Some Modified Exponential Ratio Type Estimators of Finite Population Mean in Survey Sampling

Archana Panigrahi¹, Gopabandhu Mishra²

¹,² Department of Statistics, Utkal University, Bhubaneswar-751004, Odisha, India

Abstract: Bahl and Tuteja (1991) have suggested an exponential ratio type estimator to estimate finite population mean. In this paper some modified exponential ratio type estimators of finite population mean under simple random sampling without replacement have been proposed following Searls (1964), Srivastava (1974) and Upadhyaya and Srivastava (1976a and 1976b). The efficiencies of these estimators are compared with exponential ratio type estimator as regard to bias and mean square error both theoretically and empirically.

Keywords: Simple random sampling, Exponential ratio type estimators, Bias, Mean square error, Efficiency.

I. INTRODUCTION

In survey sampling, the utilisation of auxiliary information for improving precision of the estimate is well recognised. The classical ratio estimator of Cochran (1940) is one of such estimators which make use of population mean of auxiliary variable to increase the efficiency of the estimator. Searls (1964) utilised the known coefficient of variation of study variable \( y \) for estimation of population mean \( \bar{Y} \). Srivastava (1974) developed an estimator of population mean \( \bar{Y} \) using estimated coefficient of variation of \( y \). Upadhyaya and Srivastava (1976a and 1976b) suggested an improved estimator of \( \bar{Y} \) in a symmetrical population using estimated coefficient of variation of study variable \( y \). Following Searls (1964), Srivastava (1974) and Upadhyaya and Srivastava (1976a and 1976b), some modified exponential ratio type estimators of population mean \( \bar{Y} \) under SRSWOR have been proposed.

Let there be a finite population \( U \) consisting of \( N \) units \( U_1, U_2, U_3, ..., U_N \). The \( i^{th} \) unit is indexed by a pair of real value \((y_i, x_i)\). It is assumed that the study variable \( y \) is positively correlated with the auxiliary variable \( x \) and is denoted by \( \rho \).

II. PROPOSED ESTIMATORS

A sample size ‘\( n \)’ is selected from \( U \) with simple random sampling without replacement (SRSWOR), denoting the sample mean of study variable and auxiliary variable \( \bar{Y} \) and \( \bar{x} \) respectively.

Searls (1964) proposed an estimator to estimate finite population mean \( \bar{Y} \) using known population coefficient of variation, i.e.,

\[
C_y = \frac{S_y}{Y} \quad \text{and} \quad S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2,
\]

is given by

\[
\bar{Y} = \frac{\bar{y}}{1 + \theta_1 C_y},
\]

where, \( \theta_1 = \left( \frac{1}{n} - \frac{1}{N} \right) \)

An exponential ratio estimator for estimating \( \bar{Y} \) suggested by Bahl and Tuteja (1991), which is more efficient than the conventional ratio estimator \( \bar{Y}_R = \frac{\bar{Y}}{\bar{x}} \) when there exist a low correlation between \( y \) and \( x \), is given by

\[
t_{ER1} = \bar{Y} \text{Exp} \left[ \frac{\bar{x} - \bar{x}}{\bar{x} + X} \right].
\]

Now, we suggest a modified exponential ratio type estimator of population mean when population coefficient of variation of \( y \), i.e., \( C_y \) is known in advance

\[
t_{ER2} = \frac{\bar{y}}{1 + \theta_1 C_y} \text{Exp} \left[ \frac{\bar{x} - \bar{x}}{\bar{x} + X} \right].
\]

Further in absence of known \( C_y \), considering estimated coefficient of variation i.e., \( \hat{C}_y \), from sample data, we suggest another estimator for \( \bar{Y} \)

\[
t_{ER3} = \frac{\bar{y}}{1 + \theta_1 \hat{C}_y} \text{Exp} \left[ \frac{\bar{x} - \bar{x}}{\bar{x} + X} \right].
\]
where, $C_{xy}^2 = \frac{s_x^2}{s_y^2}$ and $s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

Following Upadhyaya and Srivastava (1976a and 1976b), we suggest another modified exponential ratio type estimators for $\bar{Y}$ using estimated $\hat{C}_{xy}^2$ is given by

$$t_{\text{ER4}} = \bar{y}(1 + \theta_1 \hat{C}_{xy}^2) \exp \left[ \frac{X - \bar{x}}{X + \bar{x}} \right]$$

(2.5)

III. BIAS AND MSE OF DIFFERENT ESTIMATORS

Assuming the validity of Taylor’s series expansion of $t_{\text{ER1}}, t_{\text{ER2}}, t_{\text{ER3}}$ and $t_{\text{ER4}}$ and considering the expected value to $O\left(\frac{1}{n^2}\right)$, the bias of the different estimators are

$$B(t_{\text{ER1}}) = E(t_{\text{ER1}}) - \bar{Y} = \theta_1 \bar{Y} \left[ \frac{3}{8} C_{20} - \frac{1}{2} C_{11} \right]$$

(3.1)

$$B(t_{\text{ER2}}) = E(t_{\text{ER2}}) - \bar{Y} = \theta_1 \bar{Y} \left[ \frac{3}{8} C_{20} - C_{02} - \frac{1}{2} C_{11} \right]$$

