Journal of Statistics & Management Systems
ISSN 0972-0510 (Print), ISSN 2169-0014 (Online)
Vol. 28 (2025), No. 6, pp. 1049–1063
DOI : 10.47974/JSMS-1397

An improved calibration estimator for mean of a stratified population using two auxiliary variables

Manoj K. Chaudhary *
Nishtha Bhardwaj †
Department of Statistics
Institute of Science
Banaras Hindu University
Varanasi
Uttar Pradesh
India

Abstract

Calibration aims to minimize uncertainty in the estimation by ensuring the accuracy of the estimation procedure. The auxiliary information is usually utilized to improve the accuracy of the estimation procedure under the calibration approach. In the present article, the information on two auxiliary variables is utilized to consider a new calibration estimator for the mean of a stratified population. A set of new calibration weights is obtained by minimizing the chi-square type distance function and subject to some calibration constraints based on the auxiliary information. The Taylor linearization technique is used to obtain the expression for the variance of the proposed calibration estimator. The effectiveness of the proposed calibration estimator has been demonstrated through a simulated study. A numerical example using the real data set has also been provided to support the theoretical findings. The analysis shows that the suggested calibration estimator outperforms over the current one in terms of efficiency.

Subject Classification: *62D05.*

Keywords: *Calibration estimator, Stratified random sampling, Population mean, Two auxiliary variables, Simulation study.*

1. Introduction

In statistical surveys conducted all throughout the world, the calibration estimator is frequently utilized. To update the estimates, the

* *E-mail:* **mk15@bhu.ac.in (Corresponding Author)**
† *E-mail:* **nishthabhardwaj2050@gmail.com**

© Taru Publications

calibration estimator makes use of the data on the auxiliary variables. The common method for adjusting the initial weight under the calibration-based estimate strategy involves minimizing a given distance measurement subject to a set of calibration conditions. Under calibration approach, we utilize the auxiliary information in the estimation process by employing some calibration constraints based on the auxiliary variable(s). The calibration method defines new calibration weights to be achieved by combining calibration constraints with the distance function.

Deville and Sarndal [2] were the first who explored the word calibration in survey sampling. They constructed calibration constraints with the help of auxiliary information. Many researchers have studied calibration estimation using different calibration constraints such as ratio-type estimation, product-type estimation and combined ratio-type estimation or some well-known techniques that utilize the auxiliary information. Särndal and Estevao [12] discussed about the functional form to the calibration. Singh et al. [13] and Tracy et al. [15] proposed calibration estimators of the population mean in stratified random sampling. Kim et al. [4], Koyuncu and Kadilar [6] and Nidhi et al. [8] have recommended the various classes of calibration estimators for estimating the population mean in stratified random sampling by using the sets of calibration constraints. Koyuncu and Kadilar [7] have suggested a new calibration estimator of the population mean in stratified double sampling design. Rao et al. [10] have suggested a multivariate calibration estimator in estimating the mean of a stratified population using more than one auxiliary variable. Yadav et al. [16] have made an attempt to develop an efficient estimator for population variance by making the use of parameters of a single auxiliary variable. Ozgul [9] has propounded a ratio-type calibration estimator of the population mean in stratified sampling using the information on two auxiliary variables. Singh et al. [14] have pioneered out some calibration estimators for estimating the population mean in two-stage stratified random sampling. Grover and Kaur [3] have proposed an efficient estimator of population variance utilizing the ranks of original auxiliary variable as an additional auxiliary variable. Chaudhary and Dutta [1] have discovered a new calibration estimator of the population mean in stratified two-phase sampling utilizing the information on two auxiliary variables.

The primary objective of the present research is to enhance the stratified sampling method's ability in estimating the population mean accurately. We have developed the estimation method by suggesting a new calibration estimator under the availability of information on two

auxiliary variables. New calibration weights have been created using the chi-square type distance function and some calibration constraints based on two auxiliary variables. Additionally, an empirical study through real data and simulated data is carried out to support the effectiveness of the proposed calibration estimator. The effectiveness of the proposed calibration estimator has been elaborated over the usual estimator and the calibration estimator suggested by Ozgul [9].

2. Procedures, Notations and Definitions

Suppose a finite population consisting of N units is stratified into b strata. The o^{th} stratum consists of N_o units $(o=1,2,...,b)$ such that $\sum_{o=1}^{b} N_o = N$. Let y_{oi} be the observation on the i^{th} unit in the o^{th} stratum for the study variable Y. Let x_{oi} and z_{oi} be the observations on the i^{th} unit in o^{th} the stratum for the auxiliary variables (A. V.) X and Z respectively. Let $\overline{Y} = \sum_{o=1}^{b} p_o \overline{Y}_o$, $\overline{X} = \sum_{o=1}^{b} p_o \overline{X}_o$ and $\overline{Z} = \sum_{o=1}^{b} p_o \overline{Z}_o$ are the respective population means of the variables Y, X and Z such that $\overline{Y_o} = \frac{1}{N_o}\sum_{i=1}^{N_o} y_{oi}$, $\overline{X_o} = \frac{1}{N_o}\sum_{i=1}^{N_o} x_{oi}$, $\overline{Z_o} = \frac{1}{N_o}\sum_{i=1}^{N_o} z_{oi}$ and $p_o = \frac{N_o}{N}$.

