ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 44(2024) 1-16

Some Calibration Estimators of Finite Population Mean under Stratified Systematic Sampling in the Presence of Non-Response

Manoj K. Chaudhary[1] and Tulika Dutta[2]
[Received on November, 2022. Accepted on December, 2023]

ABSTRACT

The present research aims to discuss the problem of non-response in estimating the population mean under stratified systematic sampling utilizing the information on a single auxiliary variable. The notion of calibration has been deemed to modify the non-response rates and also the stratum weights so that the resulting calibration estimators would provide more précised estimates. The expressions for the MSE of the proposed calibration estimators have been derived using the Taylor linearization technique. An empirical study based on simulated and real data has also been carried out to check the performance of the proposed calibration estimators. The study reveals that the proposed calibration estimators outperform the existing ones.

1. Introduction

Once the sub-populations (sub-groups) within the whole population vary significantly in a statistical survey, it is useful to select a sample from each sub-population independently. Stratification is the process of dividing the whole population into homogeneous subgroups before sampling. Simple random sampling or systematic sampling can be applied to select the sample from each stratum. This mechanism certainly improves the representativeness of the sample by reducing sampling error. Systematic sampling is the simplest technique to select the required number of units with a single random start. However, many authors have used simple random sampling to select units from each stratum. Clement (2017) used systematic sampling for this purpose and proposed a calibration ratio-type estimator under stratified systematic sampling.

Nowadays, non-response is a big issue in all types of statistical surveys. There are many reasons for non-response, such as not being at home, being unable to answer the question, lack of interest, etc. Hansen and Hurwitz (1946) were the first to discuss the problem of non-response in mail surveys. The auxiliary information is usually used to compensate for the efficiency loss due to non-response. Many authors have considered the problem of non-response in estimating the parameters. Singh and Kumar (2010) have estimated the population mean in non-response presence using a two-phase sampling scheme. Kumar and Bhougal (2011) have suggested the estimators of the population mean using auxiliary information in the presence of non-response. Khare *et al.* (2012) suggested chain-type estimators for ratio of two population means using auxiliary characters in the presence of non-response. Chaudhary *et al.* (2012) proposed a general family of estimators for estimating the population mean in systematic sampling using auxiliary information under non-response. Raman *et al.*

Corresponding author: Manoj K. Chaudhary, Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, U.P.-221005, India. Email: mk15@bhu.ac.in
[2]Department of Mathematics, Faculty of Applied and Basic Sciences, SGT University, Gurugram, Haryana-122505, India.

(2016) suggested calibration approach-based product-type estimators of finite population total in single and two-phase sampling under non-response. Gautam *et al.* (2020) proposed a calibration estimator of the population mean in the presence of non-response. Recently, Chaudhary and Dutta (2023) have developed some calibration-based improvements in estimating the mean of a stratified population with a scrambled response on the second call under non-response.

In the subsequent sections, some improved calibration estimators of the population mean in stratified systematic sampling utilizing the information on a single auxiliary variable under non-response have been proposed for the first time. The expressions for the proposed calibration estimators' mean square error (MSE) have been derived using the Taylor linearization technique. A comparison of the proposed calibration estimators with the usual estimator has been presented by conducting an empirical study based on simulated and real data.

2. Sampling Strategy and Estimation Procedure

Consider a finite population of N distinct and identifiable units which consists of L strata with N_l units in the l^{th} stratum from which a systematic random sample of size n_l is selected $\left(l = 1, 2, ..., L\right)$. In order to draw the systematic sample, only the first unit is selected at random from each stratum, the rest being automatically selected by choosing k_l (sampling interval). It is further mentioned that $N_l = n_l k_l$. Let Y and X be the study and auxiliary variables with respective population means $\overline{Y}(=\sum_{l=1}^{L} w_l \overline{Y}_l)$ and $\overline{X}(=\sum_{l=1}^{L} w_l \overline{X}_l)$. Here, $\overline{Y}_l$ and $\overline{X}_l$ are respectively the population means under study and auxiliary variables for the l^{th} stratum; $w_l (= \frac{N_l}{N})$ is the weight for the l^{th} stratum. Let y_{lij} and x_{lij} be the observations on the j^{th} unit in the i^{th} systematic sample for the l^{th} stratum under study variable Y and auxiliary variable X, respectively $\left(i = 1, 2, ..., k_l; j = 1, 2, ..., n_l\right)$. Let us assume that the auxiliary variable X is free from the non-response while the study variable Y is suffering from the non-response. It is further assumed that out of n_l units of a systematic sample, n_{l1} units respond, and n_{l2} units do not respond on Y. Now, by adopting the technique given by Hansen and Hurwitz (1946), we select a sub-sample of size h_{l2} from n_{l2} non-responding units $\left(h_{l2} = \frac{n_{l2}}{g_l}; g_l > 1\right)$ and collect the information from all the h_{l2} units.

The Hansen and Hurwitz (1946) estimator for the population total under stratified systematic sampling is given as

$$T = \sum_{l=1}^{L} [\sum_{j}^{n_{l1}} d_{lij} y_{lij} + \sum_{j}^{h_{l2}} d_{lij}^{*} y_{lij}] \tag{2.1}$$

where, d_{lij} and d_{lij}^{*} are the design weights.

Here, we use systematic sampling to draw a sample from each stratum. Hence, the inclusion probabilities are defined as

$$\Pi_{lij} = \frac{1}{k_l} \text{ and } \Pi_{lij|n_{l2}} = \frac{h_{l2}}{n_{l2}}.$$

Using these values of the inclusion probabilities, the design weights become

$$d_{lij} = \frac{1}{\Pi_{lij}} = k_l \text{ and } d^*_{lij} = \frac{1}{\Pi_{lij}\Pi_{lij|n_{l2}}} = k_l \frac{n_{l2}}{h_{l2}}.$$

Putting above values of the design weights into (2.1), the estimator of the population total becomes

$$T^* = \sum_{l=1}^{L} [k_l n_l \{w_{nl1}\bar{y}_{nl1} + w_{nl2}\bar{y}_{hl2}\}]$$

where $w_{nl1} = \frac{n_{l1}}{n_l}$ and $w_{nl2} = \frac{n_{l2}}{n_l}$. $\bar{y}_{nl1}$ and $\bar{y}_{hl2}$ are the means based on n_{l1} responding units and h_{l2} non-responding units under study variable.

