Separate Ratio Estimator Using Calibration Approach for the Population mean in Stratified Random Sampling

B. B. Khare a, Sachin Singh a* and Manish Mishra a

a Department of Statistics, Banaras Hindu University, Varanasi, India.

Authors’ contributions
This work was carried out in collaboration among all authors. All authors read and approved the final manuscript.

Article Information
DOI: 10.9734/AJPAS/2022/v20i3427

Open Peer Review History:
This journal follows the Advanced Open Peer Review policy. Identity of the Reviewers, Editor(s) and additional Reviewers, peer review comments, different versions of the manuscript, comments of the editors, etc are available here: https://www.sdiarticle5.com/review-history/92665

Received 08 August 2022
Accepted 15 October 2022
Published 21 October 2022

Abstract

In this paper we investigate the problem of estimation of finite population mean using auxiliary information. The aim of this paper is to modify a separate ratio estimator by using calibration approach for the population mean using stratified random sampling. We use the calibration approach to improve the precision of estimate of the population parameter by incorporating auxiliary information. We use the chi-square distance function to minimize the distance between the original design weight and the new design weight. We use two constraints called calibration equations to find the new calibration weights. We use the Taylor’s linearization technique to find the approximate mean square (MSE) of the proposed estimator up to the first degree of approximation. We conduct a simulation study to assess the efficiency of the proposed calibration estimator based on an artificially generated data set. An empirical study has also been conducted to judge the merits of the proposed estimator, which shows that the proposed estimator performs better than the proposed estimator [9] and the usual separate ratio estimator.

Keywords: Calibration estimator; stratified random sampling; separate ratio estimator; Taylor’s linearization technique.

1 Introduction

In survey sampling, we use auxiliary information in a number of ways to improve our estimator. With the help of auxiliary information, we have some well-known techniques such as ratio-type estimation method and
regression-type estimation method. Deville and Särndal [1] first introduced the calibration approach for the population total. The calibration approach is a method that uses auxiliary information to adjust the original design weights to improve the accuracy of the estimator of the population parameter. Calibration is used to improve the estimator by modifying the sample weights, and we choose calibration weights to minimize a given distance measure while satisfying the constraints on the auxiliary variable. Some research using the calibration approach was done by Wu and Sitter [2] in (2001). Tracy et al. [3] introduced a pair of constraints that use first and second-order moments of the auxiliary variables to obtain calibration weights for estimating the population parameter. Calibration is used to improve the estimator by modifying the sample weights, and we choose calibration weights to minimize a given distance measure while satisfying the constraints on the auxiliary variable. Some research using the calibration approach was done by Wu and Sitter [2] in (2001). Tracy et al. [3] introduced a pair of constraints that use first and second-order moments of the auxiliary variables to obtain calibration weights for estimating the population mean in stratified sampling. Farrell and Singh [4] in (2005), Arnab and Singh [5] in (2005), and Estevao and Särndal [6] in (2006) have also given some calibration estimators. Kim et al. [7] proposed a calibration approach for the ratio estimators in stratified sampling design. Koyuncu and Kadilar [8] proposed estimators using the population mean calibration approach with stratified random sampling. Nidhi [9] in (2017) introduced the calibration approach estimation of mean in stratified sampling. Alam and Shabbir [10] have also proposed a calibration technique by using dual use of the auxiliary variable for estimation of the finite population mean in stratified sampling.

In this paper, we proposed a calibration approach for separate ratio estimators that uses stratified sampling in the case of a known value of the population mean of the auxiliary character in each stratum. The properties of the proposed estimator have been obtained and a comparative study of the proposed estimator with relevant estimators have been performed using a simulation study and also for a real data set.

