ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-05

(Approved and Registered with Govt. of India)

On Some Modified Exponential Ratio Type Estimators of in Two-Phase Sampling

Archana Panigrahi[#1], Gopabandhu Mishra[#2]

[#]*Department of Statistics, Utkal University, Bhubaneswar-751004, Odisha, India*

[1]archanapanigrahi10@gmail.com

[2]gmishrauu@gmail.com

Abstract— **Some modified exponential ratio type estimators of finite population mean $\bar{Y}$ under simple random sampling without replacement have been proposed in two-phase sampling using known coefficient of variation and estimated coefficient of variation of study variable (y). The efficiencies of these estimators are compared with the conventional two-phase ratio estimator, the two-phase exponential ratio type estimator suggested by Singh and Vishwakarma [7] both theoretically and empirically.**

***Keywords* -** Two-phase sampling, Study variable, Coefficient of variation, Exponential ratio type estimators, Efficiency.

I. Introduction

The auxiliary information is generally used to improve the efficiency of the estimators. Most common estimators available in sampling literature are ratio estimator [1, 2], regression estimator [3] and product estimator [4, 5]. When an estimator of a parametric function is constructed with the knowledge of advance information on some related parameters, a question may arise as to what happen in case such information is lacking. In this situation, it sometimes helps to take a large preliminary sample, if the cost of such sampling relatively cheap, to get some good estimate of the concerned parameters. Next a sub-sample from the large preliminary sample is considered to observe the study variable 'y' as well as the auxiliary variable 'x'. This procedure of drawing sample in two phases is known as two-phase sampling or double sampling.

Bahl and Tuteja [6] developed a ratio type exponential estimator to estimate finite population mean. Singh and Vishwakarma [7] suggested a ratio type exponential estimator in two-phase sampling when the information of the population mean of auxiliary variable is lacking.

In this paper following Searls [8], Srivastava [9] and Upadhyaya and Srivastava [10, 11], we developed three ratio type exponential estimators in two-phase sampling. These estimators are compared theoretically and empirically with the mean per unit estimator ($\bar{y}$), conventional two-phase ratio estimator ($t_{TR} = \frac{\bar{y}}{\bar{x}}\bar{x}'$) and two-phase ratio type exponential estimator suggested by Singh and Vishwakarma [7].

Consider a finite population U={1,2,3,.....,N}.Let y and x be two real variable assuming the value of y_i and x_i on the i^{th} unit i={1,2,3,.....,N}.Now consider y be the of study variable and x be the auxiliary variable. Further we assume that y and x is positively correlated. Here we consider simple random sampling scheme without replacement (SRSWOR) to draw samples in both phases of two-phase sampling set up. The first phase sample $s'(\, s' \subset U)$ of fixed size n' is drawn to observe 'x' only. In the second phase sample 's' of fixed size 'n' is drawn to observe y and x for given $s'(n < n')$.

Let $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$, $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i$

Now the usual two-phase ratio estimator and two-phase ratio type exponential estimator suggested by Singh and Vishwakarma [7] are respectively given by

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-05

(Approved and Registered with Govt. of India)

$$t_{TR} = \frac{\bar{y}}{\bar{x}} \bar{x}' \tag{1.1}$$

and $$t_{TER1} = \bar{y} \,\mathrm{Exp}\left[\frac{\bar{x}'-\bar{x}}{\bar{x}'+\bar{x}}\right] \tag{1.2}$$

Now the mean square errors (MSEs) of t_{TR} and t_{TER1} to $O(\frac{1}{n})$ are given by

$$\mathrm{MSE}\,(t_{TR}) = \bar{Y}^2\left[\theta_1\,(C_{02} + C_{20} - 2C_{11}) + \theta_1'(2C_{11} - C_{20})\right] \tag{1.3}$$

$$\mathrm{MSE}\,(t_{TER1}) = \bar{Y}^2\left[\theta_1\left(C_{02} + \frac{1}{4}C_{20} - C_{11}\right) + \theta_1'\left(C_{11} - \frac{1}{4}C_{20}\right)\right] \tag{1.4}$$

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

Comparing the variance of mean per unit estimator ($\bar{y}$), the MSE of two-phase ratio estimator (t_{TR}) and MSE of two-phase ratio type exponential estimator (t_{TER1}), we find t_{TER1} performs better than the estimators ($\bar{y}$) and t_{TR} if

$$\frac{1}{4}\frac{C_x}{C_y} < \rho < \frac{3}{4}\frac{C_x}{C_y} \tag{1.5}$$

where, ρ is the population correlation coefficient

II. Proposed Estimators

In two-phase (or double) sampling scheme, we proposed following modified exponential ratio type estimator to estimate population mean $\bar{Y}$

$$t_{TER2} = \frac{\bar{y}}{1 + \theta_1 C_y^2}\,\mathrm{Exp}\left[\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right] \tag{2.1}$$

where, $C_y (= \frac{S_y}{\bar{Y}})$, population coefficient of variation of y and further we assume that it is known in advance.

$$t_{TER3} = \frac{\bar{y}}{1 + \theta_1 \hat{C}_y^2}\,\mathrm{Exp}\left[\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right] \tag{2.2}$$

where, $\hat{C}_y (= \frac{s_y}{\bar{y}})$, sample coefficient of variation of y.

