STATISTICS IN TRANSITION new series, March 2024
Vol. 25, No. 1, pp. 77–91, https://doi.org/10.59170/stattrans-2024-005
Received – 13.12.2021; accepted – 29.12.2023

Improved calibration estimation of population mean in stratified sampling using two auxiliary variables

Abimibola V. Oladugba[1], **Oluwagbenga T. Babatunde**[2]

Abstract

In this paper, a new improved calibration estimator for the population mean in a stratified sampling was proposed using two auxiliary variables. A simulation study was carried out to evaluate the performance and efficiency of the proposed estimator with respect to three estimators considered in the literature for estimating the population mean in a stratified sampling using two auxiliary variables. The results showed that the new estimator proved to be more efficient than the three existing estimators considered

Key words: calibration, estimator, stratified sampling, auxiliary variables, mean square error, bias, percentage relative efficiency.

1. Introduction

Calibration estimation is a popular approach in sample survey introduced by Deville and Sarndal (1992) and meant to improve the precision of the estimated population parameter. This is achieved using additional relevant information known as auxiliary information or variable. Auxiliary variable is a variable that provides some other relevant details about the study variable (Babatunde et al. (2023)). Auxiliary variables are correlated to the study variable (Babatunde et al. (2023)) and the efficiency of an estimator depends on the level of correlation between the study and auxiliary variables (Agunbiade and Ogunyinka (2013)). Agunbiade and Ogunyinka (2013) showed that using auxiliary variable that is highly correlated to the study variable produces an estimator with smaller variance compared to when the correlation level between the auxiliary variable and study variable is medium or low. This implies that the choice of auxiliary variables is restricted only to variables that are correlated to the

[1] Department of Statistics, University of Nigeria, Nsukka, Nigeria. E-mail: abimibola.oladugba@unn.edu.ng. ORCID: https://orcid.org/0000-0002-6402-8833.
[2] Corresponding Author. Department of Statistics, University of Nigeria, Nsukka, Nigeria.
E-mail: oluwagbenga.babatunde@unn.edu.ng. ORCID: https://orcid.org/0000-0002-4761-9770.

study variable. This poses some limitations on the use of auxiliary variable as not all variables can be used as auxiliary variable.

In estimating the population mean of a stratified sampling using calibration approach, the calibration weights are used to replace the stratum weights in the estimator. The calibration weights are obtained by minimizing a distance function subject to well defined calibration constraints. Most often, calibration constraints restrict the sum of the selected sample statistics to be equal to the sum of the population parameters in the different strata (see Ozgul (2018), Alam et al. (2021), Adubi et al. (2022), Babatunde et al. (2023)).

Several calibration estimators for the population mean in a stratified sampling using different parameters of one auxiliary variable in the calibration constraints have been proposed in the literature (see Tracey, Singh and Arnab (2003), Rao, Tekabu and Khan (2016), Koyuncu and Kadilar (2016), Sisodia, Singh and Singh (2017), Alam, Singh and Shabbir (2019), Garg and Pachori (2019), Alam and Shabbir (2020), Babatunde et al. (2023), Oladugba et al. (2023) etc.). Several calibration estimators were arrived at by modifying existing estimators (see Kadilar and Cingi (2006), Garg and Pachori (2019)). Calibration estimators of the population mean have been shown to be more efficient than the general population estimator in a stratified sampling (see Rao, Khan and Khan (2012), Ozgul (2018) and Alam et al. (2019)). The use of two auxiliary variables have also been explored in the calibration estimation of the population mean in stratified sampling. Rao et al. (2012), Ozgul (2018) and Rai, Singh and Qasim (2021) proposed different calibration estimators for population mean in a stratified sampling using different calibration constraints based on two auxiliary variables.

In this paper, we propose a new improved calibration estimator for the population mean in a stratified sampling based on two auxiliary variables by modifying the estimator proposed in Ozgul (2018) with the aim of achieving a more efficient estimator. The standard deviation of the two auxiliary variables was used to define the calibration constraints.

The remainder of this paper is as follows: notations are presented in Section 2, some of the existing calibration estimators based on two auxiliary variables were discussed in Section 3. In Section 4, the proposed calibration estimator was presented. The simulation study carried out and conclusions are presented in Sections 5 and 6, respectively.

