DOI: https://doi.org/10.4038/sljas.v26i2.8168

Sri Lankan Journal of Applied Statistics, Vol (26-2)

Estimation of Population Mean in Domains of Study Under Double Sampling: A Calibration Technique for Subsampling the Nonrespondents

M. J. Iseh[1,*], E. E. Inyang[2], and E. I. Enang[3]

[1]Akwa Ibom State University, Mkpat Enin, Akwa Ibom State, Nigeria

[2]Federal Polytechnic, Ukana, Akwa Ibom State, Nigeria

[3]University of Calabar, Calabar, Nigeria

*Corresponding author: matthewiseh@aksu.edu.ng

Received: 01st February, 2025/ Revised: 14th June, 2025/ Published: 20th June, 2025

ABSTRACT

The application of calibration in domain estimation with subsampling the non-respondents is the major emphasis of this work. The detrimental impact that result from a researcher's incapacity to gather data from every elementary unit in a certain demographic or from respondents' partial or complete unwillingness to supply the necessary information is still a problem in sample surveys. In view of the mentioned problem, this work becomes imperative as it utilizes the concept of calibration in double sampling. In order to increase the accuracy of such estimates by subsampling the non-respondents, it becomes important to close the gap between the loss of survey data and establishing a reliable estimate in the research areas. This formulation is subject to two conditions, when the auxiliary variable is free from non-response (condition A) and when the auxiliary variable is not free from non-response (condition B). And it has been demonstrated that the suggested estimators are unbiased. Two populations from actual data were examined in terms of variance/MSE and percentage relative efficiency (PRE) as part of the empirical study. Additionally, two non-response scenarios-one with uniform rates and the other with non-uniform rate were taken into account. The findings demonstrated that although the two conditions yield the same population mean estimates, condition A is more efficient than condition B. The cases, on the other hand, produce different estimates of population mean and variance/MSE. Compared to the existing estimators used in this work, the proposed estimator has proven to be more effective in providing accurate estimates of the population mean *in the domains. This is evident from the efficiency comparisons and empirical investigations of the proposed estimator and the existing estimators that the proposed estimator are capable of producing efficient estimates in domains of study when subsampling the nonrespondents.*

Keywords: Auxiliary variable, Calibration, Domain, Non-response, Mean Square Error, Subsampling

1 Introduction

It is nearly impossible to reach and evaluate every member of the population in a sample survey. Depending on the population's size, data analysis might occasionally be challenging, even when the complete population is used. It is impossible to overstate the importance of sampling in any population of interest because it saves money and, of course, time. A survey must be carried out, usually on a subset of the research population, if it is big, in order to collect accurate and trustworthy information/data. For a long time, sample surveys have been a useful tool for gathering data about a community of interest. Since 1930, sample surveys have been used more frequently because they are time and money-efficient while still yielding accurate data on the population characteristics of interest using statistical inference (Sarndal et al., 1992, Rao, 2003).

The primary goal of performing a sample survey is to compute various statistics known as estimates using a particular estimator to reflect their corresponding parameter estimates. Such estimates can be for totals, averages, variances, standard deviations, proportions, and so on. The data are frequently used in analytic research or survey analyses. This often entails comparing totals, means, variances, standard deviations, and proportions for subgroups of the population. These subgroups are frequently referred to as domains of study. It is worth noting that every stage of data generation has its own set of challenges, regardless of population or sample size. Some of the challenges include: funding, coverage, illiteracy rate, lack of skilled persons to assist, timing, lack of sample frame and non-response.

In sample survey, distinct parameter estimates may be required for subpopulations into which a population is divided if they are not sampled independently. Such subpopulations with tiny sample sizes ae referred to as domains of the population(s). Domain estimation refers to the processes used to derive estimated values for certain domains. A domain is a subgroup of the survey's total study population that requires particular estimates. Estimates are made in each class into which the population is broken during sampling, for example, the focus may not just be the overall unemployment rate but also the breakdown by age, gender, and education level.

In addition, the average monthly income can be divided into three categories: high, medium, and poor. The same may apply to students' success in a certain examination, which may be viewed in terms of A-grade, B-grade, C-grade and so on. Domain units may be identified prior to sampling. In such circumstances, the domains can be considered different strata from a given sample. Stratification guarantees that the domains in the final sample are well-represented. These domains are known as planned domains.

In theory, a domain can be any clearly defined group of elements (Hively, 1974). However, for a well-defined domain, both the basic qualities of the content that the student is supposed to acquire and the behaviour by which he or she is expected to demonstrate such acquisition should be thoroughly articulated. Nitko (1984) goes on to say that a domain is well-defined if it is evident which categories of performance or types of activities are and are not prospective test items, whereas an ill-defined domain is defined solely in terms of the specific items on a single test form.

Yates (1953) discussed in detail some of the issues connected with estimating domain totals, averages, and proportions using a single-stage simple random sampling. He observed that the variance of a domain parameter estimator is raised by the fact that the number of domain elements, and thus the number of elements that may be included in a random sample of a fixed size, is unknown prior to the commencement of the survey.

Durbin (1959) gave a derivation of Yates' results in multistage sampling. The paper is one of the first attempts to unify the theory of domain estimation. Hartley (1959) provided the theory for a number of sample designs where domain estimation was of interest. The paper mostly discussed estimations that did not make use of auxiliary information. However, he considered the case of ratio estimation where population totals were known for the domains.

Scott and smith (1971) extended Yates' result to satisfied multistage sampling. The same result of Yates was extended to Bayes Estimation for domains in Stratified Sampling by (Tin and Toe, 1972). Other notable works include but not limited to Udofia (2002a) Agrawal and Midha (2007), Chhikara and Sud (2009), Pandey and Tikkiwa (2010), Aditya (2012), Chaudhary et. al. (2012), Aditya et al (2014), Clement et al (2014), Kaustav et al (2014). Also in the list is the work of Anekeya et al (2018). The paper described the theoretical estimation of domains mean using double sampling with a non-linear cost function in the presence of non-response. The estimation of domain mean is proposed using auxiliary information in which the study and auxiliary variable suffers from non-response in the second phase sampling. The expression of the biases and mean square errors of the proposed estimations are obtained. The optimal stratum sample sizes for given set of non-linear cost function are developed. Also, Ashutosh et al (2019), Alilah et al. (2020), Iseh and Enang (2021), Iseh and Bassey (2021) and Khare et al (2021) are notable authors in this area.

Non-response occurs in almost every type of single survey, causing estimate issues that cannot be eliminated by expanding the sample size. Non-response in surveys invariably affects the precision of parameter estimations and increases bias in estimates, resulting in a bigger mean square error and diminishing the parameter(s) efficiency. Estimates derived from inadequate data may be biased, especially when respondents differ from non-respondents. Hansen and Hurwitz (1946) suggested a method for correcting for non-response to address the issue of bias. The technique involves picking a subset of non-respondents. Specialized efforts are used to collect data from non-respondents in order to determine the number of non-responding units in the population. The duo created an estimate called the weighted mean of two estimators: the sample mean based on the units/elements that replied and those that did not answer El-Badry (1956) investigated the topic of non-response and proposed a strategy for making repeated attempts to obtain information from non-responsive units. After numerous trials, the non-responding units were subsampled, and the estimates were calculated by pooling the trial findings. See also the works of Kumar (2015) and Iseh and Bassey (2022).

Furthermore, this problem of non-response can be addressed using calibration (a tool for reweighting of nonresponse). Calibration estimation is a general technique of modifying the original design weights to increase the precision of estimates by using auxiliary information such as the known population mean on total of the auxiliary variables. Calibration estimator uses calibrated weights that are aimed to minimize a given measure to the original design weights while satisfying a set of constraints related to the auxiliary information. Deville and Sarndal (1992) are the pioneer authors of calibration estimation in survey sampling and since then many researchers have studied calibration estimation using different calibration weights. Often direct information is not available in sufficient quantity but auxiliary information is available at a fraction of the cost of the direct information. Sometimes no special effort is required because auxiliary data for one quantity is the direct information for another. In survey sampling, the use of auxiliary information substantially improves the precision of estimates of population total and/or means as seen in Iseh and Bassey (2021a and 2021b) and Ikot and Iseh (2023) where the calibration estimator proposed gained more efficiency even in the presence of non-response. Hence, calibration requires the formulation of suitable auxiliary variables and provides a unified treatment of the use of auxiliary information in surveys relating to domain, non-response, etc. In the presence of powerful auxiliary information, the calibration approach meets the objectives of reducing both sampling errors and nonresponse errors.

The negative effect caused by a researcher's failure to collect information from all elementary units in a given population, as well as respondents' refusal to supply the required information, either fully ro partially, remains a concern. These non-responses result in incomplete data, which always leads to estimation bias and large mean square error. This gets more difficult when the researcher wants to gather information or make a judgment on a certain field of study. Given the aforementioned challenge, this study is critical since it used a distance function to minimize the weight of the primary estimator and the calibration weight while adhering to the auxiliary variable requirements. This becomes necessary to bridge the gap in having a reliable estimate in the domain of study as an improvement on the precision of such estimate with sub-sampling the non-respondents. The objective of this work are to provide a solution to non-response problem in the domain estimation by formulating an unbiased estimator of the population mean with variance for domains of study when the auxiliary variable is/not free from non-response under double sampling.