2. Procedures and Notations

Consider a finite population \( U:\{1,2\ldots N\} \) of size \( N \). Let \( Y \) and \( X \) be the study variable and auxiliary variable. Let the population of size \( N \) be divided into \( L \) homogeneous groups called strata such that the \( h^{th} \) stratum contains \( N_h \) units and a sample of size \( n_h \) is drawn using simple random sampling without replacement (SRSWOR) method such that \( \sum_{h=1}^{L} N_h = N \) and \( \sum_{h=1}^{L} n_h = n \), where \( n \) is the total sample size. Let \( y_{hi} \) be the value of \( i^{th} \) unit of the study variable in the \( h^{th} \) stratum, where \( i = 1,2\ldots n_h \). Let \( X_{hi} \) be the value of \( i^{th} \) unit of the auxiliary variable in the \( h^{th} \) stratum. For the \( h^{th} \) stratum, \( W_h = \frac{N_h}{N} \) be the stratum weight.

Under the stratified sampling scheme, the unbiased estimator of the population mean is given by

\[
\bar{Y}_s = \sum_{h=1}^{H} W_h \bar{y}_h
\]

(1)

where \( \bar{y}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} y_{hi} \) is the sample mean of study variable \( Y \) in the \( h^{th} \) stratum.

The variance of the estimator \( \bar{Y}_s \) is given by

\[
V(\bar{Y}_s) = \sum_{h=1}^{H} W_h^2 \left( 1 - \frac{f_h}{n_h} \right) S_{hy}^2
\]

(2)

where \( S_{hy}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{n_h} (Y_{hi} - \bar{Y}_h)^2 \) is population mean square for \( h^{th} \) stratum and \( f_h = n_h / N_h \) is the sampling fraction for the \( h^{th} \) stratum.

The separate ratio estimator \( \bar{Y}_{sr} \) of population mean \( \bar{Y} \) is given by
\[
\bar{y}_{sr} = \sum_{h=1}^{H} W_h \left( \frac{\bar{y}_h}{\bar{x}_h} \right) \bar{x}_h
\]

where \( \bar{x}_h = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{hi} \) is the \( h \)th sample mean of auxiliary character \( X \) in the \( h \)th stratum and \( \bar{X}_h = \frac{1}{N_h} \sum_{h=1}^{N_h} X_{hi} \) is the population mean for the \( h \)th stratum.

The variance of \( \bar{y}_{sr} \) to the first order of approximation is given by

\[
V(\bar{y}_{sr}) \approx \sum_{h=1}^{H} W_h^2 \left( \frac{1-f_h}{n_h} \right) \bar{y}_h^2 \left[ C_{hy}^2 + C_{hx}^2 - 2 \rho_{hy} C_{hx} C_{hy} \right]
\]

where \( f_h = n_h / N_h \), \( \bar{y}_h = \frac{1}{N_h} \sum_{h=1}^{N_h} y_{hi} \), \( C_{hy}^2 = S_{hx}^2 / \bar{X}_h^2 \), \( C_{hx}^2 = S_{hy}^2 / \bar{Y}_h^2 \), and \( \rho_{hy} = S_{hy} / S_{hx} S_{hy} \) denote the population mean square of \( y \), \( x \), covariance of \( y \) and \( x \), and population correlation coefficient between \( y \) and \( x \) respectively in the \( h \)th stratum.

An estimator for the variance of the \( \bar{y}_{sr} \) is given by

\[
\hat{V}(\bar{y}_{sr}) \approx \sum_{h=1}^{H} W_h^2 \left( \frac{1-f_h}{n_h} \right) \bar{y}_h^2 \left[ s_{hy}^2 + r_h^2 s_{hx}^2 - 2 r_h s_{hx} s_{hy} \right]
\]

where \( r_h = \frac{\bar{y}_h}{\bar{x}_h} \), \( s_{hy}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (y_{hi} - \bar{y}_h)^2 \), \( s_{hx}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)^2 \), and

\[
s_{hx}^2 = \frac{1}{n_h-1} \sum_{i=1}^{n_h} (x_{hi} - \bar{x}_h)(y_{hi} - \bar{y}_h)
\]

and \( s_{hx}^2 \), \( s_{hy}^2 \), \( r_h \) denotes the sample mean square of \( x \), \( y \), covariance of \( y \) and \( x \), and the ratio of two sample mean respectively in the \( h \)th stratum.

