Austrian Journal of Statistics
2025, Volume 54, 100–120.
http://www.ajs.or.at/
doi:10.17713/ajs.v54i5.2103

Estimating the Mean of Finite Population under Double Sampling Stratification in the Presence of Non-Response

Syed Muhammad Arsalan
Department of Statistics
Federal University of Minas
Gerais (UFMG) Brazil

Javid Shabbir
Department of Statistics
University of Wah, Wah Cantt
Pakistan

Alamgir
Department of Statistics
University of Peshawar
Pakistan

Abstract

Estimating finite population mean is the primary concern in many studies, particulary when the non-response occurs for some units. This study focuses in estimating the finite population mean using the auxiliary variables under double sampling stratification (DSS) in the presence of non-response which occurs on both the study and the auxiliary variables. A regression-ratio-in-ratio type exponential strategy is proposed with non-response simultaneously. Expressions for Bias and Mean Square Errors (MSE) are derived up-to-first order of approximation. MSE and percentage relative efficiency (PRE) are computed numerically using vrious real data sets. A simulation study is also conducted to verify the performance of estimators. Results indicate that the proposed estimator has the minimum MSE and the maximum PRE compared to competitor estimators. Therefore the proposed estimator is more efficient and is recommended for practical use in future.

Keywords: DSS, non-response, auxiliary variables, bias, MSE, PRE.

1. Introduction

Survey sampling plays a vital role in nearly every field, including economic, financial, biological and social sciences. It is a valid method for selecting a representative unit from the population of interest. Whenever the population is large, obtaining information from every unit can be costly and time-consuming. To address this issue effectively, sample surveys are used as an alternative. Once the sample is selected, it can be used to estimate population parameters such as population mean, median, mode, and standard deviation etc. In the present study, the issue of non-response is addressed under the framework of double sampling stratification

(DSS). Under this strategy, a large sample is selected randomly from a population, and second sample is selected from the first sample or selecting a sample independently. The main objective of this approach is to provide reliable estimates of the population parameters. DSS is particularly effective when it is significantly cheaper and faster to collect data on the auxiliary variable.

Under DSS, non-response is serious, particularly when estimating population parameters and it occurs in almost all research studies. Ideally information should be available for all selected units in a sample, but this is often not the case. Moreover, non-response can occur on some units on both the study and the auxiliary variables. In some cases, full information may be available on the study variable but not on the auxiliary variable, or vice versa. The problem of non-response was first addressed by Hansen and Hurwitz (1946) in the context of estimating the population mean. They suggested that a fraction of those who did not response to the mail questionnaire could be contacted through personal interviews. This approach has since been adopted by many researchers in ratio and regression methods of estimation.

Several authors have addressed the issue of non-response under different situations. For instance Rao (1983) proposed a new class of unbiased product estimators. Rao (1986) introduced a ratio estimator for situations involving non-response on the outcome variable but full response on the auxiliary variable. Cochran (1940) estimated the yield of cereal trials by using the ratio of grain to total production. The issue non-response in mean estimation was tackled by Singh and Kumar (2010) through two-phase sampling. Similarly, Singh and Kumar (2008) developed a regression-based method to estimate the finite population mean under non-response. In the related study, Singh and Nigam (2020a) dealt to estimate the population mean by employing the auxiliary characteristics under non-response scenarios. Ratio method of estimation under non-response using the DSS model was addressed by Tabasum and Khan (2004) while Tabasum and Khan (2006) proposed a new estimator to enhance the performance of the estimator. Additionally, Khare and Srivastava (1993) estimated the population mean in the presence of non-response using the auxiliary attributes. The transformed ratio estimator in the occurrence of non-response for mean estimation was implemented by Khare and Srivastava (1997). Kiregyera (1984) incorporated two auxiliary variables in a regression-type estimator using the DSS model for finite population estimation. The product-type estimator was utilized by Murthy (1964).Moreover, the contributions of human population study to sampling theory was highlighted in the work of Neyman (1938). Meanwhile, Okafor and Lee (2000) applied DSS model in regression and ratio estimators under the scenario of subsampling non-respondents. Additionally, under DSS, finite population mean was estimated by Shabbir, Arsalan, and Kim (2025). Furthermore, Batool, Ali, Mohsin, Masmoudi, Kartal, and Satti (2024) explored the forecasting robustness of machine learning and statistical models across diverse and complex climate features.

In a recent contribution, Sun, Haneuse, Levis, Lee, Arterburn, Fischer, Shortreed, and Mukherjee (2025) applied the proposed approach to electronic health records (EHR) data to compare the long-term BMI reduction effects of two bariatric surgeries, emphasizing the estimation of weighted quantile treatment effects in the presence of missing outcome data via double sampling. Likewise, Pachori and Garg (2025) presented refined estimators for the population mean applicable to stratified and stratified double sampling methods. These were assessed through simulation studies involving both empirical and synthetic datasets and compared with estimators proposed by Tracy, Singh, and Arnab (2003) and Garg and Pachori (2020). Singh and Nigam (2020b) developed a ratio-in-ratio type exponential estimator under DSS to estimate the mean of a finite population, while for median estimation under DSS, estimators were utilized by Singh, Joarder, and Tracy (2001).Singh and Vishwakarma (2007) used ratio and product type estimators on the lines of Bahl and Tuteja (1991) under the presence of non-response.

In this study, different estimators are modified under DSS to account for non-response affecting both the outcome and the auxiliary variables. We compare the Bias and Mean square errors (MSE) of proposed estimator with various existing estimators under non-response conditions including the usual estimator developed by Hansen and Hurwitz (1946) classical ratio and product estimators under DSS suggested by Tripathi and Ige (1987) and the ratio-exponential and product-exponential estimators developed by Tailor, Chouhan, and Kim (2014) under DSS. Additionally, the recently developed estimators, suggested by Gupta and Tailor (2021) in DSS are modified to incorporate the effect of non-response.

2. Symbols and notations

This section contains the technical procedure in which the problem of non-response is addressed under the strategy of double sampling stratification. In this procedure, we consider two random samples from the finite population in two phases. The size of the first phase sample, larger is denoted by n' and the size of second phase small sample, which is a stratified sub sample is denoted by n drawn from the first stage large sample. The problem of non-response is considered to exist in both the main outcome (y) and the auxiliary variable (x).

Consider a finite population. Let y_i be the i^{th} observation on outcome y, where (x_i, z_i) be the i^{th} observations of the auxiliary variables (x, z), respectively. Let population of size N is categorized into data homogenous subgroups called strata. So that h^{th} stratum consists of N_h units where $h = 1, 2, .., L$ such that $\sum\limits_{h=1}^{L} N_h = N$. It is assumed that population size consist of two groups called respondent and non-respondent group. Let $N_{(1)h}$ be respondent and $N_{(2)h} = (N_h - N_{(1)h})$ be non-responding units in h^{th} stratum respectively.

- A larger sample of size n'_h is drawn from the population using simple random sampling without replacement (SRSWOR).
- Larger sample of size is distributed into L strata of units n'_h in the h^{th} stratum.
- Following the stratification, a sample of size n_h units is selected from each h^{th} stratum forming the DSS sample of size n where f is the sample fraction between larger sample n' and population N.
- The sample of size n_h of each stratum contains $n_{(1)h}$ respondent and $n_{(2)h}$ non-respondent units. Then we select a sub-sample of size $r_h = \frac{n_{(2)h}}{k_h} (k_h > 1)$ from the non-response units $n_{(2)h}$ in the h^{th} stratum non-response.
- $W_h = \frac{N_h}{N}$ is the stratum weight, where N_h is the total number of units in the h^{th} stratum. Here v_h is the ratio between the DSS sample of size n_h stratum and first stage large sample of n'_h stratum.

For Instance,

$$n = \sum_{h=1}^{L} n_h, \quad n' = \sum_{h=1}^{L} n'_h, \quad n_h = v_h n'_h \; 0 < v_h < 1, \quad f = \frac{n'_h}{N'_h} \; h = 1, 2, 3, ..., L,$$
$$w_{(2)h} = \frac{N_{(2)h}}{N}, \quad v_h = \frac{n_h}{n'_h}, \quad R_1 = \frac{\overline{Y}}{\overline{X}}, \quad R_2 = \frac{\overline{Y}}{\overline{Z}},$$

The population means, variances and the co-variances of the outcome variable y and the auxiliary variables x, z can be written and obtain as follows.

$$\bar{Y} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} y_{hi}, \quad \bar{X} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} x_{hi}, \quad \bar{Z} = \frac{1}{N} \sum_{h=1}^{L} \sum_{i=1}^{N_h} z_{hi}, \quad S_y^2 = \frac{1}{N-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left(y_{hi} - \bar{Y} \right)^2,$$
$$S_x^2 = \frac{1}{N-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left(x_{hi} - \bar{X} \right)^2, \quad S_z^2 = \frac{1}{N-1} \sum_{h=1}^{L} \sum_{i=1}^{N_h} \left(z_{hi} - \bar{Z} \right)^2, \quad S_{xh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} \left(x_{hi} - \bar{X}_h \right)^2,$$
$$S_{yh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} \left(y_{hi} - \bar{Y}_h \right)^2, \quad S_{zh}^2 = \frac{1}{N_h-1} \sum_{i=1}^{N_h} \left(z_{hi} - \bar{Z}_h \right)^2,$$

