Some Exponential Product Type Estimators Using Auxiliary Attributes.

Bulu Mahanty¹* and Gopabandu Mishra ²

¹* P.G. Department of Statistics, Utkal University, Bhubaneswar-751025, Odisha, India. ² P.G. Department of Statistics, Utkal University, Bhubaneswar, -751004, Odisha, India.

ABSTRACT

In this paper some improved exponential product type estimators of finite population mean have been suggested in presence of auxiliary attributes. For construction of estimators we have used a priori/a posteriori knowledge of coefficient variation and auxiliary attributes. The efficiencies of these estimators are compared with the exponential estimator using auxiliary attribute suggested by Singh et.al.¹ and among themselves with regard to biases and mean square errors both theoretically and numerically.

KEYWORDS: Auxiliary attributes, Population proportion, Simple random sampling, Exponential product type estimators, Bias, Mean square error, Efficiency.

*Corresponding author

Mr. Bulu Mahanty

P.G. Department of Community Medicine,

Hi-Tech Medical College,

Bhubaneswar-751025, Odisha, India.

Email. Id: bulu.mahanty@gmail.com, Mob No- +91-7978459396
1. INTRODUCTION

In theory of sampling judicious use of auxiliary information to develop efficient estimators is a long practice. Cochran developed a ratio estimator, when the study variable \( y \) and auxiliary variable \( x \) are positively correlated. However when “\( y \)” and “\( x \)” are negatively correlated the ratio estimator does not perform better. Robson and Murthy have suggested a product estimator to estimate population mean \( \bar{Y} \) which perform better than mean per unit estimator \( \bar{y} \), when \( y \) and \( x \) are highly negatively correlated.

Singh et.al. have suggested some exponential ratio type and product type estimators for positive correlation and negative correlation exit between study variable “\( y \)” and “\( a \)” (auxiliary attribute).

In this paper when \textit{a priori or a posteriori} information on population coefficient of variation and auxiliary attributes are available we suggest some improved exponential product type estimators to estimate finite population mean \( \bar{Y} \).

Let there be a finite population \( U \) consisting of \( N \) unit \( U = (U_1, U_2, U_3...U_i...U_N) \). The \( i^{th} \) unit is indexed by a pair of real value \((y_i,a_i)\) where \( y_i \) is the study variable and \( a_i \) is the auxiliary attribute. It is assumed that \( y_i \) and \( a_i \) are negatively correlated and the correlation coefficient between them is denoted by \( \rho \).

2. PROPOSED ESTIMATORS

From the finite population \( U \), a sample of size “\( n \)” is selected using simple random sampling without replacement (SRSWOR). We denote the sample mean of study variable \( \bar{y} \) and sample proportion \( \left( \frac{a}{n} \right) = p \) respectively.

Searls suggested an estimator to estimate population mean \( \bar{Y} \) using known population coefficient variation of study variable i.e. \( C_y = \frac{S_y}{\bar{y}}, S_y^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \). The suggested estimator of Searls is given by

\[
\hat{Y}_s = \frac{\bar{y}}{1 + \theta_i C_y^2},
\]

Where, \( \theta_i = \frac{1}{n} - \frac{1}{N} \).

Following Bhal and Tuteja, Singh et.al. have proposed an exponential product type estimator of population mean using population proportion (i.e. in presence of auxiliary attributes), is given by

\[
t_{EPP} = \bar{y} \exp \left[ \frac{p - \bar{p}}{p + \bar{p}} \right]
\]
Where, $P$ and $p$ are the population proportion and sample proportion respectively.

Now we proposed an improved exponential product type estimator of population mean when we have a priori knowledge of coefficient of variation of study variable i.e. $C_y$ and presence of auxiliary attributes.

$$t_{EPP2} = \frac{\bar{y}}{1 + \theta_i C_y^2} \exp \left[ \frac{p - P}{p + P} \right]$$  \hspace{1cm} (2.3)

Further, if the a priori knowledge of $C_i$ is not known, we can still construct an improved estimator by considering the estimate of population coefficient of variation of study variable $y$ from the sample. The estimator is given by

$$t_{EPP3} = \frac{\bar{y}}{1 + \theta_i \hat{C}_y^2} \exp \left[ \frac{p - P}{p + P} \right]$$  \hspace{1cm} (2.4)

Where, $\hat{C}_y^2 = \frac{s_y^2}{\bar{y}^2}$ and $s_y^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$

Following Upadhyaya and Srivastava, $a^7$, and Upadhyaya and Srivastava, $b^8$ we suggested another estimator

$$t_{EPP4} = \frac{\bar{y}(1 + \theta_i \hat{C}_y^2)}{\bar{y}} \exp \left[ \frac{p - P}{p + P} \right]$$  \hspace{1cm} (2.5)

3. BIAS AND MSE OF DIFFERENT ESTIMATORS

Assuming the validity of Taylor’s series expansion of $t_{EPP1}$, $t_{EPP2}$, $t_{EPP3}$ and $t_{EPP4}$, considering the expected value to $O\left(\frac{1}{n}\right)$, the bias of the different estimators are given as.

