Improved ratio-type estimators using maximum and minimum values under simple random sampling scheme

Mursala Khan∗, Saif Ullah†, Abdullah Y. Al-Hossain‡ and Neelam Bashir¶

Abstract

This paper presents a class of ratio-type estimators for the evaluation of finite population mean under maximum and minimum values by using knowledge of the auxiliary variable. The properties of the proposed estimators in terms of biases and mean square errors are derived up to first order of approximation. Also, the performance of the proposed class of estimators is shown theoretically and these theoretical conditions are, then, verified numerically by taking three natural populations under which the proposed class of estimators performed better than other competing estimators.

Keywords: Study variable, Auxiliary variable, Ratio estimators, Maximum and Minimum values, Simple random sampling, Mean squared error, Efficiency.

2000 AMS Classification: 62D05

1. Introduction

A lot of work has been done for the estimation of finite population mean using auxiliary information for improving the efficiency of the estimators. Das and Tripathi (1980, 1981), Uphadhyaya and Singh (1999), Singh (2004), Sisodia and Dwivedi (1981) proposed ratio estimator using coefficient of variation of an auxiliary variable. Kadilar and Cingi (2005) suggested ratio estimators in stratified random sampling. In the same way, Kadilar and Cingi (2006) proposed an improvement in estimating the population mean by using the correlation coefficient. Khan and Shabbir (2013) proposed a ratio-type estimator for the estimation of population variance using the knowledge of quartiles and their functions as auxiliary information. They proposed different modified estimators for the estimation of finite population mean using maximum and minimum values. Recently Hossain and Khan (2014) worked on the estimation of population mean using maximum and minimum values under simple random sampling by incorporating the knowledge of two auxiliary variables.

∗Department of Mathematics, COMSATS Institute of Information Technology Abbottabad, 22060, Pakistan. Email: mursala.khan@yahoo.com
†Corresponding Author.
‡Department of Mathematics, Government College University Lahore, 54000, Pakistan. Email: dr.saifullah@gcu.edu.pk
§Department of Mathematics, Faculty of Science, Jazan University, Jazan 2097, Saudi Arabia.
¶Department of Mathematics, Government College University Lahore, 54000, Pakistan.
Let us consider a finite population of size $N$ of different units $U = \{U_1, U_2, U_3, \ldots, U_N\}$. Let $y$ and $x$ be the study and the auxiliary variable with corresponding values $y_i$ and $x_i$ respectively for the $i$-th unit $i = \{1, 2, 3, \ldots, N\}$ defined on a finite population $U$. Let $\bar{Y} = (1/N) \sum_{i=1}^{N} y_i$ and $\bar{X} = (1/N) \sum_{i=1}^{N} x_i$ be the population means of the study and the auxiliary variable, respectively. Also $S_y^2 = (1/(N - 1)) \sum_{i=1}^{N} (y_i - \bar{Y})^2$ and $S_x^2 = (1/(N - 1)) \sum_{i=1}^{N} (x_i - \bar{X})^2$ be the corresponding population mean square error of the study and the auxiliary variable respectively, and let $C_y = S_y^2 / \bar{Y}$ and $C_x = S_x^2 / \bar{X}$ be the coefficients of variation of the study and the auxiliary variable respectively, and $\rho_{yx} = S_{yx} / S_y S_x$ be the population correlation coefficient between $x$ and $y$.

In order to estimate the unknown population parameters we take a random sample of size $n$ units from the finite population $U$ by using simple random sample without replacement. Let $\bar{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$ be the corresponding sample means of the study and the auxiliary variable respectively, and their corresponding sample variances are $\hat{S}_y^2 = (1/(n - 1)) \sum_{i=1}^{n} (y_i - \bar{y})^2$ and $\hat{S}_x^2 = (1/(n - 1)) \sum_{i=1}^{n} (x_i - \bar{x})^2$ respectively.

When there is no auxiliary information the usual unbiased estimator for the population mean of the study variable is:

\begin{equation}
\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n},
\end{equation}

The variance of the estimator $\bar{y}$ is given by:

\begin{equation}
\text{var} (\bar{y}) = \theta S_y^2, \quad \text{where} \quad \theta = \frac{1}{N} - \frac{1}{N}.
\end{equation}

In many populations there exist some large ($y_{\text{max}}$) or small ($y_{\text{min}}$) values and to estimate the population parameters without considering this information is very sensitive. In either case the result will be overestimated or underestimated. In order to handle this situation Sarndal (1972) suggested the following unbiased estimator for the assessment of finite population mean:

\begin{equation}
\bar{y}_s = \begin{cases} 
\bar{y} + c & \text{if sample contains } y_{\text{min}} \text{ but not } y_{\text{max}} \\
\bar{y} - c & \text{if sample contains } y_{\text{max}} \text{ but not } y_{\text{min}} \\
\bar{y} & \text{for all other samples},
\end{cases}
\end{equation}

were $c$ is a constant whose value is to be found for minimum variance.