2. Notations

Consider a situation where it is desired to estimate the population mean $\overline{Y}$ in stratified sampling using additional information from two auxiliary variables. Let M be a finite population consisting of N units, i.e. $M = (M_1, M_2, \ldots M_N)$. Let y_i, x_{1i} and

x_{2i} be the i^{th} value of the study variable, first auxiliary variable and second auxiliary variable respectively, i = 1, 2, … N. Let M be divided into Z distinct homogenous strata with each stratum containing N_h units, h = 1, 2, …, Z such that $N = \sum_{h=1}^{Z} N_h$. A sample of size n is drawn from M using simple random sampling without replacement (SRSWOR) by selecting n_h units from h^{th} stratum such that $n = \sum_{h=1}^{Z} n_h$. The mean of the sample and population of the study variable in each stratum are given as $\bar{y}_h = \frac{1}{n_h}\sum_{i=1}^{n_h} y_{hi}$ and $\bar{Y}_h = \frac{1}{N_h}\sum_{i=1}^{N_h} y_{hi}$ respectively. The mean of the sample and population of the first auxiliary variable in each stratum are given as $\bar{x}_{1h} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{1hi}$ and $\bar{X}_{1h} = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{1hi}$ respectively. The mean of the sample and population of the second auxiliary variable are given as $\bar{x}_{2h} = \frac{1}{n_h}\sum_{i=1}^{n_h} x_{2hi}$ and $\bar{X}_{2h} = \frac{1}{N_h}\sum_{i=1}^{N_h} x_{2hi}$ respectively. The sample and population standard deviation of the first auxiliary variable in each stratum are given as $s_{x_{1h}} = \sqrt{\frac{\sum_{i=1}^{n_h}\left(x_{1i} - \bar{x}_{1h}\right)^2}{n_h - 1}}$ and $S_{x_{1h}} = \sqrt{\frac{\sum_{i=1}^{N_h}\left(x_{1i} - \bar{X}_{1h}\right)^2}{N_h - 1}}$ respectively. The sample and population standard deviation of the second auxiliary variable in each stratum are given as $s_{x_{2h}} = \sqrt{\frac{\sum_{i=1}^{n_h}\left(x_{2i} - \bar{x}_{2h}\right)^2}{n_h - 1}}$ and $S_{x_{2h}} = \sqrt{\frac{\sum_{i=1}^{N_h}\left(x_{2i} - \bar{X}_{2h}\right)^2}{N_h - 1}}$ respectively. The population mean of the auxiliary variables are $\bar{X}_1 = \frac{1}{N}\sum_{i=1}^{N} x_{1i}$ and $\bar{X}_2 = \frac{1}{N}\sum_{i=1}^{N} x_{2i}$ respectively.

The population mean of a stratified sampling is estimated by:

$$\bar{y}_{st} = \sum_{h=1}^{Z} W_h \bar{y}_h \tag{2.1}$$

where $W_h = \frac{N_h}{N}$ is the h^{th} stratum weights.

The precision of the estimator in (2.1) is improved upon using the calibration approach which replaces the stratum weights W_h with calibrated weights obtained by optimizing the Chi-square distance function defined below:

$$D\left(\Omega_h, W_h\right) = \sum_{h=1}^{Z} \frac{\left(\Omega_h - W_h\right)^2}{W_h Q_h} \tag{2.2}$$

Subject to well defined calibration constraints.
where Ω_h are the calibrated weights and Q_h are defined weights for obtaining different versions of the estimator (Alam and Shabbir (2020)).

3. Some Calibration Estimators in Stratified Sampling Using Two Auxiliary Variables

Different calibration estimators have been proposed for the population mean of a stratified sampling using several known parameters of the auxiliary variables. Some of the existing calibration estimators using two auxiliary variables are reviewed below.

3.1. Rao et al. (2012)

Rao et al. (2012) proposed a calibration estimator with two auxiliary variables using the mean of the auxiliary variables in the calibration constraints as:

$$\bar{y}_R = \sum_{h=1}^{Z} \Omega_{hR} \bar{y}_h \tag{3.1}$$

where the calibrated weights Ω_{hR} are obtained by minimizing the Chi-square distance function in (2.2) subject to the calibration constraints given by:

$$\sum_{h=1}^{Z} \Omega_{hR} \bar{x}_{1h} = \bar{X}_1 \tag{3.2}$$

$$\sum_{h=1}^{Z} \Omega_{hR} \bar{x}_{2h} = \bar{X}_2 \tag{3.3}$$

By minimizing the function in (2.2) subject to (3.1) and (3.2), the optimum weights obtained are:

$$\Omega_{hR} = W_h + W_h Q_h (\lambda_1 \bar{x}_{1h} + \lambda_2 \bar{x}_{2h}) \tag{3.4}$$

where

$$\lambda_1 = \frac{-\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2 \bar{x}_{1h} \left(\bar{X}_1 - \hat{\bar{X}}_1\right) + \sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2 \bar{x}_{2h} \left(\bar{X}_2 - \hat{\bar{X}}_2\right)}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h} \bar{x}_{2h}\right)^2 - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right)} \tag{3.5}$$

$$\lambda_2 = \frac{\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h} \bar{x}_{2h}^2 \left(\bar{X}_1 - \hat{\bar{X}}_1\right) - \sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2 \bar{x}_{2h} \left(\bar{X}_2 - \hat{\bar{X}}_2\right)}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h} \bar{x}_{2h}\right)^2 - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right)} \tag{3.6}$$

$$\hat{\bar{X}}_1 = \sum_{h=1}^{Z} W_h \bar{x}_{1h} \text{ and } \hat{\bar{X}}_2 = \sum_{h=1}^{Z} W_h \bar{x}_{2h}$$