$S_{yxh} = \frac{1}{N_h-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)\left(x_{hi}-\bar{X}_h\right), \quad S_{yzh} = \frac{1}{N_h-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(y_{hi}-\bar{Y}_h\right)\left(z_{hi}-\bar{Z}_h\right),$
$S_{xzh} = \frac{1}{N_h-1}\sum_{h=1}^{L}\sum_{i=1}^{N_h}\left(x_{hi}-\bar{X}_h\right)\left(z_{hi}-\bar{Z}_h\right).$

Also the means, variances, standard deviations and co-variances of non-response are:

$\bar{Y}_{(2)h} = \frac{1}{N_{(2)h}}\sum_{i=N_1+1}^{N_{(2)h}} y_{hi}, \quad \bar{X}_{(2)h} = \frac{1}{N_{(2)h}}\sum_{i=N_{(1)h}+1}^{N_{(2)h}} x_{hi}, \quad S^2_{y(2)h} = \frac{1}{N_{(2)h}-1}\sum_{i=N_{(1)h}+1}^{N_{(2)h}}\left(y_{hi}-\bar{Y}_{(2)h}\right)^2,$
$S^2_{x(2)h} = \frac{1}{N_{(2)h}-1}\sum_{i=N_{(1)h}+1}^{N_{(2)h}}\left(x_{hi}-\bar{X}_{(2)h}\right)^2, \quad S_{yx(2)h} = \frac{1}{N_{(2)h}-1}\sum_{i=N_{(1)h}+1}^{N_{(2)h}}\left(y_{hi}-\bar{Y}_{(2)h}\right)\left(x_{hi}-\bar{X}_{(2)h}\right),$
$S_{yz(2)h} = \frac{1}{N_{(2)h}-1}\sum_{i=N_{(1)h}+1}^{N_{(2)h}}\left(y_{hi}-\bar{Y}_{(2)h}\right)\left(z_{hi}-\bar{Z}_{(2)h}\right) \quad S_{xz(2)h} = \frac{1}{N_{(2)h}-1}\sum_{i=N_{(1)h}+1}^{N_{(2)h}}\left(x_{hi}-\bar{X}_{(2)h}\right)$
$\left(z_{hi}-\bar{Z}_{(2)h}\right).$

The following error terms are:

$E({\xi_0}^*) = E({\xi_1}^*) = E({\xi_1}') = E({\xi_2}') = 0, \quad E(\xi_0^{*2}) = \frac{1}{\bar{Y}^2}[\Delta' S_y^2 + \Delta_h^* S^2_{y(2)h} + \Delta_h' S^2_{yh}],$ $E(\xi_1^{*2}) = \frac{1}{\bar{X}^2}[\Delta' S_x^2 + \Delta_h^* S^2_{x(2)h} + \Delta_h' S^2_{xh}], \quad E(\xi_1'^2) = \frac{\Delta'}{\bar{X}^2}S_x^2, \quad E(\xi_2'^2) = \frac{\Delta'}{\bar{Z}^2}S_z^2, \quad E({\xi_0}^*{\xi_1}^*) =$ $\frac{1}{\bar{X}\bar{Y}}[\Delta' S_{yx} + \Delta_h^* S_{yx(2)h} + \Delta_h' S_{yxh}], \quad E({\xi_0}^*{\xi_1}') = \frac{\Delta'}{\bar{X}\bar{Y}}S_{yx}, \quad E({\xi_1}^*{\xi_1}') = \frac{\Delta'}{\bar{X}^2}S_x^2, \quad E({\xi_0}^*{\xi_2}') =$ $\frac{\Delta'}{\bar{X}\bar{Z}}S_{yz}, \quad E({\xi_1}^*{\xi_2}') = \frac{\Delta'}{\bar{X}\bar{Z}}S_{xz}, \quad E({\xi_1}'{\xi_2}') = \frac{\Delta'}{\bar{X}\bar{Z}}S_{xz},$

Where,

$\Delta' = \left(\frac{1-f}{n'}\right), \quad \Delta_h^* = \frac{W_{(2)h}(k_h-1)}{n_h} \quad and \quad \Delta_h' = \frac{1}{n'}\sum_{h=1}^{L} w_h\left(\frac{1}{v_h}-1\right).$

Some of the existing estimators Hansen and Hurwitz (1946), Tripathi and Ige (1987), and Tailor *et al.* (2014), are reproduced and then modified by using the case of non-response which occurred on both the outcome variable y and the auxiliary variable x. Recently Gupta and Tailor (2021) developed the estimator under DSS in which we implement the technique of non-response. Also the biases and MSE are obtained up to first order of approximation. The existing estimators are as under:

- The usual unbiased estimator in DSS under the situation, where the non response are occurred on the study variable:
$$U_{ds}^* = \sum_{h=1}^{L} w_h \bar{y}_h^* \tag{1}$$
The variance of U_{ds}^* can be obtained as:
$$Var(U_{ds}^*) = \Delta' S_y^2 + \Delta_h^* S^2_{y(2)h} + \Delta_h' S^2_{yh} \tag{2}$$
- The ratio estimator by Tripathi and Ige (1987) in DSS under the presence of non-response on study variable y and auxiliary variable x is defined as:
$$R_{ds}^* = \bar{y}_{ds}^* \frac{\bar{x}'}{\bar{x}_{ds}^*} \tag{3}$$
The bias and MSE of R_{ds}^* are,
$$Bias(R_{ds}^*) = \frac{\Delta_h^*[R_1 S^2_{x(2)h} - S_{yx(2)h}] + \Delta'_h[R_1 S^2_{xh} - S_{yxh}]}{\bar{X}} \tag{4}$$
and
$$MSE(R_{ds}^*) \cong \Delta' S_y^2 + \Delta_h^*[S^2_{y(2)h} + R_1^2 S^2_{x(2)h} - 2R_1 S_{yx(2)h}] + \Delta_h'[S^2_{yh} + R_1^2 S^2_{xh} - 2R_1 S_{yxh}] \tag{5}$$

- The product estimator by Tripathi and Ige (1987) in DSS under the presence of non-response on the study variable y and the auxiliary variable x is defined as:

$$P^*_{ds} = \bar{y}^*_{ds}\frac{\bar{x}^*_{ds}}{\bar{x}'} \tag{6}$$

Similarly, the Bias and MSE for P^*_{ds} are obtained as,

$$Bias(P^*_{ds}) = \frac{[\Delta^*_h S_{yx(2)h} + \Delta'_h S_{yxh}]}{\bar{X}} \tag{7}$$

and,

$$MSE(P^*_{ds}) \cong \Delta' S^2_y + \Delta^*_h[S^2_{y(2)h} + R_1^2 S^2_{x(2)h} + 2R_1 S_{yx(2)h}] + \Delta'_h[S^2_{yh} + R_1^2 S^2_{xh} + 2R_1 S_{yxh}] \tag{8}$$

- The ratio-exponential estimator in DSS by Tailor *et al.* (2014) under non-response can be defined as:

$$R_{Exp^*_{ds}} = \bar{y}^*_{ds}\exp\left[\frac{\bar{x}' - \bar{x}^*_{ds}}{\bar{x}' + \bar{x}^*_{ds}}\right] \tag{9}$$

Expressing $R_{Exp^*_{ds}}$ in error terms, we have,

$$R_{Exp^*_{ds}} = \bar{Y}(1+\xi_0{}^*)\exp\left[\frac{\bar{X}(1+\xi_1{}') - \bar{X}(1+\xi_1{}^*)}{\bar{X}(1+\xi_1{}') + \bar{X}(1+\xi_1{}^*)}\right] \tag{10}$$

The estimator $R_{Exp^*_{ds}}$ can be expressed up-to first order of approximation as:

$$R_{Exp^*_{ds}} - \bar{Y} = \bar{Y}\left[\xi_0{}^* + \frac{\xi_1{}' - \xi_1{}^*}{2} + \frac{3\xi_1{}^{*2} - \xi_1'^2 - 2\xi_1{}'\xi_1{}^* + 4\xi_0{}^*\xi_1{}' - 4\xi_0{}^*\xi_1{}^*}{8}\right] \tag{11}$$

The bias and MSE of $R_{Exp^*_{ds}}$ can be obtained as,

$$Bias(R_{Exp^*_{ds}}) = \frac{\Delta^*_h[3R_1 S^2_{x(2)h} - 4S_{yx(2)h}] + \Delta'_h[3R_1 S^2_{xh} - 4S_{yxh}]}{8\bar{X}} \tag{12}$$

and

$$\begin{aligned} \text{MSE}(R_{Exp^*_{ds}}) \cong \Delta' S^2_y + \Delta^*_h\left[S^2_{y(2)h} + \frac{R_1^2 S^2_{x(2)h}}{4} - R_1 S_{yx(2)h}\right] \\ + \Delta'_h\left[S^2_{yh} + \frac{R_1^2 S^2_{xh}}{4} - R_1 S_{yxh}\right] \end{aligned} \tag{13}$$

- The product-exponential estimator in DSS of Tailor *et al.* (2014) under non-response can be defined as:

$$P_{Exp^*_{ds}} = \bar{y}^*_{ds}\exp\left[\frac{\bar{x}^*_{ds} - \bar{x}'}{\bar{x}^*_{ds} + \bar{x}'}\right] \tag{14}$$

