Statistics and Applications {ISSN 2454-7395 (online)}
Volume 22, No. 2, 2024 (New Series), pp 121–132
http://www.ssca.org.in/journal

Calibration Approach for Estimating Mean of a Stratified Population in the Presence of Non-response

Manoj K. Chaudhary, Anil Prajapati and Basant K. Ray
Department of Statistics, Institute of Science, Banaras Hindu University, Varanasi, India

Received: 31 December 2022; Revised: 26 November 2023; Accepted: 03 December 2023

Abstract

Calibration approach is a systematic way of including the auxiliary information in order to increase the precision of the estimates of a population parameter. In this paper, we have suggested some calibration estimators for estimating the mean of a stratified population under non-response. An efficient use of suitable auxiliary information has been elaborated to obtain a better estimate of the population mean under certain conditions. We have obtained new stratum weights for which the variance of the suggested calibration estimators would achieve its minimum. An empirical study has also been carried out to verify the theoretical outcomes.

Key words: Calibration approach; Auxiliary information; Stratified random sampling; Population mean; Non-response.

AMS Subject Classifications: 62K05, 05B05

1. Introduction

Calibration offers a methodical approach to incorporate the auxiliary information in increasing the precision of the estimates. The concept behind the calibration is to find out the new calibrated weights in such a way that the mean square error of the estimators would be minimized. To construct the new calibrated weights, the chi-square distance measure and some calibration constraints based on auxiliary information can be utilized. Deming and Stephan (1940) were the first to pick up the idea of calibration in a sample survey. Deville and Särndal (1992) adopted the idea of calibration approach in estimating the population parameters. Särndal (2007) provides a deep study of the calibration approach, including methods for avoiding extreme weights, estimation of complex parameters and estimation under a complex sampling design. Kim and Park (2010) prove that an instrumental variable calibration estimator and a functional-form calibration estimator are asymptotically equivalent.

Consider a sample 's' of size n which is drawn from a population of N units by simple random sampling without replacement (SRSWOR) scheme. The study variable Y is observed for each unit in the sample hence the observation y_i is known for all $i \in s$ since

Corresponding Author: Manoj K. Chaudhary
Email: mk15@bhu.ac.in

the values $y_1, y_2, ..., y_N$are known for the entire population. To estimate the population total $Y^* = \sum_{i=1}^{N} y_i$, Deville and Särndal (1992) have suggested the calibration estimator, which is constructed as $\widehat{Y}= \sum_{i\in S} p_i y_i$, where the calibration weights p_i's are chosen to minimize their average distance from the basic design weights $d_i = 1/\pi_i$ that are used in the Horvitz and Thompson (1952) estimator given by

$$\widehat{Y}_{HT}= \sum_{i\in S} d_i y_i \tag{1}$$

subject to the constraint

$$\sum_{i\in S} p_i x_i = X^* \tag{2}$$

where $X^*$is the known population total for the auxiliary variable Xwhich is observed for each unit in the sample hence the observation x_i is known for all $i \in s$. The most common distance measure is given as

$$\phi = \sum_{i\in S} \frac{(d_i - p_i)^2}{d_i q_i} \tag{3}$$

where q_i's are known positive weights uncorrelated withd_i. Then the resulting calibration estimator is given as follows:

$$\widehat{Y}= \sum_{i\in S} p_i y_i =\widehat{Y}_{HT} + \widehat{B} \left(X^* - \widehat{X}_{HT}\right) \tag{4}$$

where$\widehat{B}= \left[\sum_{i\in S} d_i q_i x_i^2\right]^{-1} \left[\sum_{i\in S} d_i q_i x_i y_i\right]$ and $\widehat{X}_{HT}= \sum_{i\in S} d_i x_i$. The definition of $\widehat{Y}$is equivalent to a generalized estimator with the choice of q_i.

The authors such as Rao (1994), Estevao and Särndal (2009), Sud *et al.* (2014), Han (2018), Gautam *et al.* (2020), Jaiswal *et al.* (2023), Singh *et al.* (2023) and others have contributed a lot to the survey sampling in estimating the population parameters with a view to justify the concept of calibration approach.