2 Methodology

2.1 Design for Domain Characteristics under Double Sampling Scheme

Let the finite population U under study of size N be divided into H strata such that n_h elements are considered from N_h in stratum $h,\ h = 1, 2, 3, \ldots, H$ which is further partitioned into D domains, $U_1, U_2.U_3, \ldots, U_D$ of sizes; $N_1, N_2, N_3, \ldots, N_D$ respectively. Here, the strata cut across the domains. Let Y and X denote the study and auxiliary variables respectively, such that $\bar{Y}_d$ and $\bar{X}_d$ be the population mean of the domains of study. Let y_{dh} and x_{dh} be the sample observations on the h^{th} strata in the respective variables. It is assume that non-response is observed on the study variable Y and not on the auxiliary variable X in case I, as such we have $\bar{y}_{dh1}$ and $\bar{y}_{dh2}$ as the mean of the respondents and nonrespondents respectively for the study variable in the d^{th} domain of the h^{th} stratum, while both Y and X will be affected by non-response in case II, thus, analogous to the study variable, $\bar{x}_{dh1}$ and $\bar{x}_{dh2}$ are measured as the mean of the respondents and nonrespondents respectively for the auxiliary variable in the d^{th} domain of the h^{th} stratum. It is also assumed that the domain population mean $\bar{X}_d$ of the auxiliary variable is unknown. In this case, a large preliminary sample of size n'_{dh} is selected by using simple random sampling without replacement (SRSWOR) scheme at the first phase from the population of N units. Information on X is obtained from all n'_{dh} units and use to estimate $\bar{X}_d$. A second sub-sample of size n_{dh} units $(n_{dh} < n'_{dh})$ is selected from the first phase sample units by SRSWOR at the second phase. Information from Y is obtained at the second phase sub-sample. Adopting the Hanson and Hurwitz (1946)technique of sub-sampling the non-respondents, during the first phase, out of n'_{dh} units, n'_{dh1} units respond while $n'_{dh2} = n'_{dh} - n'_{dh1}$ units do not respond on the auxiliary variable X. Again, under case II, when both Y and X are affected by non-response; for the second phase, out of n_{dh} units, n_{dh1} units respond while $n_{sh2} = n_{dh} - n_{dh1}$ units do not respond on the auxiliary and the study variables to estimate $\bar{Y}_d$. In view of the direction of non-response in the variable of interest, the classical Hanson and Hurwitz (1946) unbiased estimator with sub-sampling the non-respondent is subjected to calibration with a view to compensating for non-response bias and its attendant high variance of the estimate.

2.2 Study Notation and Definitions

For a typical d^{th} domain U_d several characteristics may be defined including the domain total, mean, variance and even covariance given respectively as; $Y_{U_d} = \sum_{U_d} y_{dh}, \bar{Y}_{U_d} = \frac{1}{N_d} \sum_{U_d} y_{dh}$,

$S^2_{U_d}(Y) = \frac{1}{N_{d-1}} \sum_{h \in U_d} (y_{dh} = \bar{Y}_{U_d})^2, S^2_{U_d}(X) = \frac{1}{N_{d-1}} \sum_{h \in U_d} (x_{dh} = \bar{X}_{U_d})^2$, $C_{U_d}(X,Y) = \frac{1}{N_{d-1}} \sum_{h \in U_d} (x_{dh} - \bar{X}_{U_d})(y_{dh} - \bar{Y}_{U_d})$. Again $W_{dh} = \frac{N_{dh}}{N_d}$ is the stratum weights.

2.3 Some Existing Estimators in Domains

There exist some domain estimators for population mean in stratified sampling technique with a single auxiliary variable in the presence of non-response.

Hansen and Hurwitz (1946) Estimator

The duo proposed what is termed as the classical estimator in domains of study with sub-sampling the non-respondent in which the study variable is influenced by non-response in the single stage sampling design defined as

$$\hat{\bar{y}}_d = \sum_{h=1}^{H} W_{dh} \bar{y}^*_{dh} \tag{1}$$

where $\bar{y}^*_{dh} = \frac{n_{dh1}\bar{y}_{dh1} + n_{dh2}\bar{y}_{dh2}}{n_{dh}}$ is unbiased estimator of the population mean for the d^{th} domain in the h^{th} stratum of the study variable with variance given as

$$V(\hat{\bar{y}}_d) = \sum_{h=1}^{H} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) W^2_{dh} S^2_{ydh} + \sum_{h=1}^{H} \frac{K_{dh} - 1}{n_{dh}} W^2_{dh} W_{dh2} S^2_{ydh2} \tag{2}$$

where W_{dh2} is the nonresponse rate of the d^{th} domain in the h^{th} stratum, S^2_{ydh} and S^2_{ydh2} are the h^{th} stratum variance of the d^{th} domain of the study variable for the response and nonresponse units respectively. $K_{dh} =$ Inverse sampling rate.

Anekeya et al (2018) Estimator

Following the Hansen and Hurwitz (1946) techniques, Anekeya et al (2018) proposed the unbiased estimators for estimating domain mean using double sampling with non-linear cost function in the presence of non-response. The unbiased estimator for estimating the domain population mean using $(n_{d_1} + r_{d_2})$ observation on y_{d1} domain study character given by;

$$\bar{y}_d = \frac{n_{d_1}}{n_d}\bar{y}_{d_1} + \frac{n_{d+2}}{n_d}\bar{y}_{rd2} = w_{d1}\bar{y}_{d1} + w_{d2}\bar{y}_{rd_2}$$

Similarly, the estimate for domain auxiliary variable is given by;

$$\bar{x}_d = \frac{n_{d_1}}{n_d}\bar{x}_{d_1} + \frac{n_{d_2}}{n_d}\bar{x}_{rd_2} = w_{d_1}\bar{x}_{d_1} + w_{d_2}\bar{x}_{rd_2} \tag{3}$$

where $\bar{y}_d$ and $\bar{x}_d$ are the sample domain means for the observation y_{d_1} and x_{d_1} respectively.

The proposed two ratio estimators Y_{dR_1} and $\hat{\bar{Y}}_{dR_2}$ thus:

$$\hat{\bar{Y}}_{dR_1} = \frac{\bar{y}_d}{x_d}\bar{x}'_d \tag{4}$$

$$\hat{\bar{Y}}_{dR_2} = \frac{\bar{y}_d}{x_d}\bar{x}_d \tag{5}$$

Bias of the Ratio Estimator $\hat{\bar{Y}}_{dR_1}$ and Y_{dR2}

The bias of ratio estimator Y_{dR1} is given by

$$B(\hat{\bar{Y}}_{dR_1}) = \bar{Y}_d\left(\frac{1}{n_d} - \frac{1}{n'_d}\right)C_{xd}(C_{xd} - \rho_{dxy}C_{yd}) + \left(\frac{K_d - 1}{n_d}\right)W_{d2}C_{xd2}$$

$(C_{xd2} - \rho_{d2xy}C_{yd2})$

The bias of the ratio estimator $\hat{\bar{Y}}_{dR_2}$ is given by:

$$B(\hat{\bar{Y}}_{dR_2}) = \bar{Y}_d\left(\frac{1}{n_d} - \frac{1}{n'_d}\right)\left(\rho_{dxy}\frac{S_{xd}}{\bar{X}_d}\frac{S_{yd}}{\bar{Y}_d}\right) + \left(\frac{K_d - 1}{n_d}\right)w_{d2}\rho_{d2xy}\frac{S_{xd2}}{\bar{X}_d}\frac{S_{yd2}}{\bar{Y}_d}$$

where $C_{xd} = \frac{S_{xd}}{\bar{X}_d}, C_{yd} = \frac{S_{yd}}{\bar{Y}_d}$, are the coefficients of variation of the domain population means for the auxiliary and study variable respectively of the respondents $C_{xd2} = \frac{S_{xd2}}{\bar{X}_d}, C_{yd2} = \frac{S_{yd2}}{\bar{Y}_d}$, are the coefficients of variation of the domain population means for the auxiliary and study variables respectively of the non-respondents. $W_{dh2} = \frac{n_{dh2}}{n_{dh}}$ is the stratum weights for the nonrespondents and ρ_{dxy} the correlation coefficient between the study and auxiliary variable $K_d =$ Inverse sampling rate.

Mean Square Error of the Ratio Estimator $\hat{\bar{Y}}_{dH_1}$

$$MSE(\hat{\bar{Y}}_{dR_1}) = \left[\left(\frac{1}{n'_d} - \frac{1}{N_d}\right) S^2_{y_d} + \left(\frac{1}{n_d} - \frac{1}{n'_d}\right) S^2_{dR_1} - \left(\frac{K_{d_1-1}}{n_d}\right) W_{d_2} S^2_{dR_2}\right] \tag{6}$$

where

$$S^2_{dR_1} = S^2_{y_d} + S^2_{x_d} R^2_d - 2\rho_{dxy} R_d S_{xd} S_{yd}$$

$$S^2_{dR_2} = S^2_{yd_2} + S^2_{x_{d2}} R^2_d - 2\rho_{d2xy} R_d S_{xd2} S_{yd2} \text{ and } R_d \frac{\bar{Y}_d}{\bar{X}_d}$$

Mean Square Error of Ratio Estimator $\hat{\bar{Y}}_{dR_2}$

$$MSE(\hat{\bar{Y}}_{dR_2}) = \left[\left(\frac{1}{n'_d} - \frac{1}{N_d}\right) S^2_{y_d} + \left(\frac{1}{n_d} - \frac{1}{n'_d}\right) S'^2_{dR_1} - \left(\frac{K_{d_1-1}}{n_d}\right) W_{d_2} S'^2_{dR_2}\right] \tag{7}$$

where

$$S'^2_{dR_1} = S^2_{y_d} + S^2_{x_d} R^2_d + 2\rho_{xdyd} R_d S_{xd} S_{yd}$$

is the variance of the domain population mean for the respondents and

$$S'^2_{dR_2} = S^2_{y_d2} + S^2_{x_d2} R^2_d + 2\rho_{xd2yd2} R_d S_{xd2} S_{yd2}$$

variance of the domain population mean for the non-respondents.