In terms of error, we have

$$P_{Exp^*_{ds}} = \bar{Y}(1+\xi_0{}^*)\exp\left[\frac{\bar{X}(1+\xi_1{}^*) - \bar{X}(1+\xi_1{}')}{\bar{X}(1+\xi_1{}^*) + \bar{X}(1+\xi_1{}')}\right] \tag{15}$$

Therefore upto the terms of order n^{-1} we have,

$$P_{Exp^*_{ds}} - \bar{Y} = \bar{Y}\left[\xi_0{}^* + \frac{\xi_1{}^* - \xi_1{}'}{2} + \frac{3\xi_1'^2 - \xi_1^{*2} - 2\xi_1{}^*\xi_1{}' + 4\xi_0{}^*\xi_1{}^* - 4\xi_0{}^*\xi_1{}'}{8}\right] \tag{16}$$

The bias and MSE of $P_{Exp^*_{ds}}$ are given by

$$Bias(P_{Exp^*_{ds}}) = \frac{\Delta^*_h[4S_{yx(2)h} - R_1 S^2_{x(2)h}] + \Delta'_h[4S_{yxh} - R_1 S^2_{xh}]}{8\bar{X}} \tag{17}$$

and

$$\begin{aligned}\mathrm{MSE}(P_{Exp^*_{ds}}) \cong \Delta' S_y^2 + \Delta_h^*\left[S_{y(2)h}^2 + \frac{R_1^2 S_{x(2)h}^2}{4} + R_1 S_{yx(2)h}\right] \\ + \Delta'_h\left[S_{yh}^2 + \frac{R_1^2 S_{xh}^2}{4} + R_1 S_{yxh}\right]\end{aligned} \tag{18}$$

- The ratio in ratio exponential estimator which were recently developed by Gupta and Tailor (2021) under the situation of non-response can be defined as:

$$RR_{Exp^*_{ds}} = \bar{y}^*_{ds}\exp\left[\frac{\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right) - \bar{x}^*_{ds}}{\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right) - \bar{x}^*_{ds}}\right] \tag{19}$$

The estimator $RR_{Exp^*_{ds}}$ can be expressed in error terms we have:

$$RR_{Exp^*_{ds}} = \bar{Y}(1+\xi_0{}^*)\exp\left[\frac{\bar{X}(1+\xi_1{}')\left(\frac{\bar{Z}}{\bar{Z}(1+\xi_2{}')}\right) - \bar{X}(1+\xi_1{}^*)}{\bar{X}(1+\xi_1{}')\left(\frac{\bar{Z}}{\bar{Z}(1+\xi_2{}')}\right) + \bar{X}(1+\xi_1{}^*)}\right] \tag{20}$$

Up-to first degree of approximation we have,

$$\begin{aligned}RR_{Exp^*_{ds}} - \bar{Y} = \bar{Y}\Big[\xi_0{}^* + \frac{1}{2}\xi_1{}' - \xi_2' - \xi_1^* + \frac{1}{8}(3\xi_1{}^{*2} - \xi_1{}'^2 + 3\xi_2{}'^2 - 4\xi_0{}^*\xi_1{}^* + 4\xi_0{}^*\xi_1{}' \\ - 4\xi_0{}^*\xi_2{}' - 2\xi_1{}^*\xi_1{}' + 2\xi_1{}^*\xi_2{}' - 2\xi_1{}'\xi_2{}')\Big]\end{aligned} \tag{21}$$

The bias and MSE of $RR_{Exp^*_{ds}}$ can be easily obtained as:

$$\begin{aligned}Bias(RR_{Exp^*_{ds}}) = \frac{\Delta'[3R_2 S_z^2 - 4S_{yz}]}{8\bar{Z}} + \frac{1}{8\bar{X}}\Big[\Delta_h^*[3R_1 S_{x(2)h}^2 - 4S_{yx(2)h}] \\ +\Delta_h'[3R_1 S_{xh}^2 - 4S_{yxh}]\Big]\end{aligned} \tag{22}$$

and

$$\begin{aligned}MSE(RR_{Exp^*_{ds}}) \cong \Delta'[S_y^2 + \frac{R_2^2 S_z^2}{4} - R_2 S_{yz}] + \Delta_h^*[S_{y(2)h}^2 + \frac{R_1^2 S_{x(2)h}^2}{4} - R_1 S_{yx(2)h}] \\ + \Delta_h{}'[S_{yh}^2 + \frac{R_1^2 S_{xh}^2}{4} - R_1 S_{yxh}]\end{aligned} \tag{23}$$

3. Proposed estimator

In this section, we have proposed an estimator under DSS in the presence of non-response for estimating the mean of finite population. The problem of non-response is a serious issue and can occurr on the study variable (y) and the auxiliary variable (x) respectively. The bias and mean square error (MSE) of the proposed estimator have been derived up to first order of approximation. The proposed estimator is given by:

$$Prop^*_{ds} = [\bar{y}^*_{ds} + d(\bar{x}' - \bar{x}^*_{ds})]\exp\left[\frac{\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right) - \bar{x}^*_{ds}}{\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right) - \bar{x}^*_{ds}}\right] \tag{24}$$

Now expressing the $Prop^*_{ds}$ in terms of ξ_{i_s}, we have

$$\begin{aligned}\mathrm{Prop}^*_{ds} = \Big[\bar{Y}(1+\xi_0{}^*) + d(\bar{X}(1+\xi_1{}') - \bar{X}(1+\xi_1^*))\Big] \times \\ \exp\left[\frac{\bar{X}(1+\xi_1{}')\left(\frac{\bar{Z}}{\bar{Z}(1+\xi_2{}')}\right) - \bar{X}(1+\xi_1{}^*)}{\bar{X}(1+\xi_1{}')\left(\frac{\bar{Z}}{\bar{Z}(1+\xi_2{}')}\right) + \bar{X}(1+\xi_1{}^*)}\right]\end{aligned} \tag{25}$$

Keeping the errors terms upto the terms of order n^{-1}, the estimator Prop^*_{ds} can be expressed as:

$$\text{Prop}^*_{ds} - \bar{Y} = \bar{Y}\Big[\frac{1}{2}(2\xi_0^* + \xi_1{}' - \xi_2{}' - \xi_0^*) + \frac{d\bar{X}}{8\bar{Y}}(4\xi_1^{*2} + 4\xi_1'^2 - 4\xi_1{}^*\xi_1{}' - 4\xi_1{}'\xi_2{}' - 4\xi_1{}^*\xi_1{}' + 4\xi_1{}^*\xi_2{}') + \frac{d\bar{X}}{\bar{Y}}(\xi_1{}' - \xi_1{}^*) + \frac{1}{8}(3\xi_1^{*2} - \xi_1'^2 + 3\xi_2'^2 + 4\xi_0{}^*\xi_1{}' - 4\xi_0{}^*\xi_1{}^* - 4\xi_0{}^*\xi_2{}' - 2\xi_1{}^*\xi_1{}' + 2\xi_1{}^*\xi_2{}' - 2\xi_1{}^*\xi_2{}')\Big] \tag{26}$$

Solving eq (26) for obtaning the bias of $Prop^*_{ds}$, we have

$$E(\text{Prop}^*_{ds} - \bar{Y}) = \bar{Y}E\Big[\frac{1}{8}(3\xi_1^{*2} - \xi_1'^2 + 3\xi_2'^2 + 4\xi_0{}^*\xi_1{}' - 4\xi_0{}^*\xi_1{}^* - 4\xi_0{}^*\xi_2{}' - 2\xi_1{}^*\xi_1{}' + 2\xi_1{}^*\xi_2{}' - 2\xi_1{}'\xi_2{}') + \frac{d\bar{X}}{8\bar{Y}}(4\xi_1^{*2} + 4\xi_1'^2 - 4\xi_1{}^*\xi_1{}' - 4\xi_1{}'\xi_2{}' - 4\xi_1{}^*\xi_1{}' + 4\xi_1{}^*\xi_2{}')\Big] \tag{27}$$

Now bias is given by

$$\begin{aligned}\text{Bias}(\text{Prop}^*_{ds}) = {} & \frac{\Delta'[3R_2S_z^2 - 4S_{yz}]}{8\bar{Z}} + \frac{\Delta'_h[3R_1S_{xh}^2 - 4S_{yxh}]}{8\bar{X}} \\ & + \frac{\Delta^*[3R_1S_{x(2)}^2 - 4S_{yx(2)}]}{8\bar{X}} + \frac{\Delta'_h[dS_{xh}^2]}{2\bar{X}} + \frac{\Delta^*[dS_{x(2)}^2]}{2\bar{X}}\end{aligned} \tag{28}$$

Similarly, ignoring higher powers more than two, taking square and apply expectation on both sides of eq (27) to get MSE of $Prop^*_{ds}$:

$$E(Prop^*_{ds} - \bar{Y})^2 = \bar{Y}^2E\left[\frac{1}{2}(2\xi_0^* + \xi_1{}' - \xi_2{}' - \xi_1^*) + \frac{d\bar{X}}{\bar{Y}}(\xi_1{}' - \xi_1^*)\right]^2 \tag{29}$$