2. Literature reviews under stratified random sampling

Consider a finite population $U = (U_1, U_2, ..., U_N)$ of size $N$and it is divided into k homogeneous groups (called strata). Let the size of i^{th} stratum be $N_i\,(i = 1, 2, ..., k)$ and hence $\sum_{i=1}^{k} N_i = N$. Let Y and X be the study and auxiliary variables with respective population means $\overline{Y}$and $\overline{X}$. A sample of size n_i is drawn by SRSWOR scheme from the i^{th} stratum such that $\sum_{i=1}^{k} n_i = n$. Let $(y_{ij,} x_{ij})$ be the observed values of (Y, X) on the j^{th} unit in the i^{th} stratum $(j = 1, 2, ..., N_i)$. The classical unbiased estimator of the population mean $\overline{Y}$ is given by

$$\overline{y}_{st} = \sum_{i=1}^{k} w_i \overline{y}_i \tag{5}$$

where $\overline{y}_i$ is the mean based on n_i units for the study variable and $w_i = \frac{N_i}{N}$.

In the availability of auxiliary information, Singh *et al.* (1998) suggested a new calibration estimator of the population mean $\overline{Y}$ as

$$\overline{y}_{c,st} = \sum_{i=1}^{k} w_i^* \overline{y}_i \tag{6}$$

where w_i^* is a new calibrated weight such that it minimizes the chi- square distance function

$$\varphi = \sum_{i=1}^{k} \frac{\left(w_i^* - w_i\right)^2}{w_i q_i} \tag{7}$$

subject to the calibration constraint

$$\sum_{i=1}^{k} w_i^* \overline{x}_i = \overline{X} \tag{8}$$

where q_i is the tuning parameter for the i^{th} stratum and $\overline{x}_i$ is the mean based on n_i units for the auxiliary variable.

The calibration constraint given in equation (8) is similar as used by Dupont (1995) and Hidiroglou and Särndal (1998) for two-phase sampling design. Minimization of chi-square distance function given in equation (7) subject to the calibration constraint (8) leads to the calibrated weights

$$w_i^* = w_i + \frac{w_i q_i \overline{x}_i}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2} \left[\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right] \tag{9}$$

Substituting the value of w_i^* from equation (9) into equation (6), one can get combined regression-type estimator given by

$$\overline{y}_{c,st} = \sum_{i=1}^{k} w_i \overline{y}_i + \frac{\sum_{i=1}^{k} w_i q_i \overline{x}_i \overline{y}_i}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2} \left[\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right] \tag{10}$$

An estimator of the variance of the calibration estimator $\overline{y}_{c,st}$ is represented as

$$\widehat{V}\left(\overline{y}_{c,st}\right) = \sum_{i=1}^{k} \frac{w_i^2 \left(1 - f_i\right) s_{ei}^2}{n_i} \tag{11}$$

where $s_{ei}^2 = \frac{1}{n_i - 1} \sum_j^{n_i} e_{ij}^2$, $f_i = \frac{n_i}{N_i}$, $e_{ij} = \left(y_{ij} - \overline{y}_i\right) - b\left(x_{ij} - \overline{x}_i\right)$and $b = \frac{\sum_{i=1}^{k} w_i q_i \overline{x}_i \overline{y}_i}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2}$.

Moreover, there are several authors who have implemented the notion of calibration approach in estimating the parameters of a stratified population. Tracy *et al.* (2003), Kim *et al.* (2007), Koyuncu and Kadilar (2013, 2014), Clement and Enang (2015), Nidhi *et al.* (2017), Rao *et al.* (2017), Ozgul (2019) and others have proposed a number of calibration estimators in stratified random sampling.

The occurrence of non-response is inherent in sample surveys. Rubin (1976) delineated three key concepts, *viz.*, (i) Missing at Random (MAR), (ii) Missing Completely at Random (MCAR) and (iii) Observed at Random (OAR). MAR method addresses non-response scenarios by assuming that missing data occur randomly and depend only on observed information. Utilizing Multiple Imputation (MI), this technique generates multiple plausible imputations, which reflect the uncertainty associated with missing values. MCAR is a category of missing data mechanism in which the likelihood of a data point being missing has no connection to either observed or unobserved data. In the context of OAR, the data adhere to this pattern if, for every conceivable missing data value, the probability of the observed missing pattern, given both observed and unobserved data, is independent of the specific values within the observed data. It is to be noted that the non-response error is not so important if the characteristics of the non-responding units are similar to those of the responding units. But, such similarity of characteristics between the responding and non-responding units is not always attained in custom. In such a situation, it is much difficult to get the précised estimates of the parameters. To deal with the problem of non-response, Hansen and Hurwitz (1946) suggested a technique of sub-sampling of non-respondents. Later on, Khare (1987), Chaudhary *et al.* (2012, 2018) have discussed the problem of non-response in estimating the parameters of a stratified population.