By substituting (3.5) and (3.6) into (3.4) and substituting (3.4) into (3.1), the obtained estimator can be expressed as:

$$\bar{y}_R = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma}_1\left(\bar{X}_1 - \hat{\bar{X}}_1\right) + \hat{\gamma}_2\left(\bar{X}_2 - \hat{\bar{X}}_2\right) \tag{3.7}$$

where

$$\hat{\gamma}_1 = \frac{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right)}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)^2 - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right)} \tag{3.8}$$

$$\hat{\gamma}_2 = \frac{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)^2 - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right)} \tag{3.9}$$

3.2. Ozgul (2018)

Ozgul (2018) proposed a calibration estimator with two auxiliary variables using the ratio of the mean of the auxiliary variables in the calibration constraints as:

$$\bar{y}_O = \sum_{h=1}^{Z} \Omega_{hO}\bar{y}_h \tag{3.10}$$

where the calibrated weights Ω_{hO} are obtained by minimizing the Chi-square distance function in (2.2) subject to the calibration constraints given by:

$$\sum_{h=1}^{Z} \Omega_{hO} = \sum_{h=1}^{Z} W_h \tag{3.11}$$

$$\sum_{h=1}^{Z} \Omega_{hO}\hat{R}_h = \sum_{h=1}^{Z} W_h R_h \tag{3.12}$$

where $\hat{R}_h = \sum_{h=1}^{Z} \frac{\bar{x}_{1h}}{\bar{x}_{2h}}$ and $R_h = \sum_{h=1}^{Z} \frac{\bar{X}_{1h}}{\bar{X}_{2h}}$

By minimizing the function in (2.2) subject to (3.11) and (3.12), the optimum weights obtained are:

$$\Omega_{hO} = W_h + W_h Q_h \left(\frac{\Delta_1 + \Delta_2 \hat{R}_h}{A}\right) \tag{3.13}$$

where

$$\Delta_1 = -\sum_{h=1}^{Z} W_h\left(R_h - \hat{R}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h\right) \tag{3.14}$$

$$\Delta_2 = \sum_{h=1}^{Z} W_h\left(R_h - \hat{R}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h\right) \tag{3.15}$$

$$A = \left(\sum_{h=1}^{Z} W_h Q_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h\right)^2 \tag{3.16}$$

By substituting (3.14) and (3.15) into (3.13) and substituting (3.13) into (3.12), the obtained estimator can be expressed as:

$$\bar{y}_O = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma}\sum_{h=1}^{Z} W_h\left(R_h - \hat{R}_h\right) \tag{3.17}$$

where

$$\hat{\gamma} = \left(\frac{\left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h \bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h\right)}{\left(\sum_{h=1}^{Z} W_h Q_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \hat{R}_h\right)^2}\right) \tag{3.18}$$

3.3. Rai et al. (2021)

Rai et al. (2021) proposed a calibration estimator with two auxiliary variables using the sample and population mean of the auxiliary variables in the calibration constraints as:

$$\bar{y}_C = \sum_{h=1}^{Z} \Omega_{hC} \bar{y}_h \tag{3.19}$$

where the calibrated weights Ω_{hC} are obtained by minimizing the Chi-square distance function in (2.2) subject to the calibration constraints given by:

$$\sum_{h=1}^{Z} \Omega_{hC} \bar{x}_{1h} = \sum_{h=1}^{Z} W_h \bar{X}_{1h} \tag{3.20}$$

$$\sum_{h=1}^{Z} \Omega_{hC} \bar{x}_{2h} = \sum_{h=1}^{Z} W_h \bar{X}_{2h} \tag{3.21}$$