2.4 The Proposed Estimator

Motivated by Hansen and Hurwitz (1946) estimator with sub-sampling the non-respondents and by extension to domain estimation, the proposed calibration estimator becomes

$$\hat{\bar{y}}_{Dcal1} = \sum_{h-1}^{H} W^*_{dh} \bar{y}^*_{dh} \tag{8}$$

where W^*_{dh} is the calibration weight chosen such that the distance function

$$\phi = \sum_{h=1}^{H} \frac{(W^*_{dh} - W_{dh})^2}{Q_{dh} W_{dh}}$$

(see Devile and Samdal, 1992)

where Q_{dh} is the tuning parameter (for the purpose of this work, Q_{dh} will be set to be 1) is minimized subject to the constraints;

$$\sum_{h=1}^{H} W_{dh}^{*}\bar{x}_{dh} = \sum_{h=1}^{H} W_{dh}\bar{x}'_{dh} \tag{9}$$

$$\sum_{h=1}^{H} W_{dh}^{*}s_{xdh}^{2} = \sum_{h=1}^{H} W_{dh}S_{xdh}^{2} \tag{10}$$

Under condition A: when the study variable is with non-response but the auxiliary variable is free from non-response, then the optimization problem is formulated such that

$$L = \sum_{h=1}^{H}\left[\left[\frac{(W_{dh}^{*}-W_{dh})^{2}}{Q_{dh}W_{dh}}\right] - 2\lambda_{1}(W_{dh}^{*}\bar{x}_{dh} - W_{dh}\bar{x}'_{dh}) - 2\lambda_{2}(W_{dh}^{*}S_{xdh}^{2}\right.$$
$$\left. -W_{dh}S_{xdh}^{2}\right)$$

and the calibration weights become

$$W_{dh}^{*} = W_{dh} + W_{dh}\bar{x}_{dh}Q_{dh}$$

$$\left[\frac{(\sum_{h=1}^{H} W_{dh}\bar{x}'_{dh} - \sum_{dh}^{H} W_{dh}\bar{x}_{dh})(\sum_{h=1}^{H} Q_{dh}W_{dh}S_{xdh}^{4}) - (\sum_{h=1}^{H} W_{dh}S_{xdh}^{2} - \sum_{h=1}^{H} W_{dh}s_{xdh}^{2})((\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2})}{(\sum_{h=1}^{H} Q_{dh}W_{dh}W_{dh}\bar{x}_{dh}^{2})(\sum_{h=1}^{H} Q_{dh}W_{dh}S_{xdh}^{4}) - [\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2}]^{2}}\right] +$$

$$W_{dh}S_{xdh}^{2}Q_{dh}$$

$$\left[\frac{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}^{1})(\sum_{h=1}^{H} W_{dh}S_{xdh}^{2} - \sum_{h=1}^{H} W_{dh}S_{xdh}^{4}) - (\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2})(\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}^{1} - \sum_{h=1}^{H} W_{dh}\bar{x}_{dh})}{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}^{2})(\sum_{h=1}^{H} Q_{dh}W_{dh}S_{xdh}^{4}) - [\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2}]^{2}}\right]$$

The calibration estimator in (7) can be written as

$$\hat{\bar{y}}_{Dcal1} = \left(\sum_{h=1}^{H} W_{dh} + \sum_{h=1}^{H} W_{dh}\bar{x}_{dh}Q_{dh}\right.$$

$$\left[\frac{(\sum_{h=1}^{H} W_{dh}\bar{x}'_{dh} - \sum_{h=1}^{H} W_{dh}\bar{x}_{dh})(\sum_{h=1}^{H} Q_{dh}W_{dh}S_{xdh}^{4}) - (\sum_{h=1}^{H} W_{dh}S_{xdh}^{2} - \sum_{h=1}^{H} W_{dh}S_{xdh}^{2})(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2})}{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}^{2})(\sum_{h=1}^{H} Q_{dh}W_{dh}S_{xdh}^{4}) - [\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2}]^{2}}\right]$$

$$+\sum_{h=1}^{H} W_{dh}S_{xdh}^{2}Q_{dh}\ \hat{\bar{y}}_{dh}^{*}$$

$$\left[\frac{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}^{2})(\sum_{h=1}^{H} W_{dh}S_{xdh}^{2} - \sum_{h=1}^{H} W_{dh}s_{xdh}^{2}) - (\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2})(\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}^{1} - \sum_{h=1}^{H} W_{dh}\bar{x}_{dh})}{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}^{2})(\sum_{h=1}^{H} Q_{dh}W_{dh}S_{xdh}^{4}) - [\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}_{dh}S_{xdh}^{2}]^{2}}\right]$$

$$\hat{\bar{y}}_{Dcal1} = \sum_{h=1}^{H} W_{dh}\bar{y}^{*}_{dh} + \hat{\beta}_1 Q_{dh}\left(\sum_{h=1}^{H} W_{dh}\bar{x}^{1}_{dh} - \sum_{h=1}^{H} W_{dh}\bar{x}^{1}_{dh}\right) + \hat{\beta}_2 Q_{dh}\left(\sum_{h=1}^{H} W_{dh}S^{2}_{xdh} - \sum_{h=1}^{H} W_{dh}S^{2}_{xdh}\right) \quad (11)$$

Let $Q_{dh} = 1$, then

$$\hat{\bar{y}}_{Dcal1} = \sum_{h=1}^{H} W_{dh}\bar{y}^{*}_{dh} + \hat{\beta}_1(\bar{X}_d - \sum_{h=1}^{H} W_{dh}\bar{x}_{dh}) + \hat{\beta}_2\left(S^2_{xd} - \sum_{h=1}^{H} W_{dh}s^2_{xdh}\right) \quad (12)$$

where

$$\hat{\beta}_1 = \frac{(\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}\bar{y}^{*}_{dh})(\sum_{h=1}^{H} W_{dh}s^4_{xdh}) - (\sum_{h=1}^{H} W_{dh}\bar{y}^{*}_{dh}s^2_{xdh})(\sum_{h=1}^{H} W_{dh}\bar{x}^2_{dh}s^2_{xdh})}{(\sum_{h=1}^{H} W_{dh}\bar{x}^2_{dh})(\sum_{h=1}^{H} W_{dh}S^4_{xdh}) - [\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}s^2_{xdh}]^2}$$

$$\hat{\beta}_2 = \frac{(\sum_{h=1}^{H} W_{dh}\bar{x}^2_{dh}\bar{y}^{*}_{dh})\,(\sum_{h=1}^{H} W_{dh}\bar{y}^{*}s^4_{xdh}) - (\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}s^2_{xdh})(\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}\bar{y}^{*}_{x}s^2_{xdh})}{(\sum_{h=1}^{H} W_{dh}\bar{x}^2_{dh})(\sum_{h=1}^{H} W_{dh}s^2_{xdh}) - [\sum_{h=1}^{H} W_{dh}\bar{x}_{dh}s^2_{xdh}]^2}$$

Unbiasedness and Variance of the Proposed Estimator

Condition A: when the auxiliary variable is free from non-response errors

Theorem 1a

The calibrated estimator being a modified version of Hansen and Hurwitz (1946) in Eq. 12 is unbiased.

Proof:

Considering the proposed estimator in Eq. 12, using the large number approximations:

Let

$$e_y = \frac{\bar{y}^{*}_{dh} - \bar{Y}_{dh}}{\bar{Y}_{dh}}, e'_x = \frac{\bar{x}_{dh} - \bar{X}_{dh}}{\bar{X}_{dh}}, e'_x = \frac{\bar{x}'_{dh} - \hat{X}_{dh}}{\bar{X}_{dh}},\ e_s = \frac{s^2_{xdh} - S^2_{xdh}}{S^2_{xdh}}$$

where

$$\bar{x}_{dh} = \frac{\sum_{h=1}^{n_{dh}} x_{dh}}{n_{dh}}, \bar{X}_d = \frac{\sum_{h=1}^{H} \bar{x}_{dh}}{N_d}, s^2_{xdh} = \frac{\sum_{h=1}^{n_{dh}}(x_{dh} - \bar{X}_{dh})^2}{n_{dh} - 1},$$

$$\bar{x}_{dh}^{1} = \frac{1}{n'_{dh}} \sum_{h=1}^{n'_{dh}} x_{dh} \quad and \quad S_{xdh}^{2} = \frac{\sum_{h=1}^{H} (X_{dh} - \bar{X}_{dh})^2}{N_{dh} - 1}, \bar{X} = \frac{\sum_{d}^{N} \bar{X}_d}{N}$$

Also

$$\hat{y}_{dh}^{*} = \bar{Y}_{dh}(1 + e_y), \bar{x}_{dh} = \bar{X}_{dh}(1 + e_x).\bar{x}'_{dh} = \bar{X}_{dh}(1 + e_{x'}), S_{xdh}^{2}(1 + e_s)$$

$$E[e_y^2] = \sum_{h=1}^{H} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^{2} + \sum_{h=1}^{H} \frac{K_{dh} - 1}{n_{dh}} W_{dh2} S_{ydh2}^{2}$$

$$E[e_x^2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \frac{S_{xdh}^{2}}{\bar{X}_{dh}^{2}} = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) C_{xdh}^{2}$$

$$E[e_s^2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^{2}$$

$$E[e_{x'}^2] = E[e_{x'} e_x] = \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) \frac{S_{xdh}^{2}}{\bar{X}_{dh}^{2}} = \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) C_{xdh}^{2}$$

$$E[e_{x'} e_y] = \frac{S_{xydh}}{\bar{X}_{dh}\bar{Y}_{dh}} \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) = \frac{1}{\bar{X}_{dh}\bar{Y}_{dh}} \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xydh} S_{xdh} S_{ydh}$$

$$E[e_x e_y] = \frac{S_{xydh}}{\bar{X}_{dh}\bar{Y}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) = \frac{1}{\bar{X}_{dh}\bar{Y}_{dh}} \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xydh} S_{xdh} S_{ydh}$$

$$E[e_y] = E[e_x] = E[e_s] = E[e_{x'}] = 0$$

$$E[e_s^2] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) (\lambda_{04} - 1), E[e_y, e_s] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \frac{S_{ydh}}{\bar{Y}_{dh}} \lambda_{12}$$

$$E[e_x e_s] = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \frac{S_{x_{dh}}}{\bar{X}_{dh}} \lambda_{03}, E[e_{x'} e_s] = \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) \frac{S_{ydh}}{\bar{X}_{dh}} \lambda_{03}$$

where

$$\lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2} \mu_{02}^{s/2}} \quad \mu_{rs} = \frac{1}{N_{dh} - 1} \sum_{i=1}^{N} (Y_{dhi} - \bar{Y}_{dh})^r (\bar{X}_{dhi} - \bar{X}_{dh})^s$$

$$\mu_{20} = S^2_{ydh}, \ \ \mu_{02} = S^2_{x_{dh}x}$$

Hence

$$\lambda_{03} = \frac{\mu_{03}}{\mu_{20}^{0/2} \quad \mu_{02}^{s/2}} \lambda_{12} = \frac{\mu_{12}}{\mu_{20}^{1/2} \quad \mu_{02}^{1}}, \lambda_{04} = \frac{\mu_{04}}{\mu_{02}^{4/2}} = \frac{\mu_{04}}{\mu_{02}^{2}}$$