Now, we have to choose a sample of n_o units from the o^{th} stratum consisting of N_o units by using simple random sampling without a replacement (SRSWOR) scheme such that $\sum_{o=1}^{b} n_o = n$. Let $\overline{y}_o = \frac{1}{n_o}\sum_{i}^{n_o} y_{oi}$, $\overline{x}_o = \frac{1}{n_o}\sum_{i}^{n_o} x_{oi}$ and $\overline{z}_o = \frac{1}{n_o}\sum_{i}^{n_o} z_{oi}$ be the sample means of o^{th} stratum for the variables Y, X and Z respectively.

The usual estimator of the population mean $\overline{Y}$ (without using auxiliary information) is given by

$$\overline{y}_{usual} = \sum_{o=1}^{b} p_o \overline{y}_o \tag{1}$$

The expression for the variance of the usual estimator $\overline{y}_{usual}$ can be represented as

$$V(\overline{y}_{usual}) = \sum_{o=1}^{b} \left(\frac{1}{n_o} - \frac{1}{N_o} \right) p_o^2 S_{Yo}^2 \tag{2}$$

where $S_{Yo}^2 = \frac{1}{N_o - 1}\sum_{i=1}^{N_o}(y_{oi} - \overline{Y}_o)^2$.

There are several authors who have utilized the suitable ancillary data to get a more précised estimate of the population mean $\overline{Y}$ under both traditional and calibration approaches. The calibrated estimator of the population mean $\overline{Y}$ (using information on two (A.V.), suggested by Ozgul [9] is represented as

$$\overline{y}_{st} = \sum_{o=1}^{b} w_o \overline{y}_o \tag{3}$$

subject to the calibration conditions

$$\sum_{o=1}^{b} w_o = 1 \tag{4}$$

$$\sum_{o=1}^{b} w_o \widehat{T}_o = \sum_{o=1}^{b} p_o T_o \tag{5}$$

where $\widehat{T}_o = \sum_{o=1}^{b} \frac{\overline{x}_o}{\overline{z}_o}$ and $T_o = \sum_{o=1}^{b} \frac{\overline{X}_o}{\overline{Z}_o}$.

Here, calibration weight $w_o;\ o = 1, 2, ..., b$ is selected so that the chi-square type distance function $\sum_{o=1}^{b} \frac{(w_o - p_o)^2}{q_o p_o}$ (q_o being the tuning parameter) would be minimized subject to the calibration conditions given in the equations (4) and (5).

The solution for w_o was obtained by using the Lagrange function. Putting the value of w_o so computed into equation (3), the calibrated estimator looks like this:

$$\overline{y}_{st} = \sum_{o=1}^{b} p_o \overline{y}_o + \beta_1 \sum_{o=1}^{b} p_o (T_o - \widehat{T}_o) \tag{6}$$

where $\beta_0 = \dfrac{\left(\sum_{o=1}^{b} p_o q_o \widehat{T}_o \overline{y}_o\right)\left(\sum_{o=1}^{b} p_o q_o\right) - \left(\sum_{o=1}^{b} p_o q_o \overline{y}_o\right)\left(\sum_{o=1}^{b} p_o q_o \widehat{T}_o\right)}{\left(\sum_{o=1}^{b} p_o q_o\right)\left(\sum_{o=1}^{b} p_o q_o \widehat{T}_o^2\right) - \left(\sum_{o=1}^{b} p_o q_o \widehat{T}_o\right)^2}$

If we assume $q_o = 1$, then the calibration estimator becomes

$$\bar{y}_{st} = \sum_{o=1}^{b} p_o \bar{y}_o + \beta_1 \sum_{o=1}^{b} p_o (T_o - \hat{T}_o) \tag{7}$$

where $\beta_0^* = \dfrac{\left(\sum_{o=1}^{b} p_o \hat{T}_o \bar{y}_o\right) - \left(\sum_{o=1}^{b} p_o \bar{y}_o\right)\left(\sum_{o=1}^{b} p_o \hat{T}_o\right)}{\left(\sum_{o=1}^{b} p_o \hat{T}_o^2\right) - \left(\sum_{o=1}^{b} p_o \hat{T}_o\right)^2}$.