By minimizing the function in (2.2) subject to (3.20) and (3.21), the optimum weights obtained are:

$$\Omega_{hC} = W_h\left(1 + \lambda_1 Q_h \bar{x}_{1h} + \lambda_2 Q_h \bar{x}_{2h}\right) \tag{3.22}$$

where

$$\lambda_1 = \frac{\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2 \hat{\bar{X}}_1 - \sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h} \bar{x}_{2h} \hat{\bar{X}}_2}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h} \bar{x}_{2h}\right)^2} \tag{3.23}$$

$$\lambda_2 = \frac{-\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\hat{\bar{X}}_1 - \sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2 \hat{\bar{X}}_2}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)^2} \tag{3.24}$$

where

$$\hat{\bar{X}}_1 = \sum_{h=1}^{Z} W_h \bar{X}_{1h} - \sum_{h=1}^{Z} W_h \bar{x}_{1h} \tag{3.25}$$

$$\hat{\bar{X}}_2 = \sum_{h=1}^{Z} W_h \bar{X}_{2h} - \sum_{h=1}^{Z} W_h \bar{x}_{2h} \tag{3.26}$$

By substituting (3.23) and (3.24) into (3.22) and substituting (3.22) into (3.19), the obtained estimator can be expressed as:

$$\bar{y}_C = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma}_1 \hat{\bar{X}}_1 + \hat{\gamma}_2 \hat{\bar{X}}_2 \tag{3.27}$$

where

$$\hat{\gamma}_1 = \frac{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)^2} \tag{3.28}$$

$$\hat{\gamma}_2 = \frac{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{y}_h\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)}{\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}^2\right)\left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{2h}^2\right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{x}_{1h}\bar{x}_{2h}\right)^2} \tag{3.29}$$

4. Proposed Calibration Estimator

The calibration estimator proposed in Ozgul (2018) was improved upon in this paper by modifying the calibration constraints. Let y_i be the i^{th} observation of the study variable and x_{1i} and x_{2i} , i = 1,2,3,…,N; h = 1,2,3,…,Z be the i^{th} observation of the two auxiliary variables. The study variable Y and the auxiliary variables X_1 and X_2 contain N observations divided into h strata with each stratum containing N_h observations such that $N = \sum_{h=1}^{Z} N_h$.

We proposed a new improved calibration estimator as:

$$\bar{y}_p = \sum_{h=1}^{Z} \Omega_{hp} \bar{y}_h \tag{4.1}$$

subject to the calibration constraints defined below:

$$\sum_{h=1}^{Z}\Omega_{hp}=\sum_{h=1}^{Z}W_h \tag{4.2}$$

$$\sum_{h=1}^{Z}\Omega_{hp}\hat{T}_h=\sum_{h=1}^{Z}W_hT_h \tag{4.3}$$

where Ω_{hp} are the calibrated weights of the proposed estimator

$$\hat{T}_h=\frac{s_{x_{1h}}+s_{x_{2h}}}{\bar{x}_{1h}+\bar{x}_{2h}} \text{ and } T_h=\frac{S_{x_{1h}}+S_{x_{2h}}}{\bar{X}_{1h}+\bar{X}_{2h}}$$

The optimum value for the proposed calibrated weight for each stratum was obtained by minimizing (2.2) subject to the constraints in (4.2) and (4.3) using Lagrange optimization method. The Lagrange function for minimizing (2.2) subject to (4.2) and (4.3) is expressed as:

$$L=\sum_{h=1}^{Z}\frac{\left(\Omega_{hp}-W_h\right)^2}{W_hQ_h}-2\lambda_1\left(\sum_{h=1}^{Z}\Omega_{hp}-\sum_{h=1}^{Z}W_h\right)-2\lambda_2\left(\sum_{h=1}^{Z}\Omega_{hp}\hat{T}_h-\sum_{h=1}^{Z}W_hT_h\right) \tag{4.4}$$

where λ_1 and λ_2 are defined as the Lagrange multipliers.

By differentiating (4.4) with respect to Ω_{hp} and equating the resultant expression to zero, we obtained the optimum calibration weight as:

$$\Omega_{hp}=W_h+W_hQ_h\left(\lambda_1+\lambda_2\hat{T}_h\right) \tag{4.5}$$

where λ_1 and λ_2 are obtained by replacing Ω_{hp} in (4.2) and (4.3) with the optimum value of Ω_{hp} in (4.5).

$$\lambda_1=\frac{-\left(\sum_{h=1}^{Z}W_h\left(T_h-\hat{T}_h\right)\right)\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h\right)}{\left(\sum_{h=1}^{Z}W_hQ_h\right)\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h^2\right)-\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h\right)^2} \tag{4.6}$$

$$\lambda_2=\frac{\left(\sum_{h=1}^{Z}W_h\left(T_h-\hat{T}_h\right)\right)\left(\sum_{h=1}^{Z}W_hQ_h\right)}{\left(\sum_{h=1}^{Z}W_hQ_h\right)\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h^2\right)-\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h\right)^2} \tag{4.7}$$