It is to be mentioned that there are two types of non-response; (i) unit non-response and (ii) item non-response. In the subsequent sections, we have tried to propose an efficient calibration method of estimation of the population mean $\overline{Y}$ in stratified random sampling utilizing the information on an auxiliary variable X under unit non-response. The calibration estimators have been pioneered out under the situation in which the knowledge about the population mean of the auxiliary variable is available in advance. It is further assumed that the study variable is suffering from the non-response, whereas the auxiliary variable does not suffer from the non-response. The theoretical facts have been demonstrated through an empirical study.

3. Proposed calibration estimators

In the presence of non-response, the sampling strategy given in section-2 has been extended to the further process. It is noted that out of n_i units, n_{i1} units respond and n_{i2}units do not respond on the study variable Y. Adopting Hansen and Hurwitz (1946) technique of sub-sampling of non-respondents, a sub-sample of $h_{i2}\left(=\frac{n_{i2}}{g_i}; g_i > 1\right)$ units is selected from the n_{i2} non-responding units using SRSWOR scheme and information is collected from all the h_{i2} units. The usual estimator of the population mean $\overline{Y}$ under non-response (without using auxiliary information) is given by

$$\overline{y}_{st}^* = \sum_{i=1}^{k} w_i \overline{y}_i^* \tag{12}$$

where $\overline{y}_i^* = \frac{n_{i1}\overline{y}_{ni1} + n_{i2}\overline{y}_{hi2}}{n_i}$. $\overline{y}_{ni1}$ and $\overline{y}_{hi2}$ are respectively the means based on n_{i1} responding units and h_{i2} non-responding units for study variable in the i^{th} stratum.

The estimate of the variance of the estimator $\overline{y}_{st}^*$ is given as

$$V\left[\overline{y}_{st}^*\right] = \sum_{i=1}^{k} \frac{w_i^2 (1 - f_i)}{n_i} s_{yi}^{*2} + \sum_{i=1}^{k} \frac{w_i^2 (g_i - 1) W_{i2}}{n_i} s_{yi(2)}^2 \tag{13}$$

where $s_{yi}^{*2} = \frac{1}{n_i^* - 1}\sum_j^{n_i^*}\left(y_{ij} - \overline{y}_i^*\right)^2$, $s_{yi(2)}^2 = \frac{1}{h_{i2} - 1}\sum_j^{h_{i2}}\left(y_{ij} - \overline{y}_{hi2}\right)^2$, $n_i^* = n_{i1} + h_{i2}$ and $W_{i2}\left(=\frac{N_{i2}}{N_i}\right)$ is the non-response rate in the population for the i^{th} stratum.

Here, we have considered the situation in which the non-response occurs on the study variable, whereas the auxiliary variable is free from the non-response. In this situation, we have suggested some calibration estimators of the population mean $\overline{Y}$ when the information about the population mean $\overline{X}$ of the auxiliary variable is known in advance. Following Singh *et al.* (1998), we now propose a calibration estimator of the population mean $\overline{Y}$ in the presence of non-response as

$$\overline{y}^*_{st(C)} = \sum_{i=1}^{k} \delta_i^* \overline{y}_i{}^* \tag{14}$$

where δ_i^* is an adjusted calibrated weight for the i^{th} stratum.

In order to get the optimum value of calibrated weight δ_i^*, we now minimize the chi-square distance function

$$\varphi^* = \sum_{i=1}^{k} \frac{(\delta_i^* - w_i)^2}{w_i q_i} \tag{15}$$

subject to the calibration constraint

$$\sum_{i=1}^{k} \delta_i^* \overline{x}_i = \overline{X} \tag{16}$$

Let us define the Lagrange function

$$L = \sum_{i=1}^{k} \frac{(\delta_i^* - w_i)^2}{w_i q_i} - 2\lambda \left(\sum_{i=1}^{k} \delta_i^* \overline{x}_i - \overline{X} \right) \tag{17}$$

where λ is the Lagrange multiplier.

Differentiating the equation (17) with respect to δ_i^* and equating the derivative to zero, we get

$$\frac{\partial L}{\partial \delta_i^*} = 2\frac{(\delta_i^* - w_i)}{w_i q_i} - 2\lambda \overline{x}_i = 0$$

$$\Rightarrow \delta_i^* = w_i + \lambda w_i q_i \overline{x}_i \tag{18}$$

Putting the value δ_i^* from equation (18) into the equation (16), we have

$$\lambda = \frac{\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2} \tag{19}$$

Substituting the value of λ form equation (19) into equation (18), we get the optimum calibrated weights as

$$\delta_i^* = w_i + \frac{w_i q_i \overline{x}_i}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2} \left[\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i \right] \tag{20}$$

Putting the value of δ_i^* from equation (20) into the equation (14), the proposed calibration estimator becomes

$$\overline{y}^*_{st(C)} = \sum_{i=1}^{k} w_i \overline{y}_i{}^* + \frac{\sum_{i=1}^{k} w_i q_i \overline{x}_i \overline{y}_i{}^*}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2} \left[\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i \right] \tag{21}$$