Now, Hansen and Hurwitz's (1946) estimator of the population mean $\bar{Y}$ under stratified systematic sampling is defined as

$$\bar{y}^* = \sum_{l=1}^{L} w_l [w_{nl1}\bar{y}_{nl1} + w_{nl2}\bar{y}_{hl2}] \tag{2.2}$$

The expression for the variance of the estimator $\bar{y}^*$ is given by

$$V(\bar{y}^*) = \sum_{l=1}^{L} w_l^2 [\frac{(N_l - 1)}{N_l n_l}(1 + (n_l - 1)\rho_{ly})S^2_{ly} + \frac{(g_l - 1)}{n_l} W_{l2} S^2_{ly2}] \tag{2.3}$$

where $\rho_{ly} \left\{ = \frac{E(y_{lij} - \bar{Y}_l)(y_{lkj} - \bar{Y}_l)}{E(y_{lij} - \bar{Y}_l)^2} \right\}$ is the intra-class correlation coefficient between the units of the same systematic sample in the l^{th} stratum under study variable. $\bar{Y}_l$ is the population mean of the study variable in the l^{th} stratum. S^2_{ly} and S^2_{ly2} are respectively the population mean squares of the entire group and non-response group in the l^{th} stratum under the study variable. W_{l2} is the population non-response rate in the l^{th} stratum.

3. Proposed Calibration Estimators

We now propose a new calibration estimator of the population total $T(= \sum_{l=1}^{L}\sum_{i=1}^{k_l}\sum_{j=1}^{n_l} y_{lij})$ in stratified systematic sampling under non-response as follows:

$$T_{cal} = \sum_{l=1}^{L} [\sum_{j}^{n_{l1}} d_{lij} y_{lij} + \sum_{j}^{h_{l2}} \Omega^*_{lij} y_{lij}] \tag{3.1}$$

where Ω^*_{lij} is the calibration weight such that it minimizes the chi-square distance

$$\sum_{j}^{h_{l2}} \frac{(\Omega^*_{lij} - d^*_{lij})^2}{q_{lij} d^*_{lij}}$$

subject to the calibration constraints

$$\sum_{j}^{h_{l2}} \Omega^*_{lij} x_{lij} = \sum_{j}^{n_{l2}} d_{lij} x_{lij} \tag{3.2}$$

$$\sum_{j}^{h_{l2}} \Omega^*_{lij} = \sum_{j}^{h_{l2}} d^*_{lij} \tag{3.3}$$

where q_{lij} is the tuning parameter.

Let us define the Lagrange function as

$$\Delta_1 = \sum_{j}^{h_{l2}} \frac{(\Omega^*_{lij} - d^*_{lij})^2}{q_{lij} d^*_{lij}} - 2\lambda_1 (\sum_{j}^{h_{l2}} \Omega^*_{lij} x_{lij} - \sum_{j}^{n_{l2}} d_{lij} x_{lij}) - 2\lambda_2 (\sum_{j}^{h_{l2}} \Omega^*_{lij} - \sum_{j}^{h_{l2}} d^*_{lij})$$

where λ_1 and λ_2 are the Lagrange multipliers.

Differentiating Δ_1 with respect to Ω^*_{lij} and equating the derivative to zero, we get

$$\Omega^*_{lij} = d^*_{lij} + q_{lij} d^*_{lij} (\lambda_1 x_{lij} + \lambda_2) \tag{3.4}$$

Putting the value of Ω^*_{lij} from (3.4) into (3.2) and (3.3), we respectively get

$$\sum_{j}^{h_{l2}} [d^*_{lij} + q_{lij} d^*_{lij} (\lambda_1 x_{lij} + \lambda_2)] x_{lij} = \sum_{j}^{n_{l2}} d_{lij} x_{lij}$$ and

$$\sum_{j}^{h_{l2}} [d^*_{lij} + q_{lij} d^*_{lij} (\lambda_1 x_{lij} + \lambda_2)] = \sum_{j}^{h_{l2}} d^*_{lij}.$$

Now, we find the values of λ_1 and λ_2 by solving the following matrix:

$$\begin{bmatrix} \sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x^2_{lij} & \sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x_{lij} \\ \sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x_{lij} & \sum_{j}^{h_{l2}} q_{lij} d^*_{lij} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix} = \begin{bmatrix} \sum_{j}^{n_{l2}} d_{lij} x_{lij} - \sum_{j}^{h_{l2}} d^*_{lij} x_{lij} \\ 0 \end{bmatrix}$$

Thus, the values of λ_1 and λ_2 are respectively given as

$$\lambda_1 = \frac{(\sum_{j}^{h_{l2}} q_{lij} d^*_{lij})[\sum_{j}^{n_{l2}} d_{lij} x_{lij} - \sum_{j}^{h_{l2}} d^*_{lij} x_{lij}]}{D}$$ and $$\lambda_2 = \frac{(-\sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x_{lij})[\sum_{j}^{n_{l2}} d_{lij} x_{lij} - \sum_{j}^{h_{l2}} d^*_{lij} x_{lij}]}{D}$$

where

$$D = (\sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x^2_{lij})(\sum_{j}^{h_{l2}} q_{lij} d^*_{lij}) - (\sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x_{lij})^2$$

Putting the values of λ_1 and λ_2 into (3.4), we find the optimum value of Ω^*_{lij} and then putting the optimum value of Ω^*_{lij} into (3.1), the calibration estimator of the population total becomes

$$T_{cal} = \sum_{l=1}^{L} [\sum_{j}^{n_{l1}} d_{lij} y_{lij} + \sum_{j}^{h_{l2}} d^*_{lij} y_{lij} + \frac{A}{D} (\sum_{j}^{n_{l2}} d_{lij} x_{lij} - \sum_{j}^{h_{l2}} d^*_{lij} x_{lij})] \tag{3.5}$$

where

$$A = (\sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x_{lij} y_{lij})(\sum_{j}^{h_{l2}} q_{lij} d^*_{lij}) - (\sum_{j}^{h_{l2}} q_{lij} d^*_{lij} y_{lij})(\sum_{j}^{h_{l2}} q_{lij} d^*_{lij} x_{lij}).$$