Then by substituting (4.6) and (4.7) into (4.5), the optimum calibrated weights are expressed as:

$$\Omega_{hp}=W_h+W_hQ_h\frac{\left(\sum_{h=1}^{Z}W_h\left(T_h-\hat{T}_h\right)\right)\left(\sum_{h=1}^{Z}W_hQ_h\right)-\left(\sum_{h=1}^{Z}W_h\left(T_h-\hat{T}_h\right)\right)\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h\right)}{\left(\sum_{h=1}^{Z}W_hQ_h\right)\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h^2\right)-\left(\sum_{h=1}^{Z}W_hQ_h\hat{T}_h\right)^2} \tag{4.8}$$

By substituting Ω_{hp} obtained in (4.8) into (4.1), we obtained the proposed calibration estimator as:

$$\bar{y}_p = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma} \sum_{h=1}^{Z} W_h \left(T_h - \hat{T}_h \right) \tag{4.9}$$

where

$$\hat{\gamma} = \frac{\left(\sum_{h=1}^{Z} W_h Q_h \hat{T}_h \bar{y}_h \right) \left(\sum_{h=1}^{Z} W_h Q_h \right) - \left(\sum_{h=1}^{Z} W_h Q_h \bar{y}_h \right) \left(\sum_{h=1}^{Z} W_h Q_h \hat{T}_h \right)}{\left(\sum_{h=1}^{Z} W_h Q_h \right) \left(\sum_{h=1}^{Z} W_h Q_h \hat{T}_h^2 \right) - \left(\sum_{h=1}^{Z} W_h Q_h \hat{T}_h \right)^2} \tag{4.10}$$

In most situations, Q_h is assumed to be equal to 1 (Ozgul, 2018). For calibration estimator involving one auxiliary variable, Q_h is assumed to be equal to the reciprocal of the sample mean and any other statistic of the auxiliary variable (see Garg and Pachori (2019) and Babatunde et al. (2023)). Since the calibration estimators considered in this paper involve two auxiliary variables, we suggested Q_h to be the reciprocal of the sum of the sample mean and sample standard deviation of the two auxiliary variables, i.e. $\frac{1}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)}$ and $\frac{1}{\left(s_{x_{1h}} + s_{x_{2h}} \right)}$.

The different values of Q_h were used to obtain different versions of the proposed calibration estimator as:

Case I: $Q_h = 1$

$$\bar{y}_{p1} = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma}_1 \sum_{h=1}^{Z} W_h \left(T_h - \hat{T}_h \right) \tag{4.11}$$

where

$$\hat{\gamma}_1 = \frac{\left(\sum_{h=1}^{Z} W_h \hat{T}_h \bar{y}_h \right) \left(\sum_{h=1}^{Z} W_h \right) - \left(\sum_{h=1}^{Z} W_h \bar{y}_h \right) \left(\sum_{h=1}^{Z} W_h \hat{T}_h \right)}{\left(\sum_{h=1}^{Z} W_h \right) \left(\sum_{h=1}^{Z} W_h \hat{T}_h^2 \right) - \left(\sum_{h=1}^{Z} W_h \hat{T}_h \right)^2} \tag{4.12}$$

Case II: $Q_h = \frac{1}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)}$

$$\bar{y}_{p2} = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma}_2 \sum_{h=1}^{Z} W_h \left(T_h - \hat{T}_h \right) \tag{4.13}$$

where

$$\hat{\gamma}_2 = \frac{\left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h \bar{y}_h}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right) \left(\sum_{h=1}^{Z} \frac{W_h}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right) - \left(\sum_{h=1}^{Z} \frac{W_h \bar{y}_h}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right) \left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right)}{\left(\sum_{h=1}^{Z} \frac{W_h}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right) \left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h^2}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right) - \left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h}{\left(\bar{x}_{1h} + \bar{x}_{2h} \right)} \right)^2} \tag{4.14}$$

Case III: $Q_h = \frac{1}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}$

$$\bar{y}_{p3} = \sum_{h=1}^{Z} W_h \bar{y}_h + \hat{\gamma}_3 \sum_{h=1}^{Z} W_h \left(T_h - \hat{T}_h\right) \tag{4.15}$$

where

$$\hat{\gamma}_3 = \frac{\left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h \bar{y}_h}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right)\left(\sum_{h=1}^{Z} \frac{W_h}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right) - \left(\sum_{h=1}^{Z} \frac{W_h \bar{y}_h}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right)\left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right)}{\left(\sum_{h=1}^{Z} \frac{W_h}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right)\left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h^2}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right) - \left(\sum_{h=1}^{Z} \frac{W_h \hat{T}_h}{\left(s_{x_{1h}} + s_{x_{2h}}\right)}\right)^2} \tag{4.16}$$