The estimators of the bias and variance of the calibration estimator $\overline{y}^*_{st(C)}$ are respectively given by

$$\widehat{B}\left(\overline{y}^*_{st(C)}\right) = \sum_{i=1}^{k} w_i b^* \overline{x}_i \left[\frac{N_i\left(N_i-n_i\right)}{\left(N_i-1\right)\left(N_i-2\right)} \cdot \frac{1}{n_i \overline{x}_i}\left\{\frac{\widehat{\mu}_{30i}}{s^2_{xi}} - \frac{\widehat{\mu}_{21i}}{s^*_{xyi}}\right\} + \frac{W_{i2}\left(g_i-1\right)}{n_i \overline{x}_i}\left\{\frac{\widehat{\mu}_{30i(2)}}{s^2_{xi}} - \frac{\widehat{\mu}_{21i(2)}}{s^*_{xyi}}\right\}\right] \tag{22}$$

$$\widehat{V}\left(\overline{y}^*_{st(C)}\right) = \sum_{i=1}^{k} \frac{w_i^2\left(1-f_i\right)}{n_i}\left(s^{*2}_{yi} + b^{*2} s^2_{xi} - 2b^* s^*_{xyi}\right) + \sum_{i=1}^{k} \frac{w_i^2\left(g_i-1\right) W_{i2}}{n_i} s^2_{yi(2)} \tag{23}$$

where $b^* = \frac{\sum_{i=1}^{k} w_i q_i \overline{x}_i \overline{y}_i^{\,*}}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2}$, $s^2_{xi} = \frac{1}{n_i-1}\sum_j^{n_i}\left(x_{ij}-\overline{x}_i\right)^2$, $s^*_{xyi} = \frac{1}{n_i^*-1}\sum_j^{n_i^*}\left(x_{ij}-\overline{x}_i^*\right)\left(y_{ij}-\overline{y}_i^*\right)$, $\overline{x}_i^* = \frac{n_{i1}\overline{x}_{ni1}+n_{i2}\overline{x}_{hi2}}{n_i}$
and $n_i^* = n_{i1} + h_{i2}$. $\overline{x}_{ni1}$ and $\overline{x}_{hi2}$ are respectively the means based on n_{i1} responding units and h_{i2} non-responding units for auxiliary variable in the i^{th} stratum.

$\widehat{\mu}_{30i} = \frac{1}{n_i-1}\sum_j^{n_i}\left(x_{ij}-\overline{x}_i\right)^3$, $\widehat{\mu}_{21i} = \frac{1}{n_i^*-1}\sum_j^{n_i^*}\left(x_{ij}-\overline{x}_i^*\right)^2\left(y_{ij}-\overline{y}_i^*\right)$,
$\widehat{\mu}_{30i(2)} = \frac{1}{h_{i2}-1}\sum_j^{h_{i2}}\left(x_{ij}-\overline{x}_{hi2}\right)^3$ and $\widehat{\mu}_{21i(2)} = \frac{1}{h_{i2}-1}\sum_j^{h_{i2}}\left(x_{ij}-\overline{x}_{hi2}\right)^2\left(y_{ij}-\overline{y}_{hi2}\right)$.

Particular cases:

(i) For instance, if $q_i = \frac{1}{\overline{x}_i}$, then the equation (21) reduces to the well-known combined ratio-type estimator of the population mean $\overline{Y}$under non-response, *i.e.*, $\overline{y}^*_{st(C)R} = \frac{\sum_{i=1}^{k} w_i \overline{y}_i^*}{\sum_{i=1}^{k} w_i \overline{x}_i}\overline{X}$.

(ii) Putting $q_i = 1$ into the equation (21), it reduces to the combined regression-type estimator of the population mean $\overline{Y}$ under non-response, *i.e.*, $\overline{y}^*_{st(C)Reg} = \sum_{i=1}^{k} w_i \overline{y}_i^{\,*} + \frac{\sum_{i=1}^{k} w_i \overline{x}_i \overline{y}_i^{\,*}}{\sum_{i=1}^{k} w_i \overline{x}_i^2}\left[\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right]$.

We now propose an improved calibration estimator of the population mean $\overline{Y}$ under non-response as follows:

$$\overline{y}^{**}_{st(C)} = \sum_{i=1}^{k} \delta_i^{**} \overline{y}_i^{\,*} \tag{24}$$

where δ_i^{**} is the new calibrated weight for the i^{th}stratum.

