OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

IMPROVED EXPONENTIAL PRODUCT TYPE ESTIMATORS OF FINITE POPULATION MEAN

Archana Panigrahi[1], Gopabandhu Mishra[2]

Department of Statistics, Utkal University, Bhubaneswar-751007, Odisha, India
archanapanigrahi10@gmail.com[1] gmishrauu@gmail.com[2]

Abstract: **In this paper some improved exponential product type estimators have been suggested for estimating finite population mean using auxiliary information under simple random sampling without replacement when there exists a negative correlation between the study and auxiliary variable. The performance of the estimators are studied both theoretically and empirically with mean per unit estimator, product estimator suggested by Robson [2] and Murthy [3] and exponential product type estimator suggested by Bahl and Tuteja [4].**

Keywords *– Simple random sampling, Auxiliary information, Exponential product type estimators, Bias, Mean square error, Efficiency.*

I INTRODUCTION

In sampling theory literature advance information of an auxiliary variable 'x' correlated with the study variable 'y' is often exploited to arrive at certain improved estimators of the population mean. Considering positive correlation between the study variable 'y' and the auxiliary variable 'x', Cochran [1] suggested a ratio estimator which is more efficient than mean per unit estimator when population correlation coefficient is more than 0.5.However, when study variable and auxiliary variable is negatively correlated then the estimator suggested by Cochran is not hold good. To overcome this, when two variables are negatively correlated Robson [2] and Murthy [3] have proposed a product estimator to estimate the population mean. Further Bahl and Tuteja [4] suggested an exponential product type estimator which performs better than usual product estimator [2, 3] when there exists a negative low correlation between the study variable and the auxiliary variable. When *a priori* and *a posteriori* information on population coefficient of variation of study variable and population mean of auxiliary variable are available, the question arises as to how far these information could be exploited to construct improved estimators over the one suggested by Bahl and Tuteja [4]. In this paper following Searls [5], Srivastava [6] and Upadhyaya and Srivastava [7, 8], some modified exponential product type estimators of population mean under SRSWOR have been proposed.

II PROPOSED ESTIMATORS

Consider a finite population of N distinct and identifiable units labeled from 1 to N. For simplicity, we let i^{th} unit is represented by U_i . Thus, we denote the finite population as U = (U_1, U_2, U_3, ..., U_N). It is assumed that the study variable y is negatively correlated with the auxiliary variable x and the correlation coefficient between them is denoted by ρ. A sample of size 'n' (n<N) is selected from U with simple random sampling without replacement (SRSWOR). Denote the sample mean of study variable and auxiliary variable and respectively.

Searls (1964) proposed an estimator to estimate finite population mean using known population coefficient of variation, i.e. (where $C_y =$ and $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y}$, is given by

$$\hat{\bar{Y}}_s = \frac{\bar{y}}{1+\theta_1 C_y^2} \qquad (2.1)$$

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

An exponential product type estimator for estimating finite population mean $\bar{Y}$ proposed by Bahl and Tuteja [4], which performs better with respect to efficiency than the product estimator ($\hat{\bar{Y}}_p = \frac{\bar{y}\bar{x}}{\bar{X}}$) when there exist a low negative correlation between y and x, is given by

$$t_{EP_1} == \bar{y}\ \mathrm{Exp}\left[\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right] \quad (2.2)$$

Now we propose a modified exponential product type estimator when population coefficient of variation of y, i.e. C_y is known in advance

$$t_{EP_2} = \frac{\bar{y}}{1+\theta_1 C_y^2}\ \mathrm{Exp}\left[\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right] \quad (2.3)$$

Further in absence of known C_y^2, considering the estimated coefficient of variation i.e. $\hat{C}_y^2$ considered from sample data, we suggest an estimator for $\bar{Y}$

$$t_{EP_3} = \frac{\bar{y}}{1+\theta_1 \hat{C}_y^2}\ \mathrm{Exp}\left[\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right] \quad (2.4)$$

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

Following Upadhyaya and Srivastava [7, 8], we propose another estimator for

$$t_{EP_4} = \bar{y}(+\theta_1\ \hat{C}_y^2)\mathrm{Exp}\left[\frac{\bar{x}-\bar{X}}{\bar{x}+\bar{X}}\right] \quad (2.5)$$

III BIAS AND MSE OF DIFFERENT ESTIMATORS

Assuming the validity of Taylor's series expansion of the estimators t_E, t_E, t_E and t_E and considering the expected value to $O($, the bias of the different estimators are

$$B(t_{EP_1}) = E(t_{EP_1}) - \bar{Y}$$
$$= \theta_1\ \bar{Y}\left[\frac{1}{2}C_{11} - \frac{1}{8}C_{20}\right]$$
$$B(t_{EP_2}) = E(t_{EP_2}) -$$
$$= \theta_1\ \bar{Y}\left[\frac{1}{2}C_{11} - \frac{1}{8}C_{20} - C_{02}\right]$$
$$B(t_{EP_3}) = E(t_{EP_3}) - \bar{Y}$$
$$= \theta_1\ \bar{Y}\left[\frac{1}{2}C_{11} - \frac{1}{8}C_{20} - C_{02}\right]$$
$$B(t_{EP_4}) = E(t_{EP_4}) -$$
$$= \theta_1\ \bar{Y}\left[\frac{1}{2}C_{11} - \frac{1}{8}C_{20} + C_{02}\right]$$

where, $C_{rs} = \frac{K_{rs}(\cdot}{\bar{X}^r \bar{Y}^s}$, $K_{rs}(x,$ being the (r, s) cumulant of x and y.