Now, we assume $q_{lij} = 1 \; \forall l, i$ and j. Thus, the calibration estimator of the population total T is reduced to

$$T_{cal} = \sum_{l=1}^{L} [k_l n_l \{w_{nl1} \bar{y}_{nl1} + w_{nl2} \bar{y}_{hl2} + \beta_{hl2}(w_{nl2} \bar{x}_{nl2} - w_{nl2} \bar{x}_{hl2})\}]$$

where

$$\beta_{hl2} = \frac{[\frac{1}{h_{l2}}(\sum_{j}^{h_{l2}} x_{lij} y_{lij}) - \bar{y}_{hl2} \bar{x}_{hl2}]}{[\frac{1}{h_{l2}}(\sum_{j}^{h_{l2}} x^2_{lij}) - \bar{x}^2_{hl2}]}.$$

Now, the calibration estimator of the population mean $\bar{Y}$ in stratified systematic sampling under non-response can be defined as

$$\bar{y}_{cal} = \sum_{l=1}^{L} \Omega_l \bar{y}'_{lsys} \tag{3.6}$$

where

$$\bar{y}'_{lsys} = w_{nl1} \bar{y}_{nl1} + w_{nl2} \bar{y}_{hl2} + \beta_{hl2}(w_{nl2} \bar{x}_{nl2} - w_{nl2} \bar{x}_{hl2})$$

Here, Ω_l is also a calibration weight such that it minimizes the chi-square distance

$$\sum_{l=1}^{L} \frac{(\Omega_l - w_l)^2}{q_l w_l}$$

where q_l is another tuning parameter.

It is to be noted that by choosing different values of q_l and different sets of calibration constraints, we can get various types of calibration estimators of the population mean $\bar{Y}$. Let us now discuss some of the cases.

Case 1:

In this case, we minimize the chi-square distance subject to the calibration constraint

$$\sum_{l=1}^{L} \Omega_l \bar{x}_{lsys} = \bar{X} \tag{3.7}$$

where $\bar{x}_{lsys}$ is the sample mean for the l^{th} stratum under auxiliary variable.

Now, we define the Lagrange function as

$$\Delta_2 = \sum_{l=1}^{L} \frac{(\Omega_l - w_l)^2}{q_l w_l} - 2\lambda_3 (\sum_{l=1}^{L} \Omega_l \bar{x}_{lsys} - \bar{X})$$

where λ_3 is the Lagrange multiplier.

Differentiating Δ_2 with respect to Ω_l and equating the derivative to zero, we get

$$\Omega_l = w_l + \lambda_3 q_l w_l \bar{x}_{lsys} \tag{3.8}$$

Putting the value of Ω_l from (3.8) into (3.7), we get the value of λ_3 as

$$\lambda_3 = \frac{\overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys}}{\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys}^{2}}.$$

Substituting the value of λ_3 into (3.8), one can get the optimum value of Ω_l and hence (3.6) provides the calibration estimator of the population mean $\overline{Y}$ as

$$\overline{y}_{cal(1)} = \sum_{l=1}^{L} w_l \overline{y}'_{lsys} + \frac{\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys} \overline{y}'_{lsys}}{\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys}^{2}} (\overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys}) \tag{3.9}$$

Particularly, if $q_l = \frac{1}{\overline{x}_{lsys}}$, the calibration estimator given in (3.9) becomes

$$\overline{y}^{*}_{cal(1)} = \frac{\sum_{l=1}^{L} w_l \overline{y}'_{lsys}}{\sum_{l=1}^{L} w_l \overline{x}_{lsys}} \overline{X} \tag{3.10}$$

Case 2:

Here, we minimize the chi-square distance subject to the calibration constraints

$$\sum_{l=1}^{L} \Omega_l \overline{x}_{lsys} = \overline{X} \tag{3.11}$$

$$\sum_{l=1}^{L} \Omega_l = \sum_{l=1}^{L} w_l \tag{3.12}$$

Let us define the Lagrange function as

$$\Delta_3 = \sum_{l=1}^{L} \frac{(\Omega_l - w_l)^2}{q_l w_l} - 2\lambda_4 (\sum_{l=1}^{L} \Omega_l \overline{x}_{lsys} - \overline{X}) - 2\lambda_5 (\sum_{l=1}^{L} \Omega_l - \sum_{l=1}^{L} w_l)$$

where λ_4 and λ_5 are the Lagrange multipliers.

Differentiating Δ_3 with respect to Ω_l and equating the derivative to zero, we get

$$\Omega_l = w_l + q_l w_l (\lambda_4 \overline{x}_{lsys} + \lambda_5) \tag{3.13}$$

Putting the value of Ω_l from (3.13) into (3.11) and (3.12), we get the following matrix:

$$\begin{bmatrix} \sum_{l=1}^{L} q_l w_l \overline{x}_{lsys}^{2} & \sum_{l=1}^{L} q_l w_l \overline{x}_{lsys} \\ \sum_{l=1}^{L} q_l w_l \overline{x}_{lsys} & \sum_{l=1}^{L} q_l w_l \end{bmatrix} \begin{bmatrix} \lambda_4 \\ \lambda_5 \end{bmatrix} = \begin{bmatrix} \overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys} \\ 0 \end{bmatrix}$$

Solving the above matrix, we respectively get the values of λ_4 and λ_5 as

$$\lambda_4 = \frac{(\sum_{l=1}^{L} q_l w_l)[\overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys}]}{D_1} \text{ and } \lambda_5 = \frac{(-\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys})[\overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys}]}{D_1}$$

where

$$D_1 = (\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys}^{2})(\sum_{l=1}^{L} q_l w_l) - (\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys})^2 .$$

Putting the values of λ_4 and λ_5 into (3.13), we can get the optimum value of Ω_l and hence, the calibration estimator given in (3.6) becomes

$$\overline{y}_{cal(2)} = \sum_{l=1}^{L} w_l \overline{y}'_{lsys} + \frac{A_1}{D_1}[\overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys}]$$

where

$$A_1 = (\sum_{l=1}^{L} q_l w_l \overline{y}'_{lsys} \overline{x}_{lsys})(\sum_{l=1}^{L} q_l w_l) - (\sum_{l=1}^{L} q_l w_l \overline{x}_{lsys})(\sum_{l=1}^{L} q_l w_l \overline{y}'_{lsys}) .$$