The new calibrated weight δ_i^{**} is chosen such that the chi-square type distance

$$\varphi^{**} = \sum_{i=1}^{k} \frac{\left(\delta_i^{**} - w_i\right)^2}{w_i q_i} \tag{25}$$

is minimum, subject to the constraints

$$\sum_{i=1}^{k} \delta_i^{**} \overline{x}_i = \overline{X} \tag{26}$$

$$\sum_{i=1}^{k} \delta_i^{**} = 1 \tag{27}$$

Let us consider the Lagrange function

$$\Delta = \sum_{i=1}^{k} \frac{(\delta_i^{**} - w_i)^2}{w_i q_i} - 2\phi_1 \left(\sum_{i=1}^{k} \delta_i^{**} \overline{x}_i - \overline{X} \right) - 2\phi_2 \left(\sum_{i=1}^{k} \delta_i^{**} - 1 \right) \tag{28}$$

where ϕ_1 and ϕ_2 are the Lagrange multipliers.

Differentiating the equation (28) with respect to δ_i^{**} and equating the derivative to zero, we get

$$\frac{\partial \Delta}{\partial \delta_i^{**}} = 2\frac{(\delta_i^{**} - w_i)}{w_i q_i} - 2\phi_1 \overline{x}_i - 2\phi_2 = 0$$

$$\left\{ Since \quad \frac{\partial}{\partial x} (F_1 \pm F_2 \pm \cdots \pm F_n) = \frac{\partial}{\partial x} F_1 \pm \frac{\partial}{\partial x} F_2 \pm \cdots \pm \frac{\partial}{\partial x} F_n \right\}$$

$$\Rightarrow \delta_i^{**} = w_i + w_i q_i (\phi_1 \overline{x}_i + \phi_2) \tag{29}$$

Let us put the value of δ_i^{**} from equation (29) into the equation (26). The resulting equation is given as

$$\phi_1 \sum_{i=1}^{k} w_i q_i \overline{x}_i^2 + \phi_2 \sum_{i=1}^{k} w_i q_i \overline{x}_i = \overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i \tag{30}$$

Let us now substitute the value of δ_i^{**} from equation (29) into the equation (27). The resulting equation becomes

$$\phi_1 \sum_{i=1}^{k} w_i q_i \overline{x}_i + \phi_2 \sum_{i=1}^{k} w_i q_i = 0 \tag{31}$$

The equations (30) and (31) can be written in the following matrix form:

$$A\phi = B \tag{32}$$

where $A = \begin{bmatrix} \sum_{i=1}^{k} w_i q_i \overline{x}_i^2 & \sum_{i=1}^{k} w_i q_i \overline{x}_i \\ \sum_{i=1}^{k} w_i q_i \overline{x}_i & \sum_{i=1}^{k} w_i q_i \end{bmatrix}$, $\phi = \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix}$ and $B = \begin{bmatrix} \overline{X} - \sum_{h=1}^{L} w_h \overline{x}_h \\ 0 \end{bmatrix}$.

The inverse of the matrix A is given as

$$A^{-1} = \frac{1}{|A|} \begin{bmatrix} \sum_{i=1}^{k} w_i q_i & -\sum_{i=1}^{k} w_i q_i \overline{x}_i \\ -\sum_{i=1}^{k} w_i q_i \overline{x}_i & \sum_{i=1}^{k} w_i q_i \overline{x}_i^2 \end{bmatrix}$$

where $|A| = \sum_{i=1}^{k} w_i q_i \overline{x}_i^2 \sum_{i=1}^{k} w_i q_i - \left(\sum_{i=1}^{k} w_i q_i \overline{x}_i \right)^2$.

The solution of the system of equation (32) is given by

$$\begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} = \frac{1}{|A|} \begin{bmatrix} \sum_{i=1}^{k} w_i q_i \left(\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i \right) \\ -\sum_{i=1}^{k} w_i q_i \overline{x}_i \left(\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i \right) \end{bmatrix} \tag{33}$$

From equation (33), we have

$$\left.\begin{aligned}\phi_1 &= \frac{\sum_{i=1}^{k} w_i q_i \left(\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right)}{|A|} \\ \phi_2 &= \frac{-\sum_{i=1}^{k} w_i q_i \overline{x}_i \left(\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right)}{|A|}\end{aligned}\right\} \tag{34}$$

Thus, the optimum weight δ_i^{**} becomes

$$\delta_i^{**} = w_i + \frac{w_i q_i \overline{x}_i \sum_{i=1}^{k} w_i q_i - w_i q_i \sum_{i=1}^{k} w_i q_i \overline{x}_i}{\sum_{i=1}^{k} w_i q_i \overline{x}_i^2 \sum_{i=1}^{k} w_i q_i - \left(\sum_{i=1}^{k} w_i q_i \overline{x}_i\right)^2} \left(\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right) \tag{35}$$

