta_{2(e)}^*\Delta_{3(e)}^*\right)\end{array}\right\} \quad (2.16)$$

When measurement error is not considered, the notations and expectations listed below will be used.

$$E(\Delta_0^{*2}) = \phi_1 \frac{S_y^{*2}}{\bar{Y}_{..}^2} + \frac{1}{Nn}\sum_{i=1}^N \phi_4 \frac{\bar{S}_{y_i}^2}{\bar{Y}_{..}^2}, \quad E(\Delta_1^{*2}) = \phi_1 \frac{S_{x_1}^{*2}}{\bar{X}_{1..}^2} + \frac{1}{Nn}\sum_{i=1}^N \phi_4 \frac{\bar{S}_{x_{1i}}^2}{\bar{X}_{1..}^2}$$

$$E(\Delta_2^{*2}) = \phi_2 \frac{S_{x_1}^{*2}}{\bar{X}_{1..}^2}, E(\Delta_3^{*2}) = \phi_1 \frac{S_{x_2}^{*2}}{\bar{X}_{2..}^2} + \frac{1}{n}\phi_3 \frac{\bar{S}_{x_2}^2}{\bar{X}_{2..}^2}$$

$$E(\Delta_0^*\Delta_1^*) = \phi_1 \frac{S_{yx_1}^*}{\bar{Y}_{..}\bar{X}_{1..}} + \frac{1}{Nn}\sum_{i=1}^N \phi_4 \frac{\bar{S}_{yx_{1i}}}{\bar{Y}_{..}\bar{X}_{1..}}, E(\Delta_0^*\Delta_2^*) = \phi_2 \frac{S_{yx_1}^*}{\bar{Y}_{..}\bar{X}_{1..}}$$

$$E(\Delta_0^*\Delta_3^*) = \phi_1 \frac{S_{yx_2}^*}{\bar{Y}_{..}\bar{X}_{2..}} + \frac{1}{n}\phi_3 \frac{\bar{S}_{yx_2}}{\bar{Y}_{..}\bar{X}_{2..}}, E(\Delta_1^*\Delta_2^*) = \phi_2 \frac{S_{x_1}^{*2}}{\bar{X}_{1..}^2} = E(\Delta_2^{*2}),$$

$$E(\Delta_1^*\Delta_3^*) = \phi_1 \frac{S_{x_1x_2}^*}{\bar{X}_{1..}\bar{X}_{2..}} + \frac{1}{n}\phi_3 \frac{\bar{S}_{x_1x_2}}{\bar{X}_{1..}\bar{X}_{2..}}, E(\Delta_2^*\Delta_3^*) = \phi_2 \frac{S_{x_1x_2}^*}{\bar{X}_{1..}\bar{X}_{2..}}$$

Similarly, for simplicity we let

$$\left.\begin{array}{c}\theta_0^* = E(\Delta_0^{*2}), \theta_1^* = E(\Delta_1^{*2}), \theta_2^* = E(\Delta_2^{*2}) = E(\Delta_1^*\Delta_2^*), \\ \theta_3^* = E(\Delta_3^{*2}), \theta_4^* = E(\Delta_0^*\Delta_1^*), \theta_5^* = E(\Delta_0^*\Delta_2^*), \\ \theta_6^* = E(\Delta_0^*\Delta_3^*), \theta_7^* = E(\Delta_1^*\Delta_3^*), \theta_8^* = E(\Delta_2^*\Delta_3^*)\end{array}\right\} \quad (2.17)$$

Where

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

$\phi_1 = \left(\frac{1}{n} - \frac{1}{N}\right), \phi_2 = \left(\frac{1}{n\prime} - \frac{1}{N}\right), \phi_3 = \left(\frac{1}{m} - \frac{1}{M}\right), \phi_4 = \left(\frac{1}{mq_i + 2p_i} - \frac{1}{M}\right)$

Taking expectation on both sides of (2.15) and applying the results of (2.16) we obtain the bias of t_p^{*}as:

$$B^{*}(t_p^{*})_I = \frac{\bar{r}}{m}(\bar{Y}_{..} - \alpha_{1(e)}^{*}\bar{X}_{1..})\left[(\delta_X^2 + \delta_X + 1)\theta_{3(e)}^{*} - (\delta_X + 1)(\theta_{6(e)}^{*} - \theta_{7(e)}^{*})\right] \quad (2.18)$$