5. Simulation Study

We demonstrated the efficiency of the proposed calibration estimator over the estimators proposed in Rao et al. (2012), Ozgul (2018) and Rai et al. (2021) through a simulation study. A simulation study establishes the consistency of the obtained result under different scenarios. The study population is MU284 obtained from Sarndal, Swensson and Wretman (1992, pp. 652–659) consisting of 284 municipalities from Sweden partitioned by geographical region into eight strata. The study variable is the 1985 populations (in thousands) while the first and second auxiliary variables are the 1975 populations (in thousands) and total number of seats in municipal council respectively. The population parameters in each stratum are presented in Table 1. Using SRSWOR, a random sample of size n; $n_1 = 57$, $n_2 = 71$, $n_3 = 85$, $n_4 = 99$ and $n_5 = 114$, which correspond to 20%, 25%, 30%, 35% and 40% of the population units respectively, were drawn. The sample size for each stratum $n_h = nW_h$ was obtained using proportional allocation.

For each sample size, we simulated $K = 50{,}000$ samples and computed both the proposed and existing estimators considered in this work for all the simulated samples. The performance and efficiency of the calibration estimators were assessed using the absolute relative bias (ARB), empirical mean square error (MSE) and percentage relative efficiency (PRE) expressed in (5.1), (5.2) and (5.3) respectively. The results obtained are presented in Tables 2, 3 and 4.

$$ARB\left(\bar{y}_j\right) = \left| \frac{\frac{1}{K}\sum_{k=1}^{K}\left(\bar{y}_j\right)_k - \bar{Y}}{\bar{Y}} \right| \tag{5.1}$$

$$MSE\left(\bar{y}_j\right) = \frac{1}{K}\sum_{k=1}^{K}\left(\left(\bar{y}_j\right)_k - \bar{Y}\right)^2 \tag{5.2}$$

$$PRE\left(\bar{y}_l, \bar{y}_P\right) = \frac{MSE\left(\bar{y}_l\right)}{MSE\left(\bar{y}_P\right)} \times 100 \tag{5.3}$$

where $j = R, O, C$ and P and $l = R, O$ and C.

Table 1: Population parameters of the study and auxiliary variables

Strata	Mean			Standard Deviation		
	Y	X_1	X_2	Y	X_1	X_2
1	62.4400	59.5200	51.1600	122.0685	126.1038	13.7860
2	29.6042	29.1667	47.6667	35.9547	34.6791	12.7628
3	24.0625	23.9375	50.2500	20.7710	20.5790	10.1704
4	31.0000	30.6316	48.4737	38.6775	40.9373	8.9406
5	29.4107	28.7143	46.3571	56.2348	59.1731	9.8060
6	20.8293	20.9756	46.5610	17.5359	17.1343	8.1272
7	26.6667	26.6000	54.2000	23.8038	23.2975	11.0224
8	17.5172	17.1379	40.1724	21.4164	19.7968	9.7912

Table 2: Absolute Relative Bias of the Calibrated Estimators Using Two Auxiliary Variables

		$Q_h = 1$		
Sample size	$ARB(\bar{y}_P)$	$ARB(\bar{y}_O)$	$ARB(\bar{y}_R)$	$ARB(\bar{y}_C)$
n_1	0.003880	0.031470	0.035461	0.035462
n_2	0.002748	0.029486	0.032508	0.032509
n_3	0.002305	0.028617	0.030898	0.030899
n_4	0.001492	0.027078	0.029139	0.029140
n_5	0.002690	0.024627	0.026226	0.026227
		$Q_h = 1/(\bar{x}_1 + \bar{x}_2)$		
n_1	0.008747	0.029921	0.033910	0.033911
n_2	0.002528	0.027947	0.030941	0.030942
n_3	0.001439	0.027323	0.029546	0.029547
n_4	0.004879	0.024616	0.026268	0.026269
n_5	0.000527	0.023686	0.025184	0.025185
		$Q_h = 1/(s_{x_{1h}} + s_{x_{2h}})$		
n_1	0.009412	0.033364	0.037549	0.037550
n_2	0.001385	0.033441	0.036775	0.036776
n_3	0.002847	0.033369	0.035985	0.035986
n_4	0.000918	0.031900	0.034010	0.034011
n_5	0.008633	0.030618	0.032596	0.032598

Table 3: Mean Square Error of the Calibrated Estimators Using Two Auxiliary Variables