In particular, if $q_l = 1 \ \forall l$, the calibration estimator reduces to

$$\overline{y}^{*}_{cal(2)} = \sum_{l=1}^{L} w_l \overline{y}'_{lsys} + \hat{\beta}[\overline{X} - \sum_{l=1}^{L} w_l \overline{x}_{lsys}] \tag{3.14}$$

where

$$\hat{\beta} = \frac{A_1^*}{D_1^*}, \ A_1^* = (\sum_{l=1}^{L} w_l \overline{y}'_{lsys} \overline{x}_{lsys}) - (\sum_{l=1}^{L} w_l \overline{x}_{lsys})(\sum_{l=1}^{L} w_l \overline{y}'_{lsys}) \text{ and}$$

$$D_1^* = (\sum_{l=1}^{L} w_l \overline{x}_{lsys}^{2}) - (\sum_{l=1}^{L} w_l \overline{x}_{lsys})^2 .$$

4. Properties of the Proposed Calibration Estimators

In order to obtain the variance/mean square errors (MSEs) of $\overline{y}^*$ and proposed calibration estimators $\overline{y}^*_{cal(1)}$ and $\overline{y}^*_{cal(2)}$, we use the Taylor linearization technique. Let

$$\overline{y}_{hhl} = w_{nl1}\overline{y}_{nl1} + w_{nl2}\overline{y}_{hl2}$$

and hence

$$V(\overline{y}_{hhl}) = \frac{(N_l - 1)}{N_l n_l}[1 + (n_l - 1)\rho_{ly}]S_{ly}^2 + W_{l2}\frac{(g_l - 1)}{n_l}S_{ly2}^2 .$$

Moreover, we have

$$V(w_{nl2}\overline{x}_{nl2}) = V[E(w_{nl2}\overline{x}_{nl2} \mid n_{l2})] + E[V(w_{nl2}\overline{x}_{nl2} \mid n_{l2})]$$

$$= V(w_{nl2}\overline{x}_{nl2}) = W_{l2}^2 \frac{(N_{l2} - 1)}{N_{l2} n_{l2}}[1 + (n_{l2} - 1)\rho_{lx2}]S_{lx2}^2$$

$$V(w_{nl2}\bar{x}_{hl2}) = V[E(w_{nl2}\bar{x}_{hl2} \mid n_{l2})] + E[V(w_{nl2}\bar{x}_{hl2} \mid n_{l2})]$$
$$= V\left(w_{nl2}\bar{x}_{nl2}\right) + E\left[w_{nl2}^2\left(\frac{1}{h_{l2}} - \frac{1}{n_{l2}}\right)s_{lx2}^2\right]$$
$$= W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{lx2}]S_{lx2}^2 + W_{l2}\left(\frac{g_l-1}{n_l}\right)S_{lx2}^2 ,$$

$$C\text{ov}(w_{nl2}\bar{x}_{nl2}, w_{nl2}\bar{x}_{hl2})$$
$$= C\text{ov}[E(w_{nl2}\bar{x}_{nl2} \mid n_{l2}), E(w_{nl2}\bar{x}_{hl2} \mid n_{l2})] + E[C\text{ov}(w_{nl2}\bar{x}_{nl2}, w_{nl2}\bar{x}_{hl2} \mid n_{l2})]$$
$$= W_{l2}^2 C\text{ov}(\bar{x}_{nl2}, \bar{x}_{nl2}) = W_{l2}^2 V(\bar{x}_{nl2})$$
$$= W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{lx2}]S_{lx2}^2 ,$$

$$C\text{ov}(\bar{y}_{hhl}, w_{nl2}\bar{x}_{nl2}) = C\text{ov}(w_{nl2}\bar{y}_{hl2}, w_{nl2}\bar{x}_{nl2})$$
$$= \left(C\text{ov}[E(w_{nl2}\bar{y}_{hl2} \mid n_{l2}), E(w_{nl2}\bar{x}_{nl2} \mid n_{l2})] + E[C\text{ov}(w_{nl2}\bar{y}_{hl2}, w_{nl2}\bar{x}_{nl2} \mid n_{l2})]\right)$$
$$= W_{l2}^2 C\text{ov}(\bar{y}_{nl2}, \bar{x}_{nl2})$$
$$= W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{ly2}]^{1/2}[1+(n_{l2}-1)\rho_{lx2}]^{1/2}\rho_{l2}S_{ly2}S_{lx2}$$

and

$$C\text{ov}(\bar{y}_{hhl}, w_{nl2}\bar{x}_{hl2}) = C\text{ov}(w_{nl2}\bar{y}_{hl2}, w_{nl2}\bar{x}_{hl2})$$
$$= \left(C\text{ov}[E(w_{nl2}\bar{y}_{hl2} \mid n_{l2}), E(w_{nl2}\bar{x}_{hl2} \mid n_{l2})] + E[C\text{ov}(w_{nl2}\bar{y}_{hl2}, w_{nl2}\bar{x}_{hl2} \mid n_{l2})]\right)$$
$$= W_{l2}^2 C\text{ov}(\bar{y}_{nl2}, \bar{x}_{nl2}) + E\left[w_{nl2}^2\left(\frac{1}{h_{l2}} - \frac{1}{n_{l2}}\right)s_{lxy2}\right]$$
$$= W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{ly2}]^{1/2}[1+(n_{l2}-1)\rho_{lx2}]^{1/2}\rho_{l2}S_{ly2}S_{lx2} + W_{l2}\left(\frac{g_l-1}{n_l}\right)S_{lxy2}$$

where S_{lx2}^2 is the population mean square of the non-response group in the l^{th} stratum under the auxiliary variable. ρ_{lx2} is the intra-class correlation coefficient between the units of the same systematic sample of the non-response group in the l^{th} stratum under auxiliary variable. ρ_{l2} is the correlation coefficient between the study and auxiliary variables for the non-response group in the l^{th} stratum. $S_{lxy2} = \rho_{l2}S_{ly2}S_{lx2}$.

Let us now calculate $MSE(\sum_{l=1}^{L} w_l \bar{y}'_{lsys})$. We have

$$\bar{y}'_{lsys} = w_{nl1}\bar{y}_{nl1} + w_{nl2}\bar{y}_{hl2} + w_{nl2}\beta_{hl2}(\bar{x}_{nl2} - \bar{x}_{hl2})$$

$$= \bar{y}_{hhl} + \frac{[\frac{1}{h_{l2}}(\sum_{j=1}^{h_{l2}} x_{lij} y_{lij}) - \bar{y}_{hl2}\bar{x}_{hl2}]}{[\frac{1}{h_{l2}}(\sum_{j=1}^{h_{l2}} x_{lij}^2) - \bar{x}_{hl2}^2]}(w_{nl2}\bar{x}_{nl2} - w_{nl2}\bar{x}_{hl2}).$$

Let

$$PAR = (\bar{Y}_l, \bar{Y}_{l2}, \bar{X}_{l2}, W_{l2}\bar{X}_{l2}, \sum_{i,j\in S_2} x_{lij} y_{lij}, \sum_{i,j\in S_2} x_{lij}^2).$$

where $\bar{Y}_{l2} = \frac{1}{N_{l2}} \sum_{i,j\in S_2} y_{lij}$ and $\bar{X}_{l2} = \frac{1}{N_{l2}} \sum_{i,j\in S_2} x_{lij}$. S_2 is the set of all the non-responding units in the l^{th} stratum. N_{l2} is the number of non-responding units in the l^{th} stratum.