The mean square error (MSE) of t_p^{*} is obtained by taking expectation and square on both sides of (2.15) and applying the results of (2.16)

$$MSE^{*}(t_p^{*})_I = \bar{Y}_{..}^2\theta_{0(e)}^{*} + \alpha_{1(e)}^{*2}\bar{X}_{1..}^2(\theta_{1(e)}^{*} + \theta_{2(e)}^{*}) + \frac{var(r) + \bar{r}}{m^2}(\delta_X + 1)^2(\bar{Y}_{..} - \alpha_{1(e)}^{*}\bar{X}_{1..})^2\theta_{3(e)}^{*} - 2\alpha_{1(e)}^{*}\bar{Y}_{..}\bar{X}_{1..}(\theta_{4(e)}^{*} - \theta_{5(e)}^{*}) - 2\frac{\bar{r}}{m}(\delta_X + 1)\bar{Y}_{..}(\bar{Y}_{..} - \alpha_{1(e)}^{*}\bar{X}_{1..})\theta_{6(e)}^{*} - 2\alpha_{1(e)}^{*2}\bar{X}_{1..}^2\theta_{2(e)}^{*} + 2\frac{\bar{r}}{m}(\delta_X + 1)\alpha_{1(e)}^{*}\bar{X}_{1..}(\bar{Y}_{..} - \alpha_{1(e)}^{*}\bar{X}_{1..})(\theta_{7(e)}^{*} - \theta_{8(e)}^{*}) \quad (2.19)$$

To obtain expression for that minimize $MSE^{*}(t_p^{*})_I$, differentiate (2.19) partially with respect to $\alpha_{1(e)}^{*}$ and equate the result to zero.

$$\alpha_{1opt(e)}^{*} = \frac{\bar{Y}_{..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta_{3(e)}^{*} + \theta_{4(e)}^{*} - \theta_{5(e)}^{*} - \frac{\bar{r}}{m}(\delta_X+1)(\theta_{6(e)}^{*} + \theta_{7(e)}^{*} - \theta_{8(e)}^{*})\right]}{\bar{X}_{1..}\left[\theta_{1(e)}^{*} - \theta_{2(e)}^{*} + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta_{3(e)}^{*} - 2\frac{\bar{r}}{m}(\delta_X+1)(\theta_{7(e)}^{*} - \theta_{8(e)}^{*})\right]} \quad (2.20)$$

Substituting the value of $\alpha_{1opt(e)}^{*}$ in (2.19), gives the minimum value of $MSE^{*}(t_p^{*})_I$as:

$$MSE_{min}^{*}(t_p^{*})_I = \bar{Y}_{..}^2\left[\theta_{0(e)}^{*} + \frac{var(r)+\bar{r}}{m^2}(\delta_X + 1)^2\theta_{3(e)}^{*} - 2\frac{\bar{r}}{m}(\delta_X + 1)\theta_{6(e)}^{*}\right] - 2\alpha_{1opt(e)}^{*}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X + 1)^2\theta_{3(e)}^{*} + \theta_{4(e)}^{*} - \theta_{5(e)}^{*} - \frac{\bar{r}}{m}(\delta_X + 1)(\theta_{6(e)}^{*} + \theta_{7(e)}^{*} -\right.$$

$$\theta^*_{8(e)}\Big)\Big] + \alpha^{*2}_{1(e)}\bar{X}^2_{1..}\left[\theta^*_{1(e)} - \theta^*_{2(e)} + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_{3(e)} - 2\frac{\bar{r}}{m}(\delta_X+1)\Big(\theta^*_{7(e)} - \theta^*_{8(e)}\Big)\right] \quad (2.21)$$

The mean square error without measurement error is given by:

$$MSE_{min}\big(t^*_p\big)_I = \bar{Y}^2_{..}\left[\theta^*_0 + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3 - 2\frac{\bar{r}}{m}(\delta_X+1)\theta^*_6\right] - 2\alpha^*_{1opt}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3 + \theta^*_4 - \theta^*_5 - \frac{\bar{r}}{m}(\delta_X+1)(\theta^*_6+\theta^*_7-\theta^*_8)\right] + \alpha^{*2}_1\bar{X}^2_{1..}\left[\theta^*_1 - \theta^*_2 + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3 - 2\frac{\bar{r}}{m}(\delta_X+1)(\theta^*_7-\theta^*_8)\right] \quad (2.22)$$

where,

$$\alpha^*_{1opt} = \frac{\bar{Y}_{..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3+\theta^*_4-\theta^*_5-\frac{\bar{r}}{m}(\delta_X+1)(\theta^*_6+\theta^*_7-\theta^*_8)\right]}{\bar{X}_{1..}\left[\theta^*_1-\theta^*_2+\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3-2\frac{\bar{r}}{m}(\delta_X+1)(\theta^*_7-\theta^*_8)\right]} \quad (2.23)$$