3. Proposed Calibration Estimator

The proposed calibrated estimator of the population mean using two A. V. X and Z in stratified random sampling is given by

$$\bar{y}_{st}^* = \sum_{o=1}^{b} w_o^* \bar{y}_o \tag{8}$$

subject to calibration constraints

$$\sum_{o=1}^{b} w_o^* \widehat{T}_o = \sum_{o=1}^{b} p_o T_o \tag{9}$$

$$\sum_{o=1}^{b} w_o^* \widehat{U}_o = \sum_{o=1}^{b} p_o U_o \tag{10}$$

where $\widehat{U}_o = \sum_{o=1}^{b} \frac{s_{xo}^2}{s_{zo}^2}$, $U_o = \sum_{o=1}^{b} \frac{S_{xo}^2}{S_{zo}^2}$, $S_{Xo}^2 = \frac{1}{N_o - 1} \sum_{i=1}^{N_o} (x_{oi} - \overline{X}_o)^2$, $S_{Zo}^2 = \frac{1}{N_o - 1} \sum_{i=1}^{N_o} (z_{oi} - \overline{Z}_o)^2$, $s_{xo}^2 = \frac{1}{n_o - 1} \sum_{i}^{n_o} (x_{oi} - \overline{x}_o)^2$ and $s_{zo}^2 = \frac{1}{n_o - 1} \sum_{i}^{n_o} (z_{oi} - \overline{z}_o)^2$. w_o^* is the new calibrated weight for the o^{th} stratum.

One can derive a new calibration weight solution by minimizing the chi-square type distance function $\sum_{o=1}^{b} \left(\frac{w_o^* - p_o}{q_o p_o} \right)^2$ subject to the given calibration conditions. Let us define the Lagrange function as

$$L = \sum_{o=1}^{b} \left(\frac{w_o^* - p_o}{q_o p_o} \right)^2 - 2\lambda_1 \left(\sum_{o=1}^{b} w_o^* \hat{T}_o - \sum_{o=1}^{b} p_o T_o \right) - 2\lambda_2 \left(\sum_{o=1}^{b} w_o^* \widehat{U}_o - \sum_{o=1}^{b} p_o U_o \right) = 0 \tag{11}$$

where λ_1 and λ_2 are the Lagrangian multipliers.

On differentiating the equation (11) with respect to (w. r. t.) w_o^* and equalizing the derivative to zero, we get

$$\frac{\partial L}{\partial w_o^*} = 2\frac{(w_o^* - p_o)}{q_o p_o} - 2\lambda_1\left(\sum_{o=1}^{b}\hat{T}_o\right) - 2\lambda_2\left(\sum_{o=1}^{b}\hat{U}_o\right) = 0$$

$$w_o^* = p_o + q_o p_o\left(\lambda_1\sum_{o=1}^{b}\hat{T}_o + \lambda_2\sum_{o=1}^{b}\hat{U}_o\right) \tag{12}$$

Substituting the values of w_o^* from equation (12) into the equations (9) and (10), we get the following system of equations in the form of matrix

$$\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} = \begin{bmatrix}\sum_{o=1}^{b} q_o p_o \hat{T}_o^2 & \sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\\ \sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o & \sum_{o=1}^{b} q_o p_o \hat{U}_o^2\end{bmatrix}^{-1}\begin{bmatrix}\sum_{o=1}^{b} p_o T_o - \sum_{o=1}^{b} p_o \hat{T}_o\\ \sum_{o=1}^{b} p_o U_o - \sum_{o=1}^{b} p_o \hat{U}_o\end{bmatrix}$$

The values of λ_1 and λ_2 can be obtained by solving the above matrix. Thus, we have

$$\lambda_1 = \frac{p}{q},\quad \lambda_2 = \frac{m}{q}$$

where $p = \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right)\left[\sum_{o=1}^{b} p_o T_o - \sum_{o=1}^{b} p_o \hat{T}_o\right] - \left(\sum_{o=1}^{b} q_o p_o \hat{T}_o \hat{U}_o\right)\left[\sum_{o=1}^{b} p_o U_o - \sum_{o=1}^{b} p_o \hat{U}_o\right]$,

$$m = \left(-\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)\left[\sum_{o=1}^{b} p_o T_o - \sum_{o=1}^{b} p_o \hat{T}_o\right] - \left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right)\left[\sum_{o=1}^{b} p_o U_o - \sum_{o=1}^{b} p_o \hat{U}_o\right]$$

and $q = \left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)^2$

Substituting the values of λ_1 and λ_2 into the equation (12), we get the optimum value of w_o^* as

$$p_o + \left[\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)\right]\left[\left(\sum_{o=1}^{b} p_o\left(T_o - \hat{T}_o\right)\right)\right] +$$

$$w_o^* = \frac{\left[\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o\right)\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{T}_o\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)\right]\left[\sum_{o=1}^{b} p_o\left(U_o - \hat{U}_o\right)\right]}{\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)^2}.$$