The calibration estimator defined by [9], is given by

\[
\bar{y}_{sr}(n) = \sum_{h=1}^{H} \Omega_h \bar{y}_h
\]

where \( \Omega_h \) are calibrated weights used to minimize the chi-square distance function, given as

\[
\sum_{h=1}^{H} \frac{(\Omega_h - W_h)^2}{W_h \Omega_h}
\]
where \( Q_h \) are suitably chosen positive quantity which determines the type of the estimator, subject to the following constraints

\[
\sum_{h=1}^{H} \Omega_h \bar{x}_h = \sum_{h=1}^{H} W_h \bar{X}_h \tag{7}
\]

and \( \sum_{h=1}^{H} \Omega_h = 1 \tag{8} \)

The Lagrange function utilizing calibration constraints and chi-square distance function is given by

\[
\Phi = \sum_{h=1}^{H} \left( \frac{(\Omega_h - W_h)^2}{W_h \Omega_h} \right) - 2\lambda_1 \left( \sum_{h=1}^{H} \Omega_h \bar{x}_h - \sum_{h=1}^{H} W_h \bar{X}_h \right) - 2\lambda_2 \left( \sum_{h=1}^{H} \Omega_h - 1 \right) \tag{9}
\]

where \( \lambda_1 \) and \( \lambda_2 \) are Lagrange multipliers.

Now differentiate (9) with respect to \( \Omega_h \) and equating to zero, we get the calibrated weight given as

\[
\Omega_h = W_h + W_h \Omega_h \frac{\left( \sum_{h=1}^{H} W_h \bar{x}_h \right) - \left( \sum_{h=1}^{H} W_h \bar{X}_h \right)}{\left( \sum_{h=1}^{H} W_h \Omega_h \bar{x}_h \right) - \left( \sum_{h=1}^{H} W_h \Omega_h \bar{X}_h \right)}; \text{ for } h=1,2,\ldots,L \tag{10}
\]

The calibration approach provides an estimator of variance given by

\[
\hat{V}[\bar{y}_h(n)] \approx \sum_{h=1}^{H} \Omega_h^2 \left( \frac{1-f_h}{n_h} \right) \frac{1}{(n_h-1)} \sum_{i=1}^{n_h} \hat{e}_{ih}^2 \tag{11}
\]

where \( \hat{e}_{ih} = (y_{ih} - \bar{y}_h) - \hat{f}_h(x_{ih} - \bar{x}_h) \) denotes the residual term in the \( h^{th} \) stratum and \( \hat{f}_h = \frac{s_{xy}}{s_{xx}} \).

### 3 Proposed Estimators

We propose separate ratio estimator using calibration approach for the population mean using stratified random sampling which is given by

\[
\bar{y}_o(c) = \sum_{h=1}^{H} W_h \left( \frac{y_h}{\bar{x}_h} \right) \bar{X}_h \tag{12}
\]

where the new weights \( W_h \) are chosen such that the chi-square distance function
\[ \sum_{h=1}^{H} \left( W_h^* - W_h \right)^2 \]

is minimum depending on the condition

\[ \sum_{h=1}^{H} W_h^* \bar{x}_h = \sum_{h=1}^{H} W_h \bar{x}_h \quad \text{and} \quad \sum_{h=1}^{H} W_h^* = 1 \]  

(13)

The choice of quantity \( Q_h \) determines the form of the estimator.

Now, let us consider the Lagrange function given by

\[ \Psi = \sum_{h=1}^{H} \left( \frac{(W_h^* - W_h)^2}{W_h Q_h} \right) - 2 \nu_1 \left( \sum_{h=1}^{H} W_h^* \bar{x}_h - \sum_{h=1}^{H} W_h \bar{x}_h \right) - 2 \nu_2 \left( \sum_{h=1}^{H} Q_h - 1 \right) \]  