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

$E\big(\Delta^*_{0(e)}\Delta^*_{2(e)}\big) = E\big(\Delta^*_{1(e)}\Delta^*_{2(e)}\big) = E\big(\Delta^*_{2(e)}\Delta^*_{3(e)}\big) = 0$, 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 t^*_p as:

$$MSE^*_{min}\big(t^*_p\big)_{II} = \bar{Y}^2_{..}\left[\theta^*_{0(e)} + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_{3(e)} - 2\frac{\bar{r}}{m}(\delta_X+1)\theta^*_{6(e)}\right] - 2\alpha^*_{1opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_{3(e)} + \theta^*_{4(e)} - \frac{\bar{r}}{m}(\delta_X+1)\Big(\theta^*_{6(e)}+\theta^*_{7(e)}\Big)\right] + \alpha^{*2}_{1(e)}\bar{X}^2_{1..}\left[\theta^*_{1(e)} + \theta^*_{2(e)} + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_{3(e)} - 2\frac{\bar{r}}{m}(\delta_X+1)\theta^*_{7(e)}\right] \quad (2.24)$$

Where

$$\alpha^*_{1opt(e)} = \frac{\bar{Y}_{..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_{3(e)}+\theta^*_{4(e)}-\frac{\bar{r}}{m}(\delta_X+1)\Big(\theta^*_{6(e)}+\theta^*_{7(e)}\Big)\right]}{\bar{X}_{1..}\left[\theta^*_{1(e)}+\theta^*_{2(e)}+\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_{3(e)}-2\frac{\bar{r}}{m}(\delta_X+1)\theta^*_{7(e)}\right]} \quad (2.25)$$

The mean square error without measurement error is given by:

$$MSE_{min}\big(t^*_p\big)_{II} = \bar{Y}^2_{..}\left[\theta^*_0 + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3 - 2\frac{\bar{r}}{m}(\delta_X+1)\theta^*_6\right] - 2\alpha^*_{1opt}\bar{Y}_{..}\bar{X}_{1..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3 + \theta^*_4 - \frac{\bar{r}}{m}(\delta_X+1)(\theta^*_6+\theta^*_7)\right] + \alpha^{*2}_{1(e)}\bar{X}^2_{1..}\left[\theta^*_1 + \theta^*_2 + \frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta^*_3 - 2\frac{\bar{r}}{m}(\delta_X+1)\theta^*_7\right] \quad (2.26)$$

Where

$$\alpha_{1opt}^{*} = \frac{\bar{Y}_{..}\left[\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta_3^*+\theta_4^*-\frac{\bar{r}}{m}(\delta_X+1)(\theta_6^*+\theta_7^*)\right]}{\bar{X}_{1..}\left[\theta_1^*+\theta_2^*+\frac{var(r)+\bar{r}}{m^2}(\delta_X+1)^2\theta_3^*-2\frac{\bar{r}}{m}(\delta_X+1)\theta_7^*\right]} \quad (2.27)$$

1. Efficiency comparison

To assess the efficiency of the suggested estimators, we compare them to the standard mean per unit estimator without further information, as well as the [29] population mean estimators in a two-stage cluster sampling scheme, employing the technique mentioned in Section 2.1.1.

The mean per unit estimator t_0 and its variation in the presence of measurement error are provided by:

$$t_0 = \bar{y}_{nm(e)} \quad (2.28)$$

$$V^*(t_0) = \phi_1\left(S_b^2 + S_{b(U)}^2\right) + \frac{1}{n}\phi_3\left(S_w^2 + S_{w(U)}^2\right) \quad (2.29)$$

The variance in the absence of measurement error is given by:

$$V(t_0) = \phi_1 S_b^2 + \frac{1}{n}\phi_3 S_w^2 \quad (2.30)$$

[29] provided the following population estimator and its mean square error for cases A and B in the absence of measurement error.

$$t_{MSB} = \bar{y}_{n(m-r_i)}^{*} + B_1\left(\bar{x}_{1n'M} - \bar{x}_{1n(m-r_i)}^{*}\right) \quad (2.31)$$

The $MSE_{min}(t_{MSB})$ for both Case A and Case B are given by:

Case A

$$MSE_{min}(t_{MSB})_I = \bar{Y}_{..}^2\left(\theta_0^* + \frac{1}{4}\theta_3^* - \theta_6^*\right) - 2B_{1opt}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\theta_3^* - \frac{1}{2}\theta_6^* + \frac{1}{2}\theta_8^* - \frac{1}{2}\theta_7^* - \theta_5^* + \theta_4^*\right) + B_{1opt}^2\bar{X}_{1..}^2\left(\frac{1}{4}\theta_3^* + \theta_1^* - \theta_2^* + \theta_8^* - \theta_7^*\right) \quad (2.32)$$

Where,

$$B_{1opt} = \frac{\bar{Y}_{..}\left(\frac{1}{4}\theta_3^* - \frac{1}{2}\theta_6^* + \frac{1}{2}\theta_8^* - \frac{1}{2}\theta_7^* - \theta_5^* + \theta_4^*\right)}{\bar{X}_{1..}\left(\frac{1}{4}\theta_3^* + \theta_1^* - \theta_2^* + \theta_8^* - \theta_7^*\right)} \quad (2.33)$$

Case B

$$MSE_{min}(t_{MSB})_{II} = \bar{Y}_{..}^2\left(\theta_0^* + \frac{1}{4}\theta_3^* - \theta_6^*\right) - 2B_{1opt}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\theta_3^* - \frac{1}{2}\theta_6^* - \frac{1}{2}\theta_7^* + \theta_4^*\right) + B_{1opt}^2\bar{X}_{1..}^2\left(\frac{1}{4}\theta_3^* + \theta_1^* - \theta_2^* - \theta_7^*\right) \quad (2.34)$$

Where,

$$B_{1opt} = \frac{\bar{Y}_{..}\left(\frac{1}{4}\theta_3^* - \frac{1}{2}\theta_6^* - \frac{1}{2}\theta_7^* + \theta_4^*\right)}{\bar{X}_{1..}\left(\frac{1}{4}\theta_3^* + \theta_1^* - \theta_2^* - \theta_7^*\right)} \quad (2.35)$$

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

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

Case A

$$MSE_{min}^*(t_{MSB})_I = \bar{Y}_{..}^2\left(\theta_{0(e)}^* + \frac{1}{4}\theta_{3(e)}^* - \theta_{6(e)}^*\right) - 2B_{1opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\theta_{3(e)}^* - \frac{1}{2}\theta_{6(e)}^* + \frac{1}{2}\theta_{8(e)}^* - \frac{1}{2}\theta_{7(e)}^* - \theta_{5(e)}^* + \theta_{4(e)}^*\right) + B_{1opt(e)}^2\bar{X}_{1..}^2\left(\frac{1}{4}\theta_{3(e)}^* + \theta_{1(e)}^* - \theta_{2(e)}^* + \theta_{8(e)}^* - \theta_{7(e)}^*\right) \quad (2.36)$$

where,

$$B_{1opt(e)} = \frac{\bar{Y}_{..}\left(\frac{1}{4}\theta_{3(e)}^* - \frac{1}{2}\theta_{6(e)}^* + \frac{1}{2}\theta_{8(e)}^* - \frac{1}{2}\theta_{7(e)}^* - \theta_{5(e)}^* + \theta_{4(e)}^*\right)}{\bar{X}_{1..}\left(\frac{1}{4}\theta_{3(e)}^* + \theta_{1(e)}^* - \theta_{2(e)}^* + \theta_{8(e)}^* - \theta_{7(e)}^*\right)} \quad (2.37)$$

Case B

$$MSE^{*}_{min}(t_{MSB})_{II} = \bar{Y}^{2}_{..}\left(\theta^{*}_{0(e)} + \frac{1}{4}\theta^{*}_{3(e)} - \theta^{*}_{6(e)}\right) - 2B_{1opt(e)}\bar{Y}_{..}\bar{X}_{1..}\left(\frac{1}{4}\theta^{*}_{3(e)} - \frac{1}{2}\theta^{*}_{6(e)} - \frac{1}{2}\theta^{*}_{7(e)} + \theta^{*}_{4(e)}\right) + B^{2}_{1opt(e)}\bar{X}^{2}_{1..}\left(\frac{1}{4}\theta^{*}_{3(e)} + \theta^{*}_{1(e)} - \theta^{*}_{2(e)} - \theta^{*}_{7(e)}\right) \tag{2.38}$$