Eq. 12 could be written as

$$\hat{\bar{y}}_{Dcal1} = \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}(1+e_y)+\hat{\beta}_1 \left[\sum_{h=1}^{H} W_{dh}\bar{X}_{dh}(1+e_{x'}) - \sum_{h=1}^{H} W_{dh}\bar{X}_{dh}(1+e_x)\right] +$$

$$\hat{\beta}_2 \left[\sum_{h=1}^{H} W_{dh}S^2_{xdh} - \sum_{h=1}^{H} W_{dh}S^2_{xdh}(1+e_s)\right] \qquad (13)$$

$$E[\hat{\bar{y}}_{Dcal1}] = \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}+\sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}E[e_y]+\hat{\beta}_1 \left[\sum_{h=1}^{H} W_{dh}\bar{X}_{dh}E[e_{x'}] - \sum_{h=1}^{H} W_{dh}\bar{X}_{dh}[e_x]\right]$$

$$-\beta_2 \left[\sum_{h=1}^{H} W_{dh}S^2_{xdh}E[e_x]\right]$$

As sample size increases $(n \to \infty)$

$$E(\hat{\bar{y}}_{Dcal1}) \simeq \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}$$

but $\bar{Y}_d = \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}$

$$\Rightarrow E(\hat{\bar{y}}_{Dcal1}) \simeq \bar{Y}_d$$

This confirms that the proposed estimator is asymptotically unbiased.

Theorem 1b
The variance of unbiased calibrated estimator with sub-sampling the non-respondents is given as

$$Min[V(\hat{\bar{y}}_{Dcal1})] = \sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^2 + \frac{K_{dh}-1}{n_{dh}} W_{dh2} S_{ydh2}^2\right]$$

$$-\frac{\left[\sum_{dh}^{H} W_{dh}^2 \left(\frac{1}{n'_{dh}} - \frac{1}{n_{dh}}\right) \rho_{xdh} y_{dh} S_{ydh} S_{xdh}\right]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right) S_{xdh}^2} - \frac{\left[\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^2 S_{ydh} \lambda_{12}\right]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^4 (\lambda_{04}-1)}$$

Proof
The prove for the above theorem being the variance of the asymptotically unbiased calibration estimator with sub-sampling the non-respondents is presented thus:
Recall Eq. 13 and obtain $E[\hat{\bar{y}}_{Dcal1} - \bar{Y}_d]^2 = V(\hat{\bar{y}}_{Dcal1})$
using Large Sample Approximation technique

$$\begin{aligned} V(\hat{\bar{y}}_{Dcal1}) \quad &= \quad \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh}^2 E[e_{y^*}^2] + 2\hat{\beta}_1 \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh} \bar{X}_{dh} E[e_{x'} e_{y^*}] - 2\hat{\beta}_1 \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh} \bar{X}_{dh} E[e_x e_{y'}] \\ &- \quad 2\hat{\beta}_2 \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh} S_{xdh}^2 E[e_s e_{y^*}] + \hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh}^2 E[e_{x'}^2] - 2\hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh}^2 E[e_x e_{x'}] \\ &- \quad \hat{\beta}_1 \hat{\beta}_2 \left(\sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh} S_{xdh}^2 e_s e_{x'}\right) + \hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh}^2 E[e_x^2] + 2\hat{\beta}_1 \hat{\beta}_2 \left(\sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh} S_{xdh}^2 e_s e_x\right) \\ &+ \quad \hat{\beta}_2^2 \sum_{h=1}^{H} W_{dh}^2 S_{xdh}^4 E[e_s^2] \end{aligned}$$

Let

$$F_1 = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right), F_2 = \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right), F_3 = F_1 - F_2 = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) - \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) = \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right)$$

and

$$F_4 = \frac{K_{dh}-1}{n_{dh}}, F'_4 = \left(\frac{K_{dh}-1}{n_{dh}}\right) - \left(\frac{K_{dh}-1}{n'_{dh}}\right)$$

$$= \sum_{h=1}^{H} W_{dh}^2 [F_1 S_{ydh}^2 + F_4 W_{dh2} S_{ydh2}^2] + 2\hat{\beta}_1 \sum_{h=1}^{H} W_{dh}^2 F_2 \rho_{xdhydh} S_{ydh} S_{xdh} - 2\hat{\beta}_1 \sum_{h=1}^{H} W_{dh}^2 F_1 \rho_{xdhydh} S_{ydh} S_{xdh} -$$

$$2\hat{\beta}_2 \sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^3 S_{ydh} \lambda_{12} + \hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 F_2 S_{xdh}^2 - 2\hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 F_2 S_{xdh}^2 - \hat{\beta}_1 \hat{\beta}_2 (\sum_{h=1}^{H} W_{dh}^2 F_2 S_{xdh}^3 \lambda_{03}) +$$

$$\hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^2 + 2\hat{\beta}_1 \hat{\beta}_2 (\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^2 \lambda_{03}) + \hat{\beta}_2^2 \sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^4 (\lambda_{04} - 1) \qquad (14)$$

Minimizing the variance, Eq. 14 is differentiated with respect to β_1 and β_2

$$\begin{aligned} Min[V(\hat{\bar{y}}_{Dcal1})] &= \sum_{h=1}^{H} W_{dh}^2[F_1 S_{ydh}^2 + F_4 W_{dh2} S_{ydh2}^2] - 2\frac{[\sum_{h=1}^{H} W_{dh}^2 F_3 \rho_{xdhydh} S_{ydh} S_{xdh}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_3 S_{xdh}^2} \\ &- 2\frac{[\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^2 S_{ydh} \lambda_{12}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^4 (\lambda_{04} - 1)} + \frac{[\sum_{h=1}^{H} W_{dh}^2 F_3 \rho_{xdhydh} S_{ydh} S_{xdh}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_3 S_{xdh}^2} \\ &+ \frac{[\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^2 S_{ydh} \lambda_{12}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^4 (\lambda_{04} - 1)} \end{aligned}$$

$$= \sum_{h=1}^{H} W_{dh}^2[F_1 S_{ydh}^2 + F_4 W_{dh2} S_{ydh2}^2] - \frac{[\sum_{h=1}^{H} W_{dh}^2 F_3 \rho_{xdhydh} S_{ydh} S_{xdh}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_3 S_{xdh}^2} - \frac{[\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^2 S_{ydh} \lambda_{12}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^4 (\lambda_{04} - 1)}$$

substituting for F_1, F_2, F_3 and F_4 gives

$$= \sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^2 + \frac{K_{dh}-1}{n_{dh}} W_{dh2} S_{ydh2}^2\right] - \frac{[\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right) \rho_{xdhydh} S_{ydh} S_{xdh}]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right) S_{xdh}^2} -$$

$$\frac{[\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^2 S_{ydh} \lambda_{12}]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^4 (\lambda_{04} - 1)} \tag{15}$$

Eq. 15 is the variance when the study variable is not free from non-response error and the auxiliary variable is free.

Condition B:
When the study and auxiliary variables are not free from non-response error. Analogous to Eq. 1, the estimator of the auxiliary variable with sub-sampling the non-respondents is given as

$$\hat{\bar{x}}_d = \sum_{h=1}^{H} W_{dh} \bar{x}_{dh}^* \tag{16}$$

where

$$\bar{x}_{dh}^* = \frac{n_{dh}\bar{x}_{dh} + n_{dh2}\bar{x}_{dh2}}{n_{dh}}$$

Then to obtain the calibration weights for Eq. 7 with sub-sampling the non-respondents in both auxiliary and study variables, we minimized the distance function $\phi = \sum_{h=1}^{H} \frac{(W_{dh}^* - W_{dh})^2}{Q_{dh} W_{dh}}$ subject to the following constraints

$\sum_{h=1}^{H} W_{dh}^* \bar{x}_{dh}^* = \sum_{h=1}^{H} W_{dh} \bar{x}'_{dh}$

$$\sum_{h=1}^{H} W_{dh}^* s_{xdh}^2 = \sum_{h=1}^{H} W_{dh} S_{xdh}^2$$

and then, the calibration estimator in Eq. 7 becomes

$$\hat{\bar{y}}_{Dcal2} = \sum_{h=1}^{H} W_{dh}\bar{y}^*_{dh} + \hat{\beta}_1(\bar{X}_d - \sum_{h=1}^{H} W_{dh}\bar{x}^*_{dh}) + \hat{\beta}_2(S^2_{xd} - \sum_{h=1}^{H} W_{dh}s^2_{xdh}) \quad (17)$$

where

$$\hat{\beta}_1 = \frac{(\sum_{h=1}^{H} W_{dh}\bar{x}^*_{dh}\bar{y}^*_{dh})(\sum_{h=1}^{H} Q_{dh}W_{dh}s^4_{xdh}) - (\sum_{h=1}^{H} W_{dh}\bar{y}^*_{dh}s^2_{xdh})(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^*_{dh}s^2_{xdh})}{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^{*2}_{dh})(\sum_{h=1}^{H} Q_{dh}W_{dh}s^4_{xdh}) - [\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^*_{dh}s^2_{xdh}]^2}$$

$$\hat{\beta}_2 = \frac{\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^{*2}_{dh}(\sum_{h=1}^{H} W_{dh}\bar{y}^*_{dh}s^2_{xdh}) - (\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^*_{dh}s^2_{xdh})(\sum_{h=1}^{H} W_{dh}\bar{x}^*_{dh}\bar{y}_{dh})}{(\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^{*2}_{dh})(\sum_{h=1}^{H} Q_{dh}W_{dh}s^4_{xdh}) - [\sum_{h=1}^{H} Q_{dh}W_{dh}\bar{x}^*_{dh}s^2_{xdh}]^2}$$

Theorem 2a

The calibrated estimator in Eq. 17 is asymptotically unbiased under double sampling when the study and auxiliary variables are affected with non-response error.