(14)

where \( \nu_1 \) and \( \nu_2 \) are Lagrange’s multipliers. Now differentiating (14) with respect to \( W_h^* \) and equal to zero, we get the new calibration weight as

\[
W_h^* = W_h + W_h Q_h \left[ \sum_{h=1}^{H} W_h (\bar{x}_h - \bar{x}_h) \left( \sum_{h=1}^{H} W_h Q_h \right) \right] \\
+ W_h Q_h \left[ \sum_{h=1}^{H} W_h (\bar{x}_h - \bar{x}_h) \left( \sum_{h=1}^{H} W_h Q_h \bar{x}_h \right) \right] \\
\times \left( \sum_{h=1}^{H} W_h Q_h \bar{x}_h \right) - \left( \sum_{h=1}^{H} W_h Q_h \bar{x}_h \right)^2 
\]

(15)

Now substituting the value of \( W_h^* \) in the equation (12), we get the calibration estimator as

\[ \bar{y}_{sr}(c) = \sum_{h=1}^{H} W_h \left( \frac{y}{x_h} \right) \bar{x}_h + \hat{\beta}_w^* (\bar{x} - \sum_{h=1}^{H} W_h \bar{x}_h) \]  

(16)

Where

\[
\hat{\beta}_w^* = \frac{\left( \sum_{h=1}^{H} Q_h W_h \right) \left( \sum_{h=1}^{H} Q_h W_h \bar{x}_h \right) - \left( \sum_{h=1}^{H} Q_h W_h \left( \frac{y}{x_h} \right) \bar{x}_h \right) \left( \sum_{h=1}^{H} Q_h W_h \bar{x}_h \right)}{\left( \sum_{h=1}^{H} Q_h W_h \right) \left( \sum_{h=1}^{H} Q_h W_h \bar{x}_h^2 \right) - \left( \sum_{h=1}^{H} Q_h W_h \bar{x}_h \right)^2} 
\]

(17)

3.1 Approximate mean square error (MSE) of the proposed estimator

The variance estimate is necessary to estimate the accuracy of the survey estimates. In most scenarios, the unbiased estimates of the nonlinear parametric functions are not available. So we use the Taylor linearization method to find the approximate mean square error (MSE) of nonlinear estimators that can be used for complicated survey designs. To find the variance of estimators, we define the following notations and for simplicity we take \( Q_h = 1 \). Let us assume
\[
\begin{align*}
t_1 &= \sum_{h=1}^{H} W_h \left( \frac{y_h}{x_h} \right) \bar{X}_h, \quad t_2 = \sum_{h=1}^{H} Q_h W_h \bar{y}_h, \quad t_3 = \sum_{h=1}^{H} Q_h W_h \left( \frac{y_h}{x_h} \right) \bar{X}_h, \\
\quad t_4 &= \sum_{h=1}^{H} Q_h W_h \bar{x}_h, \quad t_5 = \sum_{h=1}^{H} Q_h W_h \bar{x}_h^2, \quad \text{and} \quad t_6 = \sum_{h=1}^{H} W_h \bar{x}_h
\end{align*}
\] (18)

By using the values of \( t_1, t_2, t_3, t_4, t_5, \) and \( t_6 \) from (18), equation (16) becomes
\[
\bar{y}_{sr}(c) = t_1 + \left( \frac{t_2 - t_4}{t_5 - t_6^2} \right) (\bar{X} - t_6)
\] (19)

Let, \( \bar{y}_{sr}(c) = T = F(t_1, t_2, t_3, t_4, t_5, t_6) \) (20)

For further calculation, we use some approximations for the \( h^{th} \) stratum. For large sample approximation, we assume
\[
\bar{y}_h = \bar{Y}_h(1 + \epsilon_{h0}) \quad \text{and} \quad \bar{x}_h = \bar{X}_h(1 + \epsilon_{h1}) \quad \text{such that} \quad E(\epsilon_{h0}) = E(\epsilon_{h1}) = 0
\]

Then \( E(\epsilon^2_{h0}) = \left( \frac{1 - f_h}{n_h} \right) C_{yy}^2, \quad E(\epsilon^2_{h1}) = \left( \frac{1 - f_h}{n_h} \right) C_{xx}^2, \quad E(\epsilon_{h0} \epsilon_{h1}) = \left( \frac{1 - f_h}{n_h} \right) \rho_{xy} C_{xx} C_{yy} \) and
\[
f_h = \frac{1}{n_h} - \frac{1}{N_h}
\] (21)