Thus, we have

$$\left(\frac{\partial \bar{y}'_{lsys}}{\partial \bar{y}_{hhl}}\right)_{PAR} = 1,$$

$$\left(\frac{\partial \bar{y}'_{lsys}}{\partial w_{nl2}\bar{x}_{nl2}}\right)_{PAR} = [\frac{1}{N_{l2}}(\sum_{j=1}^{N_{l2}} x_{lij} y_{lij}) - \bar{Y}_{l2}\bar{X}_{l2}][\frac{1}{N_{l2}}(\sum_{j=1}^{N_{l2}} x_{lij}^2) - \bar{X}_{l2}^2]^{-1}$$

$$= \frac{[(\sum_{j=1}^{N_{l2}} x_{lij} y_{lij}) - N_{l2}\bar{Y}_{l2}\bar{X}_{l2}]}{[(\sum_{j=1}^{N_{l2}} x_{lij}^2) - N_{l2}\bar{X}_{l2}^2]} = \frac{S_{lxy2}}{S_{lx2}^2} = \delta_l,$$

$$\left(\frac{\partial \bar{y}'_{lsys}}{\partial w_{nl2}\bar{x}_{hl2}}\right)_{PAR} = -[\frac{1}{N_{l2}}(\sum_{j=1}^{N_{l2}} x_{lij} y_{lij}) - \bar{Y}_{l2}\bar{X}_{l2}][\frac{1}{N_{l2}}(\sum_{j=1}^{N_{l2}} x_{lij}^2) - \bar{X}_{l2}^2]^{-1}$$

$$= -\frac{[(\sum_{j=1}^{N_{l2}} x_{lij} y_{lij}) - N_{l2}\bar{Y}_{l2}\bar{X}_{l2}]}{[(\sum_{j=1}^{N_{l2}} x_{lij}^2) - N_{l2}\bar{X}_{l2}^2]} = -\frac{S_{lxy2}}{S_{lx2}^2} = -\delta_l$$

and

$$\left(\frac{\partial \bar{y}'_{lsys}}{\partial \bar{y}_{hl2}}\right)_{PAR} = \left(\frac{\partial \bar{y}'_{lsys}}{\partial \bar{x}_{hl2}}\right)_{PAR} = \left(\frac{\partial \bar{y}'_{lsys}}{\partial(\sum_{j=1}^{h_{l2}} x_{lij} y_{lij})}\right)_{PAR} = \left(\frac{\partial \bar{y}'_{lsys}}{\partial(\sum_{j=1}^{h_{l2}} x_{lij}^2)}\right)_{PAR} = 0$$

Therefore, the MSE of $\bar{y}'_{lsys}$ becomes

$$MSE(\bar{y}'_{lsys}) = V(\bar{y}_{hhl}) + \delta_l^2 V(w_{nl2}\bar{x}_{nl2}) + \delta_l^2 V(w_{nl2}\bar{x}_{hl2}) + 2\delta_l C\text{ov}(\bar{y}_{hhl}, w_{nl2}\bar{x}_{nl2}) -$$
$$2\delta_l C\text{ov}(\bar{y}_{hhl}, w_{nl2}\bar{x}_{hl2}) - 2\delta_l^2 C\text{ov}(w_{nl2}\bar{x}_{nl2}, w_{nl2}\bar{x}_{hl2})$$

$$= \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{ly}]S_{ly}^2 + W_{l2}\frac{(g_l-1)}{n_l}S_{ly2}^2 + \delta_l^2 W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{lx2}]S_{lx2}^2 +$$
$$\delta_l^2[W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}\{1+(n_{l2}-1)\rho_{lx2}\}S_{lx2}^2 + W_{l2}\left(\frac{g_l-1}{n_l}\right)S_{lx2}^2]+$$
$$2\delta_l W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{ly2}]^{1/2}[1+(n_{l2}-1)\rho_{lx2}]^{1/2}\rho_{l2}S_{ly2}S_{lx2} -$$

$$2\delta_l[W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}\{1+(n_{l2}-1)\rho_{ly2}\}^{1/2}\{1+(n_{l2}-1)\rho_{lx2}\}^{1/2}\rho_{l2}S_{ly2}S_{lx2} +$$
$$W_{l2}\left(\frac{g_l-1}{n_l}\right)S_{lxy2}] - 2\delta_l^2 W_{l2}^2\frac{(N_{l2}-1)}{N_{l2}n_{l2}}[1+(n_{l2}-1)\rho_{lx2}]S_{lx2}^2 .$$

Consequently, we have

$$MSE(\bar{y}'_{lsys}) = \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{ly}]S_{ly}^2 + \frac{(g_l-1)}{n_l}W_{l2}(1-\rho_{l2}^2)S_{ly2}^2 \tag{4.1}$$

$$V(\bar{x}_{lsys}) = \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{lx}]S_{lx}^2 \tag{4.2}$$

$$Cov(\bar{y}'_{lsys}, \bar{x}_{lsys}) = \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{ly}]^{1/2}[1+(n_l-1)\rho_{lx}]^{1/2}\rho_l S_{ly}S_{lx} \tag{4.3}$$

$$MSE(\sum_{l=1}^{L} w_l \bar{y}'_{lsys}) = \sum_{l=1}^{L} w_l^2[\frac{(N_l-1)}{N_l n_l}\{1+(n_l-1)\rho_{ly}\}S_{ly}^2 + \frac{(g_l-1)}{n_l}W_{l2}(1-\rho_{l2}^2)S_{ly2}^2] \tag{4.4}$$

$$V(\sum_{l=1}^{L} w_l \bar{x}_{lsys}) = \sum_{l=1}^{L} w_l^2\frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{lx}]S_{lx}^2 \tag{4.5}$$