Substituting the value of δ_i^{**} from equation (35) into the equation (24), the proposed calibration estimator becomes

$$\overline{y}_{st(C)}^{**} = \sum_{i=1}^{k} w_i \overline{y}_i^* + \hat{\beta} \left(\overline{X} - \sum_{i=1}^{k} w_i \overline{x}_i\right) \tag{36}$$

where $\hat{\beta} = \frac{\left(\sum_{i=1}^{k} w_i q_i \overline{x}_i \overline{y}_i^*\right)\left(\sum_{i=1}^{k} w_i q_i\right) - \left(\sum_{i=1}^{k} w_i q_i \overline{y}_i^*\right)\left(\sum_{i=1}^{k} w_i q_i \overline{x}_i\right)}{\left(\sum_{i=1}^{k} w_i q_i \overline{x}_i^2\right)\left(\sum_{i=1}^{k} w_i q_i\right) - \left(\sum_{i=1}^{k} w_i q_i \overline{x}_i\right)^2}$.

Now, the estimators of the bias and variance of the proposed calibration estimator $\overline{y}_{st(C)}^{**}$ are respectively represented as

$$\begin{aligned}\widehat{B}\left(\overline{y}_{st(C)}^{**}\right) = \sum_{i=1}^{k} w_i \hat{\beta} \overline{x}_i &\left[\frac{N_i (N_i - n_i)}{(N_i - 1)(N_i - 2)} \cdot \frac{1}{n_i \overline{x}_i} \left\{\frac{\widehat{\mu}_{30i}}{s_{xi}^2} - \frac{\widehat{\mu}_{21i}}{s_{xyi}^*}\right\}\right. \\ &\left. + \frac{W_{i2}(g_i - 1)}{n_i \overline{x}_i} \left\{\frac{\widehat{\mu}_{30i(2)}}{s_{xi}^2} - \frac{\widehat{\mu}_{21i(2)}}{s_{xyi}^*}\right\}\right]\end{aligned} \tag{37}$$

$$\widehat{V}\left(\overline{y}_{st(C)}^{**}\right) = \sum_{i=1}^{k} \frac{w_i^2 (1 - f_i)}{n_i} \left(s_{yi}^{*2} + \hat{\beta}^2 s_{xi}^2 - 2\hat{\beta} s_{xyi}^*\right) + \sum_{i=1}^{k} \frac{w_i^2 (g_i - 1) W_{i2}}{n_i} s_{yi(2)}^2 \tag{38}$$

Note: The equation (37) can provide the non-response versions of a number of combined-type estimators of the population mean $\overline{Y}$ for the suitable choices of q_i.

4. Simulation study

In this section, a simulation study has been carried out with a view to verify the performance of the proposed calibration estimators. We have considered a hypothetical data set which is generated using R software under the condition of normal distribution. Here, we first define the two random variables Y^* and X^* *i.e.*, $Y^* \sim N(0,1)$ and $X^* \sim N(0,1)$. Now, we generate a set of correlated random variables with correlation coefficient ρ using the transformations $Y^{**} = Y^*$ and $X^{**} = \rho Y^* + \sqrt{1-\rho^2} X^*$ [See Reddy *et al.* (2010)]. Finally, we define the random variables Y and X using the transformations $Y = \mu_Y + \sigma_Y Y^{**}$ and $X = \mu_x + \sigma_X X^{**}$. The above transformations constitute the random variables which are normally distributed with some meansμ_Y, μ_X and variancesσ_Y^2, σ_X^2. In this data set, a population of 15000 units has been shaped out. The population is divided into four strata with respective sizes 6000, 3000, 1500 and 4500. The sample size has been fixed as 3000. The sample size for each stratum has been determined under proportional allocation. To carry out the simulation analysis, the number of runs has been considered as 1000. The summary of the data set is given in Table 1.

Table 1: Distribution of population

Stratum No.(i)	Stratum size(N_i)	Sample size(n_i)	Distribution of of Y *i.e.* $Y \sim N(\mu_Y, \sigma_Y)$	Distribution of X *i.e.* $X \sim N(\mu_X, \sigma_X)$	Correlation coefficient between Y and X
1	6000	1200	$N(200, 20)$	$N(100, 10)$	0.78
2	3000	600	$N(230, 17)$	$N(120, 15)$	0.82
3	1500	300	$N(240, 22)$	$N(145, 22)$	0.8
4	4500	900	$N(235, 23)$	$N(135, 19)$	0.75

Table 2 depicts the estimate of the variance of the usual estimator $\overline{y}_{st}^*$ and proposed calibration estimators $\overline{y}_{st(C)}^*$ and $\overline{y}_{st(C)}^{**}$ at the different levels of non-response rate W_{i2} and inverse sampling rate g_i. The percentage relative efficiency (PRE) of the proposed calibration estimators $\overline{y}_{st(C)}^*$ and $\overline{y}_{st(C)}^{**}$ with respect to the usual estimator $\overline{y}_{st}^*$ has also been computed.