Now we have,
\[
\begin{align*}
E(t_1) &= \sum_{h=1}^{H} W_h \bar{y}_h \left( 1 + E(\epsilon^2_1) - E(\epsilon_{01}) \right) = \theta_1, \quad E(t_2) = \sum_{h=1}^{H} W_h Q_h \bar{y}_h \bar{x}_h = \theta_2 \\
E(t_3) &= \sum_{h=1}^{H} W_h Q_h \bar{y}_h \left( 1 + E(\epsilon^2_1) - E(\epsilon_{01}) \right) = \theta_3, \quad E(t_4) = \sum_{h=1}^{H} W_h Q_h \bar{y}_h \bar{x}_h = \theta_4 \\
E(t_5) &= \sum_{h=1}^{H} W_h Q_h \bar{x}_h \left( 1 + E(\epsilon^2_1) \right) = \theta_5, \quad E(t_6) = \sum_{h=1}^{H} W_h \bar{x}_h = \theta_6
\end{align*}
\] (22)

Now differentiating the function \( \bar{y}_{sr}(c) = F(t_1, t_2, t_3, t_4, t_5, t_6) \) with respect to \( t_1, t_2, t_3, t_4, t_5, \) and \( t_6 \), putting their expected value in the function, we get
\[
\begin{align*}
g_1 &= \frac{\partial}{\partial t_1} F(t_1, t_2, t_3, t_4, t_5, t_6) \bigg|_{\theta=0} = 1, \quad g_2 = \frac{\partial}{\partial t_2} F(t_1, t_2, t_3, t_4, t_5, t_6) \bigg|_{\theta=0} = 0, \\
g_3 &= \frac{\partial}{\partial t_3} F(t_1, t_2, t_3, t_4, t_5, t_6) \bigg|_{\theta=0} = 0, \quad g_4 = \frac{\partial}{\partial t_4} F(t_1, t_2, t_3, t_4, t_5, t_6) \bigg|_{\theta=0} = 0
\end{align*}
\]
\[ g_5 = \frac{\partial}{\partial t_5} F(t_1, t_2, t_3, t_4, t_5, t_6) \bigg|_{t_1=t_2=t_3=t_4=0} = 0, \quad g_6 = \frac{\partial}{\partial t_6} F(t_1, t_2, t_3, t_4, t_5, t_6) \bigg|_{t_1=t_2=t_3=t_4} = (-1) \left( \frac{t_2 - t_4}{t_2 - t_4} \right) \]  

(23)

By using (23), we can write equation (19) as

\[ T = \bar{Y} + g_1(t_1 - \theta_1) + g_2(t_2 - \theta_2) + g_3(t_3 - \theta_3) + g_4(t_4 - \theta_4) + g_5(t_5 - \theta_5) + g_6(t_6 - \theta_6) \]  

(24)

Putting the values of calculated \( g_1, g_2, g_3, g_4, g_5, \) and \( g_6 \) in equation (24), we get

\[ T = \bar{Y} + g_1(t_1 - \theta_1) + g_6(t_6 - \theta_6) \]

Now the approximate mean square error (MSE) of \( T \) is given as

\[ MSE(T) \approx V(\bar{Y}) + V(g_1(t_1 - \theta_1)) + V(g_6(t_6 - \theta_6)) , \]

\[ MSE(T) \approx V(\bar{Y}) + g_1^2V(t_1 - \theta_1) + g_6^2V(t_6 - \theta_6) , \]

\[ MSE(T) \approx V(t_1) + g_6^2V(t_6) + 2g_6Cov(t_1, t_6) \]  

(25)

where \( V(t_1) = \sum_{h=1}^{H} W_h^2 \left[ \frac{1-f_h}{n_h} \right] S_{hy}^2 + R_h^2 S_{hc}^2 - 2R_h S_{hxy} \] , \( V(t_6) = \sum_{h=1}^{H} W_h^2 \left[ \frac{1-f_h}{n_h} \right] S_{hc}^2 \) , and

\[ Cov(t_1, t_6) = \sum_{h=1}^{H} W_h^2 \bar{y}_h \bar{x}_h \left[ E(\varepsilon_0 \varepsilon_1) - E(\varepsilon_1^2) \right] \]

After substituting the values of the above expressions into an equation (25), we get the approximate mean square error (MSE) of the proposed calibrated estimator. We take \( Q_h = 1 \) throughout our study for simplicity.