The minimum variance of the estimator $\bar{y}_s$ up to first order of approximation is given as under:

\begin{equation}
\text{var} (\bar{y}_s)_{\text{min}} = \text{var} (\bar{y}) - \frac{\theta (y_{\text{max}} - y_{\text{min}})^2}{2(N - 1)},
\end{equation}

where the optimum value of $c_{\text{opt}}$ is

\begin{equation}
c_{\text{opt}} = \frac{(y_{\text{max}} - y_{\text{min}})}{2n}.
\end{equation}
The classical ratio estimator for finding the population mean of the study variable is given by:

\[ \tilde{Y}_R = \bar{Y} \frac{X}{x} \]  

(1.5)

The bias and mean square errors of the estimator \( \tilde{Y}_R \) up to first order of approximation are given by:

\[ \text{Bias}(\tilde{Y}_R) = \frac{\theta}{X} (RS_x^2 - S_{yx}) \]  

(1.6)

\[ \text{MSE}(\tilde{Y}_R) = \theta (S_y^2 + R^2 S_x^2 - 2RS_{yx}) \]  

(1.7)

Similarly, Sisodia and Dwivedi (1981) suggested the following ratio estimator using the knowledge of coefficient of variation of the auxiliary variable:

\[ \tilde{Y}_{SD} = \bar{Y} \left( \frac{X + C_x}{x + C_x} \right) \]  

(1.8)

The bias and mean square errors of the estimator \( \tilde{Y}_{SD} \) up to first order of approximation are as follows:

\[ \text{Bias}(\tilde{Y}_{SD}) = \frac{\theta \alpha_1}{X} (R\alpha_1 S_x^2 - S_{yx}), \]  

(1.9)

\[ \text{MSE}(\tilde{Y}_{SD}) = \theta (S_y^2 + \alpha_1^2 S_x^2 - 2\alpha_1 S_{yx}), \]  

(1.10)

where \( \alpha_1 = \frac{\bar{Y}}{X + C_x} \).

2. The proposed class of estimators

On the lines of Sarndal (1972), we propose a class of ratio-type estimators for the estimation of finite population mean using knowledge of the coefficient of variation and coefficient of correlation of an auxiliary variable. Usually when the correlation between the study variable \( (y) \) and the auxiliary variable \( (x) \) is positive, then the selection of the larger value of the auxiliary variable \( (x) \), the larger value of study variable \( (y) \) is to be expected, and the smaller the value of auxiliary variable \( (x) \), the smaller the value of study variable \( (y) \). Using such type of information, we propose the following class of estimators given by:

\[ \tilde{Y}_{P1} = \bar{Y}_{c_1} \left( \frac{X + C_x}{x + C_x} \right), \]  

(2.1)

\[ \tilde{Y}_{P2} = \bar{Y}_{c_1} \left( \frac{X + \rho_{yx}}{x + \rho_{yx}} \right), \]  

(2.2)

\[ \tilde{Y}_{P3} = \bar{Y}_{c_1} \left( \frac{XC_x + \rho_{yx}}{x + \rho_{yx}} \right), \]  

(2.3)

\[ \tilde{Y}_{P4} = \bar{Y}_{c_1} \left( \frac{X\rho_{yx} + C_x}{x + \rho_{yx}} \right), \]  

(2.4)
where \( \bar{y}_{c_1} = \bar{y} + c_1, \; \bar{x}_{c_2} = \bar{x} + c_2 \), also \( c_1 \) and \( c_2 \) are unknown constants.