Finally, MSE of $Prop^*_{ds}$ is obtained as:

$$\begin{aligned}MSE(Prop^*_{ds}) = {} & \Delta'[S_y^2 + \frac{R_2^2S_z^2}{4} - R_2S_{yz}] + \Delta'_h[S_{yh}^2 + \frac{R_2^2S_{xh}^2}{4} - R_1S_{yxh}] + \Delta^*_h[S_{y(2)h}^2 + \\ & \frac{R_2^2S_{x(2)h}^2}{4} - R_1S_{yx(2)h}] + \Delta'_h[d^2S_{xh}^2] + \Delta^*_h[d^2S_{x(2)h}^2] - 2\Delta'_h[dS_{yxh}] \\ & - 2\Delta^*_h[dS_{yx(2)h}] + \Delta'_h[dR_1S_{xh}^2] + \Delta^*_h[dR_1S_{x(2)h}^2]\end{aligned} \tag{30}$$

where,

$$d_{opt} = \frac{\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta^*_h\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)}{\Delta_h{}'S_{xh}^2 + \Delta^*_hS_{x(2)h}^2}$$

The MSE of $Prop^*_{ds}$ at the optimum value of data is as follows

$$\begin{aligned}MSE_{\min}(Prop^*_{ds}) \cong {} & \Delta'[S_y^2 + \frac{R_2^2S_z^2}{4} - R_2S_{yz}] + \Delta'_h[S_{yh}^2 + \frac{R_2^2S_{xh}^2}{4} - R_1S_{yxh}] + \Delta^*_h[S_{y(2)h}^2 + \\ & \frac{R_2^2S_{x(2)h}^2}{4} - R_1S_{yx(2)h}] - \\ & \frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta^*_h\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta^*_hS_{x(2)h}^2}.\end{aligned} \tag{31}$$

4. Theoritical comparison

1. $MSE_{\min}(Prop^*_{ds}) < Var(U^*_{ds})$ if,

$$\Delta'[R_2S_{yz} - \frac{R_2^2S_z^2}{4}] + \Delta_h'[R_1S_{yxh} - \frac{R_1^2S_{xh}^2}{4}] + \Delta_h^*[R_1S_{yx(2)h} - \frac{R_1^2S_{x(2)h}^2}{4}] + \frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta_h^*\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta_h^*S_{x(2)h}^2} > 0 \tag{32}$$

2. $MSE_{\min}(Prop^*_{ds}) < MSE(R^*_{ds})$ if,

$$\Delta'[R_2S_{yz} - \frac{R_2^2S_z^2}{4}] + \Delta_h'[\frac{3R_1^2S_{xh}^2}{4} - R_1S_{yxh}] + \Delta_h^*[\frac{3R_1^2S_{x(2)h}^2}{4} - R_1S_{yx(2)h}] + \frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta_h^*\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta_h^*S_{x(2)h}^2} > 0 \tag{33}$$

3. $MSE_{\min}(Prop^*_{ds}) < MSE(P^*_{ds})$ if,

$$\Delta'[R_2S_{yz} - \frac{R_2^2S_z^2}{4}] + \Delta_h'[\frac{3R_1^2S_{xh}^2}{4} + 3R_1S_{yxh}] + \Delta_h^*[\frac{3R_1^2S_{x(2)h}^2}{4} + 3R_1S_{yx(2)h}] + \frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta_h^*\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta_h^*S_{x(2)h}^2} > 0 \tag{34}$$

4. $MSE_{\min}(Prop^*_{ds}) < MSE(R_{Exp^*_{ds}})$ if,

$$\Delta'[R_2S_{yz} - \frac{R_2^2S_z^2}{4}] + \frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta_h^*\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta_h^*S_{x(2)h}^2} > 0 \tag{35}$$

5. $MSE_{\min}(Prop^*_{ds}) < MSE(P_{Exp^*_{ds}})$ if,

$$\Delta'[R_2S_{yz} - \frac{R_2^2}{4}S_z^2] + 2\Delta_h'R_1S_{yxh} + 2\Delta_h^*R_1S_{yx(2)h} + \frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta_h^*\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta_h^*S_{x(2)h}^2} > 0 \tag{36}$$

6. $MSE_{\min}(Prop^*_{ds}) < MSE(RR_{Exp^*_{ds}})$ if,

$$+\frac{\left[\Delta_h{}'\left(S_{yxh} - \frac{R_1S_{xh}^2}{2}\right) + \Delta_h^*\left(S_{yx(2)h} - \frac{R_1S_{x(2)h}^2}{2}\right)\right]^2}{\Delta_h{}'S_{xh}^2 + \Delta_h^*S_{x(2)h}^2} > 0 \tag{37}$$

5. Numerical illustration and simulation study

To evaluate the theoretical performance of proposed and existing estimators, the data based on Populations I and II are given in Tables 1 and 2. MSEs values are given in Tables 3, 4 and

5 under different values of k PRE-values are given in Tables 6, 7 and 8. Simulation results are given in Tables 9 and 10.

Population I Source: (Singh 2003)

Let y be the outcome: the number of fish caught in 1995 and x: number of fish caught in 1994 and z: number of fishes caught in 1993 be the auxiliary variables respectively, x is stratified as: $32 \leq x_1 \leq 993$, $994 \leq x_2 \leq 2173$, $2174 \leq x_3 \leq 4929$, and $4930 \leq x_4 \leq 38007$ for using the proportional allocation $n_h = n\left(\frac{N_h}{N}\right)$ and obtained the sample sizes, which are in Table 1 below.

Table 1: Summary statistics of Population I

	Stratum$_h$			
Parameter	1	2	3	4
N_h	18	18	17	16
n'_h	15	15	13	13
n_h	12	12	11	10
$\bar{Y}_h$	472.7778	1453.667	3930.294	13127.31
$\bar{X}_h$	509	1553.444	3965.824	14832.06
$\bar{Z}_h$	574.7778	1778	3773.176	13143.12
S_{yh}	360.41	422.8567	1022.806	7537.544
S_{xh}	339.2797	397.934	978.1423	9043.681
S_{zh}	6315.21	830.3542	1537.887	8348.291
S_{yxh}	101840.7	110424.6	732165.7	60923018
S_{yzh}	95027.48	152164.4	557113.3	57207317
S_{xzh}	79857.88	205439.7	496194.6	71834783
S_y^2	37199578			
S_x^2	49829270			
S_z^2	39881874			
$W_2 = 10\%$ **Non-Response**				
$S_{y(2)h}$	36.062	36.062	81.317	7.778
$S_{x(2)h}$	7.778	7.778	45.25	8.485
$S_{yx(2)h}$	1300.5	280.5	-3680	-66
$W_2 = 20\%$ **Non-Response**				
$S_{y(2)h}$	58.8387	58.838	69.787	41.073
$S_{x(2)h}$	44.093	44.0936	40.857	29.484
$S_{yx(2)h}$	3462	-2319.33	-2844.66	1142
$W_2 = 30\%$ **Non-Response**				
$S_{y(2)h}$	51.983	51.9836	162.62	383.362
$S_{x(2)h}$	54.755	54.755	59.1540	58.453
$S_{yx(2)h}$	2702.3	-1335.85	-7288.8	12641.5

Population II Source: (Kadilar and Cingi 2005)

Let y denote the outcome of apple production amount in 1999: x denote the apple production amount in 1998: and z denote the number of trees in 1999 from 204 villages in black

Table 2: Summary statistics of Population II

	Stratum$_h$					
Parameter	1	2	3	4	5	6
N_h	106	106	85	171	204	173
n'_h	77	71	53	109	136	111
n_h	39	37	29	73	91	65
$\bar{Y}_h$	1536.77	2212.59	9384.3	5588.01	966.95	404.39
$\bar{X}_h$	24711.81	26840.04	72723.76	73191.2	26833.75	9903.3
$\bar{Z}_h$	24375.59	27421.7	72409.95	74364.68	26441.72	9843.82
S_{yh}	6425.08	11551.53	29907.48	28643.42	2389.77	945.74
S_{xh}	49134.76	53978.71	161109.5	262495.6	45174.26	18977.28
S_{zh}	49189.08	57460.61	160757.3	285603.1	45402.78	18793.96
S_{yxh}	257508714	521230984	4323018145	7379640297	76727855	15609193
S_{yzh}	257778692	568176176	4332446622	8065108356	77372777	15883145
S_{xzh}	2415109169	3097324685	25875782559	74820940586	2042075789	354134505
S_y^2	292606036					
S_x^2	18384575563					
S_z^2	20965219913					
$W_2 = 10\%$ **Non-Response**						
$S_{y(2)h}$	2063.54	2063.548	175.73	28456.04	2756.02	827.9202
$S_{x(2)h}$	33216.81	33216.81	4306.80	102881.5	44873.26	14855.76
$S_{yx(2)h}$	4258229	68137801	661814.3	2922247655	105979309	10107195
$W_2 = 20\%$ **Non-Response**						
$S_{y(2)h}$	1594.56	1594.56	277.43	20097.87	2004.98	19812.84
$S_{x(2)h}$	31257.31	31257.31	27053.91	76937.31	34691.78	75818.2
$S_{yx(2)h}$	2542622	44134399	6753340	1454729421	59649102	1416064014
$W_2 = 30\%$ **Non-Response**						
$S_{y(2)h}$	1338.85	1338.85	240.08	16514.17	15073.96	16352.59
$S_{x(2)h}$	31582.87	31582.87	22072.73	69482.71	61999.6	68895.42
$S_{yx(2)h}$	1792531	32956184	4585109	1018930092	840738153	998242499

sea region of turkey. Description of the overall population are: $N = 854$, $\bar{Y} = 2930.126$, $\bar{X} = 37600.12$, $\bar{Z} = 37474.92$, while the detail description by stratum is provided in Table 2.