4 Empirical Study

4.1 Simulation study

In the sampling theory, researchers typically check the estimator's performance by comparing its approximate mean square error (MSE) with respect to the other estimators. The approximate mean square error (MSE) of the estimators relies entirely on the population parameters assumed to be known. But in the practical situation, parametric values are rarely available to be known. To avoid such practical difficulties, one can perform a simulation analysis based on the actual samples drawn from the population. In simulation analysis, a large number of independent samples are drawn from the population and the mean square error (MSE) is calculated to compare the efficiency of the estimators. We therefore attempted to conduct a simulation study to verify the performance of the proposed calibration estimator \( \bar{Y}_c \). We perform a simulation study by generating four artificial population from normal distribution with different assumed means and standard deviations. To get different level of correlations between study variables and auxiliary variable, we use transformation from Reddy et al. [11]. Each population consists of four strata. We select a sample from each stratum by using simple random sampling. Here we take a population of size 2650. Table 1 shows the distribution of generated population to perform the study. Table 2 represents the transformations which is used to generate correlated random variables.
Table 1. Population parameters for the \( h^{th} \) stratum

<table>
<thead>
<tr>
<th>Strata no.</th>
<th>Stratum Size ( (N_h) )</th>
<th>Sample size ( (n_h) )</th>
<th>Distribution of study variable ( (y_{hi}) )</th>
<th>Distribution of auxiliary variable ( (x_{hi}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>600</td>
<td>120</td>
<td>( N(22,2) )</td>
<td>( N(20,2) )</td>
</tr>
<tr>
<td>II</td>
<td>700</td>
<td>140</td>
<td>( N(33,3) )</td>
<td>( N(25,3) )</td>
</tr>
<tr>
<td>III</td>
<td>550</td>
<td>110</td>
<td>( N(44,4) )</td>
<td>( N(30,4) )</td>
</tr>
<tr>
<td>IV</td>
<td>800</td>
<td>160</td>
<td>( N(55,5) )</td>
<td>( N(40,5) )</td>
</tr>
</tbody>
</table>

Table 2. Number of strata, correlation coefficient and the transformation

<table>
<thead>
<tr>
<th>Strata no.</th>
<th>Correlation between ( (y_{hi}) ) and ( (x_{hi}) )</th>
<th>Transformed auxiliary variable ( (x_{hi}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>( r_{y1} = 0.85 )</td>
<td>( y_1 = r_{y1}x_{21} + \sqrt{1 - r_{y1}^2} y_{21} )</td>
</tr>
<tr>
<td>II</td>
<td>( r_{y2} = 0.88 )</td>
<td>( y_2 = r_{y2}x_{22} + \sqrt{1 - r_{y2}^2} y_{22} )</td>
</tr>
<tr>
<td>III</td>
<td>( r_{y3} = 0.80 )</td>
<td>( y_3 = r_{y3}x_{23} + \sqrt{1 - r_{y3}^2} y_{23} )</td>
</tr>
<tr>
<td>IV</td>
<td>( r_{y4} = 0.75 )</td>
<td>( y_4 = r_{y4}x_{24} + \sqrt{1 - r_{y4}^2} y_{24} )</td>
</tr>
</tbody>
</table>