To obtain the properties of \( \hat{Y}_{P_i} \) in terms of Bias and Mean square error, we define the following relative error terms and their expectations:

\[
\zeta_0 = \frac{y_{c_1} - \bar{y}}{\bar{y}}, \quad \zeta_1 = \frac{x_{c_2} - \bar{x}}{\bar{x}}, \text{ such that } E(\zeta_0) = E(\zeta_1) = 0,
\]

Also,

\[
E(\zeta_0^2) = \frac{\theta}{\bar{y}} \left( S_y^2 - \frac{2nc_1}{N-1}(y_{max} - y_{min} - nc_1) \right), \quad E(\zeta_1^2) = \frac{\theta}{\bar{x}} \left( S_x^2 - \frac{2nc_2}{N-1}(x_{max} - x_{min} - nc_2) \right)
\]

and

\[
E(\zeta_0 \zeta_1) = \frac{\theta}{\bar{y} \bar{x}} \left( S_{yx} - \frac{n}{N-1}(c_2(y_{max} - y_{min}) + c_1(x_{max} - x_{min}) - 2nc_1c_2) \right).
\]

Where \( \theta = \frac{1}{n} - \frac{1}{N}, \quad R = \frac{\bar{Y}}{\bar{X}}, \quad \alpha_{P_1} = \frac{\bar{X}}{\bar{X} + C_x}, \quad \alpha_{P_2} = \frac{\bar{X}}{\bar{X} + \rho_{yx}}, \quad \alpha_{P_3} = \frac{\bar{X} C_x}{\bar{X} C_x + \rho_{yx}}, \quad \alpha_{P_4} = \frac{\bar{X} \rho_{yx}}{\bar{X} \rho_{yx} + C_x}, \quad k_{P_1} = \frac{\bar{Y}}{\bar{X} + C_x}, \quad k_{P_2} = \frac{\bar{Y}}{\bar{X} + \rho_{yx}}, \quad k_{P_3} = \frac{\bar{Y} C_x}{\bar{X} C_x + \rho_{yx}}, \quad k_{P_4} = \frac{\bar{Y} \rho_{yx}}{\bar{X} \rho_{yx} + C_x}\]

Rewriting \( \hat{Y}_{P_i} \) in terms of \( \zeta_i \)'s, we have

\[
\hat{Y}_{P_i} = Y \left( 1 + \zeta_0 \right) \left( 1 + \alpha_{P_i} \zeta_1 \right)^{-1},
\]

where \( \hat{Y}_{P_i} \) represent the proposed class of estimators for \( i=1,2,3,4 \).

Expanding the right hand side of the equation given above and including terms up to second powers of \( \zeta_i \)'s i.e., up to first order of approximation, we have:

\[
(2.5) \quad \hat{Y}_{P_i} - Y = Y \left( \zeta_0 - \alpha_{P_i} \zeta_1 + \alpha_{P_i}^2 \zeta_1^2 - \alpha_{P_i} \zeta_0 \zeta_1 \right)
\]

Taking expectation on both sides of (2.5), we get bias up to first order of approximation which is given as:

\[
\text{Bias}(\hat{Y}_{P_i}) = \frac{\theta}{\bar{y}} k_{P_i} \left[ k_{P_i} \left( S_y^2 - \frac{2nc_1}{N-1}(x_{max} - x_{min} - nc_2) \right) \right] - S_{yx}
\]

(2.6) \quad + \frac{n}{N-1}(c_2(y_{max} - y_{min}) + c_1(x_{max} - x_{min}) - 2nc_1c_2)

On squaring both sides of (2.5), and keeping \( \zeta_i \)'s powers up to first order of approximation, we get:

\[
(2.7) \quad \left( \hat{Y}_{P_i} - Y \right)^2 = Y^2 \left( \zeta_0^2 + \alpha_{P_i}^2 \zeta_1^2 - 2\alpha_{P_i} \zeta_0 \zeta_1 \right)
\]
Taking expectation on both sides of (2.7), we get mean square error up to first order of approximation, given as under:

\[
\text{MSE}\left(\hat{Y}_{P_i}\right) = \theta\left[\left(S_y^2 + k_P^2 S_x^2 - 2k_P S_{yx}\right) - \frac{2n}{N-1}\left\{\left(c_1 - c_2 k_P\right)(y_{\text{max}} - y_{\text{min}})
\right.\right.
\vspace{0.5cm}
\left.\left. - n(c_1 - c_2 k_P) - k_P(x_{\text{max}} - x_{\text{min}})\right)\right\}\right]
\]

(2.8)

The optimum values of \(c_1\) and \(c_2\) are given in the following lines:

\[
\begin{align*}
&\quad c_1 = \frac{(y_{\text{max}} - y_{\text{min}})}{2n} \\
&\quad c_2 = \frac{(x_{\text{max}} - x_{\text{min}})}{2n}
\end{align*}
\]

(2.9)

On substituting the optimum value of \(c_1\) and \(c_2\) in (2.8), we get the minimum mean square error of the proposed estimators as follows:

\[
\text{MSE}\left(\hat{Y}_{P_i}\right)_{\text{min}} = \theta\left[\left(S_y^2 + k_P^2 S_x^2 - 2k_P S_{yx}\right)
\vspace{0.5cm}
\left.\left.- \frac{1}{2(N-1)}\left((y_{\text{max}} - y_{\text{min}}) - k_P(x_{\text{max}} - x_{\text{min}})\right)^2\right)\right]\]

(2.10)

3. Comparison of estimators

In this section, we compare the proposed class of estimators with other existing estimators and some of their efficiency comparison conditions have been carried out under which the proposed class of estimators perform better than the other existing estimators discussed in the literature above.