Putting the optimum value w_o^* in the equations (8), we can write the suggested calibration estimator as follows:

$$\bar{y}_{st}^* = \sum_{0=1}^{b} p_o y_o + \beta_1 \sum_{o=1}^{b} p_o (T_o - \hat{T}_o) + \beta_2 \sum_{o=1}^{b} p_o (U_o - \hat{U}_o) \tag{13}$$

where $\beta_1 = \dfrac{\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right)\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o \bar{y}_o\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \bar{y}_o\right)}{\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)^2}$ and

$$\beta_2 = \frac{\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \bar{y}_o\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o \bar{y}_o\right)}{\left(\sum_{o=1}^{b} q_o p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} q_o p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} q_o p_o \hat{U}_o \hat{T}_o\right)^2}$$

Assuming $q_o = 1,$ the suggested calibrated estimator reduces to

$$\bar{y}_{st}^* = \sum_{o=1}^{b} p_o y_o + \beta_1^* \sum_{o=1}^{b} p_o (T_o - \hat{T}_o) + \beta_2^* \sum_{o=1}^{b} p_o (U_o - \hat{U}_o) \tag{14}$$

where $\beta_1^* = \dfrac{\left(\sum_{o=1}^{b} p_o \hat{U}_o^2\right)\left(\sum_{o=1}^{b} p_o \hat{T}_o \bar{y}_o\right) - \left(\sum_{o=1}^{b} p_o \hat{U}_o \hat{T}_o\right)\left(\sum_{o=1}^{b} p_o \hat{U}_o \bar{y}_o\right)}{\left(\sum_{o=1}^{b} p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} p_o \hat{U}_o \hat{T}_o\right)^2}$ and

$$\beta_2^* = \frac{\left(\sum_{o=1}^{b} p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} p_o \hat{U}_o \bar{y}_o\right) - \left(\sum_{o=1}^{b} p_o \hat{U}_o \hat{T}_o\right)\left(\sum_{o=1}^{b} p_o \hat{T}_o \bar{y}_o\right)}{\left(\sum_{o=1}^{b} p_o \hat{T}_o^2\right)\left(\sum_{o=1}^{b} p_o \hat{U}_o^2\right) - \left(\sum_{o=1}^{b} p_o \hat{U}_o \hat{T}_o\right)^2}.$$

4. Properties of Proposed Calibration Estimator

We use the Taylor linearization technique to get the expressions for variance of the suggested calibration estimator and calibration estimator suggested by Ozgul [9].

Let $PAR = \left(\sum_{o=1}^{b} p_o \overline{Y}_o, \sum_{o=1}^{b} p_o R_o \overline{Y}_o, \sum_{o=1}^{b} p_o R_o^2, \sum_{o=1}^{b} p_o R_o \right)$. Thus, we have

$$\left| \frac{\partial\left(\overline{y}_{st}\right)}{\partial\left(\sum_{o=1}^{b} p_o \overline{y}_o\right)} \right|_{PAR} = 1, \left| \frac{\partial\left(\overline{y}_{st}\right)}{\partial\left(\sum_{o=1}^{b} p_o \widehat{T}_o\right)} \right|_{PAR} = \frac{-\left(\sum_{o=1}^{b} p_o T_o \overline{Y}_o\right) - \left(\sum_{o=1}^{b} p_o T_o\right)\left(\sum_{o=1}^{b} p_o \overline{Y}_o\right)}{\left(\sum_{o=1}^{b} p_o T_o^2\right) - \left(\sum_{o=1}^{b} p_o T_o\right)^2} = \gamma_0$$

$$\text{and} \left| \frac{\partial\left(\overline{y}_{st}\right)}{\partial\left(\sum_{o=1}^{b} p_o \widehat{T}_o \overline{y}_o\right)} \right|_{PAR} = \left| \frac{\partial\left(\overline{y}_{st}\right)}{\partial\left(\sum_{o=1}^{b} p_o \widehat{T}_b^2\right)} \right|_{PAR} = 0$$

Now, Let

$$PAR^* = \left(\sum_{o=1}^{b} p_o \overline{Y}_o, \sum_{o=1}^{b} p_o T_o, \sum_{o=1}^{b} p_o U_o, \sum_{o=1}^{b} p_o U_o T_o, \sum_{o=1}^{b} p_o \overline{Y}_o T_o, \sum_{o=1}^{b} p_o U_o \overline{Y}_o, \sum_{o=1}^{b} p_o U_o^2, \sum_{o=1}^{b} p_o T_o^2 \right)$$