The mean square error (MSE) of different estimators to $O($ are derived as

$$MSE\ (t_{EP_1}) = \bar{Y}^2[\theta_1\ (C_{02} + \frac{1}{4}C_{20} + C_{11})$$
$$+\left(\theta_2 - \frac{3\theta_1}{N}\right)\left(\frac{1}{4}C_{21} - \frac{1}{8}C_{30} + C_{12}\right)$$
$$+\ \theta_1^2\ (\frac{7}{64}C_{20}^2 - \frac{5}{8}C_{11}C_{20}\ )]$$

$$MSE\ (t_{EP_2}) = MSE\ (t_{EP_1})$$
$$-\ \theta_1^2\ \bar{Y}^2\left(C_{02}^2 + 3C_{11}C_{02} + \frac{1}{4}C_{02}C_{20}\right)$$
$$MSE\ (t_{EP_3}) = MSE\ (t_{EP_1})$$
$$-\ \theta_1^2\ \bar{Y}^2\left(C_{11}C_{02} + \frac{1}{4}C_{02}C_{20} + 2C_{03} + C_{12} - 3C_{02}^2\right)$$
$$MSE\ (t_{EP_4}) = MSE\ (t_{EP_1})$$
$$-\ \theta_1^2\bar{Y}^2\left(C_{02}^2 - \frac{1}{4}C_{02}C_{20} - C_{11}C_{02} - 2C_{03} - C_{12}\right)$$

IV COMPARISION OF BIASES AND MEAN SQUARE ERRORS

The biases of t_{EP1}, t_{EP2}, t_{EP3} and t_{EP4} are of order $O($ and hence, are negligible when sample size is large. The mean square errors of t_E, t_E, t_E and t_E to $O($ are same. Thus for the purpose of comparison, the mean square error of estimators are considered to $O($.

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

i. t_E is more efficient than t_E if
Case (a)
$$C_{11} > -\frac{1}{12}(C_{20} + 4C_{02})$$
$$\text{i.e. } \rho > -\frac{1}{12Z}(Z^2 + 4)$$
Case (b) same as above condition

where, $Z = \left(\frac{C_{20}}{C_{02}}\right.$

ii. t_E is more efficient than t_E if
Case (a) C_{11}
$$\frac{1}{C_{02}}\left(3C_{02}^2 - \frac{1}{4}C_{20}C_{02} - 2C_{03} - C_{12}\right)$$
Case (b)
$$\rho > \frac{1}{47}(12 - Z^2)$$

iii. t_E is more efficient than t_E if
Case (a) C_{11}
$$\frac{1}{C_{02}}\left(C_{02}^2 - \frac{1}{4}C_{20}C_{02} - 2C_{03} - C_{12}\right)$$
Case (b)
$$\rho < \frac{1}{47}(4 - Z^2)$$

iv. t_E is more efficient than t_E if
Case (a)
$$C_{11} < \frac{1}{2C_{02}}(C_{12} + 2C_{03} - 4C_{02}^2)$$
Case (b)
$$\rho < -\frac{2}{7}$$

v. t_E is more efficient than t_E if
Case (a) C_{11}

$\frac{1}{4C_{02}}\left(-\frac{1}{2}C_{20}C_{02} - 2C_{03} - C_{12}\right)$ (4.9)

Case (b)

$\rho < -\frac{Z}{8}$

vi. t_E is more efficient than t_E if

Case (a) C_{11} $\frac{1}{2C_{02}}\left(4C_{02}^2 - \frac{1}{2}C_{20}C_{02} - 4C_{03} - 2C_{12}\right)$ (4.11)

Case (b) $\rho < \frac{1}{27}(8 - Z^2)$

V EMPIRICAL STUDY

To study the efficiency of different estimators we consider six natural populations collected from different textbooks. The comparison is based on exact mean square errors. For this purpose all possible $({}^N C$ samples are drawn without replacement from the given populations and the exact mean square errors are calculated. Table 1 gives the descriptions of the populations, size of the population (N), Correlation Coefficient and the Coefficient of Variations of x and y and The Table 2 gives the exact MSE of different estimators i.e. mean per unit estimator ( ), the conventional product estimator ($\hat{\bar{Y}}_P = \frac{\bar{y}}{\bar{x}}$, t_E, t_E, t_E and t_E.)