(i) By (1.2) and (2.10),

\[
\left[\text{MSE}\left(\bar{Y}\right) - \text{MSE}\left(\hat{Y}_{P_i}\right)_{\text{min}}\right] \geq 0, \text{ if}
\]

\[
\left[\frac{1}{2(N-1)}\left((y_{\text{max}} - y_{\text{min}}) - k_P(x_{\text{max}} - x_{\text{min}})\right)^2 - k_P^2 S_x^2 + 2k_P S_{yx}\right] \geq 0.
\]

(ii) By (1.4) and (2.10),

\[
\left[\text{MSE}\left(\bar{Y}_s\right) - \text{MSE}\left(\hat{Y}_{P_i}\right)_{\text{min}}\right] \geq 0, \text{ if}
\]

\[
k_P\left[\frac{(x_{\text{max}} - x_{\text{min}})^2}{2(N-1)} - S_x^2\right] - \left(\frac{(y_{\text{max}} - y_{\text{min}})(x_{\text{max}} - x_{\text{min}})}{N-1} - 2S_{yx}\right) \geq 0.
\]

(iii) By (1.7) and (2.10),

\[
\left[\text{MSE}\left(\bar{Y}_R\right) - \text{MSE}\left(\hat{Y}_{P_i}\right)_{\text{min}}\right] \geq 0, \text{ if}
\]

\[
\left[\frac{1}{2(N-1)}\left((y_{\text{max}} - y_{\text{min}}) - k_P(x_{\text{max}} - x_{\text{min}})\right)^2 + S_x^2(R - k_P)(R + k_P - 2\delta)\right] \geq 0.
\]
(iv) By (1.10) and (2.10),
\[
\left[ \text{MSE} \left( \hat{Y}_{SD} \right) - \text{MSE} \left( \hat{Y}_{P_i} \right) \right]_{\min} \geq 0, \quad \text{if}
\]
\[
\left[ \frac{1}{2(N-1)} \left( (y_{max} - y_{min}) - k_{P_i}(x_{max} - x_{min}) \right)^2 + S^2_x (R\alpha_1 - k_{P_i})(R\alpha_1 + k_{P_i} - 2\delta) \right] \geq 0.
\]
Where
\[
\delta = \frac{\rho_{yx} S_y}{S_x}.
\]

4. Numerical illustration

In this section, we illustrate the performance of the proposed class of estimators in comparison with various other existing estimators through three natural populations. The description and the necessary data statistics are given by:

**Population-1:** [Source: Singh and Mangat (1996), p.193]

Y: be the milk yield in kg after new food, and

X: be the yield in kg before new yield.

\[
N = 27, \quad n = 12, \quad \bar{X} = 10.4111, \quad \bar{Y} = 11.2519, \quad y_{max} = 14.8, \quad y_{min} = 7.9, \quad x_{max} = 14.5,
\]
\[
x_{min} = 6.5, \quad S^2_y = 4.103, \quad S^2_x = 4.931, \quad S_{yx} = 4.454, \quad \rho_{yx} = 0.990.
\]

**Population-2:** [Source: Murthy (1967), p.399]

Y: be the area under wheat crop in 1964, and

X: be the area under wheat crop in 1963.

\[
N = 34, \quad n = 12, \quad \bar{X} = 208.882, \quad \bar{Y} = 199.441, \quad y_{max} = 634, \quad y_{min} = 6, \quad x_{max} = 564,
\]
\[
x_{min} = 5, \quad S^2_y = 22564.56, \quad S^2_x = 22652.05, \quad S_{yx} = 22158.05, \quad \rho_{yx} = 0.980.
\]

**Population-3:** [Source: Cochran (1977), p.152]

Y: be the population size in 1930 (in 1000), and

X: be the population size in 1920 (in 1000).

\[
N = 49, \quad n = 12, \quad \bar{X} = 103.1429, \quad \bar{Y} = 127.7959, \quad y_{max} = 634, \quad y_{min} = 46, \quad x_{max} = 507,
\]
\[
x_{min} = 2, \quad S^2_y = 15158.83, \quad S^2_x = 10900.42, \quad S_{yx} = 12619.78, \quad \rho_{yx} = 0.98.
The mean squared error of the proposed class and the existing estimators are shown in Table-1.