This leads to

$$\left| \frac{\partial\left(\overline{y}_{st}^*\right)}{\partial\left(\sum_{o=1}^{b} p_o \overline{y}_o\right)} \right|_{PAR^*} = 1,$$

$$\left| \frac{\partial\left(\overline{y}_{st}^*\right)}{\partial\left(\sum_{o=1}^{b} p_o \widehat{T}_0\right)} \right|_{PAR^*} = -\frac{\left(\sum_{o=1}^{b} p_o U_o^2\right)\left(\sum_{o=1}^{b} p_o T_o \overline{Y}_o\right) - \left(\sum_{o=1}^{b} p_o U_o T_o\right)\left(\sum_{o=1}^{b} p_o U_o \overline{Y}_o\right)}{\left(\sum_{o=1}^{b} p_o T_o^2\right)\left(\sum_{o=1}^{b} p_o U_o^2\right) - \left(\sum_{o=1}^{b} p_o U_o T_o\right)^2} = \gamma_1$$

$$\left| \frac{\partial\left(\overline{y}_{st}^*\right)}{\partial\left(\sum_{o=1}^{b} p_o \widehat{U}_o\right)} \right|_{PAR^*} = -\frac{\left(\sum_{o=1}^{b} p_o T_o^2\right)\left(\sum_{o=1}^{b} p_o U_o \overline{Y}_o\right) - \left(\sum_{o=1}^{b} p_o U_o T_o\right)\left(\sum_{o=1}^{b} p_o T_o \overline{Y}_o\right)}{\left(\sum_{o=1}^{b} p_o T_o^2\right)\left(\sum_{o=1}^{b} p_o U_o^2\right) - \left(\sum_{o=1}^{b} p_o U_o T_o\right)^2} = \gamma_2,$$

$$\left|\frac{\partial\left(\overline{y}_{st}^{*}\right)}{\partial\left(\sum_{o=1}^{b}p_o\widehat{U}_o^2\right)}\right|_{PAR^*}=\left|\frac{\partial\left(\overline{y}_{st}^{*}\right)}{\partial\left(\sum_{o=1}^{b}p_o\widehat{T}_o^2\right)}\right|_{PAR^*}=0 \text{ and}$$

$$\left|\frac{\partial\left(\overline{y}_{st}^{*}\right)}{\partial\left(\sum_{o=1}^{b}p_o\widehat{T}_o\overline{y}_o\right)}\right|_{PAR^*}=\left|\frac{\partial\left(\overline{y}_{st}^{*}\right)}{\partial\left(\sum_{o=1}^{b}p_o\widehat{U}_o\widehat{T}_o\right)}\right|_{PAR^*}=\left|\frac{\partial\left(\overline{y}_{st}^{*}\right)}{\partial\left(\sum_{o=1}^{b}p_o\widehat{U}_o\overline{y}_o\right)}\right|_{PAR^*}=0.$$

Using Taylor linearization technique, the equations for the variance of the existing calibration estimator $\overline{y}_{st}$ and suggested calibration estimator $\overline{y}_{st}^{*}$ can respectively be obtained as

$$V\left(\overline{y}_{st}\right)=V\left(\sum_{o=1}^{b}p_0\overline{y}_0\right)+\gamma_0^2V\left(\sum_{o=1}^{b}p_0\hat{R}_0\right)+2\gamma_0Cov\left(\sum_{o=1}^{b}p_0y_0,\sum_{o=1}^{b}p_0\hat{R}_0\right) \tag{15}$$

$$\begin{aligned}V\left(\overline{y}_{st}^{*}\right)=&V\left(\sum_{o=1}^{b}p_o\overline{y}_o\right)+\gamma_1^2V\left(\sum_{o=1}^{b}p_o\widehat{T}_o\right)+\gamma_2^2V\left(\sum_{o=1}^{b}p_o\widehat{U}_o\right)\\&+2\gamma_1Cov\left(\sum_{o=1}^{b}p_oy_o,\sum_{o=1}^{b}p_o\widehat{T}_o\right)+2\gamma_1Cov\left(\sum_{o=1}^{b}p_o\overline{y}_o,\sum_{o=1}^{b}p_o\widehat{T}_o\right)\\&+2\gamma_2Cov\left(\sum_{o=1}^{b}p_oy_o,\sum_{o=1}^{b}p_o\widehat{U}_o\right)+2\gamma_1\gamma_2Cov\left(\sum_{o=1}^{b}p_o\widehat{T}_o,\sum_{o=1}^{b}p_o\widehat{U}_o\right)\end{aligned} \tag{16}$$

where

$$V\left(\sum_{o=1}^{b}p_o\overline{y}_o\right)=\sum_{o=1}^{b}p_o^2\lambda_o\overline{Y}_o^2C_{Y_o}^2,V\left(\sum\nolimits_{o=1}^{b}p_o\hat{R}_o\right)=\sum_{o=1}^{b}p_o^2\lambda_o\frac{\overline{X}_o^2}{\overline{Z}_o^2}\left(C_{X_o}^2+C_{Z_o}^2-2\rho_{XZ_o}C_{X_o}C_{Z_o}\right),$$

$$Cov\left(\sum_{o=1}^{b}p_o\overline{y}_o,\sum_{o=1}^{b}p_o\hat{R}_o\right)=\sum\nolimits_{o=1}^{b}p_o^2\lambda_o\frac{\overline{Y}_o\overline{X}_o}{\overline{Z}_o}\left(\rho_{YX_o}C_{Y_o}C_{X_o}-\rho_{YZ_o}C_{Y_o}C_{Z_o}\right),$$