**B (t_{EPP1})**

$$B(t_{EPP1}) = E(t_{EPP1}) - \bar{Y} = \theta_i \bar{Y} \left[ \frac{1}{2} C_{11} - \frac{1}{8} C_{20} \right]$$  \hspace{1cm} (3.1)

**B (t_{EPP2})**

$$B(t_{EPP2}) = E(t_{EPP2}) - \bar{Y} = \theta_i \bar{Y} \left[ \frac{1}{2} C_{11} - \frac{1}{8} C_{20} - C_{02} \right]$$  \hspace{1cm} (3.2)

**B (t_{EPP3})**

$$B(t_{EPP3}) = E(t_{EPP3}) - \bar{Y} = \theta_i \bar{Y} \left[ \frac{1}{2} C_{11} - \frac{1}{8} C_{20} - C_{02} \right]$$  \hspace{1cm} (3.3)

**B (t_{EPP4})**

$$B(t_{EPP4}) = E(t_{EPP4}) - \bar{Y} = \theta_i \bar{Y} \left[ \frac{1}{2} C_{11} - \frac{1}{8} C_{20} + C_{02} \right]$$  \hspace{1cm} (3.4)

Where, $C_{rs} = \frac{\mu_{rs}(p, y)}{p' \bar{y}^r}$

Where, $\mu_{rs}(p, y)$ in the $(r, s)^{th}$ bivariate moments of $p$ and $y$. 
The Mean Square Error (MSE) of different estimators to $O\left(\frac{1}{n^2}\right)$ are given as

$$
\text{MSE} \left( I_{EPP1} \right) = \mathbb{E} \left( I_{EPP1} - \bar{Y} \right)^2
$$

$$
= \bar{Y}^2 \left[ \theta_1 (C_{02} + \frac{1}{4} C_{20} + C_{11}) + \theta_2 (C_{12} + \frac{1}{4} C_{21} - \frac{1}{8} C_{30} - \frac{5}{8} C_{11} C_{20} + \frac{7}{64} C_{20}^2) \right] \quad (3.5)
$$

Where $\theta_2 = \left( \frac{1}{n^2} - \frac{1}{N^2} \right)$

$$
\text{MSE} \left( I_{EPP2} \right) = \mathbb{E} \left( I_{EPP2} - \bar{Y} \right)^2
$$

$$
= \bar{Y}^2 \left[ \theta_1 (C_{02} + \frac{1}{4} C_{20} + C_{11}) + \theta_2 (C_{12} + \frac{1}{4} C_{21} - \frac{1}{8} C_{30} \\
- \frac{5}{8} C_{11} C_{20} + \frac{7}{64} C_{20}^2 - 3 C_{11} C_{02} - \frac{1}{4} C_{20} C_{02} - C_{02}^2) \right] \quad (3.6)
$$

$$
\text{MSE} \left( I_{EPP3} \right) = \mathbb{E} \left( I_{EPP3} - \bar{Y} \right)^2
$$

$$
= \bar{Y}^2 \left[ \theta_1 (C_{02} + \frac{1}{4} C_{20} + C_{11}) + \theta_2 (C_{12} + \frac{1}{4} C_{21} - \frac{1}{8} C_{30} \\
- \frac{5}{8} C_{11} C_{20} + \frac{7}{64} C_{20}^2 - C_{11} C_{02} - \frac{1}{4} C_{20} C_{02} + C_{02}^2 - 2 C_{02} C_{03} - C_{02} C_{12}) \right] \quad (3.7)
$$

$$
\text{MSE} \left( I_{EPP4} \right) = \mathbb{E} \left( I_{EPP4} - \bar{Y} \right)^2
$$

$$
= \bar{Y}^2 \left[ \theta_1 (C_{02} + \frac{1}{4} C_{20} + C_{11}) + \theta_2 (C_{12} + \frac{1}{4} C_{21} - \frac{1}{8} C_{30} \\
- \frac{5}{8} C_{11} C_{20} + \frac{7}{64} C_{20}^2 + \frac{1}{4} C_{20} C_{02} + 3 C_{11} C_{02} + 3 C_{02}^2 + 2 C_{02} C_{03} + C_{02} C_{12}) \right] \quad (3.8)
$$

4. COMPARISON OF BIASES AND MEAN SQUARED ERRORS

When the sample is large enough the biases of the estimators $I_{EPP1}, I_{EPP2}, I_{EPP3}$ and $I_{EPP4}$ of $O\left(\frac{1}{n}\right)$ are negligible.