Table 2: Estimate of variance and PRE of the estimators $\overline{y}_{st}^*$, $\overline{y}_{st(C)}^*$, and $\overline{y}_{st(C)}^{**}$

$W_{i2} \forall i$	$g_i \forall i$	Estimate of Variance			PRE		
		$\overline{y}_{st}^*$	$\overline{y}_{st(C)}^*$	$\overline{y}_{st(C)}^{**}$	$\overline{y}_{st}^*$	$\overline{y}_{st(C)}^*$	$\overline{y}_{st(C)}^{**}$
0.1	1	0.04522	0.03329	0.02214	100	135.843	204.295
	2	0.04794	0.03592	0.02482	100	133.47	193.16
	2.5	0.05055	0.03857	0.02745	100	131.076	184.169
	3	0.05325	0.04123	0.03013	100	129.164	176.754
0.2	1.5	0.04789	0.03591	0.02479	100	133.376	193.223
	2	0.05325	0.04128	0.03015	100	128.988	176.627
	2.5	0.05861	0.04667	0.03551	100	125.572	165.059
	3	0.06385	0.05186	0.04075	100	123.124	156.708
0.3	1.5	0.05057	0.03858	0.02746	100	131.069	184.145
	2	0.05838	0.04648	0.03534	100	125.61	165.184
	2.5	0.06642	0.05437	0.04329	100	122.161	153.414
	3	0.07448	0.06241	0.05133	100	119.352	145.1
0.4	1.5	0.05224	0.04027	0.02915	100	129.712	179.227
	2	0.06175	0.04984	0.03869	100	123.902	159.61
	2.5	0.07157	0.05951	0.04843	100	120.266	147.784
	3	0.08125	0.06921	0.0581	100	117.403	139.848

From the Table 2, it is revealed that the estimates of the variance of the proposed calibration estimators $\overline{y}^*_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ are much smaller than the usual estimator $\overline{y}^*_{st}$ and hence the PRE of the proposed calibration estimators $\overline{y}^*_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ is much higher as compared to the usual estimator$\overline{y}^*_{st}$. It is further revealed that the estimates of the variance of the proposed calibration estimators $\overline{y}^*_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ increase with the increase in non-response rate W_{i2} and inverse sampling rate g_i as well. Such kind of outcomes is intuitively anticipated. Table 3 represents the estimate of the bias of the proposed calibration estimators $\overline{y}^*_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ at the different levels of non-response rate W_{i2} and inverse sampling rate g_i.

Table 3: Estimate of bias of estimators $\overline{y}^*_{st(C)}$and $\overline{y}^{**}_{st(C)}$

$W_{i2} \forall i$	$g_i \forall i$	Estimate of Bias	
		$\overline{y}^*_{st(C)}$	$\overline{y}^{**}_{st(C)}$
0.1	1	-0.00119	-0.0005
	2	-0.00128	-0.0006
	2.5	-0.00134	-0.0006
	3	-0.00139	-0.0006
0.2	1	-0.00115	-0.0005
	2	-0.00137	-0.0006
	2.5	-0.00163	-0.0007
	3	-0.00168	-0.0008
0.3	1	-0.00116	-0.0005
	2	-0.00153	-0.0007
	2.5	-0.0017	-0.0008
	3	-0.00186	-0.0008
0.4	1	-0.00115	-0.0005
	2	-0.00171	-0.0008
	2.5	-0.00192	-0.0009
	3	-0.00209	-0.0009

The Table 3 reveals that both calibration estimators $\overline{y}^*_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ provide negative bias of very less magnitude. A negative bias in the estimator of the finite population mean suggests that on an average, the estimator leads to underestimate the true mean of the population. Alternatively, if one has to take multiple samples from the population and compute the mean using the estimator, the average of these computations would be below the true population mean.