Where,

$$B_{1opt(e)} = \frac{\bar{Y}_{..}\left(\frac{1}{4}\theta^{*}_{3(e)} - \frac{1}{2}\theta^{*}_{6(e)} - \frac{1}{2}\theta^{*}_{7(e)} + \theta^{*}_{4(e)}\right)}{\bar{X}_{1..}\left(\frac{1}{4}\theta^{*}_{3(e)} + \theta^{*}_{1(e)} - \theta^{*}_{2(e)} - \theta^{*}_{7(e)}\right)} \tag{2.39}$$

To demonstrate the performance of our proposed estimators, we compared their percentage relative efficiency (PRE) to the traditional mean per unit estimator, which is based on the standard two-stage design technique without supplemental information, and [29] estimators. The empirical investigation was conducted using simulated population data sets.

The PRE of an estimator t relative to the natural mean per unit estimator t_0 is defined as:

$$PRE = \frac{\left(V(t_0)\right)}{MSE_{min}(t)} \times 100$$

Study Using Artificially Generated Population

A key part of simulation is creating a simulation model that replicates the actual system. Simulation enables the comparison of analytical approaches and helps determine whether a newly discovered technique is superior to existing ones. Motivated by [6] and [29-30], who employed artificial population generation techniques.

$$y = \mu_y + \sigma_y\left(\rho_{x_1y} \times x'_1 + \sqrt{1-\rho^2_{x_1y}} \times y'\right) + U$$

$$x_1 = \mu_{x_1} + \sigma_{x_1} \times x'_1 + V$$

$$x_2 = \mu_{x_2} + \sigma_{x_2}\left(\rho_{x_1x_2} \times x'_1 + \sqrt{1-\rho^2_{x_1x_2}} \times y'\right) + W$$

$y' = pop[,1], \; x'_1 = pop[,2]$

$U \sim N(0,3), V \sim N(0,8), W \sim N(0,10)$

$\rho_{x_1 y} = 0.7, \rho_{x_1 x_2} = 0.6, \sigma_y^2 = 6, \sigma_{x_1}^2 = 12, \sigma_{x_2}^2 = 9, \mu_y = 20$, $\mu_{x_1} = 50$, $\mu_{x_2} = 40$,

$N = 10, M = 10, n' = 9, n = 4, m = 7$

Numerical Illustration using Artificial Population

Table 1. Percentage Relative Efficiency (PRE) of Estimators without and with Measurement Error under Case A .

Estimators	**Auxiliary Parameters**	**PRE without Measurement Error.** **P1=0.05, p2=0.1, p3=0.15, p4=0.2**	**PRE with Measurement Error.** **P1=0.05, p2=0.1, p3=0.15, p4=0.2**
t_0	Not applicable	100	100
t_{MSB}	Not applicable	0.01	0.01
t_1^*	$A_{x_2} = 1 , B_{x_2} = 0$	257.25	307.93
t_2^*	$A_{x_2} = 1, B_{x_2} = B_1(X_2)$	257.25	307.94
t_3^*	$A_{x_2} = 1, B_{x_2} = B_2(X_2)$	257.25	307.96
t_4^*	$A_{x_2} = 1, B_{x_2} = C_{X_2}$	257.25	307.93
t_5^*	$A_{x_2} = 1, B_{x_2} = S_{X_2}$	257.25	308.11
t_6^*	$A_{x_2} = B_1(X_2), B_{x_2} = B_2(X_2)$	257.25	308.44
t_7^*	$A_{x_2} = B_1(X_2), B_{x_2} = C_{X_2}$	257.25	307.94
t_8^*	$A_{x_2} = B_1(X_2), B_{x_2} = S_{X_2}$	257.25	308.99
t_9^*	$A_{x_2} = B_2(X_2), B_{x_2} = B_1(X_2)$	257.25	307.93
t_{10}^*	$A_{x_2} = B_2(X_2), B_{x_2} = C_{X_2}$	257.25	307.93
t_{11}^*	$A_{x_2} = B_2(X_2), B_{x_2} = S_{X_2}$	257.31	308.03
t_{12}^*	$A_{x_2} = C_{X_2}, B_{x_2} = B_1(X_2)$	257.34	308.13

t_{13}^{*}	$A_{x_2} = C_{X_2}, B_{x_2} = B_2(X_2)$	257.43	308.39
t_{14}^{*}	$A_{x_2} = C_{X_2}, B_{x_2} = S_{X_2}$	257.84	309.23
t_{15}^{*}	$A_{x_2} = S_{X_2}, B_{x_2} = B_1(X_2)$	257.25	307.93
t_{16}^{*}	$A_{x_2} = S_{X_2}, B_{x_2} = B_2(X_2)$	257.25	307.93
t_{17}^{*}	$A_{x_2} = S_{X_2}, B_{x_2} = C_{X_2}$	257.25	307.93

Table 2. Percentage Relative Efficiency (PRE) of Estimators without and with Measurement Error under Case B.