$$V\left(\sum_{o=1}^{b}p_o\widehat{T}_o\right)=\sum_{o=1}^{b}p_o^2\lambda_o\frac{\overline{X}_o^2}{\overline{Z}_o^2}\left(C_{X_o}^2+C_{Z_o}^2-2\rho_{XZ_o}C_{X_o}C_{Z_o}\right),$$

$$V\left(\sum_{o=1}^{b}p_o\widehat{U}_o\right)=\sum_{o=1}^{b}p_o^2\lambda_o\frac{S_{X_o}^4}{S_{Z_o}^4}\left(\lambda_{40_o}+\lambda_{04_o}-\lambda_{22_o}\right),$$

$$Cov\left(\sum_{o=1}^{b} p_o \bar{y}_o, \sum_{o=1}^{b} p_o \hat{T}_o\right) = \sum_{o=1}^{b} p_o^2 \lambda_o \bar{Y}_o \frac{\bar{X}_o}{\bar{Y}_o}\left(\rho_{YX_o} C_{Y_o} C_{X_o} - \rho_{YZ_o} C_{Y_o} C_{Z_o}\right),$$

$$Cov\left(\sum_{o=1}^{b} p_o \bar{y}_o, \sum_{o=1}^{b} p_o \hat{U}_o\right) = \sum_{o=1}^{b} p_o^2 \lambda_o \bar{Y}_o \frac{S_{X_o}^2}{S_{Y_o}^2}\left(C_{Y_o} \lambda_{30_o} - C_Y \lambda_{12_o} + \lambda_{04_o} - \lambda_{22_o}\right),$$

$$Cov\left(\sum_{o=1}^{b} p_o \hat{T}_o, \sum_{o=1}^{b} p_o \hat{U}_o\right) = \sum_{o=1}^{b} p_o^2 \lambda_o \bar{Y}_o \frac{\bar{X}_o}{\bar{Z}_o} \frac{S_{X_o}^2}{S_{Y_o}^2}\left[C_{X_o} \lambda_{30_o} + C_{Z_o}\left(\lambda_{13_o} + \lambda_{31_o}\right)\right] -$$
$$\left[C_{Z_o}^2\left(\lambda_{04_o} + C_{X_o} \lambda_{03_o}\right) + \rho_{XZ_o} C_{Z_o} C_{X_o}^2 \lambda_{03_o}\right],$$

$$\lambda_o = \left(\frac{1}{n_o} - \frac{1}{N_o}\right), \lambda_{rs_o} = \frac{\mu_{rs_o}}{\mu_{20_o}^{r/2} \mu_{02_o}^{s/2}} \text{ and } \mu_{rs_o} = \frac{1}{N_o - 1} \sum_{i=1}^{N_o} \left(X_{oi} - \bar{Y}_o\right)^r \left(X_{oi} - \bar{Y}_o\right)^s.$$

Here, ρ_{XY_o}, ρ_{YZ_o} and ρ_{YZ_o} are the correlation coefficients between X and Y; Y and Z; X and Z respectively for the o^{th} stratum. C_{X_o}, C_{Y_o} and C_{Z_o} are respectively the coefficients of variation for X, Y and Z in the o^{th} stratum.

5. Empirical Study

It is much difficult to arrive at a conclusion about the efficacy of the suggested calibration estimator because the results derived in the previous sections involve a number of parameters. Thus, it is essential to illustrate the theoretical facts with some numerical examples in order to assess the behaviour of the suggested calibration estimator. In this section, we have conducted an empirical investigation using both real data and artificially created data.

5.1 *Real Data Set*

To investigate the conceptual outcomes, we have considered the data used by Koyuncu and Kadilar [5]. This data set consists of 923 districts of Turkey that are stratified into six regions, viz., Marmara, Aegean, Mediterranean, Central Anatolia, Black Sea, and East and Southeast Anatolia. Here, Y = Number (No.) of teachers, X = No. of students (both primary and secondary schools), and Z = No. of classes (both primary and secondary schools). Table 1 shows the description of population.