From the above equations (3.2) and (3.3) both the estimators to $O\left(\frac{1}{n}\right)$ the biases are same. i.e.

$$
B \left( I_{EPP2} \right) = B \left( I_{EPP3} \right)
$$

However, the estimators $I_{EPP2}, I_{EPP3}$ and $I_{EPP4}$ are more biased than $I_{EPP1}$.
The mean square errors $E_{PP1}$, $E_{PP2}$, $E_{PP3}$ and $E_{PP4}$ to $O\left(\frac{1}{n}\right)$ are same. Thus for the purpose of comparison of efficiencies, the MSE of the estimators are considered up to $O\left(\frac{1}{n^2}\right)$.

The comparison of efficiencies of different estimators are made under two cases.

Case I: Under general condition

Case II: Under the Bivariate Symmetrical Distribution.

I. $E_{PP2}$ is more efficient than $E_{PP1}$ if

Case I: $C_{11} > -\frac{1}{12} \left( C_{20} + 4 C_{02} \right)$  \hspace{1cm} (4.1)

Case II: $\rho < -\frac{1}{12W} (W^2 + 4)$  \hspace{1cm} (4.2)

Where, $W = \left( \frac{C_{02}}{C_{20}} \right)^{\frac{1}{2}}$

II. $E_{PP3}$ is more efficient than $E_{PP1}$ if

Case I: $C_{11} < \frac{1}{C_{02}} \left( -\frac{1}{4} C_{20} C_{02} - C_{02}^2 - 2 C_{03} - C_{12} \right)$  \hspace{1cm} (4.3)

Case II: $\rho > -\frac{1}{4W} (W^2 - 4)$  \hspace{1cm} (4.4)

III. $E_{PP4}$ is more efficient than $E_{PP1}$ if

Case I: $C_{11} < \frac{1}{3C_{02}} \left( \frac{1}{4} C_{20} C_{02} + 3 C_{02}^2 + 2 C_{02} C_{03} + C_{02} C_{12} \right)$  \hspace{1cm} (4.5)

Case II: $\rho < \frac{1}{12W} (W^2 + 4)$  \hspace{1cm} (4.6)

IV. $E_{PP3}$ is more efficient than $E_{PP2}$ if

Case I: $C_{11} < \frac{1}{2C_{02}} \left( 2 C_{02}^2 + 2 C_{03} + C_{12} \right)$  \hspace{1cm} (4.7)

Case II: $\rho < \frac{1}{W}$  \hspace{1cm} (4.8)

V. $E_{PP4}$ is more efficient than $E_{PP2}$ if

Case I: $C_{11} < \frac{1}{4C_{02}} \left( \frac{1}{2} C_{02} C_{20} + 2 C_{02}^2 - 4 C_{03} - 2 C_{12} \right)$  \hspace{1cm} (4.9)

Case II: $\rho < \frac{1}{8W} (W^2 + 16)$  \hspace{1cm} (4.10)
VI. \( t_{EPP4} \) is more efficient than \( t_{EPP3} \) if

\[
C_{11} < \frac{1}{6C_{02}} \left( \frac{1}{2} C_{30} C_{20} + 4 C_{22}^2 - 2 C_{03} - 2 C_{12} \right) \tag{4.11}
\]

Case II : \( \rho < \frac{1}{12W}(W^2 + 8) \tag{4.12} \)

5. NUMERICAL ILLUSTRATION

For comparison of biases and mean square errors, we consider four natural populations. One population is considered from Draper and Smith, one population considered from Swain, and two populations are consider from Daniel and Cross. Biases are calculated considering terms up to first order of approximation and mean square errors are calculated considering second order of approximations. For calculations we consider four datasets showing \( N, n, \bar{Y}, P \) and \( C_{rs}(p, y) \)

Where, \( C_{rs}(p, y) = \frac{\mu_{rs}(p, y)}{P' \bar{Y}'^r} \)

**Data Set-1**

The data for the empirical analysis are taken from Natural Population dataset was considered by Draper and Smith [1998, P.40], Appendix 1A, stream plant data.