**Table-1: MSE of the Competing and the Proposed Class of Estimators**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Population 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$MSE(\cdot)$</td>
<td>$MSE(\cdot)$</td>
<td>$MSE(\cdot)$</td>
</tr>
<tr>
<td>Existing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{y}$</td>
<td>0.1900</td>
<td>1220.9455</td>
<td>953.4904</td>
</tr>
<tr>
<td>$\bar{y}_s$</td>
<td>0.1476</td>
<td>898.8652</td>
<td>726.9560</td>
</tr>
<tr>
<td>$\hat{\bar{y}}_s$</td>
<td>0.0109</td>
<td>48.5723</td>
<td>39.0823</td>
</tr>
<tr>
<td>$\hat{\bar{y}}_{SD}$</td>
<td>0.0092</td>
<td>48.6879</td>
<td>37.8472</td>
</tr>
<tr>
<td>Proposed</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\bar{y}}_{P_1}$</td>
<td>0.0070</td>
<td>46.8878</td>
<td>37.1675</td>
</tr>
<tr>
<td>$\hat{\bar{y}}_{P_1}$</td>
<td>0.0044</td>
<td>41.2479</td>
<td>37.2200</td>
</tr>
<tr>
<td>$\hat{\bar{y}}_{P_5}$</td>
<td>0.0085</td>
<td>41.2162</td>
<td>37.2378</td>
</tr>
<tr>
<td>$\hat{\bar{y}}_{P_4}$</td>
<td>0.0070</td>
<td>41.1896</td>
<td>37.1704</td>
</tr>
</tbody>
</table>

For the percent relative efficiencies (PREs) of the proposed class and the existing estimators, we use the following expression for efficiency comparison. The results are, then, shown in Table-2.

\[
\text{PRE}(\hat{\bar{y}}_g, \bar{y}) = \frac{\text{MSE}(\bar{y})}{\text{MSE}(\hat{\bar{y}}_g)} \times 100, \quad \text{where} \quad g = S, R, SD, P_1, P_2, P_3 \text{ and } P_4.
\]
5. Conclusion

In this study, we have developed some ratio-type estimators under maximum and minimum values using knowledge of the coefficient of variation and coefficient of correlation of the auxiliary variable. We have found some theoretical possibilities under which the proposed class of estimators have smaller mean squared errors than the usual unbiased estimator; the classical ratio estimator; and the other competing estimators suggested by statisticians. Theoretical results are also verified with the help of three natural populations and their statistics are shown in table 1 and table 2, which clearly indicates that the proposed estimators have smaller mean squared errors and larger percent relative efficiency than the other estimators discussed in the literature. Thus the proposed estimators under maximum and minimum values may be preferred over the existing estimators for the use of practical applications.

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Population 1</th>
<th>Population 2</th>
<th>Population 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{y}$</td>
<td>100.00</td>
<td>100.00</td>
<td>100.00</td>
</tr>
<tr>
<td>$\hat{y}_2$</td>
<td>128.7263</td>
<td>135.8319</td>
<td>131.1621</td>
</tr>
<tr>
<td>$\hat{y}_{12}$</td>
<td>1743.1193</td>
<td>2513.6662</td>
<td>2439.6988</td>
</tr>
<tr>
<td>$\hat{y}_{CD}$</td>
<td>2065.2174</td>
<td>2507.6414</td>
<td>2519.3156</td>
</tr>
<tr>
<td>$\hat{y}_{A}$</td>
<td>2714.2857</td>
<td>2603.9727</td>
<td>2565.3875</td>
</tr>
<tr>
<td>$\hat{y}_{22}$</td>
<td>4318.1818</td>
<td>2960.0186</td>
<td>2561.7689</td>
</tr>
<tr>
<td>$\hat{y}_{2B}$</td>
<td>2235.2941</td>
<td>2962.2952</td>
<td>2560.5444</td>
</tr>
<tr>
<td>$\hat{y}_{24}$</td>
<td>2714.2857</td>
<td>2964.2082</td>
<td>2565.1874</td>
</tr>
</tbody>
</table>

Table-2: PRE of Different Estimators with Respect to $y$
6. Acknowledgements

The authors wish to thank the reviewers for their careful assessment, constructive comments and suggestions, which significantly improved this paper.

References