Proof
Considering the fact that both the study and auxiliary variables are subjected to sub-sampling the nonrespondents, the large number approximations become

$$e^*_x = \frac{\bar{x}^*_{dh} - \bar{X}_{dh}}{\bar{X}_{dh}}, \bar{x}^*_{dh} = \bar{X}_{dh}(1 + e_{x^*})$$

$$E[e^2_{x'}] = \sum_{h=1}^{H}\left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) S^2_{xdh} + \sum_{h=1}^{H} \frac{K_{dh}-1}{n_{dh}} W_{dh2}S^2_{xdh2}$$

$$E[e^2_{x^*}] = \sum_{h=1}^{H}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S^2_{xdh} + \sum_{h=1}^{H} \frac{K_{dh}-1}{n_{dh}} W_{dh2}S^2_{xdh2}$$

$$E[e_{x'}e_{x^*}] = \sum_{h=1}^{H}\left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}}\right) S^2_{xdh} + \sum_{h=1}^{H} \frac{K_{dh}-1}{n_{dh}} W_{dh2}S^2_{xdh2}$$

$$E[e_{x^*}e_y] = \sum_{h=1}^{H}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) \rho_{xydh}S_{xdh}S_{ydh} + \sum_{h=1}^{H} \frac{K_{dh}-1}{n_{dh}} W_{dh2}\rho_{xydh2}S_{xdh2}S_{ydh2}$$

The, Eq. 17 can be written as

$$\hat{\bar{y}}_{Dcal2} = \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}(1+e_y)+\hat{\beta}_1[\sum_{h=1}^{H} W_{dh}\bar{X}_{dh}(1+e_{x'})-\sum_{h=1}^{H} W_{dh}\bar{X}_{dh}(1+e_x)]+$$

$$\hat{\beta}_2[\sum_{h=1}^{H} W_{dh}S^2_{xdh} - \sum_{h=1}^{H} W_{dh}S^2_{xdh}(1+e_s)] \tag{18}$$

$$E[\hat{\bar{y}}_{Dcal2}] = \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}+\sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}E[e_y]+\hat{\beta}_1[\sum_{h=1}^{H} W_{dh}\bar{X}_{dh}E[e_{x'}]-\sum_{h=1}^{H} W_{dh}\bar{X}_{dh}[e_x]]-\beta_2[\sum_{h=1}^{H} W_{dh}S^2_{xdh}E[e_x]]$$

Again as sample size increases the large sample approximation applies as $n \to \infty$,

$$E(\hat{\bar{y}}_{Dcal2}) \simeq \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}$$

but $\bar{Y}_d = \sum_{h=1}^{H} W_{dh}\bar{Y}_{dh}$

$$\Rightarrow E(\hat{\bar{y}}_{Dcal2}) = \bar{Y}_d$$

Theorem 2b
The variance of the calibrated estimator with sub-sampling the non-respondents when the study and auxiliary variables are plague with non-response error is given as

$$Min[V(\hat{\bar{y}}_{Dcal2})] = \sum_{h=1}^{H} W^2_{dh}\left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S^2_{ydh} + F_4 W_{dh2}S^2_{ydh2}\right]$$
$$- \frac{\left[\sum_{dh}^{H} W^2_{dh}\left(\left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right)\rho_{xdhydh}S_{ydh}S_{xdh} + \left[\frac{K_{dh}-1(n'_{dh}-n_{dh})}{n'_{dh}n_{dh}}\right]W_{dh2}\rho_{xdh2ydh2}S_{ydh2}S_{xdh2}\right)\right]^2}{\sum_{h=1}^{H} W^2_{dh}\left[\left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right)S^2_{xdh} + \left[\frac{K_{dh}-1(n'_{dh}-n_{dh})}{n'_{dh}n_{dh}}\right]W_{dh2}S^2_{sdh2}\right]}$$
$$- \frac{\left[\sum_{h=1}^{H} W^2_{dh}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S^2_{xdh}S_{ydh}\lambda_{12}\right]^2}{\sum_{h=1}^{H} W^2_{dh}\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right)S^4_{xdh}(\lambda_{04}-1)}$$

Proof

Considering the large sample approximations with sub-sampling the non-respondents for the study and auxiliary variables and let $V(\hat{\bar{y}}_{Dcal2})$ be the variance of Eq. 17.

Thus $V(\hat{\bar{y}}_{Dcal2}) = E[\hat{\bar{y}}_{Dcal2} - \hat{Y}_d]^2$

$$
\begin{aligned}
V(\hat{\bar{y}}_{Dcal2}) &= \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh}^2 E[e_{y^*}^2] + 2\hat{\beta}_1 \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh} \bar{X}_{dh} E[e_{x'^*} e_{y^*}] - 2\hat{\beta}_1 \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh} \bar{X}_{dh} E[e_{x^*} e_{y^*}] \\
&- 2\hat{\beta}_2 \sum_{h=1}^{H} W_{dh}^2 \bar{Y}_{dh} S_{xdh}^2 E[e_s e_{y^*}] + \hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh}^2 E[e_{x'^*}^2] - 2\hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh}^2 E[e_{x^*} e_{x'^*}] \\
&- \hat{\beta}_1 \hat{\beta}_2 \left(\sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh} S_{xdh}^2 e_s e_{x'^*} \right) + \hat{\beta}_1^2 \sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh}^2 E[e_{x^*}^2] + 2\hat{\beta}_1 \hat{\beta}_2 \left(\sum_{h=1}^{H} W_{dh}^2 \bar{X}_{dh} S_{xdh}^2 e_s e_{x^*} \right)
\end{aligned}
$$

$$
+\hat{\beta}_2^2 \sum_{h=1}^{H} W_{dh}^2 S_{xdh}^4 E[e_s^2] \tag{19}
$$

Let

$$
F_1 = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right),\ F_2 = \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}} \right),\ F_3 = F_1 - F_2 = \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) - \left(\frac{1}{n'_{dh}} - \frac{1}{N_{dh}} \right) = \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}} \right)
$$

and

$$
F_4 = \frac{K_{dh} - 1}{n_{dh}}, F'_4 = \left(\frac{K_{dh} - 1}{n_{dh}} \right) - \left(\frac{K_{dh} - 1}{n'_{dh}} \right) \left[\frac{K_{dh} - 1(n'_{dh} - n_{dh})}{n'_{dh} n_{dh}} \right]
$$

$$
Min[V(\hat{\bar{y}}_{Dcal2})] =
$$

$$
\begin{aligned}
& \sum_{h=1}^{H} W_{dh}^2 [F_1 S_{ydh}^2 + F_4 W_{dh2} S_{ydh2}^2] - \frac{[\sum_{h=1}^{H} W_{dh}^2 (F_3 \rho_{xdhydh} S_{ydh} S_{xdh} + F_{4'} W_{dh2} \rho_{xdh2ydh2} S_{xdh2})]^2}{\sum_{h=1}^{H} W_{dh}^2 [F_3 S_{xdh}^2 + F'_4 W_{dh2} S_{xdh2}^2]} \\
- & \frac{[\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^2 S_{ydh} \lambda_{12}]^2}{\sum_{h=1}^{H} W_{dh}^2 F_1 S_{xdh}^4 (\lambda_{04} - 1)}
\end{aligned}
$$

Substituting for $F_1, F_2, F_3 F_4$ and F'_4

$$
\begin{aligned}
= & \sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{ydh}^2 + \frac{K_{dh} - 1}{n_{dh}} W_{dh2} S_{ydh2}^2 \right] \\
- & \frac{\left[\sum_{h=1}^{H} W_{dh}^2 \left(\left(\frac{1}{n'_{dh}} - \frac{1}{n_{dh}} \right) \rho_{xdhydh} S_{ydh} S_{xdh} + \left[\frac{K_{dh} - 1(n'_{dh} - n_{dh})}{n'_{dh} n_{dh}} \right] W_{dh2} \rho_{xdh2ydh2} S_{ydh2} S_{xdh2} \right) \right]^2}{\sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{n_{dh}} \right) S_{xdh}^2 + \left[\frac{K_{dh} - 1(n'_{dh} - n_{dh})}{n'_{dh} n_{dh}} \right] W_{dh2} S_{dh2}^2 \right]}
\end{aligned}
$$

$$
- \frac{\left[\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^2 S_{ydh} \lambda_{12} \right]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}} \right) S_{xdh}^4 (\lambda_{04} - 1)} \tag{20}
$$

Eq. 20 is the variance when the study and the auxiliary variables are not free from nonresponse error.

3 Theoretical / Empirical Results and Discussion

To validate the theoretical derivations, this section considers efficiency comparison of the proposed and the existing estimators and in addition, empirical analysis is also carried and using real-life data. Furthermore, discussion of results is also done.

3.1 Efficiency Comparison of the Estimators

The efficiency comparison of the existing Anekeya et al (2018) estimators with the proposed estimators is treated under these conditions

Condition 1: $Min[V(\hat{\bar{y}}_{Dcal1})] - MSE(\hat{\bar{Y}}_{dR_1}) < 0$ if

$$= \sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{ydh}^2 + \frac{K_{dh}-1}{n_{dh}} W_{dh2} S_{ydh2}^2\right] - \frac{[\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right) \rho_{xdhydh} S_{ydh} S_{xdh}]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{n'_{dh}}\right) S_{xdh}^2}$$

$$- \frac{[\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^2 S_{ydh} \lambda_{12}]^2}{\sum_{h=1}^{H} W_{dh}^2 \left(\frac{1}{n_{dh}} - \frac{1}{N_{dh}}\right) S_{xdh}^4 (\lambda_{04} - 1)}$$

$$- \left[\left(\frac{1}{n_d} - \frac{1}{n_d}\right) S_{yd}^2 + \left(\frac{1}{n'_d} - \frac{1}{N_d}\right) S_{dR_1}^2 - \left(\frac{K_{d_1} - 1}{n_d}\right) W_{d_2} S_{dR2}^2\right] < 0$$

If condition 1 is satisfied, then the proposed estimator in double sampling under case 1 is more efficient than the existing estimator by Anekeya et al (2018).