Table 3 compares the mean square error (MSE) values for different estimators, including the proposed estimator, under two population scenarios (Population I and Population II) with a 10% non-response rate. Each population is assessed across varying values of k (from $k = 2$ to $k = 5$). For both populations and across all k values, the proposed estimator consistently shows the lowest MSE compared to other estimators. In Population I, the MSE values of the proposed estimator ranges from 50742.60 at $k = 2$ to 51556.55 at $k = 5$, significantly lower than the MSEs of other estimators. In Population II, the proposed estimator also achieves the lowest MSE, starting from 713550.55 at $k = 2$ and increasing to 2180287.76 at $k = 5$, maintaining its superior performance relative to the competing estimators. This pattern demonstrates the estimator's robustness and effectiveness even under increased stratification levels. Furthermore, the consistently lower MSE values suggest improved precision and reduced variability, making the proposed estimator a reliable choice for practical applications in the presence of non-response.

Table 3: MSE of various estimators for different values of K for Population I and II with 10% of non-response rate

	Population I $W_2 = 10\%$ **Non-Response**				**Population II** $W_2 = 10\%$ **Non-Response**			
Estimators	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
U^*_{ds}	197109.48	197291.65	197473.82	197655.99	2543374.62	4588600.35	6633826.08	8679051.82
R^*_{ds}	143047.01	143345.25	143643.50	143941.75	1344301.23	2415407.79	3486514.36	4557620.92
P^*_{ds}	421791.13	421930.75	422070.38	422210.01	4424185.86	7908847.77	11393509.67	14878171.57
$R_{Exp^*_{ds}}$	148750.84	148981.86	149212.88	149443.90	1858620.69	3358622.21	4858623.74	6358625.26
$P_{Exp^*_{ds}}$	288122.91	288274.61	288426.32	288578.03	3398563.01	6105342.20	8812121.39	11518900.59
$RR_{Exp^*_{ds}}$	59307.73	59538.75	59769.76	60000.78	1768044.58	3268046.10	4768047.62	6268049.15
$\textbf{Prop}^*_{ds}$	**50742.60**	**51014.03**	**51285.35**	**51556.55**	**713550.55**	**1209397.44**	**1696308.61**	**2180287.76**

Table 4: MSE of various estimators for different values of k for Population I and II with 20% of non-response rate

	Population I $W_2 = 20\%$ **Non-Response**				**Population II** $W_2 = 20\%$ **Non-Response**			
Estimators	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
U^*_{ds}	197166.00	197404.69	197643.38	197882.07	2383658.43	4269167.98	6154677.53	8040187.07
R^*_{ds}	143102.43	143456.09	143809.76	144163.43	1292152.96	2311111.25	3330069.55	4349027.84
P^*_{ds}	421962.76	422274.02	422585.28	422896.53	4143428.10	7347332.23	10551236.36	13755140.49
$R_{Exp^*_{ds}}$	148792.56	149065.30	149338.04	149610.77	1754372.67	3150126.18	4545879.68	5941633.18
$P_{Exp^*_{ds}}$	288222.73	288474.26	288725.79	288977.33	3180010.24	5668236.66	8156463.09	10644689.51
$RR_{Exp^*_{ds}}$	59349.45	59622.19	59894.92	60167.66	1663796.56	3059550.06	4455303.57	5851057.07
$\textbf{Prop}^*_{ds}$	**50789.69**	**51108.15**	**51426.43**	**51744.53**	**772394.46**	**1348113.40**	**1917561.94**	**2484939.97**

Table 4 displays MSE values for multiple estimators, including the proposed estimator, across two populations (Population I and Population II) with a 20% non-response rate. Each population is evaluated at increasing k values ($k = 2$ to $k = 5$). In both populations and for all k values, the proposed estimator achieves the lowest MSE. For Population I, its MSE ranges from 50789.69 at $k = 2$ to 51744.53 at $k = 5$, outperforming alternative estimators. Similarly, in Population II, the proposed estimator consistently exhibits the lowest MSE, from 772394.46 at $k = 2$ to 2484939.97 at $k = 5$, confirming its advantage across conditions. Notably, at $k = 2$ in Population I, it achieves 50789.69, outperforming the next best estimator, which records significantly higher error. This gap is even more pronounced in Population II, where its MSE remains well below competitors, demonstrating strength in high non-response rates. These consistent margins indicate both efficiency and resilience across diverse stratification levels.

Table 5: MSE of various estimators for different values of k for Population I and II with 30% of non-response rate

	Population I $W_2 = 30\%$ **Non-Response**				**Population II** $W_2 = 30\%$ **Non-Response**			
Estimators	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
U^*_{ds}	199773.24	202619.18	205465.12	208311.05	2032413.48	540902.7	5100942.67	6635207.27
R^*_{ds}	145567.68	148386.60	151205.52	154024.44	1102396.00	298687.3	2760798.66	3589999.99
P^*_{ds}	424863.17	428074.85	431286.52	434498.19	3575320.56	1218100.00	8846913.75	11482710.35
$R_{Exp^*_{ds}}$	151309.92	154100.00	156890.09	159680.18	1490793.54	365422.2	3755142.28	4887316.66
$P_{Exp^*_{ds}}$	290957.66	293944.13	296930.59	299917.06	2727255.82	825128.8	6798199.83	8833671.83
$RR_{Exp^*_{ds}}$	61866.80	64656.89	67446.98	70237.07	1400217.43	274846.1	3664566.17	4796740.54
$\mathbf{Prop}^*_{ds}$	**53269.63**	**56068.17**	**58866.68**	**61665.14**	**694598.88**	**207408.5**	**1716860.75**	**2224108.65**

Table 5 gives a detail of the MSE values for each estimator under a 30% non-response rate across varying k values. For Population I, the proposed estimator shows the lowest MSEs, starting at 53269.63 for $k = 2$ and gradually increasing to 61665.14 at $k = 5$, maintaining a consistently lower error than the alternatives. In Population II, the proposed estimator also outperforms, starting with an MSE of 694598.88 for $k = 2$ and increasing to 2224108.65 by $k = 5$. This pattern confirms that the proposed estimator delivers good performance in estimating results, particularly as k values and non-response levels increase, consistently outperforming all other estimators in terms of MSE across both populations. The proposed estimator's MSE at $k = 2$ in Population I is notably the smallest at 53269.63, a clear indicator of its effectiveness even under higher non-response rates. In Population II, the estimator maintains a distinct advantage, with the MSE rising more slowly compared to the others, peaking at 2224108.65 at $k = 5$. These results demonstrate the estimator's ability to handle increased complexity in both populations, ensuring stable and accurate estimates regardless of the rising non-response rate.

Table 6: PRE with $\boldsymbol{W_2 = 10\%}$ non-response for Population I and II

	Population I $W_2 = 10\%$ **Non-Response**				**Population II** $W_2 = 10\%$ **Non-Response**			
Estimators	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
U^*_{ds}	100	100	100	100	100	100	100	100
R^*_{ds}	137.79	137.63	137.47	137.31	189.19	189.97	190.27	190.42
P^*_{ds}	46.731	46.7592	46.78	46.81	57.48	58.01	58.22	58.33
$R_{Exp^*_{ds}}$	132.509	132.426	132.34	132.26	136.84	136.62	136.53	136.49
$P_{Exp^*_{ds}}$	68.4115	68.4387	68.46	68.49	74.83	75.15	75.28	75.34
$RR_{Exp^*_{ds}}$	332.350	331.366	330.39	329.42	143.85	140.40	139.13	138.46
$\mathbf{Prop}^*_{ds}$	**388.44**	**386.73**	**385.04**	**383.37**	**356.43**	**379.41**	**391.07**	**398.069**

Table 6 illustrates the percentage relative efficiency of the proposed estimator for both Population I and Population II, each with a 10% non-response rate at varying values of k. The results reveal that the proposed estimator is achieving remarkable efficiencies for all values of k. Notably, the proposed estimator consistently demonstrates superior performance when compared to the other existing estimators across all values of k. This substantial advantage indicates that the proposed estimator not only meets but exceeds expectations in terms of efficiency across both populations. Consequently, these findings underscore the proposed estimator's effectiveness in mitigating the challenges posed by non-response, solidifying its position as a leading choice in the analysis. In comparison, the second-best estimator $RR_{Exp^*_{ds}}$, performs well, with efficiencies such as 332.350 at $k = 2$ in Population I, but it still fails to reach the impressive results of the proposed estimator. Despite performing admirably compared to other estimators, $RR_{Exp^*_{ds}}$, consistently fall short in beating the proposed estimator's efficiency, as seen in its highest value of 138.46 at $k = 5$ for Population II.