We finally calculate the Average MSE (AMSE) of the estimators using the following formulas

\[
AMSE(\overline{y}_{st}(n)) = \frac{1}{5000} \sum_{l=1}^{5000} (\overline{y}_{st}(n) - \overline{Y})^2; l = 1,2,...,5000
\]

\[
AMSE(\overline{y}_{sr}) = \frac{1}{5000} \sum_{l=1}^{5000} (\overline{y}_{sr} - \overline{Y})^2; l = 1,2,...,5000
\]

\[
AMSE(\overline{y}_{sr}(c)) = \frac{1}{5000} \sum_{l=1}^{5000} (\overline{y}_{sr}(c) - \overline{Y})^2; l = 1,2,...,5000
\]

The formula for the percent relative efficiency (PRE) is given as the ratio of the average mean square error (AMSE) of two estimators to be compared.

\[
PRE = \frac{AMSE(\overline{y}_{sr})}{AMSE(\overline{y}_{sr}(c))} \times 100
\]

Table 3. Average mean square error (AMSE) and percent relative efficiency (PRE) of estimator in case of stratified sampling

<table>
<thead>
<tr>
<th>Estimator</th>
<th>AMSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \overline{y}_{st}(n) )</td>
<td>0.03259475</td>
<td>100</td>
</tr>
<tr>
<td>( \overline{y}_{sr} )</td>
<td>0.02264014</td>
<td>203.6183</td>
</tr>
<tr>
<td>( \overline{y}_{sr}(c) )</td>
<td>0.01111891</td>
<td>287.6743</td>
</tr>
</tbody>
</table>
4.2 Real data application

To examine the performance of the proposed calibration estimator, we conduct a simulation study on real data set to support the theoretical result. The population data was obtained from the Kaggle website. https://www.kaggle.com/datasets/darshansavaliya/gujratcropdataanalysis is the link provided. We are attempting to determine the average production of Gujarat in 2011 as it varies over the various agricultural seasons in Gujarat. In this dataset, we take into account X as an area in (hectares) and Y as a production in (tones). Based on the harvest season, our population is split into four groups: Kharif, Rabi, Summer, and Whole Year. Table 4 provides more information about the population’s statistical parameters.

<table>
<thead>
<tr>
<th>Strata No.</th>
<th>Ni</th>
<th>$\bar{Y}_h$</th>
<th>$\bar{X}_h$</th>
<th>$\rho_{xyh}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>292</td>
<td>61016.1</td>
<td>27080.82</td>
<td>0.91</td>
</tr>
<tr>
<td>2</td>
<td>163</td>
<td>43096.93</td>
<td>13474.23</td>
<td>0.78</td>
</tr>
<tr>
<td>3</td>
<td>112</td>
<td>19500.89</td>
<td>8243.75</td>
<td>0.92</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
<td>169716</td>
<td>5560</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Table 4. Values of the parameter for the $h^{th}$ stratum

<table>
<thead>
<tr>
<th>Estimator</th>
<th>AMSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{Y}_{st}(n)$</td>
<td>1151285003</td>
<td>100</td>
</tr>
<tr>
<td>$\bar{Y}_{sr}$</td>
<td>715864702</td>
<td>163.381</td>
</tr>
<tr>
<td>$\bar{Y}_{sr}(c)$</td>
<td>438156732</td>
<td>287.6743</td>
</tr>
</tbody>
</table>

Table 5. Average mean square error (AMSE) and percent relative efficiency (PRE) of the estimator in case of stratified sampling

5 Conclusions

A new calibration estimator has been proposed to estimate the population mean in a stratified sample design using the information about auxiliary variables. In this paper, we obtained some new optimal stratum weights for which the estimator would achieve a minimum approximate mean square error (MSE). The performance of the proposed calibration estimator has been compared with respect to $\bar{Y}_{st}(n)$ and $\bar{Y}_{sr}$. The theoretical facts are supported by the empirical study. The empirical study was carried out using both an artificially generated data set and a real data set. Table 3 and Table 5 shows that the proposed calibration estimator $\bar{Y}_{sr}(c)$ gives better estimates as compared to the existing calibration estimator $\bar{Y}_{st}(n)$ and $\bar{Y}_{sr}$.

Acknowledgements

Authors are thankful to the anonymous reviewers for their constructive remarks.

Competing Interests

Authors have declared that no competing interests exist.

References