$$Cov(\sum_{l=1}^{L} w_l \bar{y}'_{lsys}, \sum_{l=1}^{L} w_l \bar{x}_{lsys}) = \sum_{l=1}^{L} w_l^2\frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{ly}]^{1/2}[1+(n_l-1)\rho_{lx}]^{1/2}\rho_l S_{ly}S_{lx} \tag{4.6}$$

where S_{lx}^2 is the population mean square of the entire group in the l^{th} stratum under auxiliary variable. $\rho_{lx}\left(=\frac{E(x_{lij}-\bar{X}_l)(x_{lkj}-\bar{X}_l)}{E(x_{lij}-\bar{X}_l)^2}\right)$ is the intra-class correlation coefficient between the units of the same systematic sample in the l^{th} stratum under auxiliary variable. $\bar{X}_l$ is the population mean of the auxiliary variable in the l^{th} stratum. $\rho_l = \frac{E(y_{lij}-\bar{Y}_l)(x_{lij}-\bar{X}_l)}{\left(E(y_{lij}-\bar{Y}_l)^2 E(x_{lij}-\bar{X}_l)^2\right)^{1/2}}$ is the correlation coefficient between study and auxiliary variables in the l^{th} stratum.

Now, we have

$$\left(\frac{\partial \bar{y}^{*}_{cal(1)}}{\partial(\sum_{l=1}^{L} w_l \bar{y}'_{lsys})}\right)_{\bar{X},\bar{Y}} = 1 \quad \text{and} \quad \left(\frac{\partial \bar{y}^{*}_{cal(1)}}{\partial(\sum_{l=1}^{L} w_l \bar{x}_{lsys})}\right)_{\bar{X},\bar{Y}} = -\frac{\bar{Y}}{\bar{X}} = -R$$

where

$$R = \frac{\bar{Y}}{\bar{X}}.$$

Thus, the expression for the MSE of the proposed calibration estimator $\bar{y}^{*}_{cal(1)}$ is represented as

$$MSE(\bar{y}^{*}_{cal(1)}) = MSE(\sum_{l=1}^{L} w_l \bar{y}'_{lsys}) + R^2 V(\sum_{l=1}^{L} w_l \bar{x}_{lsys}) - 2RCov(\sum_{l=1}^{L} w_l \bar{y}'_{lsys}, \sum_{l=1}^{L} w_l \bar{x}_{lsys})$$

$$= \sum_{l=1}^{L} w_l^2 [\frac{(N_l-1)}{N_l n_l}\{1+(n_l-1)\rho_{ly}\}S_{ly}^2 + \frac{(g_l-1)}{n_l} W_{l2}(1-\rho_{l2}^2)S_{ly2}^2] + R^2 \sum_{l=1}^{L} w_l^2 \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{lx}]S_{lx}^2 -$$

$$2R\sum_{l=1}^{L} w_l^2 \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{ly}]^{1/2}[1+(n_l-1)\rho_{lx}]^{1/2}\rho_l S_{ly} S_{lx} \qquad (4.7)$$

Now, let

$PAR^{*} = (\bar{Y}, \bar{X}, \sum_{l=1}^{L} w_l \bar{Y}_l \bar{X}_l, \sum_{l=1}^{L} w_l \bar{X}_l^2)$. Therefore, we have

$$\left(\frac{\partial \bar{y}^{*}_{cal(2)}}{\partial(\sum_{l=1}^{L} w_l \bar{y}'_{lsys})}\right)_{PAR^{*}} = 1,$$

$$\left(\frac{\partial \bar{y}^{*}_{cal(2)}}{\partial(\sum_{l=1}^{L} w_l \bar{x}_{lsys})}\right)_{PAR^{*}} = -[(\sum_{l=1}^{L} w_l \bar{Y}_l \bar{X}_l) - 2(\sum_{l=1}^{L} w_l \bar{Y}_l)(\sum_{l=1}^{L} w_l \bar{X}_l) + (\sum_{l=1}^{L} w_l \bar{Y}_l)\bar{X}][(\sum_{l=1}^{L} w_l \bar{X}_l^2) - (\sum_{l=1}^{L} w_l \bar{X}_l)^2]^{-1}$$

$$= -[(\sum_{l=1}^{L} w_l \bar{Y}_l \bar{X}_l) - (\sum_{l=1}^{L} w_l \bar{Y}_l)(\sum_{l=1}^{L} w_l \bar{X}_l)][(\sum_{l=1}^{L} w_l \bar{X}_l^2) - (\sum_{l=1}^{L} w_l \bar{X}_l)^2]^{-1} = \gamma$$

and

$$\left(\frac{\partial \bar{y}^{*}_{cal(2)}}{\partial(\sum_{l=1}^{L} w_l \bar{y}'_{lsys} \bar{x}_{lsys})}\right)_{PAR^{*}} = \left(\frac{\partial \bar{y}^{*}_{cal(2)}}{\partial(\sum_{l=1}^{L} w_l \bar{x}_{lsys}^2)}\right)_{PAR^{*}} = 0.$$

Thus, the expression for the MSE of the proposed calibration estimator $\bar{y}^*_{cal(2)}$ is given by

$$MSE(\bar{y}^*_{cal(2)}) = MSE(\sum_{l=1}^{L} w_l \bar{y}'_{lsys}) + \gamma^2 V(\sum_{l=1}^{L} w_l \bar{x}_{lsys}) + 2\gamma Cov(\sum_{l=1}^{L} w_l \bar{y}'_{lsys}, \sum_{l=1}^{L} w_l \bar{x}_{lsys})$$

$$= \sum_{l=1}^{L} w_l^2 [\frac{(N_l-1)}{N_l n_l}\{1+(n_l-1)\rho_{ly}\}S_{ly}^2 + \frac{(g_l-1)}{n_l} W_{l2}(1-\rho_{l2}^2)S_{ly2}^2] + \gamma^2 \sum_{l=1}^{L} w_l^2 \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{lx}]S_{lx}^2 +$$

$$2\gamma \sum_{l=1}^{L} w_l^2 \frac{(N_l-1)}{N_l n_l}[1+(n_l-1)\rho_{ly}]^{1/2}[1+(n_l-1)\rho_{lx}]^{1/2}\rho_l S_{ly} S_{lx} \quad (4.8)$$

5. Empirical Study

5.1 Through Real Data

To examine the performance of the proposed calibration estimators, we have performed an empirical study through real data which were collected in a pilot survey for estimating the extent of cultivation and production of fresh fruits in three districts of Uttar Pradesh in the year 1976-77 [Daroga and Chaudhary (1986), Page 162]. Table 1 and Table 2 depict the particulars of parameters and statistics.