Condition 2: $V(\hat{\bar{y}}_{Dcal2}) - MSE(\hat{\bar{Y}}_{dR_1}) < 0$ if

$$\Rightarrow \sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{N_d}\right) S_{ydh}^2 + \frac{K_{dh}-1}{n_{dh2}} W_{dh2} S_{ydh2}^2\right]$$

$$- \frac{\left[\sum_{h=1}^{H} W_{dh}^2 \left(\left(\frac{1}{n'_{dh}} - \frac{1}{n_{dh}}\right) \rho_{xdhydh} S_{ydh} S_{xdh} + \left[\frac{K_{dh}-1(n'_{dh}-n_{dh})}{n'_{dh} n_{dh}}\right] W_{dh2} \rho_{xdh2ydh2} S_{ydh2} S_{xdh2}\right)\right]^2}{\sum_{h=1}^{H} W_{dh}^2 \left[\left(\frac{1}{n_{dh}} - \frac{1}{n_{dh}}\right) S_{xdh}^2 + \left[\frac{K_{dh}-1(n'_{dh}-n_{dh})}{n'_{dh} n_{dh}}\right] W_{dh2} S_{dh2}^2\right]}$$

$$-\frac{\left[\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{2} S_{ydh} \lambda_{12}\right]^{2}}{\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{4}(\lambda_{04}-1)}$$

$$-\left[\left(\frac{1}{n_{d}'}-\frac{1}{N_{d}}\right) S_{yd}^{2}+\left(\frac{1}{n_{d}}-\frac{1}{n_{d}'}\right) S_{dR_{1}}^{2}-\left(\frac{K_{d_{1}}-1}{n_{d}}\right) W_{d_{2}} S_{dR_{2}}^{2}\right]<0$$

If condition 2 is satisfied, then the proposed estimator in double sampling under case II is more efficient than the existing estimator by Anekeya et al (2018).

Condition 3: $V(\hat{\bar{y}}_{Dcal1})-MSE(\hat{\bar{Y}}_{dR_{2}})<0$ if

$$=\sum_{h=1}^{H} W_{dh}^{2}\left[\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{ydh}^{2}+\frac{K_{dh}-1}{n_{dh}} W_{dh2} S_{ydh2}^{2}\right]-\frac{\left[\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{n_{dh}'}\right) \rho_{xdhydh} S_{ydh} S_{xdh}\right]^{2}}{\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{n_{dh}'}\right) S_{xdh}^{2}}$$

$$-\frac{\left[\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{2} S_{ydh} \lambda_{12}\right]^{2}}{\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{4}(\lambda_{04}-1)}$$

$$-\left[\left(\frac{1}{n_{d}'}-\frac{1}{N_{d}}\right) S_{yd}^{2}+\left(\frac{1}{n_{d}}-\frac{1}{n_{d}'}\right) S_{dR_{1}}^{2}+\left(\frac{K_{d_{1}}-1}{n_{d}}\right) W_{d_{2}} S_{dR_{2}}^{2}\right]<0$$

If condition 3 is satisfied, then the proposed estimator in double sampling under case 1 is more efficient than the existing estimator by Anekeya et al (2018).

Condition 4: $V(\hat{\bar{y}}_{Dcal2})-MSE(\hat{\bar{Y}}_{dR_{2}})<0$ if

$$=\sum_{h=1}^{H} W_{dh}^{2}\left[\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{ydh}^{2}+\frac{K_{dh}-1}{n_{dh}}, W_{dh2} S_{ydh2}^{2}\right]$$

$$-\frac{\left[\sum_{h=1}^{H} W_{dh}^{2}\left(\left(\frac{1}{n_{dh}'}-\frac{1}{n_{dh}}\right) \rho_{xdhydh} S_{ydh} S_{xdh}+\left[\frac{K_{dh}-1(n_{dh}'-n_{dh})}{n_{dh}' n_{dh}}\right] W_{dh2} \rho_{xdh2ydh2} S_{ydh2} S_{xdh2}\right)\right]^{2}}{\sum_{h=1}^{H} W_{dh}^{2}\left[\left(\frac{1}{n_{dh}}-\frac{1}{n_{dh}}\right) S_{xdh}^{2}+\left[\frac{K_{dh}-1(n_{dh}'-n_{dh})}{n_{dh}' n_{dh}}\right] W_{dh2} S_{dh2}^{2}\right]}$$

$$-\frac{\left[\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{2} S_{ydh} \lambda_{12}\right]^{2}}{\sum_{h=1}^{H} W_{dh}^{2}\left(\frac{1}{n_{dh}}-\frac{1}{N_{dh}}\right) S_{xdh}^{4}(\lambda_{04}-1)}$$

$$-\left[\left(\frac{1}{n'_d}-\frac{1}{N_d}\right)S^2_{yd}+\left(\frac{1}{n_d}-\frac{1}{n'_d}\right)S^2_{dR_1}+\left(\frac{K_{d_1}-1}{n_d}\right)W_{d_2}S^2_{dR_2}\right]<0$$

If condition 4 is satisfied, then the proposed estimator in double sampling under case II is more efficient than the existing estimator by Anekeya et al. (2018).

3.2 Empirical Study

The municipalities of MU284 in Sweden are taken into consideration (Samdal et al 1992), appendix B). The population is separated into eight groups (domains) based on geography. The domain sizes are 56, 48, 41, 38, 32, 29, 25, and 15, respectively, and they are labeled as such. The first four domains that were considered larger than the others were taken into consideration in this investigation. Calibration estimation techniques are the suggested estimators. Each domain is then conveniently divided into two homogeneous groups (strata): stratum 1 (value of at most 1500 million kronor) and stratum 2 (value of more than 1500 million kronor). In the second population, four out of the first five large domains were considered randomly. The domains in this case are conveniently classified into two homogeneous groups (strata); value of at most 100 millions of kronor (stratum 1) and above 100 millions of kronor (stratum 2). In both populations 1 and 2 we examine cases 1 and II of non-response.

Case 1: If a consistent percentage of nonrespondents (about 30%) are available in both stratum (1 and 2) nd the domains.

Case II: If there are varying percentages of nonrespondents in stratum 1 and 2, roughly 20% and 40% respectively.

Quantities like n_{dh1}, n_{dh} and n'_{dh} were computed based on existing information from the populations. In the case of double sampling, we computed the first phase sample size (n'_{dh}) such that it will be less than the second phase (n_{dh}). Hence, we assumed n'_{dh} to be 80% of the stratum size (N_{dh}) in both populations.

In this work, two existing estimators excluding the classical estimators are compared independently with the proposed estimator subject to the two conditions mentioned.

Population 1

Y: Real Estate Values according to 1984 Assessment (in millions of kronor).
X: Total number of municipal employees in 1984.

Computation of Parameter Estimates

The data set for population 1 was used to calculate the parameter estimates per stratum for every domain. Both the uniform and non-uniform non-response rates were taken into consideration while carrying out the computation. In Tables 1 and 2, these are indicated. The second population was similarly subjected to the same characteristics and cases, as indicated in Tables 3 and 4.

Table 1: The parameter values of Strata 1 and 2 for domains (Population 1)

	DOMAIN							
Domain Parameter	D_1		D_2		D_3		D_4	
Domain Size	56		48		41		38	
Stratum	1	2	1	2	1	2	1	2
N_{dh}	25	31	30	18	19	22	15	23
W_{dh}	0.446	0.554	0.625	0.375	0.463	0.537	0.395	0.605
$\bar{Y}_{dh}$	961.84	4727.613	1195.10	5930.722	1060.947	3137.727	1216.933	4023.261
$\bar{X}_{dh}$	430.920	3168.516	569.833	3473.444	487.105	1628.864	576.467	2825.478
S^2_{Ydh}	65851.31	44482456	161392	15624549	102762.2	3282835	32261.78	12834142
S^2_{Xdh}	4523.99	67662210	63455.18	7035367	21797.21	1292120	19051.41	24786285
S_{XYdh}	21561.94	53488810	81508.67	9893731	36389.01	2001120	17316.75	170581.98
ρ_{XYdh}	0.697	0.975	0.805	0.943	0.769	0.972	0.698	0.956

Source: Statistical Computation from Original Data 2025

Table 2: The Parameter Values of Strata 1 and 2 for domains in cases I and II (Population 1)

Domain	Strata	S^2_{Ydh2}	S^2_{Xdh2}	S_{XYdh2}	K_{dh}	n_{dh2}	w_{dh2}	n_{dh1}	n_{dh}	n'_{dh}
Case I										
1	1	19755.4	4357.2	6468.6	2	4	0.3	9	13	20
	2	13344736.8	20298663	16046643	3	5	0.3	12	17	25
2	1	48417.6	19036.6	24452.3	3	5	0.3	12	17	24
	2	4687365	2110610.1	2968119.3	2	3	0.3	7	10	14
3	1	30828.7	6539.2	10916.7	2	3	0.3	7	10	15
	2	984850.5	387636.0	600336	2	3	0.3	7	10	18
4	1	9678.53	5715.42	5195.03	2	3	0.3	7	10	12
	2	3850242.6	7435885.5	5117459.4	2	4	0.3	9	13	18
Case II										
1	1	13170.3	2904.8	4312.4	2	3	0.2	12	15	20
	2	17792982.4	27064884	4	6	0.4	9	17	15	25
2	1	32278.4	12691	16301.5	2	3	0.2	12	15	24
	2	6249819.6	2814146.8	3957492.4	3	4	0.4	6	10	14
3	1	20552.4	4359.4	7277.8	2	2	0.2	8	10	15
	2	1313134	516848	800448	3	5	0.4	8	13	18
4	1	6452.4	3810.3	3463.4	2	2	0.2	8	10	12
	2	5133656.8	9914514	6823279.2	3	5	5	0.4	8*13	18

Source: Statistical computation from original data 2025.