Table 7: PRE with $\boldsymbol{W_2 = 20\%}$ non-response for Population I and II

	Population I $W_2 = 20\%$ **Non-Response**				**Population II** $W_2 = 20\%$ **Non-Response**			
Estimators	$k = 2$	$k = 3$	$k = 4$	$k = 5$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
U^*_{ds}	100	100	100	100	100	100	100	100
R^*_{ds}	137.77	137.60	137.43	137.26	184.47	184.72	184.82	184.87
P^*_{ds}	46.72	46.74	46.77	46.79	57.52	58.10	58.33	58.45
$R_{Exp^*_{ds}}$	132.51	132.42	132.34	132.26	135.86	135.52	135.39	135.31
$P_{Exp^*_{ds}}$	68.40	68.43	68.45	68.47	74.95	75.31	75.45	75.53
$RR_{Exp^*_{ds}}$	332.21	331.09	329.98	328.88	143.26	139.53	138.14	137.41
$\mathbf{Prop}^*_{ds}$	**388.20**	**386.24**	**384.32**	**382.42**	**308.60**	**316.67**	**320.96**	**323.55**

Table 7, the percentage relative efficiency of the proposed estimator for both Population I and Population II, under a 20% non-response rate, is examined across various values of k. The results clearly highlight the exceptional efficiency of the proposed estimator, with values starting at 388.20 for $k = 2$ and steadily decreasing to 382.42 for $k = 5$ in Population I. Meanwhile, for Population II, the proposed estimator also maintains a strong performance, ranging from 308.60 at $k = 2$ to 323.55 at $k = 5$. While the second-best estimator, $RR_{Exp^*_{ds}}$, shows reasonable efficiency values such as 332.21 for Population I at $k = 2$ and 328.88 for Population II at $k = 5$, it does not surpass the performance of the proposed estimator at any stage. The proposed estimator consistently outperforms $RR_{Exp^*_{ds}}$ making it the superior choice in terms of efficiency. The proposed estimator efficiencies dominate the existing estimators across all values of k. The proposed estimator's efficiency is substantially higher, further emphasizing its robustness and suitability for handling non-response situations. These results strongly suggest that the proposed estimator offers a more reliable and efficient alternative compared to existing methods, particularly in managing varying non-response rates in both populations.

Table 8 presents the percentage relative efficiency of the proposed estimator for both Population I and Population II under a 30% non-response rate. The data demonstrates that the proposed estimator consistently performs better compared to other estimators, as reflected across different k values. For Population I, the proposed estimator's percentage relative efficiency starts at 375.02 for $k = 2$ and decreases steadily to 337.81 at $k = 5$. Similarly, for Population II, the proposed estimator maintains a strong performance, with values ranging from 292.60 at $k = 2$ to 298.33 at $k = 5$. In comparison, the second-best estimator $RR_{Exp^*_{ds}}$, shows relatively lower efficiency values such as 322.90 for Population I at $k = 2$ and 138.32 for Population II at $k = 5$. While $RR_{Exp^*_{ds}}$, performs decently, its efficiencies remain significantly below those of the proposed estimator across all k values, further supporting the superior performance of the proposed method. Primarily, the proposed estimator surpasses

Table 8: PRE with $\boldsymbol{W_2 = 30\%}$ non-response for Population I and II

	Population I $W_2 = 30\%$ **Non-Response**				**Population II** $W_2 = 30\%$ **Non-Response**			
Estimators	$k=2$	$k=3$	$k=4$	$k=5$	$k=2$	$k=3$	$k=4$	$k=5$
U_{ds}^*	100	100	100	100	100	100	100	100
R_{ds}^*	137.23	136.54	135.88	135.24	184.36	184.64	184.76	184.82
P_{ds}^*	47.02	47.33	47.64	47.94	56.84	57.42	57.65	57.78
$R_{Exp_{ds}^*}$	132.02	131.48	130.96	130.45	136.33	135.97	135.83	135.76
$P_{Exp_{ds}^*}$	68.66	68.93	69.19	69.45	74.52	74.88	75.03	75.11
$RR_{Exp_{ds}^*}$	322.90	313.37	304.63	296.58	145.14	140.84	139.19	138.32
Prop$_{ds}^*$	**375.02**	**361.37**	**349.03**	**337.81**	**292.60**	**295.24**	**297.10**	**298.33**

the efficiencies of other estimators, whose performances are relatively lower. These results suggest that the proposed estimator is highly efficient, particularly in scenarios involving a 30% non-response rate, making it an optimal choice for such conditions.

Simulation study

We created two populations using the R language software, both drawn from a normal distribution. The first population was generated with equal strata, while the second was generated with unequal strata.

Population I (with equal strata)

$N_1 = 2000$, $X_1 = round(rnorm(N_1, 120, 15), 0)$, $Z_1 = round(rnorm(N_1, 200, 25), 0)$, $e_1 = rnorm(N_1, 0, 1)$, $Y_1 = round(0.50 + \rho_{xz} \times X_1 \times Z_1(\sqrt{1-\rho_{xz}^2})), 0)$, $N_2 = 2000$, $X_2 = round(rnorm(N_2, 240, 30), 0)$, $Z_2 = round(rnorm(N_2, 300, 50), 0)$, $e_2 = rnorm(N_2, 0, 1)$, $Y_2 = round(0.50 + \rho_{xz} \times X_2 \times Z_2(\sqrt{1-\rho_{xz}^2})), 0)$, $n_1' = 1500$, $n_2' = 1500$, $n_1 = 750$, $n_2 = 750$.

Population II with (unequal strata)

$N_1 = 1200$, $X_1 = round(rnorm(N_1, 4, 7), 0)$, $Z_1 = round(rnorm(N_1, 5, 8), 0)$, $e_1 = rnorm(N_1, 0, 1)$, $Y_1 = round(0.50 + \rho_{xz} \times X_1 \times Z_1(\sqrt{1-\rho_{xz}^2})), 0)$, $N_2 = 1600$, $X_2 = round(rnorm(N_2, 6, 10), 0)$, $Z_2 = round(rnorm(N_2, 7, 11), 0)$, $e_2 = rnorm(N_2, 0, 1)$, $Y_2 = round(0.50 + \rho_{xz} \times X_2 \times Z_2(\sqrt{1-\rho_{xz}^2})), 0)$, $n_1' = 900$, $n_2' = 900$, $n_1 = 300$, $n_2 = 350$

Hypothetical values are generated by taking the sample of sizes as described above in both populations. Mean square errors (MSE) and percentage relative efficiencies (PRE) of estimators are obtained by simulated 20,000 times. Results for both Population I and II are given in Table 9 and Table 10.

Table 9 gives results on simulation study for equal strata at 10% and 20% non-response rates clearly highlight the superior performance of the proposed Estimator with an MSE consistently around 3059327.45 to 3060060.75. The Proposed Estimator demonstrates remarkable stability and efficiency across varying k values. Its PRE, ranging from 943.36 to 939.56, underscores its dominant performance, significantly surpassing all other estimators. At a 20% non-response rate, the proposed estimator continues to excel with MSE values around 3548115.73 to 3551016.22 and exceptionally high PRE-values from 984.88 to 984.08. This robust efficiency indicates that even with increased non-response, the Proposed Estimator maintains its strong performance advantage.

Among the other estimators, $R_{Exp_{ds}^*}$ and $RR_{Exp_{ds}^*}$ are approximately similar in performance, exhibiting respectable PRE-values such as 679.29 and 679.67 at $k = 2$ and 678.40 and 678.97 at $k = 6$ respectively under 10% non-response rate. For 20% non-response R_{ds}^* attains a

Table 9: MSE and PRE values for Population I with equal strata with 10% and 20% non-response for different values of k_h

	Mean Square Error MSE $W_2 = 10\%$ **Non-Response**			**Percentage Relative Efficiency PRE** $W_2 = 10\%$ **Non-Response**		
	$\boldsymbol{k_h}$			$\boldsymbol{k_h}$		
Estimators	2	4	6	2	4	6
U^*_{ds}	2886073 1.02	2883897 7.89	2887364 9.78	100	100	100
R^*_{ds}	516255 5.67	517184 3.80	516133 6.51	559.03	557.61	559.42
P^*_{ds}	105866 645.92	105824 830.40	105883 117.90	27.26	27.25	27.26
$R_{Exp^*_{ds}}$	424864 5.72	424173 1.75	425612 3.95	679.29	679.88	678.40
$P_{Exp^*_{ds}}$	6492668 3.91	6489333 0.59	6494232 4.84	44.45	44.44	44.46
$RR_{Exp^*_{ds}}$	424624 9.57	423932 6.72	425251 5.07	679.67	680.27	678.97
Prop$^*_{ds}$	**305932 7.45**	**306534 0.93**	**306006 0.75**	**943.36**	**940.80**	**939.56**
	Mean Square Error MSE $W_2 = 20\%$ **Non-Response**			**Percentage Relative Efficiency PRE** $W_2 = 20\%$ **Non-Response**		
	$\boldsymbol{k_h}$			$\boldsymbol{k_h}$		
Estimators	2	4	6	2	4	6
U^*_{ds}	3494494 8.74	3496541 8.02	3494495 3.38	100	100	100
R^*_{ds}	438055 6.85	437536 9.28	438578 1.16	797.72	799.14	796.77
P^*_{ds}	117491 023.91	117511 224.63	117472 396.37	29.74	29.75	29.74
$R_{Exp^*_{ds}}$	649975 4.68	651314 4.71	650688 9.51	537.63	536.84	537.04
$P_{Exp^*_{ds}}$	7419373 7.02	7421557 1.66	7418438 6.94	47.09	47.11	47.10
$RR_{Exp^*_{ds}}$	649778 3.31	651038 5.29	650486 1.58	537.79	537.07	537.21
Prop$^*_{ds}$	**354811 5.73**	**354450 8.98**	**355101 6.22**	**984.88**	**986.46**	**984.08**