Table 1: Particulars of Parameter.

Stratum No.	Total no. of villages (N_l)	Total area (in hect.) under orchard (X_l)	No. of villages in sample (n_l)	Area under orchards in hect. (X_l)	Total no. of trees (Y_l)
1	985	11253	6	10.63, 9.90, 1.45, 3.38, 5.17, 10.35	747, 719, 78, 201, 311, 448
2	2196	25115	8	14.66, 2.61, 4.35, 9.87, 2.42,5.60, 4.70, 36.75	580, 103, 316, 739, 196, 235, 212, 1646
3	1020	18870	11	11.60, 5.29, 7.94, 7.29, 8.00, 1.20, 11.50, 7.96, 23.15, 1.70, 2.01	488, 227, 374, 491, 499, 50, 455, 47, 879, 115, 115

From the above data, we calculate the following statistic and due to the unavailability of raw data, we considered the assumed values for ρ_{lx} and ρ_{ly}.

Table 2: Particulars of Statistic.

Str. No.	$\bar{x}_l$	$\bar{y}_l$	s_{lx}^2	s_{ly}^2	s_{lxy}	$s_{ly2}^2 = \frac{3}{4} s_{ly}^2$	ρ_{lx}	ρ_{ly}	$\rho_{l2} = \frac{3}{4}\rho_{ly}$
1	6.81	417.33	15.97	74775.47	1007.05	56081.60	0.82	0.90	0.68
2	10.12	503.38	132.66	259113.40	5709.16	194335	0.89	0.84	0.63
3	7.97	340.00	38.44	65885.60	1404.71	49414.2	0.78	0.80	0.60

Table 3 shows the VAR/MSE of the estimators $\bar{y}^*$, $\bar{y}^*_{cal(1)}$ and $\bar{y}^*_{cal(2)}$ for the different values of non-response rate W_{l2} and sub-sample rate g_l. The percentage relative efficiency (PRE) of the proposed calibration estimators $\bar{y}^*_{cal(1)}$ and $\bar{y}^*_{cal(2)}$ with respect to the estimator $\bar{y}^*$ has also been given. The PRE has been computed using the following formula:

$$PRE\left(\bar{y}^*_{cal(j)}\right) = \frac{VAR(\bar{y}^*)}{MSE\left(\bar{y}^*_{cal(j)}\right)} \times 100; j = 1,2$$

Table 3: VAR/**MSE and PRE of Estimators** $\bar{y}^*$ **,** $\bar{y}^*_{cal(1)}$ **and** $\bar{y}^*_{cal(2)}$**.**

(Parentheses figures show the PRE.)

W_{l2} $\forall l$	g_l $\forall l$	VAR/MSE		
		$\bar{y}^*$	$\bar{y}^*_{cal(1)}$	$\bar{y}^*_{cal(2)}$
0.1	2.0	68289.72 (100)	7785.06 (877.19)	6228.01 (1096.49)
	2.5	68660.51 (100)	8007.66 (857.44)	6450.61 (1064.40)
	3.0	69031.29 (100)	8230.26 (838.75)	6673.21 (1034.45)
0.2	2.0	69031.29(100)	8230.26 (838.75)	6673.21 (1034.45)
	2.5	69772.86 (100)	8675.46 (804.26)	7118.41 (980.18)
	3.0	70514.44 (100)	9120.66 (773.13)	7563.61 (932.29)
0.3	2.0	69772.86 (100)	8675.46 (804.26)	7118.41 (980.18)
	2.5	70885.22(100)	9343.26 (758.68)	7786.21 (910.39)
	3.0	71997.58 (100)	10011.07 (719.18)	8454.01 (851.64)
0.4	2.0	70514.44 (100)	9120.66 (773.13)	7563.61 (932.29)
	2.5	71997.58 (100)	10011.07 (719.18)	8454.01 (851.64)
	3.0	73480.73 (100)	10901.47 (674.04)	9344.42 (786.36)

5.2 Through Simulated Data

To get some idea about the efficiency of the proposed calibration estimators, we have generated an artificial data set by using the procedure given by Reddy et al. (2010). Here, the population consists of six strata with respective sizes 800, 600, 300, 500, 200 and 400. The data under study variable Y for each stratum have been generated from Normal distribution with certain mean and standard deviation (S. D.). We have generated the data under a dummy variable Z for each stratum with the same distribution as that of Y. Finally, we have generated the data under the auxiliary variable X for each stratum using the transformation $X_l = \rho_l Y_l + \sqrt{1-\rho_l^2}\, Z_l$. The assumed means, standard deviations and correlation coefficients are given in Table 4.

Table 4: Particulars of Population.

Stratum No. (l)	Stratum Population for Study Variable Y	Stratum Size (N_l)	Correlation Coefficient (ρ_l)
1	$N(Mean = 21, S.D. = 6.5)$	800	0.8903
2	$N(Mean = 30, S.D. = 6.6)$	600	0.9161
3	$N(Mean = 24, S.D. = 6.7)$	300	0.8522
4	$N(Mean = 33, S.D. = 6.8)$	500	0.8927
5	$N(Mean = 27, S.D. = 6.9)$	200	0.8980
6	$N(Mean = 36, S.D. = 7.0)$	400	0.8956

Here N =2800, n =840. To determine the sample size for each stratum, proportional allocation has been used. Now, we select a sample of the specified size from each stratum using a systematic sampling scheme and then select a sub-sample from the non-responding units according to the sub-sample rate $g_l (= 2.0, 2.5, 3.0)$. The number of responding units in each stratum has been fixed according to the non-response rate W_{l2} $(= 0.1, 0.2, 0.3, 0.4)$. Subsequently, the estimates of the population mean $\overline{Y}$ have been obtained using the estimators $\overline{y}^*$, $\overline{y}^*_{cal(1)}$ and $\overline{y}^*_{cal(2)}$. We simulate the process of selecting the sample/sub-sample and obtaining the estimate 5000 times. Ultimately, we have computed the approximate VAR/MSE (AVAR/AMSE) of the estimators $\overline{y}^*$, $\overline{y}^*_{cal(1)}$ and $\overline{y}^*_{cal(2)}$ using the following formulae:

$$AVAR\left(\overline{y}^*\right) = \frac{1}{5000}\sum_{t=1}^{5000}\left(\overline{y}^*_t - \overline{Y}\right)^2 \; ; \; t = 1, 2, ..., 5000$$

$$AMSE\left(\overline{y}^*_{cal(1)}\right) = \frac{1}{5000}\sum_{t=1}^{5000}\left(\overline{y}^*_{cal(1)t} - \overline{Y}\right)^2 \; ; \; t = 1, 2, ..., 5000$$

$$AMSE\left(\overline{y}^*_{cal(2)}\right) = \frac{1}{5000}\sum_{t=1}^{5000}\left(\overline{y}^*_{cal(2)t} - \overline{Y}\right)^2 \; ; \; t = 1, 2, ..., 5000$$