\( Y = \)Response

\( P = \) Predictors

\[ \bar{Y} = 52.6, \quad P = 0.92, \quad N = 25, \quad n = 6, \quad \rho = -0.3416, \quad C_{20} = 0.0869, \quad C_{02} = 0.1034, \quad C_{11} = -0.0324, \]

\[ C_{30} = -0.0793, \quad C_{03} = -0.0036, \quad C_{12} = -0.0032, \quad C_{21} = 0.0295 \]

**Data Set-2**

The data for the empirical analysis are taken from Natural Population dataset considered by Swain A.K.P.C. [2003, P.274],

\( Y = \) No. of Milk Cows, 1956

\( P = \) No. of Milk Cows in Rainy Season

\[ \bar{Y} = 67.3684, \quad P = 0.4210, \quad N = 19, \quad n = 7, \quad \rho = -0.5899, \quad C_{20} = 1.375, \quad C_{02} = 1.2385, \quad C_{11} = -0.7699, \]

\[ C_{30} = 0.5156, \quad C_{03} = 2.2061, \quad C_{12} = -0.6252, \quad C_{21} = 0.2887 \]
Data Set-3

The data for the empirical analysis are taken from Natural Population dataset considered by Daniel and Cross [2015, P.758], Tab No. 14.3.2

\[ Y = \text{Time (Months)} \]

\[ P = \text{Vital Status (Censored and Dead)} \]

\[ \bar{Y} = 63.0256, P = 0.5641, \ N = 39, \ n = 14, \ \rho = -0.5362, C_{20} = 0.7727, \ C_{o2} = 1.2629, \ C_{11} = -0.5297, C_{30} = -0.1756, C_{o3} = 2.4308, C_{12} = -0.4395, C_{21} = 0.1204 \]

Data Set-4

The data for the empirical analysis are taken from Natural Population dataset considered by Daniel and Cross [2015, P.757], Tab No. 14.3.1

\[ Y = \text{Time (Months)} \]

\[ P = \text{Tumor Grade (Low Grade & High Grade)} \]

\[ \bar{Y} = 63.025, P = 0.3589, \ N = 39, \ n = 20, \ \rho = -0.4715, C_{20} = 1.7857, \ C_{o2} = 1.2629, C_{11} = -0.7087, \ C_{30} = 1.4030, C_{o3} = 2.4308, C_{12} = -0.5378, C_{21} = -0.5568 \]

<table>
<thead>
<tr>
<th>Data set No.</th>
<th>( t_{EPP1} )</th>
<th>( t_{EPP2} )</th>
<th>( t_{EPP3} )</th>
<th>( t_{EPP4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0312</td>
<td>0.1510</td>
<td>0.1517</td>
<td>0.0871</td>
</tr>
<tr>
<td>2</td>
<td>0.1844</td>
<td>0.5674</td>
<td>1.3146</td>
<td>0.1679</td>
</tr>
<tr>
<td>3</td>
<td>0.0804</td>
<td>0.3602</td>
<td>0.4631</td>
<td>0.1737</td>
</tr>
<tr>
<td>4</td>
<td>0.0898</td>
<td>0.2843</td>
<td>0.3190</td>
<td>0.0955</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Data set No.</th>
<th>( t_{y_0} = \bar{y} )</th>
<th>( t_{EPP1} )</th>
<th>( t_{EPP2} )</th>
<th>( t_{EPP3} )</th>
<th>( t_{EPP4} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>36.2490</td>
<td>33.2524</td>
<td>33.1240</td>
<td>32.8448</td>
<td>34.0557</td>
</tr>
<tr>
<td>2</td>
<td>507.1931</td>
<td>336.6183</td>
<td>369.9062</td>
<td>68.9115</td>
<td>608.7238</td>
</tr>
<tr>
<td>3</td>
<td>229.7066</td>
<td>167.9355</td>
<td>169.3361</td>
<td>102.4355</td>
<td>224.1438</td>
</tr>
<tr>
<td>4</td>
<td>122.2039</td>
<td>97.5004</td>
<td>98.7411</td>
<td>78.4485</td>
<td>121.3179</td>
</tr>
</tbody>
</table>

6. CONCLUSIONS

1. It is observed that the suggested estimators \( t_{EPP2}, t_{EPP3}, \) and \( t_{EPP4} \) are more biased than the estimator \( t_{EPP1} \). However the biases are negligible if sample size is large.

2. Comparing the biases of \( t_{EPP1}, t_{EPP2}, t_{EPP3} \) and \( t_{EPP4} \) we observed
B(EPP_t1) < B(EPP_t2) < B(EPP_t3) < B(EPP_t4) for all populations.

3. Considering the value of MSE of mean per unit estimator (y), EPP_t1, EPP_t2, EPP_t3 and EPP_t4 we observed that the MSE of EPP_t3 is most efficient for the all populations.

ACKNOWLEDGEMENT

We acknowledge Dr. P.R. dash HOD of P.G. department of statistics, Utkal University for giving some fruitful suggestion to improve the article.

REFERENCES