Table 1

Different Parameters of Population

N_o	n_o	$\overline{Y}_o$	$\overline{X}_o$	$\overline{Z}_o$	C_{Y_o}	C_{X_o}	C_{Z_o}	ρ_{YX_o}	ρ_{YZ_o}	ρ_{XZ_o}
127	31	703.74	20804.59	498.28	1.26	1.47	1.11	0.936	0.979	0.94
117	21	413	9211.79	318.33	1.56	1.65	1.15	0.996	0.976	0.97
103	29	573.17	14309.3	431.36	1.8	1.93	1.42	0.994	0.984	0.977
170	38	424.66	9478.85	311.32	1.91	1.92	1.47	0.983	0.983	0.964
205	22	267.03	5569.95	227.2	1.51	1.53	1.15	0.989	0.964	0.968
201	39	393.84	12997.59	313.71	1.81	1.78	1.27	0.965	0.983	0.996

Here, Table 2 depicts the variance of the estimators $\overline{y}_{usual}$, $\overline{y}_{st}$ and $\overline{y}^{*}_{st}$. The percentage relative efficiency (PRE) of the calibration estimator $\overline{y}_{st}$ and $\overline{y}^{*}_{st}$ w. r. t the usual estimator $\overline{y}_{usual}$ is also depicted. The gain in terms of precision of the calibrated estimators $\overline{y}_{st}$ and $\overline{y}^{*}_{st}$ w. r. t the usual estimator $\overline{y}_{usual}$ is computed as well. The PRE, precision and gain in precision are calculated using the following formulae:

$$PRE\left(\overline{y}_{st} \text{ or } \overline{y}^{*}_{st}\right) = \frac{V\left(\overline{y}_{usual}\right)}{V\left(\overline{y}_{st} or \overline{y}^{*}_{st}\right)} * 100, \text{ Precision}\left(\overline{y}_{usual} \text{ or } \overline{y}_{st} \text{ or } \overline{y}^{*}_{st}\right) =$$

$$\frac{1}{V\left(\overline{y}_{usual} or \overline{y}_{st} or \overline{y}^{*}_{st}\right)} \text{ and } \% \text{ Gain}\left(\overline{y}_{st} or \overline{y}^{*}_{st}\right) =$$

$$\frac{\text{Precision}(\overline{y}_{st} \text{ or } \overline{y}^{*}_{st}) - \text{Precision}(\overline{y}_{usual})}{\text{Precision}(\overline{y}_{usual})} \times 100.$$

Table 2

Variance, PRE and Gain in Terms of Precision based on Real Data Set

Estimator	AMSE	PRE	Precision	%Gain
$\overline{y}_{usual}$	2231.421	100	0.00044	_____
$\overline{y}_{st}$	1322.762	168.694	0.00075	70.454
$\overline{y}^{*}_{st}$	368.327	605.826	0.00271	515.909

5.2 *Simulation Analysis*

The performance of the suggested calibration estimator has also been shown by a simulation examination. The simulation analysis was carried out using the R programming language. We now consider a stratified population which is divided into six strata. The data under the study variable Y for each stratum are produced using the Normal distribution with predetermined mean (μ) and standard deviation (σ). We followed the instructions provided by Reddy et al. [11] to create the data associated with the auxiliary variables X and Z for each stratum under the assumption that they would have specific correlations with the study variable Y. Therefore, we first generate the data under a dummy variable T for each stratum with the same distribution as that of Y. Then we have generated the data under the A.V. X and Z for each stratum using the transformation $AV = \rho Y + \sqrt{1-\rho^2}T$. Here, ρ indicates the correlation coefficient between study variable Y and A.V. The particulars of the population are given in Table 3.

Table 3

Particulars of Population

Strat. No.	Stratum Population under Y	Stratum Size (N_o)	ρ_{YZ_o}	ρ_{YZ_o}	ρ_{XZ_o}
1.	$N(\mu = 6, \sigma = 1)$	500	0.6471	0.8889	0.9244
2.	$N(\mu = 14, \sigma = 2)$	400	0.5896	0.8355	0.9365
3.	$N(\mu = 10, \sigma = 3)$	600	0.6710	0.8928	0.9330
4.	$N(\mu = 18, \sigma = 4)$	800	0.5371	0.8036	0.9337
5.	$N(\mu = 15, \sigma = 5)$	900	0.5464	0.8095	0.9340
6.	$N(\mu = 23, \sigma = 6)$	300	0.6640	0.8995	0.9240

Here, $N = 3500$. The sample size is fixed as $n = 538$. The proportional allocation is used to determine the sample size for each stratum. Now, we have selected the sample from each stratum using SRSWOR technique and the sample information has been utilized to compute the approximate variance (AVAR). The AVAR of the estimators $\overline{y}_{usual}$, $\overline{y}_{st}$ and $\overline{y}^*_{st}$ has been computed using the following formulae:

$$AVAR(\bar{y}_{usual}) = \frac{1}{5000}\sum_{j=1}^{5000}\left(\bar{y}_{usual_j} - \bar{Y}\right)^2;\ j = 1,2,...,5000;$$

$$AVAR(\bar{y}_{st}) = \frac{1}{5000}\sum_{j=1}^{5000}\left(\bar{y}_{st_j} - \bar{Y}\right)^2;\ j = 1,2,...,5000$$

$$AVAR(\bar{y}^*_{st}) = \frac{1}{5000}\sum_{j=1}^{5000}\left(\bar{y}^*_{st_j} - \bar{Y}\right)^2;\ j = 1,2,...,5000$$

Table 4 shows the approximate variance of the estimators $\bar{y}_{usual}$, $\bar{y}_{st}$ and $\bar{y}^*_{st}$. The PRE along with gain in precision of the calibration estimators $\bar{y}_{st}$ and $\bar{y}^*_{st}$ w. r. t the usual estimator $\bar{y}_{usual}$ has also been shown.