Sample size	$MSE(\bar{y}_P)$	$MSE(\bar{y}_O)$	$MSE(\bar{y}_R)$	$MSE(\bar{y}_C)$
		$Q_h = 1$		
n_1	648.97	42693.33	54207.70	54211.06
n_2	325.51	37478.80	45554.96	45558.00
n_3	229.09	35302.94	41153.64	41156.49
n_4	95.95	31608.20	36603.02	36605.67
n_5	311.98	26144.38	29650.31	29652.67
		$Q_h = 1/(\bar{x}_1 + \bar{x}_2)$		
n_1	3298.22	38593.30	49568.96	49572.17
n_2	275.58	33670.15	41268.60	41271.48
n_3	89.27	32182.77	37631.68	37634.39
n_4	1026.02	26122.38	29745.02	29747.40
n_5	11.97	24184.35	27340.07	27342.33
		$Q_h = 1/(s_{x_{1h}} + s_{x_{2h}})$		
n_1	5002.00	51977.03	65263.01	65266.74
n_2	85.05	48265.29	58361.29	58365.28
n_3	334.90	48173.62	56008.55	56011.90
n_4	36.29	43866.74	49862.30	49865.42
n_5	3212.79	40411.47	45803.84	45806.80

Table 4: Percentage Relative Efficiency of the Proposed Estimator

Sample size	$PRE(\bar{y}_O, \bar{y}_P)$	$PRE(\bar{y}_R, \bar{y}_P)$	$PRE(\bar{y}_C, \bar{y}_P)$
		$Q_h = 1$	
n_1	65.79	83.53	83.53
n_2	115.14	139.95	139.96
n_3	154.10	179.64	179.65
n_4	329.42	381.48	381.51
n_5	83.80	95.04	95.05
		$Q_h = 1/(\bar{x}_1 + \bar{x}_2)$	
n_1	11.70	15.03	15.03
n_2	122.18	149.75	149.76
n_3	360.51	421.55	421.58
n_4	25.46	28.99	28.99
n_5	2020.41	2284.05	2284.24
		$Q_h = 1/(s_{x_{1h}} + s_{x_{2h}})$	
n_1	10.39	13.05	13.05
n_2	567.49	686.20	686.25
n_3	143.84	167.24	167.25
n_4	1208.78	1374.00	1374.08
n_5	12.58	14.26	14.26

Table 2 presents the absolute relative bias for all the calibration estimators. The proposed estimator has the least absolute relative bias followed by the estimators proposed by Ozgul (2018), Rao et al. (2012) and Rai et al. (2021) for all the cases of Q_h considered. This implies that the estimates of the population mean obtained from the proposed estimator are closer to the population mean compared to the estimates obtained from the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021). For all the different sample sizes considered, the ARB values of the proposed estimators are smaller compared to the ARB values of the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021). For $Q_h = 1$, the ARB values of the proposed estimator reduced as the sample size increases except for $n_5 = 114$ where the ARB value increased but for $Q_h = 1/(\bar{x}_1 + \bar{x}_2)$ and $Q_h = 1/\left(s_{x_{1h}} + s_{x_{2h}}\right)$, the ARB values are not consistent. For $Q_h = 1$ and $Q_h = 1/\left(s_{x_{1h}} + s_{x_{2h}}\right)$, the least ARB value is observed when the sample size $n_4 = 99$ while for $Q_h = 1/(\bar{x}_1 + \bar{x}_2)$ the least ARB value is observed when the sample size $n_5 = 114$. This suggest that the proposed estimator performed better with a large sample size.

The results in Table 3 are the mean square errors for all the calibration estimators. The proposed estimator has the least mean square error followed by the estimators proposed by Ozgul (2018), Rao et al. (2012) and Rai et al. (2021) for all the cases of Q_h considered. For all the different sample sizes considered, the MSE values of the proposed estimators are smaller compared to the MSE values of the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021). For $Q_h = 1$, the MSE values of the proposed estimator reduced as the sample size increases except for $n_5 = 114$ where the MSE value increased but for $Q_h = 1/(\bar{x}_1 + \bar{x}_2)$ and $Q_h = 1/\left(s_{x_{1h}} + s_{x_{2h}}\right)$, the MSE values are not consistent. For $Q_h = 1$ and $Q_h = 1/\left(s_{x_{1h}} + s_{x_{2h}}\right)$ the least MSE value is observed when the sample size $n_4 = 99$ while for $Q_h = 1/(\bar{x}_1 + \bar{x}_2)$ the least MSE value is observed when the sample size $n_5 = 114$.

From Table 4, all the percentage relative efficiencies obtained are greater than 100% implying that the proposed estimator is more efficient when compared to the estimators proposed by Ozgul (2018), Rao et al. (2012) and Rai et al. (2021) for all the cases of Q_h considered. For all the different sample sizes considered, the proposed estimator is more efficient compared to the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021). However, the result of this study shows that the proposed estimator is more efficient when compared to the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021) for large sample sizes. For example, for $Q_h = 1$ and $Q_h = 1/\left(s_{x_{1h}} + s_{x_{2h}}\right)$, the efficiency of the proposed estimator compared to the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021) was higher for sample size $n_4 = 99$ while for $Q_h = 1/(\bar{x}_1 + \bar{x}_2)$ the efficiency of the proposed estimator compared to the estimators proposed in Ozgul (2018), Rao et al. (2012) and Rai et al. (2021) was higher for sample size $n_5 = 114$.