(3.2)

$$B(t_{\text{ER3}}) = E(t_{\text{ER3}}) - \bar{Y} = \theta_1 \bar{Y} \left[ \frac{3}{8} C_{20} - C_{02} - \frac{1}{2} C_{11} \right]$$

(3.3)

$$B(t_{\text{ER4}}) = E(t_{\text{ER4}}) - \bar{Y} = \theta_1 \bar{Y} \left[ \frac{3}{8} C_{20} + C_{02} - \frac{1}{2} C_{11} \right]$$

(3.4)

where, $C_{rs} = \frac{K_{rs}(x,y)}{X^r Y^s}$

$K_{rs}(x,y)$ being the $(r,s)^{th}$ cumulant of $x$ and $y$.

The mean square errors (MSEs) of different estimators to $O\left(\frac{1}{n^2}\right)$ are derived as

$$\text{MSE}(t_{\text{ER1}}) = \bar{Y}^2 \left[ \theta_1 \left( C_{02} + \frac{1}{4} C_{20} - C_{11} \right) + (\theta_2 - \frac{3}{8} \theta_1 \frac{3}{4} C_{21} - \frac{3}{8} C_{30} - C_{12}) + \theta_1^2 (C_{02} C_{20} + 2 C_{21} - \frac{31}{8} C_{11} C_{20} + \frac{79}{64} C_{20}^2) \right]$$

(3.5)

where, $\theta_2 = (\frac{1}{n} - \frac{1}{N^2})$

$$\text{MSE}(t_{\text{ER2}}) = \text{MSE}(t_{\text{ER1}}) + \theta_1 \bar{Y}^2 \left( 3 C_{11} C_{20} - \frac{5}{4} C_{02} C_{20} - C_{20} \right)$$

(3.6)

$$\text{MSE}(t_{\text{ER3}}) = \text{MSE}(t_{\text{ER1}}) + \theta_1 \bar{Y}^2 \left( C_{11} C_{20} - \frac{5}{4} C_{02} C_{20} + 3 C_{02}^2 - 2 C_{03} + C_{12} \right)$$

(3.7)

$$\text{MSE}(t_{\text{ER4}}) = \text{MSE}(t_{\text{ER1}}) + \theta_1 \bar{Y}^2 \left( -C_{11} C_{20} + \frac{5}{4} C_{02} C_{20} - C_{02}^2 + 2 C_{03} - C_{12} \right)$$

(3.8)

IV. COMPARISON OF BIASES AND MEAN SQUARE ERRORS

The biases of $t_{\text{ER1}}, t_{\text{ER2}}, t_{\text{ER3}}$ and $t_{\text{ER4}}$ are of order $O\left(\frac{1}{n}\right)$ and hence, are negligible when sample size is large. From (3.2) and (3.3), the biases of modified estimators $t_{\text{ER2}}$ and $t_{\text{ER4}}$ are same i.e.

$$B(t_{\text{ER2}}) = B(t_{\text{ER4}})$$

(4.1)

However the estimators $t_{\text{ER2}}, t_{\text{ER3}}$ and $t_{\text{ER4}}$ are more biased than $t_{\text{ER1}}$.

The mean square errors of $t_{\text{ER1}}, t_{\text{ER2}}, t_{\text{ER3}}$ and $t_{\text{ER4}}$ to $O\left(\frac{1}{n}\right)$ are same. Thus for the purpose of comparison of efficiencies, the mean square error of the estimators are considered up to $O\left(\frac{1}{n^2}\right)$.

The comparisons of efficiencies of different estimators are made (a) under general conditions and (b) under bivariate symmetrical distribution.

1) $t_{\text{ER2}}$ is more efficient than $t_{\text{ER1}}$ if

Case (a) $C_{11} < \frac{1}{3} \left( \frac{5}{4} C_{20} + C_{02} \right)$

(4.2)
i.e. \( \rho < \frac{1}{12Z}(5Z^2 + 4) \)  

Case (b) same as above condition

where, \( Z = \left( \frac{C_{20}}{C_{c2}} \right)^{\frac{1}{2}} \)

2) \( t_{ER3} \) is more efficient than \( t_{ER1} \) if

Case (a) \( C_{11} < \frac{1}{C_{c2}} \left( \frac{5}{4} C_{20} C_{02} - 2C_{03} - C_{12} \right) \)

Case (b) \( \rho < \frac{1}{4Z}(5Z^2 - 12) \)  

3) \( t_{ER4} \) is more efficient than \( t_{ER1} \) if

Case (a) \( C_{11} > \frac{1}{C_{c2}} \left( \frac{5}{4} C_{20} C_{02} - 2C_{03} - C_{12} \right) \)

Case (b) \( \rho > \frac{1}{4Z}(5Z^2 - 4) \)  

4) \( t_{ER3} \) is more efficient than \( t_{ER2} \) if

Case (a) \( C_{11} > \frac{1}{2C_{c2}} (4C_{02} - 2C_{03} + C_{12}) \)

Case (b) \( \rho > \frac{2}{Z} \)  