Table 4

AVAR, PRE and Precision in Gain based on Simulation

Estimator	AMSE	PRE	Precision	Gain
$\bar{y}_{usual}$	0.02191	100	45.641	____
$\bar{y}_{st}$	0.01758	124.601	56.869	24.6006
$\bar{y}^*_{st}$	0.01185	184.839	84.388	84.8915

6. Concluding Remarks and Interpretations

In the current study, we have taken into account some new calibration constraints based on two auxiliary variables that establish a new calibration estimator of the mean of a stratified population. Using the Taylor linearization method, the variance expression for the proposed calibration estimator is derived. In order to evaluate the effectiveness of the recommended calibration estimator, an empirical study has been carried out. A natural data set has been used to complete the empirical study. By creating an artificial data set in the R software, a simulation analysis has also been done. The efficiency of the proposed calibration estimator has been compared with that of the usual estimator and the calibration estimator suggested by Ozgul [9]. The comparison has also been made in terms of gain in precision. The Table 2 and Table 4 reveal that because of retaining smaller values of variance/AVAR and greater values of PRE, the

suggested calibration estimator outperforms over the usual estimator and the calibration estimator offered by Ozgul [9]. Therefore, the suggested calibration estimator can be preferred over the current one.

References

[1] M. K. Chaudhary and T. Dutta, "Calibration estimator of population mean under stratified two-phase sampling using two auxiliary variables," *Communications in Statistics - Simulation and Computation*, vol. 52, no. 11, pp. 5423-5433 (2023).

[2] J. C. Deville and C. E. Särndal, "Calibration estimators in survey sampling," *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 376-382 (1992).

[3] L. K. Grover and A. Kaur, "On improvement in variance estimation using ranks of original auxiliary variable as an additional auxiliary variable," *Journal of Statistics and Management Systems*, vol. 24, no. 5, pp. 1041-1057 (2021).

[4] J.-M. Kim, E. A. Sungur, and T.-Y. Heo, "Calibration approach estimators in stratified sampling," *Statistics & Probability Letters*, vol. 77, no. 1, pp. 99–103 (2007).

[5] N. Koyuncu and C. Kadilar, "Ratio and product estimators in stratified random sampling," *Journal of Statistical Planning and Inference*, vol. 139, no. 8, pp. 2552-2558 (2009).

[6] N. Koyuncu and C. Kadilar, "Calibration estimator using different distance measures in stratified random sampling," *International Journal of Modern Engineering Research*, vol. 3, no. 1, pp. 415-419 (2013).

[7] N. Koyuncu and C. Kadilar, "A new calibration estimator in stratified double sampling," *Hacettepe Journal of Mathematics and Statistics*, vol. 43, no. 2, pp. 337-346 (2014).

[8] B. V. S. Sisodia, S. Singh, S. K. Singh, and Nidhi, "Calibration approach estimation of the mean in stratified sampling and stratified double sampling," *Communications in Statistics - Theory and Methods*, vol. 46, no. 10, pp. 4932-4942 (2017).

[9] N. Ozgul, "New calibration estimator based on two auxiliary variables in stratified sampling," *Communications in Statistics - Theory and Methods*, vol. 48, no. 6, pp. 1481-1492 (2019).

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

[11] M. K. Reddy, K. R. Rao, and N. K. Boiroju, "Comparison of ratio estimators using Monte Carlo simulation," *International Journal of Agriculture and Statistical Sciences*, vol. 6, no. 2, pp. 517-527 (2010).

[12] C. E. Särndal and V. M. Estevao, "A functional form to calibration," *Journal of Official Statistics*, vol. 16, no. 4, pp. 379-399 (2000).

[13] S. Singh, S. Horn, and F. Yu, "Estimation of variance of the general regression estimator: Higher level calibration approach," *Survey Methodology*, vol. 24, pp. 41–50 (1998).

[14] D. Singh, B. V. S. Sisodia, Nidhi, and S. Pundir, "Some calibration estimators for finite population mean in two-stage stratified random sampling," *Communications in Statistics - Theory and Methods*, vol. 49, no. 17, pp. 4234-4247 (2020).

[15] D. S. Tracy, S. Singh, and A. R. Arnab, "Note on calibration in stratified and double sampling," *Survey Methodology*, vol. 29, no. 1, pp. 99–104 (2003).

[16] S. K. Yadav, S. S. Mishra, and S. Gupta, "An efficient estimator for population variance using parameters of an auxiliary variable," *Journal of Statistics and Management Systems*, vol. 22, no. 6, pp. 1005-1013 (2019).

Received April, 2024