Table 1 : Description of Population

Popl. No.	N	Y	X	Ref.
1	25	Avg.atm. temp. (in ^{o}F)	Number of startups	[9] p.352
2.	13	Amount of Tricalcium Aluminate	Amount of Dicalcium Silicate	[9] p.366
3	19	Percentage of migrants	Median income	[10] p.459
4	12	Contract constructio n	Transportation and public utilities	[10] p.481
5	12	Percentage of fruits worms	Hundreds of fruits on a tree	[11] p.252
6	10	Price (in EUR)	Presence of sale assistant (in hours)	[12] p.336

TABLE 2 : POPULATION PARAMETER

Popl. No.	ρ	C_x	C_y	Ref.
1	-0.236	0.201	0.345	[9] p.352
2.	-0.245	0.581	0.820	[9] p.366
3	-0.247	0.171	0.282	[10] p.459
4	-0.372	0.037	0.054	[10] p.481
5	-0.542	0.447	0.372	[11] p.252
6	-0.464	0.158	0.157	[12] p.336

Table 3: MSE of Estimators

Popl. No.	$t_0 = \bar{y}$	$\hat{\bar{Y}}_P = \frac{\bar{y}\bar{x}}{\bar{X}}$	t_{EP1}	t_{EP2}	t_{EP3}	t_{EP4}
1	65.413	70.203	61.494	58.534	66.112	57.913
2	5.997	7.099	5.711	4.581	4.496	6.560
3	4.463	4.739	4.179	4.057	4.105	4.361
4	3754.5	3474.2	3084.0	3081.0	3106.7	3062.8
5	21.680	25.165	15.447	14.759	15.773	15.628
6	36.815	104.437	29.064	28.849	28.862	29.368

VI CONCLUSION

a. For the above six populations, the estimators, t_{EP_1}, t_{EP_2}, t_{EP_3} and t_{EP_4} are more efficient than the usual product estimators.

b. For all six populations, $MSE(t_{EP_2}) < MSE(t_{EP_1})$.

c. For populations 3, 4, 5 and 6 the estimators $t_{EP_1}, t_{EP_2}, t_{EP_3}$ and t_{EP_4} are more efficient than the mean per unit estimator.

d. For populations 1 and 4, $MSE(t_{EP_4}) < MSE(t_{EP_2}) < MSE(t_{EP_1})$.

e. For populations 3 and 6, $MSE(t_{EP_2}) < MSE(t_{EP_3}) < MSE(t_{EP_1})$.

f. For populations 2, 3 and 6, $MSE(t_{EP_3}) < MSE(t_{EP_1})$.

g. For populations 1, 2 and 4, $MSE(t_{EP_4}) < MSE(t_{EP_1})$.

h. For populations 1 and 4, MSE of t_{EP_4} is the least.

i. For population 2, MSE of t_{EP_3} is the least.

j. For populations 3, 5 and 6, MSE of t_{EP_2} is the least.

REFERENCES

[1] W.G. Cochran, “The estimation of the yield of cereal experiments by sampling for the ratio of grain to total procedure”, *Jour. Agric. Sci.*, vol. 30, pp. 262-275, 1940.

[2] D. S. Robson, “Application of multivariate polykays to the theory of unbiased ratio-type estimation”. *Journal of the American Statistical Association,* vol. 52, pp. 511-522, 1957.

[3] M. N. Murthy, “Product method of estimation”, *Sankhya A*, vol. 26, pp. 69-74, 1964.

[4] S. Bahl and R. K. Tuteja., “Ratio and product type exponential estimator”, *Information and Optimization Sciences*, vol. 12, pp. 159-163, 1991.

[5] D.T. Searls, “The utilization of known Coefficient of Variation in estimation procedure,” *Journal of the American Statistical Association.*, vol. 59, pp. 1225-26, 1964.

[6] V. K. Srivastava, “On use of Coefficient of Variation in estimating mean,” *Journal of Indian Society of Agricultural Statistics,* vol. 26, pp. 33-36, 1974.

[7] L. N. Upadhyaya and S. R. Srivastava, “A note on use of Coefficient of Variation in estimating mean,” *Journal of Indian Society of Agricultural Statistics,* vol. 28, pp. 97-98, 1976.

[8] L. N. Upadhyaya and S. R. Srivastava, “An efficient estimator of mean when population variance is known,” *Journal of Indian Society of Agricultural Statistics,* vol. 28, pp. 9-10, 1976.

[9] N. R. Draper and H. Smith, *Applied Regression Analysis*, John Wiley & Sons, New York, US, 1966.

[10] F.E . Croxton, D.J. Cowden, and S.K. Klein, *Applied General Statistics*, 3rd Edition, Prentice-Hall of India Pvt. Ltd., New Delhi, India, 1988.

[11] J.F. Kenney, & E.S. Keeping, *Mathematics of Statistics-Part One*, 3rd Edition, D.Van Nostrand Company, New York, US, 1954.

[12] W.W. Hardle and Z. Hlavka, *Multivariate Statistics: Exercise and Solutions*, Springer Publications, New York, US, 2007.