5) \( t_{ER4} \) is more efficient than \( t_{ER2} \) if

Case (a) \( C_{11} > \frac{1}{4C_{c2}} \left( \frac{5}{4} C_{20} C_{02} + 2C_{03} - C_{12} \right) \)

Case (b) \( \rho > \frac{5}{8}Z \)  

6) \( t_{ER4} \) is more efficient than \( t_{ER3} \) if

Case (a) \( C_{11} > \frac{1}{2C_{c2}} \left( \frac{5}{4} C_{20} C_{02} - 4C_{03} + 4C_{12} \right) \)

Case (b) \( \rho > \frac{1}{4Z}(5Z^2 - 8) \)

V. EMPIRICAL STUDY

To study the efficiency of different estimators we have considered eight natural populations from different textbooks. The comparison is based on exact mean square errors. We have drawn all possible \( \binom{N}{4} \) samples of size four without replacement from given populations and the exact mean square errors are calculated. Table 1 gives the descriptions of population with Correlation Coefficient \( \rho \) and the Coefficient of Variation \( C_k \) and \( C_y \). Table 2 gives the exact MSE of different estimators i.e. mean per unit estimator \( t_0(= \bar{y}) \), \( t_{ER1}, t_{ER2}, t_{ER3} \) and \( t_{ER4} \).

**Table 1: Description of Population**

<table>
<thead>
<tr>
<th>Population No.</th>
<th>Description</th>
<th>N</th>
<th>Y</th>
<th>X</th>
<th>( \rho )</th>
<th>( C_k )</th>
<th>( C_y )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Cochran (1977) p.325</td>
<td>10</td>
<td>Persons</td>
<td>Rooms</td>
<td>0.651</td>
<td>0.135</td>
<td>0.153</td>
</tr>
<tr>
<td>2</td>
<td>Cochran (1977) p.34</td>
<td>17</td>
<td>Food Cost</td>
<td>Family Size</td>
<td>0.466</td>
<td>0.393</td>
<td>0.319</td>
</tr>
<tr>
<td>3</td>
<td>Drapper &amp; Smith (1966) p.352</td>
<td>25</td>
<td>Response Vector</td>
<td>Operating days per month</td>
<td>0.536</td>
<td>0.149</td>
<td>0.173</td>
</tr>
<tr>
<td>4</td>
<td>Drapper &amp; Smith (1966) p.352</td>
<td>25</td>
<td>Response Vector</td>
<td>Average wind velocity</td>
<td>0.474</td>
<td>0.276</td>
<td>0.173</td>
</tr>
<tr>
<td>5</td>
<td>Drapper &amp; Smith</td>
<td>25</td>
<td></td>
<td>Pounds of</td>
<td>0.305</td>
<td>0.181</td>
<td>0.173</td>
</tr>
</tbody>
</table>
TABLE 2: MSE OF DIFFERENT ESTIMATORS

<table>
<thead>
<tr>
<th>Population No.</th>
<th>$t_0 = \bar{y}$</th>
<th>$t_{ER1}$</th>
<th>$t_{ER2}$</th>
<th>$t_{ER3}$</th>
<th>$t_{ER4}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>35.953</td>
<td>22.325</td>
<td>22.170</td>
<td>22.169</td>
<td>22.461</td>
</tr>
<tr>
<td>2</td>
<td>13.844</td>
<td>10.914</td>
<td>10.500</td>
<td>10.465</td>
<td>11.625</td>
</tr>
<tr>
<td>3</td>
<td>0.558</td>
<td>0.402</td>
<td>0.396</td>
<td>0.388</td>
<td>0.417</td>
</tr>
<tr>
<td>4</td>
<td>0.558</td>
<td>0.474</td>
<td>0.468</td>
<td>0.465</td>
<td>0.486</td>
</tr>
<tr>
<td>5</td>
<td>0.558</td>
<td>0.531</td>
<td>0.525</td>
<td>0.517</td>
<td>0.594</td>
</tr>
<tr>
<td>6</td>
<td>0.337</td>
<td>0.1819</td>
<td>0.1817</td>
<td>0.1803</td>
<td>0.1837</td>
</tr>
<tr>
<td>7</td>
<td>11.228</td>
<td>3.916</td>
<td>3.860</td>
<td>4.084</td>
<td>3.800</td>
</tr>
<tr>
<td>8</td>
<td>511.017</td>
<td>254.651</td>
<td>163.896</td>
<td>154.842</td>
<td>481.965</td>
</tr>
</tbody>
</table>

VI. CONCLUSION

A. For populations 1, 2, 3, 4, 6, 7 and 8, the estimators, $t_{ER1}$, $t_{ER2}$, $t_{ER3}$ and $t_{ER4}$ are more efficient than the mean per unit estimator $t_0 = \bar{y}$.

B. For populations 1, 2, 3, 4, 5, 6 and 8, the estimator $t_{ER3}$ is most efficient.

C. For population 7, the estimator $t_{ER4}$ is most efficient.

As the estimator $t_{ER3}$ perform better than other estimators in most of populations considered here, so it may be used as an alternative estimator of $t_{ER1}$.

REFERENCES