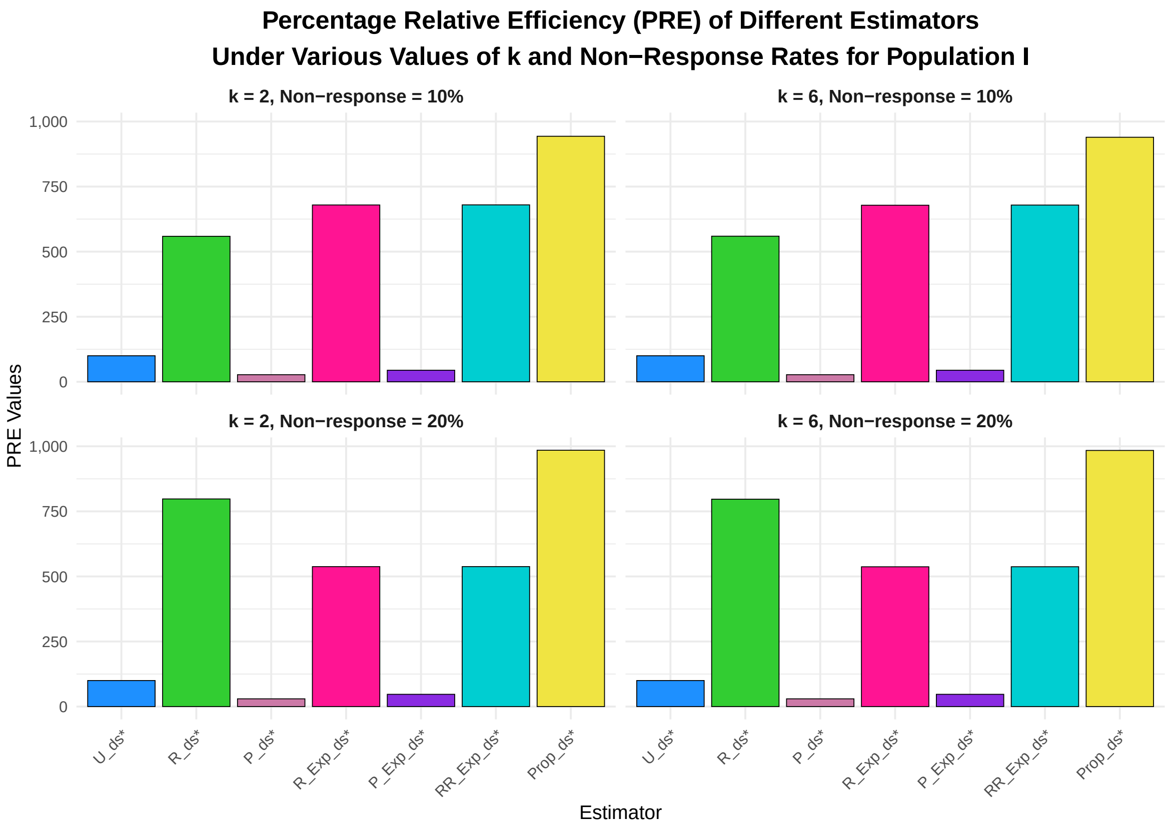

Figure 1: Visualization of percentage relative efficiency (PRE) of different estimators under varying values of k and non-response rates for Population I with equal strata

PRE-value of 797.72 at $k = 2$ and 796.77 at $k = 6$. Although these estimators demonstrate relatively competitive efficiencies, their performance still falls noticeably short of the proposed estimator across all values of k. This comparison further emphasizes the superiority of the proposed method in delivering minimum MSE and maximizing relative efficiency regardless of the non-response intensity or the value of k.

Figure 1 depicts the results of Population I based on the simulation study. In the figure, scenarios with $k = 2$ and $k = 6$ under 10% and 20% non-response rates are used to compare estimator performance visually. The graphical representation clearly supports the numerical findings, as the proposed estimator stands out with the lowest MSE and highest PRE-across all configurations, reinforcing its consistency and effectiveness in handling non-response scenarios.

Table 10, under simulation study, the context of unequal strata with non-response rates of 10% and 20%, the proposed estimator demonstrates strong advantages over all other estimators. When facing a 10% non-response rate, the proposed estimator yields exceptionally low MSE values, spanning from 7.31 to 8.20. This estimator's efficiency is further underscored by its high PRE-values, which range between 519.69 and 575.98. Such figures signify that it achieves a high level of precision and stability, markedly surpassing than other estimators. For the 20% non-response scenario, the proposed estimator continues to show robust performance, with MSE values between 7.41 and 9.35. The PRE-values remain impressive, ranging from 491.19 to 599.20. This indicates that even with increased non-response, the proposed estimator maintains efficiency and consistently outperforms existing estimators. In conclusion, across both non-response rates, the proposed estimator shows the highest efficiency and the lowest error, affirming its superior performance and effectiveness as a reliable choice for handling unequal strata under varied non-response conditions.

Table 10: MSE and PRE values for Population II with unequal strata with 10% and 20% non-response for different values of k_h

	Mean Square Error MSE $W_2 = 10\%$ **Non-Response**			**Percentage Relative Efficiency PRE** $W_2 = 10\%$ **Non-Response**		
	$\boldsymbol{k_h}$			$\boldsymbol{k_h}$		
Estimators	2	4	6	2	4	6
U^*_{ds}	42.12	42.17	42.62	100	100	100
R^*_{ds}	14.00	14.88	15.75	300.82	283.34	270.55
P^*_{ds}	102.44	102.28	102.61	41.11	41.23	41.53
$R_{Exp^*_{ds}}$	17.79	18.04	18.58	236.74	233.68	229.37
$P_{Exp^*_{ds}}$	71.89	71.82	72.22	58.58	58.71	59.01
$RR_{Exp^*_{ds}}$	17.10	17.37	17.90	246.23	242.78	238.04
$\mathbf{Prop}^*_{ds}$	**7.31**	**7.722**	**8.20**	**575.98**	**546.11**	**519.69**
	Mean Square Error MSE $W_2 = 20\%$ **Non-Response**			**Percentage Relative Efficiency PRE** $W_2 = 20\%$ **Non-Response**		
	$\boldsymbol{k_h}$			$\boldsymbol{k_h}$		
Estimators	2	4	6	2	4	6
U^*_{ds}	44.40	45.24	45.96	100	100	100
R^*_{ds}	13.43	15.29	17.37	330.56	295.74	264.59
P^*_{ds}	105.88	106.34	106.75	41.93	42.54	43.05
$R_{Exp^*_{ds}}$	18.85	19.93	20.91	235.53	226.91	219.78
$P_{Exp^*_{ds}}$	74.93	75.57	76.13	59.26	59.86	60.37
$RR_{Exp^*_{ds}}$	18.13	19.24	20.24	244.93	235.07	227.08
$\mathbf{Prop}^*_{ds}$	**7.41**	**8.33**	**9.35**	**599.20**	**542.87**	**491.19**

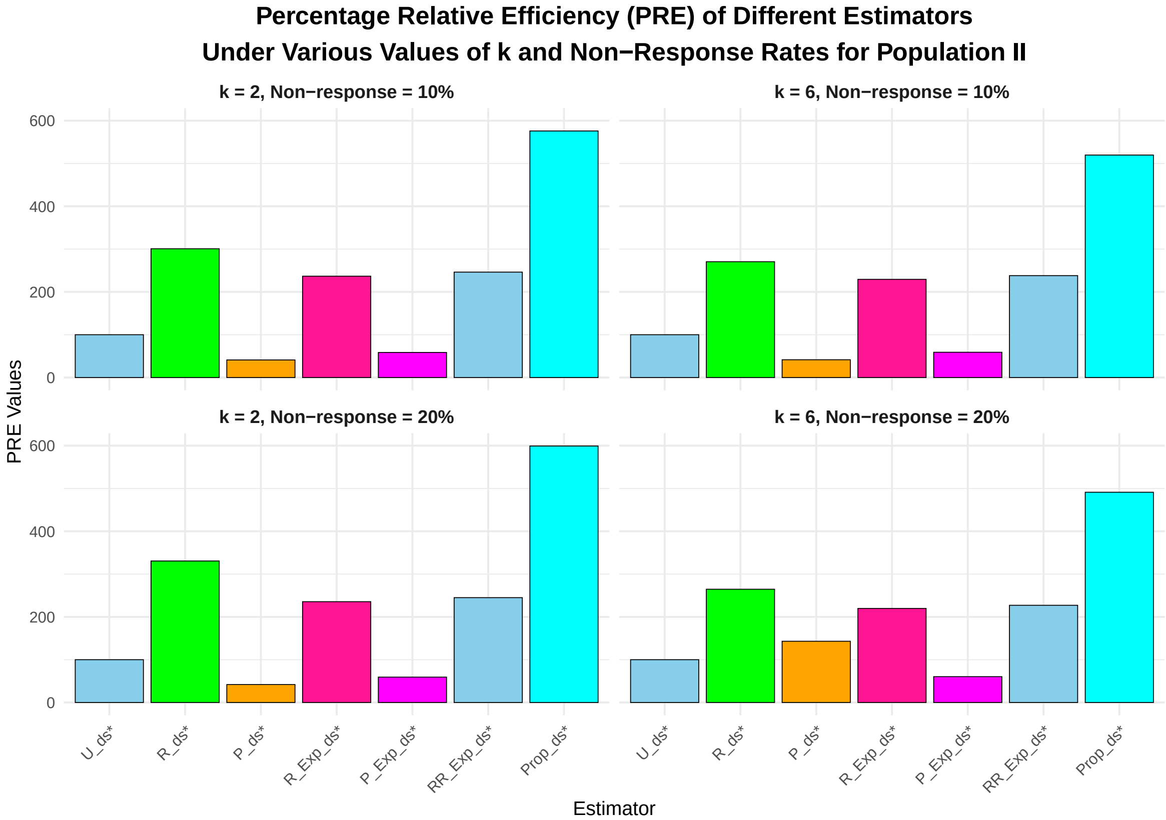

Figure 2: Graphical illustration of PRE versus different estimators for different values of k and non-response rates for Population II with unequal strata