6. Conclusion

The standard estimator of the population mean in a stratified sampling was improved in this paper through calibration estimation approach using two auxiliary variables. The calibration estimator proposed in Ozgul (2018) was modified by defining a new set of calibration constraints. Through a simulation study, the efficiency of the proposed calibration estimator was assessed and compared to the estimators proposed in Rao et al. (2012), Ozgul (2018) and Rai et al. (2021). The proposed estimator has the least absolute relative bias and mean square error for all the cases of Q_h considered. These results are consistent with the results obtained in Ozgul (2018), where the absolute relative bias and mean square error of the estimators proposed by Ozgul (2018) and Rao et al. (2018) were compared for $Q_h = 1$. Also, the proposed estimator is more efficient when compared to the estimators proposed in Rao et al. (2012), Ozgul (2018) and Rai et al. (2021) for all the cases of Q_h considered. Furthermore, the efficiency of the proposed estimator was found to be higher for large sample sizes.

References

Adubi, A. S., Babatunde, O. T., and Ude, I. O., (2022). Modified ratio estimators for population means with two auxiliary parameters using calibration weights. *Asian Journal of Probability and Statistics,* 20(4), pp. 169–183.

Agunbiade, D., Ogunyinka, P., (2013). Effect of correlation level on the use of auxiliary variable in double sampling for regression estimation. *Open Journal of Statistics,* 3(5), pp. 312–318.

Alam, S., Singh, S., and Shabbir, J., (2019). New methodology of calibration in stratified sampling. Proceedings of JSM-Survey Research Methods Section.

Alam, S., Shabbir, J., (2020). Calibration estimation of mean by using double use of auxiliary information. *Communications in Statistics – Simulations and Computation,* https://doi.org/10.1080/03610918.2020.1749660.

Babatunde, O. T., Oladugba, A. V., Ude, I. O., and Adubi, A. S., (2023) Calibration estimation of population mean in stratified sampling using standard deviation. *Quality & Quantity,* https://doi.org/10.1007/s11135-023-01737-1.

Deville, J. C., Sarndal, C., -E., (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association,* 87(418), pp. 376–382.

Garg, N., Pachori, M., (2019). Use of coefficient of variation in calibration estimation of population mean in stratified sampling. *Communications in Statistics - Theory and Methods,* https://doi.org/10.1080/03610926.2019.1622729.

Kadilar, C., Cingi, H., (2006). Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters,* 19, pp. 75–79.

Koyuncu, N., Kadilar, C., (2016). Calibration weighting in stratified random sampling. *Communications in Statistics: Simulation and Computation,* 45(7), pp. 2267–2275.

Oladugba, A. V., Adubi, A. S., Okafor, F. C., Babatunde, O. T. and Adubi, P. C., (2023). Calibrated estimators for population means using standard deviation of auxiliary variable. *Journal of the Indian Society for Probability and Statistics,* https://doi.org/10.1007/s41096-023-00167-4.

Ozgul, N., (2018). New calibration estimator based on two auxiliary variables in stratified sampling. *Communication in Statistics – Theory and Methods,* https://doi.org/10.1080/03610926.2018.1433852.

Rao, D., Khan, M., and Khan, S., (2012). Mathematical programming on multivariate calibration estimation in stratified sampling. *World Academy of Science, Engineering and Technology,* 72, pp. 12–27.

Rao, D., Tekabu, T., and Khan, M., (2016). New calibration estimators in stratified sampling. In Computer Science and Engineering (APWC on CSE), 3rd Asia-Pacific World Congress Nadi, Fijion, December 5–6, pp. 66–70, IEEE.

Rai, P. K., Singh, A., and Qasim, M., (2021). Calibration-based estimators using different distance measures under two auxiliary variables: a comparative study. *Journal of Modern Applied Statisticsl Methods,* 19(1), pp. 1–20.

Sarndal, C. -E., Swensson, B., and Wretman, J., (1992). *Model assisted survey sampling.* New York: Springer-Verlag Publishing, https://doi.org/10.1007/978-1-4612-4378-6.

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(10), pp. 4932–4942.

Tracey, D. S., Singh, S., and Arnab, R., (2003). Note on calibration in stratified and double Sampling. *Survey Methodology,* 29(1), pp. 99–104.