Population 2:
Y: Revenue from the Municipal taxation for 1985
X: The 1985 population

Table 3: The parameter values of strata 1 and 2 for domains (Population 2)

	DOMAIN							
Domain Parameter	D_1		D_2		D_3		D_4	
Domain Size	48		32		38		56	
Stratum	1	2	1	2	1	2	1	2
N_{dh}	22	26	14	18	14	24	29	27
W_{dh}	0.458	0.542	0.436	0.564	0.368	0.632	0.518	0.482
$\bar{Y}_{dh}$	65.314	376.153	67.5	260.61	75.857	376.5	63.448	498.704
$\bar{X}_{dh}$	9.318	46.846	10.642	34.5	12.357	41.875	9.862	50.370
S^2_{Ydh}	4175	129916.4	275.9615	4120084	143.362	466644.2	304.113	1565813
S^2_{Xdh}	7.465	1801.415	544.971	4.863	2133.853	5.695	5924.088	24786285
S_{XYdh}	53.846	10243.34	32.038	4559.147	23.747	30428.41	39.385	95437.92
ρ_{XYdh}	0.964	0.670	0.904	0.962	0.899	0.964	0.966	0.991

Source: Statistical Computation from Original Data 2025

Table 4: The Parameter Values of Strata 1 and 2 for domains in cases I and II (Population 2)

Domain	Strata	S^2_{Ydh2}	S^2_{Xdh2}	S_{XYdh2}	K_{dh}	n_{dh2}	w_{dh2}	n_{dh1}	n_{dh}	n'_{dh}
Case I										
1	1	125.33	2.234	16.154	2	3	0.3	7	10	18
	2	38974.92	540.424	3073.002	2	4	0.3	9	13	21
2	1	82.79	1.366	9.6114	2	2	0.3	5	7	11
	2	12360.25	163.49	1367.744	2	3	0.3	7	10	14
3	1	43.01	1.4589	7.1241	2	2	0.3	5	7	11
	2	139993.26	640.156	9128.523	2	4	0.3	9	13	19
4	1	91.234	1.7085	11.815	3	5	0.3	12	17	25
	2	469743.90	1777.226	28631.376	2	4	0.3	9	13	22
Case II										
1	1	83.55	1.495	10.7692	2	2	0.2	8	10	18
	2	51966.56	720.566	4097.336	3	5	0.4	8	13	21
2	1	55.192	0.9110	6.4076	2	2	0.2	8	10	11
	2	16480.34	217.988	1823.658	2	4	0.4	6	10	14
3	1	28.67	0.9726	4.7494	2	2	0.2	8	10	11
	2	186657.68	853.54	12171.364	3	5	0.4	8	13	19
4	1	60.82	1.139	7.877	2	3	0.2	12	15	23
	2	626325.20	2369.635	38175.168	3	3	6	0.4	89	22

Source: Statistical computation from original data 2025.

Estimation of Mean Square Error

The computation of MSE and Average MSE for Population 1 and 2 where carried out with the help of the Minitab software. The results for the existing and proposed estimators are presented in Table 5 while Table 6 is for the Percentage Relative Efficiency (PRE).

Table 5: MSE for the existing and proposed estimators

Estimator	D_1	D_2	D_3	D_4	AMSE
Case I (Population I)					
$\hat{\hat{Y}}_{DR1}$	6213149	1110078	878593	1588785	2447651
$\hat{\hat{Y}}_{DR2}$	15608763	3667830	1716088	4341202	6333471
$\hat{\hat{Y}}_{Dcal1}$	845316	167401	81740	265836	340073
$\hat{\hat{Y}}_{Dcal2}$	410451	93159	32457	147814	170970
Case II (Population 1)					
$\hat{\hat{Y}}_{DR1}$	16478334	3161536	1081041	4272688	6248400
$\hat{\hat{Y}}_{DR2}$	28544349	5217464	1441470	6508007	9752823
$\hat{\hat{Y}}_{Dcal1}$	3683404	1294946	215809	855860	1512505
$\hat{\hat{Y}}_{Dcal2}$	697078	144965	45788	233384	280304
Case I (Population 2)					
$\hat{\hat{Y}}_{DR1}$	8141.60	2323.30	18327.20	57725.00	21630
$\hat{\hat{Y}}_{DR2}$	20141.00	7551.00	5252.00	202860.00	70769
$\hat{\hat{Y}}_{Dcal1}$	1232.63	355.99	3652.60	5789.12	2758
$\hat{\hat{Y}}_{Dcal2}$	1423	540	5679	9534	4294
Case II (Population 2)					
$\hat{\hat{Y}}_{DR1}$	14691.00	2554.90	33222.60	92800.00	35817
$\hat{\hat{Y}}_{DR2}$	32503.00	8706.00	86750.00	253634.00	95398
$\hat{\hat{Y}}_{Dcal1}$	1907.40	446.80	6949.90	10963.20	5067
$\hat{\hat{Y}}_{Dcal2}$	2118	609	9111	14824	6666

Source: Statistical computation from original data 2025

Table 6: PRE for the existing and proposed estimators

Estimator	D_1	D_2	D_3	D_4
Case 1 (Population 2)				
$\hat{\hat{Y}}_{DR1}$	100	100	100	100
$\hat{\hat{Y}}_{DR2}$	40	30	51	37
$\hat{\hat{Y}}_{Dcal1}$	735	633	1075	598
$\hat{\hat{Y}}_{Dcal2}$	1514	1192	2707	1075
Case II (Population 1)				
$\hat{\hat{Y}}_{DR1}$	100	100	100	100
$\hat{\hat{Y}}_{DR2}$	64	61	75	66
$\hat{\hat{Y}}_{Dcal1}$	447	244	501	499
$\hat{\hat{Y}}_{Dcal2}$	2364	2181	2361	1831
Case 1 (Population 2)				
$\hat{\hat{Y}}_{DR1}$	100	100	100	100
$\hat{\hat{Y}}_{DR2}$	40	31	35	28
$\hat{\hat{Y}}_{Dcal1}$	661	653	502	997
$\hat{\hat{Y}}_{Dcal2}$	572	430	323	605
Case II (Population 2)				
$\hat{\hat{Y}}_{DR1}$	100	100	100	100
$\hat{\hat{Y}}_{DR2}$	45	29	38	37
$\hat{\hat{Y}}_{Dcal1}$	770	572	478	846
$\hat{\hat{Y}}_{Dcal2}$	694	420	365	626

Source: Statistical computation from original data 2025.

Estimation of population means using estimators

The mean estimate for the classical, existing and proposed estimators were computed using real data set a contained in population I, both cases of uniform and non-uniform non-response error per stratum and domain were considered. The results for both cases are displayed in Table 7.

Table 7: Population mean estimates of the estimators in both conditions and cases

	Domain				
Estimator	D_1	D_2	D_3	D_4	Total
Case 1					
$\hat{\hat{Y}}_{DR1}$	3231.93	2855.59	2201.02	4301.37	12589.91
$\hat{\hat{Y}}_{DR2}$	4429.72	2061.36	1454.55	1325.03	9280.66
$\hat{\hat{Y}}_{Dcal1}$	1424.11	3085.46	2281.68	9134.98	15926.23
$\hat{\hat{Y}}_{Dcal2}$	1424.11	3085.46	2281.68	9134.98	15926.23
Case II					
$\hat{\hat{Y}}_{DR1}$	3115.87	2869.52	2142.65	4301.37	12429.41
$\hat{\hat{Y}}_{DR2}$	4945.61	2074.65	1839.60	1325.03	10184.89
$\hat{\hat{Y}}_{Dcal1}$	3321.00	202133.00	2192.00	9135.00	216781
$\hat{\hat{Y}}_{Dcal2}$	3321.00	202133.00	2192.00	9135.00	216781

3.3 Discussion of Findings

Both analytical and empirical analysis, with results shown in Tables, served as the foundation for this discussion. It is apparent from Table 5 (Populations 1 and 2) that the suggested estimators are more efficient than the existing estimators ($\hat{\hat{Y}}_{DR1}$ and $\hat{\hat{Y}}_{DR2}$) across all domains. This is evident in both situations where the nonresponse rate was non-uniform as indicated by the data and when it was uniform throughout the strata. Additionally, it is noted that in case I, where the nonresponse rate was consistent throughout the strata across the domains, the variance or MSE was lower than in case II, where the nonresponse rate was not consistent. In contrast to Anekeya et al. (2018), who just proposed estimators with MSE without distinguishing the estimators with the smallest or largest variance or MSE within the domains, this is an improvement. The results further demonstrated that in both cases I and II of population 1, the variance of the suggested estimators under condition A performed better in terms of gains in efficiency and more effectively than that under condition B. In Population 2, the situation is the opposite. This demonstrates that in population I, the nonresponse's impact on the auxiliary variable was minimal. The idea of double sampling, which lessened the impact of nonresponse on the auxiliary variable, is largely responsible for this. However, in population 2, the suggested estimator under condition A performed better than that under condition B, demonstrating the impact of nonresponse on the auxiliary variable.

It is also apparent from Table 6 (Populations 1 and 2) that the suggested estimators' percentage relative efficiency (PRE) for the two conditions exceeded

100% which is the PRE of the existing estimator $(\hat{\hat{Y}}_{DR1})$ for every domain. According to the data, this occurs in both nonresponse cases. Compared to case II, where the nonresponse rate was not uniform, case I, where the nonresponse rate was uniform across the strata in the domains, exhibited a higher PRE. Once more, this is an improvement above the work of Anekeya et al. (2018), who just suggested an estimator with theoretical mean square error derivations without conducting any further research to determine the extent of efficiency gains.