Among the alternative estimators, the R^*_{ds} estimator emerges as the next best performer, displaying reasonably strong PRE-values such as 300.82 and 270.55 at $k = 2$ and $k = 6$ under a 10% non-response rate, and 330.56 and 264.59 at $k = 2$ and $k = 6$ under a 20% non-response rate. Although this estimator delivers decent efficiency, its performance still remains notably behind that of the Proposed Estimator across all conditions. Furthermore, Figure 2 visually summarizes these findings for Population II, illustrating the comparative efficiency of different estimators at $k = 2$ and $k = 6$ under both 10% and 20% non-response rates. This graphical representation reinforces the numerical outcomes, providing a clear depiction of how dominantly the proposed estimator outperforms the rest.

6. Discussion and conclusion

Estimating population parameters is the primary concern for researchers across all fields. In the present study, a regression-ratio-in-ratio type exponential estimator is modified to account for the issue of non-response, in order to estimate the finite population mean under double sampling stratification (DSS) by using two auxiliary variables. The conditional comparison of the proposed estimator was caried out against other competitor estimators under DSS in the presence of non-response problems, such as usual mean estimator, ratio estimator, product estimator, ratio-exponential estimator, product-exponential estimator and ratio-in-ratio exponential estimator. Following the conditional comparision, numerical results were obtained to assess the performance of proposed estimator relative to the competing estimators. The comparison was made in terms of mean square error (MSE) and percentage relative efficiency (PRE) under varying rate of non-response. The results based on population data sets indicate that the proposed estimator consistently outperformed the exisitng estimators in all scenario. Furthermore, to validate the performance of proposed estimator, a simulation study was conducted. In this study, two populations were generated with 10% and 20% non-response rates with different values of a constant. Based on both the numerical and simulation results, we conclude that the performance of our proposed estimator demonstrates superior performance compared to the existing estimators. Therefore, we recommend the use of this in practical applications for estimating the finite population mean under DSS in the presence of non-response.

R-Codes

The interested readers can get the R-codes from the authors if required.

Acknowledgements

Authors are thankful to the learned referees for their valuable suggesstions which helped to improve the manuscript.

References

Bahl S, Tuteja RK (1991). "Ratio and Product Type Exponential Estimators." *Journal of Information and Optimization Sciences*, **12**(1), 159–164.

Batool A, Ali Z, Mohsin M, Masmoudi A, Kartal V, Satti S (2024). "Assessing the Generalization of Forecasting Ability of Machine Learning and Probabilistic Models for Complex Climate Characteristics." *Stochastic Environmental Research and Risk Assessment*, **38**(8), 2927–2947.

Cochran WG (1940). "The Estimation of the Yields of Cereal Experiments by Sampling for the Ratio of Grain to Total Produce." *The Journal of Agricultural Science*, **30**(2), 262–275.

Garg N, Pachori M (2020). "Use of Coefficient of Variation in Calibration Estimation of Population Mean in Stratified Sampling." *Communications in Statistics-Theory and Methods*, **49**(23), 5842–5852.

Gupta A, Tailor R (2021). "Ratio in Ratio Type Exponential Strategy for the Estimation of Population Mean." *Journal of Reliability and Statistical Studies*, pp. 551–564.

Hansen MH, Hurwitz WN (1946). "The Problem of Non-response in Sample Surveys." *Journal of the American Statistical Association*, **41**(236), 517–529.

Kadilar C, Cingi H (2005). "A New Ratio Estimator in Stratified Random Sampling." *Communications in Statistics—Theory and Methods*, **34**(3), 597–602.

Khare BB, Srivastava S (1993). "Estimation of Population Mean Using Auxiliary Character in Presence of Nonresponse." *National Academy Science Letters-India*, **16**(3), 111–114.

Khare BB, Srivastava S (1997). "Transformed Ratio Type Estimators for the Population Mean in the Presence of Nonresponse." *Communications in Statistics-Theory and Methods*, **26**(7), 1779–1791.

Kiregyera B (1984). "Regression-type Estimators Using Two Auxiliary Variables and the Model of Double Sampling from Finite Populations." *Metrika*, **31**(1), 215–226.

Murthy MN (1964). "Product Method of Estimation." *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 69–74.

Neyman J (1938). "Contribution to the Theory of Sampling Human Populations." *Journal of the American Statistical Association*, **33**(201), 101–116.

Okafor FC, Lee H (2000). "Double Sampling for Ratio and Regression Estimation with Subsampling the Non-respondents." *Survey Methodology*, **26**(2), 183–188.

Pachori M, Garg N (2025). "Improved Calibration Estimator of Population Mean in Stratified and Stratified Double Sampling." *Life Cycle Reliability and Safety Engineering*, pp. 1–12.

Rao PSRS (1986). "Ratio Estimation with Sub Sampling the Non-respondents." *Survey Methodology*, **12**(2), 217–230.

Rao TJ (1983). "A New Class of Unbiased Product Estimators." *Stat Math. Tech. Report*, **15**, 83.

Shabbir J, Arsalan SM, Kim JM (2025). "On Use of Double Sampling Strategy for the Estimation of Finite Population Mean." *Measurement: Interdisciplinary Research and Perspectives*, pp. 1–10.

Singh HP, Kumar S (2008). "A Regression Approach to the Estimation of the Finite Population Mean in the Presence of Non-response." *Australian & New Zealand Journal of Statistics*, **50**(4), 395–408.

Singh HP, Kumar S (2010). "Estimation of Mean in Presence of Non-response Using Two Phase Sampling Scheme." *Statistical Papers*, **51**, 559–582.

Singh HP, Nigam P (2020a). "Ratio-ratio-type Exponential Estimator of Finite Population Mean in Double Sampling for Stratification." *International Journal of Agricultural and Statistical Science*, **16**(1), 251–257.

Singh HP, Nigam P (2020b). "Ratio-ratio-type Exponential Estimator of Finite Population Mean in Double Sampling for Stratification." *International Journal of Agricultural and Statistical Science*, **16**(1), 251–257.

Singh HP, Vishwakarma GK (2007). "A General Procedure for Estimating the Mean Using Double Sampling for Stratification." *Model Assisted Statistics and Applications*, **2**(4), 225–237.

Singh S (2003). *Advanced Sampling Theory with Applications: How Michael"" Selected"" Amy*, volume 2. Springer Science & Business Media.

Singh S, Joarder AH, Tracy DS (2001). "Median Estimation Using Double Sampling." *Australian & New Zealand Journal of Statistics*, **43**(1), 33–46.

Sun S, Haneuse S, Levis AW, Lee C, Arterburn DE, Fischer H, Shortreed S, Mukherjee R (2025). "Estimating Weighted Quantile Treatment Effects with Missing Outcome Data by Double Sampling." *Biometrics*, **81**(2), ujaf038.

Tabasum R, Khan IA (2004). "Double Sampling for Ratio Estimation with Non-response." *Journal of the Indian Society of Agricultural Statistics*, **58**(3), 300–306.

Tabasum R, Khan IA (2006). "Double Sampling Ratio Estimator for the Population Mean in Presence of Non-response." *Assam Statistical Review*, **20**(1), 73–83.

Tailor R, Chouhan S, Kim JM (2014). "Ratio and Product Type Exponential Estimators of Population Mean in Double Sampling for Stratification." *CSAM (Communications for Statistical Applications and Methods)*, **21**(1), 1–9.

Tracy DS, Singh S, Arnab R (2003). "Note on Calibration in Stratified and Double Sampling." *Survey Methodology*, **29**(1), 99–104.

Tripathi TP, Ige AF (1987). "On Double Sampling for Stratification and Use of Auxiliary Information." *Journal of the Indian Society of Agricultural Statistics*, **39**(2), 191–201.

Affiliation:

Syed Muhammad Arsalan
Department of Statistics
Federal University of Minas Gerais (UFMG), Brazil
E-mail: syedma@ufmg.br

Javid Shabbir
Department of Statistics
University of Wah, Wah Cantt, Pakistan
E-mail: javid.shabbir@uow.edu.pk

Alamgir
Department of Statistics
University of Peshawar, Pakistan
E-mail: alamgir_khalil@uop.edu.pk

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume 54 *Submitted:* 2024-12-04
2025 *Accepted:* 2025-06-10