Table 5 depicts the AVAR/AMSE of the estimators $\overline{y}^*$, $\overline{y}^*_{cal(1)}$ and $\overline{y}^*_{cal(2)}$ for the different choices of non-response rate W_{l2} and sub-sample rate g_l. The percentage relative efficiency (PRE) of the proposed calibration estimators $\bar{y}^*_{cal(1)}$ and $\bar{y}^*_{cal(2)}$ with respect to the estimator $\bar{y}^*$ has also been given. The PRE has been computed using the following formula:

$$PRE\left(\bar{y}^*_{cal(j)}\right) = \frac{AVAR(\bar{y}^*)}{AMSE(\bar{y}^*_{cal(j)})} \times 100; \, j = 1, 2$$

Table 5: AVAR/AMSE and PRE of Estimators $\bar{y}^*$, $\bar{y}^*_{cal(1)}$ and $\bar{y}^*_{cal(2)}$.

W_{l2} $\forall l$	g_l $\forall l$	AVAR/AMSE		
		$\bar{y}^*$	$\bar{y}^*_{cal(1)}$	$\bar{y}^*_{cal(2)}$
0.1	2.0	0.046484 (100)	0.005487 (847.18)	0.005094 (912.54)
	2.5	0.049151 (100)	0.005856 (839.34)	0.005479 (897.04)
	3.0	0.052342 (100)	0.006882 (760.54)	0.006552 (798.91)
0.2	2.0	0.052638 (100)	0.007102 (741.15)	0.006612 (796.07)
	2.5	0.055034 (100)	0.007555 (728.45)	0.007078 (777.53)
	3.0	0.063449 (100)	0.008888 (713.91)	0.008423 (753.30)
0.3	2.0	0.057096 (100)	0.008517 (670.40)	0.008015 (712.35)
	2.5	0.064169 (100)	0.010254 (625.78)	0.009732 (659.35)
	3.0	0.073473 (100)	0.012344 (595.19)	0.011798 (622.75)
0.4	2.0	0.062038 (100)	0.009478 (654.53)	0.008934 (694.43)
	2.5	0.069505 (100)	0.011523 (603.19)	0.010947 (634.94)
	3.0	0.080141 (100)	0.013929 (575.37)	0.013302 (602.46)

(Parentheses figures show the PRE)

6. Analysis of Tables

Table 3 and Table 5 show that as we increase the non-response rate W_{l2}, MSE/AMSE increases and PRE decreases and if we decrease the sub-sample size h_{l2} (increases sub-sample rate g_l) MSE/AMSE increases, and hence PRE decreases. Also, we have

$MSE/AMSE(\bar{y}^*_{cal(2)}) < MSE/AMSE(\bar{y}^*_{cal(1)}) < MSE/AMSE(\bar{y}^*)$ and

$PRE(\bar{y}^*_{cal(2)}) > PRE(\bar{y}^*_{cal(1)}) > PRE(\bar{y}^*)$

So, in between proposed calibration estimators $\bar{y}^*_{cal(2)}$ and $\bar{y}^*_{cal(1)}$ perform better than $\bar{y}^*$.

7. Concluding Remarks

In this paper, we have proposed some improved calibration estimators of the population mean under stratified systematic sampling in the presence of non-response. The information on an auxiliary variable is utilized to compensate for reduced precision due to non-response. The expressions for the MSE of the proposed calibration estimators have been derived using the Taylor linearization technique. An empirical study based on simulated and real data has been carried out to check the efficiency of the proposed calibration estimators. The MSE/AMSE has been used as a tool to check the precision of the proposed calibration estimators over the usual Hansen-Hurwitz (1946) estimator under stratified systematic sampling. The study reveals that the proposed calibration estimators provide better results as compared to the existing one and hence the proposed calibration estimators can play an important role in the real-life situations.

References

Chaudhary M. K., Malik S., Singh J. and R. Singh (2012): A general family of estimators for estimating population mean in systematic sampling using auxiliary information in the presence of missing observations, *Journal of Rajasthan Statistical Association (JRSA)*, **1(2)**, 1-8.

Chaudhary M. K. and T. Dutta (2023): Calibration-based improvements in estimating the mean of a stratified population with scrambled response on second call under non-response, *Journal of Applied Mathematics and Statistics*, **62(1)**, 22-38.

Clement E. P. (2017): Estimation of population mean in calibration ratio-type estimator under systematic sampling, *Elix Stat*, **106**, 46480-46486.

Gautam A. K., Sharma M. K. and B. V. S. Sisodia (2020): Development of calibration estimator of population mean under non-response, *International Journal of Chemical Studies*, **8(5)**, 1360-1371.

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

Khare B. B., Srivastava U. and K. Kumar (2012): Chain type estimators for ratio of two population means using auxiliary characters in the presence of non-response, *Journal of Scientific Research*, **56**, 183-96.

Kumar S. and S. Bhougal, (2011): Estimation of the population mean in presence of non-response, *Communications for Statistical Applications and Methods*, **18(4)**, 537-548.

Raman R. K., Sud U. C. and H. Chandra (2016): Calibration approach-based product estimator of finite population total with sub-sampling of non-respondents under single and two-phase sampling, *Communications in Statistics-Simulation and Computation*, **45(8)**, 2965-2980.

Reddy M. K., Rao K. R. and N. K. Boiroju (2010): Comparison of ratio estimators using Monte Carlo simulation, *International Journal of Agriculture and Statistical Sciences*, **6(2)**, 517-527.

Singh D. and F. S. Chaudhary (1986): Theory and analysis of sample survey designs, *John Wiley & Sons,* Inc. New York.

Singh H. P. and S. Kumar (2010): Estimation of mean in presence of non-response using two phase sampling scheme, *Statistical Papers*, **51(3)**, 559-582.