This implies that the estimator(s) will become more efficient if the nonresponse rate is low and consistent across all subgroups, which may not be practical in real-world experience. As a result, case I estimators are more efficient than case II estimators. The PRE of the suggested estimator under condition B, however, performed better in population I than the one under condition A, indicating that the impact of nonresponse on the auxiliary variable was minimal. Since the estimator under condition A became more efficient than that under condition B, the opposite incidence was seen in population 2. This was noted in every domain. Given that the two phase sampling strategy has reduced the non-response effect on the auxiliary variable, it can be concluded that the effect of nonresponse on the auxiliary variable was minimal in population I.

Table 7 shows the estimated population means for cases I and II. The population mean estimate is found to be unaffected with either the study variable or both the study and auxiliary variables being exposed to the influence of nonresponse. This is basically due to the existing Hansen and Hurwitz (1946) unbiased estimator, which through the sub-sampling technique had increased the estimator's precision of the population mean.

According to the theoretical derivations of unbiasedness, this is evident in conditions A and B, where the nonrespondents' subsampling had an impact on the study variable and both the study and auxiliary variables, respectively. This suggests that any of these criteria will be undoubtedly result in the same estimate any of the estimators that was developed from the classical unbiased Hansen and Hurwitz (1946). Nonetheless, as the suggested estimator shows, improvement on the variance is desired. Therefore, the most efficient estimator is thought to be the domain estimator with the lower MSE.

The suggested estimators are superior to Anekeya et al. (2018) estimators when nonresponse is observed in the study variable and when it is observed in both the study and auxiliary variables, according to the analytical and empirical results. Once more, the idea of calibration with two constraints has proven to be effective in generating estimators that are not only unbiased but also have higher efficiency benefits, especially in cases when the auxiliary variable exhibits nonresponse.

4 Conclusion

The detrimental impact caused by a researcher's incapacity to gather data from every elementary unit in a certain demographic or by respondents' partial or complete unwillingness to supply the necessary information is still an issue. These non-responses add up to incomplete data, which inevitably produces an untrustworthy outcome. When the researcher wants to gather information or decide on a specific field of study, this becomes more difficult. Because it used a minimizing distance function between the calibration weight and the weight of the main (classical) estimator, subject to the established limitations on the auxiliary variable, this study becomes essential in light of the aforementioned dilemma. This becomes necessary to bridge the gap between loss of information from the survey data and having reliable estimates in the domain of study as an improvement on the precision of such estimates with sub-sampling the non-respondents as well as obtaining a more efficient estimator for population mean in domains of study. The study contributes immensely to the theory of domain estimation of the population mean in stratified random sampling with sub-sampling the non-respondents, when there is nonresponse in the study variable and auxiliary variable is free from non-response. It also contributes in a situation where both the study and auxiliary variables are affected by non-response error.

When the population parameter of the auxiliary variable is unknown, the suggested class of estimators offers the possibility of incorporating various sample estimates of the domain population parameters of the auxiliary variable into estimators with subsampling the nonrespondents using the calibration concept. Putting an estimator under conditions where the study variable is affected by nonresponse while the auxiliary is free, or putting both variables under the influence of nonresponse, did not influence the population mean estimate; according to the study. The second constraint, which relates to the stratum variance of the auxiliary variable, also did not contribute in any way to the auxiliary variable.

It is clear from efficiency comparisons and empirical research that the calibration technique has been effective in estimating the population mean by subsampling the non-respondents which yields higher efficiency gains than the current estimator. Therefore, in more complex sample designs, such as domain designs and small area estimation, the calibration approach should be used to compensate for non-response. Again simulation study will be considered in future study to ascertain which probability distribution will perform better.

References

Aditya, K., Sud, U. C. and Chandra, H. (2012). Estimation of domain total for unknown domain size in the presence of non-response statistics and applications. 10, 1&2, (New Series), 13-25.

Aditya, K., Sud, U. C. and Ghrade, Y. (2014). Estimation of domain mean using two stage sampling with subsampling of non-respondents. *Journal of Indian Social Agricultural Statistics.* 68(1), 39-54.

Alilah, D., Onyango, C. and Ombaka, E. (2020). Efficiency of domain mean estimators in the presence of non-response using two-stage sampling with non-linear and linear cost function. *Annals of Data Science.* Vol. 10.

Anekeya, A. D., Onyango, O. C., Nyongesa, K. (2018). Domain mean estimation using double sampling with non-linear cost function int he presence of non response. *Science Journal of Applied Mathematics and Statistics.* Vol. 6(1): 28 - 42 http://www.sciencepublishinggroup.com/j/sjams dio: 10.11648/j.sjams.20180601.14.

Ashutosh, Khare, B. B. and Khare, S. (2019). Generalized synthetic estimator for domain mean in two phase sampling using single auxiliary character. *Journal of Reliability and Statistical Studies.* Vol. 12(2), 139-15.

Agrawal, M. C. and Midha, C. K. (2007). Some Efficient Estimators of the Domain Parameters, Statistics & Probability Letters, Volume 77(7), 704-709.

Chaudhary, M. K., Singh, V. K., Shukla, R. K. (2012). Combined-Type Family of Estimators of Population Mean in Stratified Random Sampling under Non-response. *Journal of Reliability and Statistical Studies (JRSS),* Vol. 5(2), 133-142.

Chhikara, R. and Sud, U. (2009). Estimation of population and domain totals under two-phase sampling in the presence of non-response. *Journal of the Indian Society of Agricultural Statistics.* Vol. 63.

Clement, E. P., Udofia, G. A. and Enang, E. I. (2014). Estimation for domain in stratified sampling design in the presence of non-response. *American Journal of Mathematics & Statistics.* 4(2), 65-71.

Cochran, W. G. (1977). *Sampling Techniques,* 3^{rd} Edition. New York: John Wiley & Sons, Inc.

Deville, J. C., and Sarndal, C. E., (1992). Calibration estimators in survey sampling. *Journal of American Statistics Association.* 87, 376-382.

Durbin, J. (1958). Sampling theory for estimates based on fewer indivisuals than the number selected. Bull. Int. Stat. 36(3) 113-s119.

Hansen, M. JH. and Hurwitz, W. N. (1946). The Problem of non-response in sample surveys. *Journal of American Statistics Association.* 41, 517-529.

Hartley, H. O. (1959). Analytical studies of survey data. Rome: *Instituto Di Statistica.*

Hively, W. (1974). Introduction to domain-reference testing. *Educational TEchnology,* 14(6), 5-10. https://www.learntechlib.org/p/162913/

Ikot, E. E. and Iseh, M. J. (2024). Calibration for Efficiency of Ratio Estimator in Domains of Study with Sub-sampling the Nonrespondents. *Biom Biostat Int. J.* 13(2): 42-50. DOI: 10.1540s6/bbij.2024.13.00413.

Iseh, M. J. and Bassey, K. J. (2021a). A new calibration estimator of population mean for small area with non-response. *Asian Journal of Probability and statistics.* 12(2): 14:51.

Iseh, M. J. and Bassey, K. J. 2021b). Calibration estimator for population mean in small sample size with Non-response. European Journal of Statistics and Probability. Vol. 9(1), 32-42.

Iseh, M. and Enang, E. (2021). A Calibrated Synthetic Estimator for Small Area Estimation Statistics in Transition New Series Vol. 22(3): 15-30. DOI - 1021307/stattrans-2021-025.

Iseh, M. J. and Bassey M. O. (2022). Calibration estimators for population mean with sub-sampling the non-respondents under stratified sampling. *Science Journal of Applied Mathematics and Statistics.* Vol. 10(4), 45-56. DOI: 10.11s648/j.sjams.20221004.11.

Kaustav A, Sud, U. C. and Hukum C. (2014). Estimation of domain mean using two state sampling with sub-sampling of non-respondents. *Journal of the Indian Society of Agricultural Statistics.* Vol.68(1), 39-54.

Khare, B. B. and Srivastava, S. (1995). Study of conventional and alternative two-phase sampling ratio, product and regression estimators in presence of non-response. *Proc. Indian Nat. Sci. Acad.* Vol. 65, 195-203.

Khare, B. B., Ashutosh and Rai, P. K. (2021). A comparative study of a class of direct estimators for domain mean with a direct ratio estimator for domain mean using Auxiliary character. *Statistics in Transition new series,* Vol. 22(2), 189-200. DOI 10.21307/statrans-2021-023.

Kumar, S. (2015). Efficient use of auxiliary information in estimating the population ratio, product and mean in the presence of non-response. *Journal for advanced computing.* Vol. 4, 68-87.

Nitko, A. J. (1984). Item analysis appropriate for domain - referenced classroom testing. (Project Technical Report Number I) Paper presented at the Annual Meeting of the American Educational Research Association (68th, New Orleans).

Pandey, K. K. and Tikkiwal, G. C. (2010). The generalized class of composite method of estimation for crop acreage in small domain, *International Journal of Statistics and Systems.* Vol. 5(3), 319-333.

Rao, P. S.R. S. (1986). Ratio estimation with sub-sampling the non-respondents. Survey methodology, Vol. 12(2): 217-230.

Rao. J. N. K. (2003). Small area estimation. Wiley Inter-Science, John Wiley and sons. new Jersey.

Sarndal, C. E., Swensson, B. and Wretman, J. (1992). Model assisted survey sampling. Vol. 24(2), 167-191.

Tin, M. and Toe, T. (1972). Estimation of domain in multistage sampling. *Journal of American Statistical Association.* Vol. 67, 913-916.

Torabi, M. Datta, G. and Rao, J. N. K. (2009). Empirical Bayes estimation of small area means under nested error linear regression model with measurement errors in the covariates. Scandinavian Journal of Statistics. Vol. 36, 355-368.

Udofia, G. A. (2002). Estimation for Domains in Double Sampling for Probabilities Proportional to Size. Sankhya, B. Vol. 36, 355-368.

Yates, F. (1953). Sampling Methods for Censuses and Surveys. London: Charles W. Griffin.