5. Concluding remarks

We have suggested some calibration estimators for estimating the mean of a stratified population in the presence of non-response. The information on a single auxiliary variable has been utilized to develop the calibration estimators. The chi-square distance measure has been used in obtaining the new stratum weights under the given constraints. The calibration estimators have been proposed under the situation in which the non-response occurs on study variable, whereas the auxiliary variable is free from the non-response. The basic properties of the proposed calibration estimators have been discussed in detail. The expressions for the estimators of the bias and variance of the proposed calibration estimators have been derived. To examine the behavior of the proposed calibration estimators, a simulation study has been carried out by generating an artificial data set. The Table 2 shows that the proposed calibration estimators $\overline{y}^{*}_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ perform very well as compared to the usual estimator$\overline{y}^{*}_{st}$. From Table 3, it is also revealed that both calibration estimators $\overline{y}^{*}_{st(C)}$ and $\overline{y}^{**}_{st(C)}$ confer bias of very less extent.

References

Chaudhary, M. K., Kumar, A., and Prajapati, A. (2018). Estimating the mean of a stratified population using two auxiliary variables under double sampling scheme in the presence of non-response. *International Journal of Statistics and Economics*, **19**, 123-134.

Chaudhary, M. K., Singh, V. K., and Shukla, R. K. (2012). Combined-type family of estimators of population mean in stratified random sampling under non-response. *Journal of Reliability and Statistical Studies (JRSS)*, **5**, 133-142.

Clement, E. P. and Enang, E. I. (2015). Calibration approach alternative ratio estimator for population mean in stratified sampling. *International Journal of Statistics and Economics*, **16**, 83-93.

Deming, W. E. and Stephan, F. F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *The Annals of Mathematical Statistics*, **11**, 427-444.

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

Dupont, F. (1995). Alternative adjustments where there are several levels of auxiliary information. *Survey Methodology*, **21**, 125–135.

Estevao, V. M. and Särndal, C. E. (2009). A new face on two-phase sampling with calibration estimators. *Survey Methodology*, **35**, 3-14.

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

Han, P. (2018). Calibration and multiple robustness when data are missing not at random. *Statistica Sinica*, **28**, 1725-1740.

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

Hidiroglou, M. A. and Särndal, C. E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, **24**, 11-20.

Horvitz, D. G. and Thompsom, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, **47**, 663-685.

Jaiswal, A. K., Usman, M., and Singh, G. N. (2023). New calibration estimation procedure in the presence of unit non response. *Ain Shams Engineering Journal*, **14**, 101910.

Khare, B. B. (1987). Allocation in stratified sampling in presence of non-response. *Metron*, **45**, 213-221.

Kim, J. K. and Park, M. (2010). Calibration estimation in survey sampling. *International Statistical Review*, **78**, 21-39.

Kim, J.-M., Sungur, E. A., and Heo, T. Y. (2007). Calibration approach estimators in stratified sampling. *Statistics & Probability Letters*, **77**, 99-103.

Koyuncu, N. and Kadilar, C. (2013). Calibration estimator using different distance measures in stratified random sampling. *International Journal of Modern Engineering Research*, **3**, 415-419.

Koyuncu, N. and Kadilar, C. (2014). A new calibration estimator in stratified double sampling. *Hacettepe Journal of Mathematics and Statistics*, **43**, 337-346.

Nidhi, Sisodia, B. V. S., Singh, S., and Singh, S. K. (2017). Calibration approach estimation of the mean in stratified sampling and stratified double sampling. *Communications in Statistics-Theory and Methods*, **46**, 4932-4942.

Ozgul, N. (2019). New calibration estimator based on two auxiliary variables in stratified sampling. *Communications in Statistics-Theory and Methods*, **48**, 1481-1492.

Rao, D. K., Khan, M. G. M., and Singh, G. K. (2017). On calibrated weights in stratified sampling. *ANZIAM Journal*, **59**, C190-C204.

Rao, J. N. K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, **10**, 153.

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

Rubin, B. D. (1976). Inference and missing data. *Biometrika*, **63**, 581-592.

Särndal, C. E. (2007). The calibration approach in survey theory and practice. *Survey Methodology*, **33**, 99-119.

Singh, G. N., Bhattacharyya, D., and Bandyopadhyay, A. (2023). A general class of calibration estimators under stratified random sampling in presence of various kinds of non-sampling errors. *Communications in Statistics-Simulation and Computation*, **52**, 320-333.

Singh, S., Horn, S., and Yu, F. (1998). Estimation of variance of general regression estimator: Higher level calibration approach. *Survey Methodology*, **24**, 41-50.

Sud, U.C., Chandra, H., and Gupta, V.K. (2014). Calibration-based product estimator in single- and two phase sampling. *Journal of Statistical Theory and Practice*, **8**, 1-11.

Tracy, D., Singh, S., and Arnab, R. (2003). Note on calibration in stratified and double sampling. *Survey Methodology*, **29**, 99–104.