Estimators	**Auxiliary Parameters**	**PRE without Measurement Error.** **P1=0.05, p2=0.1, p3=0.15, p4=0.2**	**PRE with Measurement Error.** **P1=0.05, p2=0.1, p3=0.15, p4=0.2**
t_0	Not applicable	100	100
t_{MSB}	Not applicable	289.37	252.69
t_1^{*}	$A_{x_2} = 1 , B_{x_2} = 0$	288.41	254.22
t_2^{*}	$A_{x_2} = 1, B_{x_2} = B_1(X_2)$	288.42	254.22
t_3^{*}	$A_{x_2} = 1, B_{x_2} = B_2(X_2)$	288.44	254.22
t_4^{*}	$A_{x_2} = 1, B_{x_2} = C_{X_2}$	288.41	254.22
t_5^{*}	$A_{x_2} = 1, B_{x_2} = S_{X_2}$	288.54	254.22
t_6^{*}	$A_{x_2} = B_1(X_2), B_{x_2} = B_2(X_2)$	288.33	254.22
t_7^{*}	$A_{x_2} = B_1(X_2), B_{x_2} = C_{X_2}$	288.41	254.22
t_8^{*}	$A_{x_2} = B_1(X_2), B_{x_2} = S_{X_2}$	288.73	254.22
t_9^{*}	$A_{x_2} = B_2(X_2), B_{x_2} = B_1(X_2)$	288.42	254.22
t_{10}^{*}	$A_{x_2} = B_2(X_2), B_{x_2} = C_{X_2}$	288.41	254.22
t_{11}^{*}	$A_{x_2} = B_2(X_2), B_{x_2} = S_{X_2}$	288.48	254.29
t_{12}^{*}	$A_{x_2} = C_{X_2}, B_{x_2} = B_1(X_2)$	288.49	254.26
t_{13}^{*}	$A_{x_2} = C_{X_2}, B_{x_2} = B_2(X_2)$	288.96	254.85
t_{14}^{*}	$A_{x_2} = C_{X_2}, B_{x_2} = S_{X_2}$	289.57	255.54
t_{15}^{*}	$A_{x_2} = S_{X_2}, B_{x_2} = B_1(X_2)$	288.41	254.22
t_{16}^{*}	$A_{x_2} = S_{X_2}, B_{x_2} = B_2(X_2)$	288.42	254.23
t_{17}^{*}	$A_{x_2} = S_{X_2}, B_{x_2} = C_{X_2}$	288.41	254.22

Discussion

The results presented in Tables 1 and 2 confirm the superiority of the proposed regression-type estimators. Table 1 shows that the proposed estimators which utilized the second auxiliary variable parameter both in the presence and absence of measurement error, is more efficient, with higher percentage relative efficiencies (PREs) than the usual mean per unit estimator (t_0) without auxiliary information and the Maji *et al.* (2018) estimator (t_{MSB}) in case A for all choices of probabilities. From table 2, it can be observed that the suggested estimator (t_{14}^*), is more efficient, that the usual mean per unit estimator (t_0) without auxiliary information and the Maji *et al.* (2018) estimator (t_{MSB}), while Maji *et al.* (2018) estimator (t_{MSB}) performed better than other estimators in the same scenario when measurement error is absence for all non-response probability selections. In another scenario, when measurement is presence, the family members of the proposed estimator are more efficient for all non-response probabilities.

These results demonstrate the robustness and applicability of the proposed technique in real-world survey conditions involving non-response and measurement error.

References

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

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

[3] D B Rubin. Inference and missing data. Biometrika, 1976; 63(3): 581–592.

[4] D F Heitjan and S Basu. Distinguishing 'missing at random' and 'missing completely at random. American Statistician, 1996; 50 (3): 207–213.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

[24] R Singh, S Malik, and M Kohshnivesan. 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, 2014b; 2: 107-111.

[25] S Singh and A H Joarder. Estimation of finite population variance using random non-response in survey sampling. Metrika, 1998; 47(1): 241–249.

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

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

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

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

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