$$t_{TER4} = \bar{y}\left(1 + \theta_1 \hat{C}_y^2\right)\mathrm{Exp}\left[\frac{\bar{x}' - \bar{x}}{\bar{x}' + \bar{x}}\right] \tag{2.3}$$

where, $S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$ and $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

III. Bias and MSE of Different Estimators

Assuming the validity of Taylor's series expansion of t_{TER1}, t_{TER2}, t_{TER3} and t_{TER4} and considering the expected value to $O(\frac{1}{n})$, the bias of the different estimators are

$$B(t_{TER1}) = E(t_{TER1}) - \bar{Y} = \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{3}{8}C_{20} - \frac{1}{2}C_{11}\right)\right] \tag{3.1}$$

$$B(t_{TER2}) = E(t_{TER2}) - \bar{Y} = \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{3}{8}C_{20} - \frac{1}{2}C_{11}\right) - \theta_1 C_{02}\right] \tag{3.2}$$

$$B(t_{TER3}) = E(t_{TER3}) - \bar{Y} = \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{3}{8}C_{20} - \frac{1}{2}C_{11}\right) - \theta_1 C_{02}\right] \tag{3.3}$$

$$B(t_{TER4}) = E(t_{TER4}) - \bar{Y} = \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{3}{8}C_{20} - \frac{1}{2}C_{11}\right) + \theta_1 C_{02}\right] \tag{3.4}$$

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

$\mu_{rs}(x, y)$ being the $(r, s)^{th}$ bivariate moment of x and y.

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

$$\mathrm{MSE}(t_{TER1}) = \bar{Y}^2[(\theta_1 - \theta_1')\{\left(\frac{1}{4}C_{20} - C_{11}\right) + \theta_1 C_{02}\} + \{\left(\theta_2 - \frac{3\theta_1}{N}\right) - \left(\theta_2' - \frac{3}{N}\theta_1'\right)\} (\frac{5}{4}C_{21} - \frac{3}{8}C_{30} - C_{12}) + (\theta_1^2 - \theta_1'^2)(C_{20}C_{02} + 2C_{11}^2 - \frac{31}{8}C_{11}C_{20} + \frac{79}{64}C_{20}^2)] \quad (3.5)$$

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

$$\mathrm{MSE}(t_{TER2}) = \mathrm{MSE}(t_{TER1}) - \bar{Y}^2[\theta_1^2 C_{02}^2 - \theta_1(\theta_1 - \theta_1')\left(3C_{11}C_{02} - \frac{5}{4}C_{02}C_{20}\right)] \quad (3.6)$$

$$\mathrm{MSE}(t_{TER3}) = \mathrm{MSE}(t_{TER1}) - \bar{Y}^2[\theta_1^2(2C_{03} - 3C_{02}^2) - \theta_1(\theta_1 - \theta_1')\left(C_{11}C_{02} - \frac{5}{4}C_{02}C_{20} + C_{12}\right)] \quad (3.7)$$

$$\mathrm{MSE}(t_{TER4}) = \mathrm{MSE}(t_{TER1}) - \bar{Y}^2[\theta_1^2(C_{02}^2 - 2C_{03}) - \theta_1(\theta_1 - \theta_1')\left(\frac{5}{4}C_{02}C_{20} - C_{11}C_{02} - C_{12}\right)] \quad (3.8)$$

IV. Comparison of Efficiency

From (3.5), (3.6), (3.7) and (3.8) it is seen that the mean square errors of t_{TER1}, t_{TER2}, t_{TER3} and t_{TER4} to $O(\frac{1}{n})$ are same. So the mean square error of the estimators are considered up to $O(\frac{1}{n^2})$ for comparison of efficiency.

i. The estimator t_{TER2} is more efficient than t_{TER1} if

$$C_{11} < \frac{1}{12(\theta_1 - \theta_1')}[4\theta_1 C_{02} + 5(\theta_1 - \theta_1')C_{20}] \quad (4.1)$$

Assuming symmetrical bivariate distribution of (x, y) the inequality (4.1) reduces to

$$\rho < \frac{1}{12(\theta_1 - \theta_1')Z}[5(\theta_1 - \theta_1')Z^2 + 4\theta_1] \quad (4.2)$$

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

ii. The estimator t_{TER3} is more efficient than t_{TER1} if

$$C_{11} < \frac{1}{4C_{02}(\theta_1 - \theta_1')}[4\theta_1(2C_{03} - 3C_{02}^2) - (\theta_1 - \theta_1')(4C_{12} - 5C_{20}C_{02})] \quad (4.3)$$

Assuming symmetrical bivariate distribution of (x, y) the inequality (4.3) reduces to

$$\rho < \frac{1}{4(\theta_1 - \theta_1')Z}[5(\theta_1 - \theta_1')Z^2 - 12\theta_1] \quad (4.4)$$

iii. The estimator t_{TER4} is more efficient than t_{TER1} if

$$C_{11} > \frac{1}{4(\theta_1 - \theta_1')C_{02}}[4\theta_1(2C_{03} - C_{02}^2) + (\theta_1 - \theta_1')(5C_{20}C_{02} - 4C_{12})] \quad (4.5)$$

Assuming symmetrical bivariate distribution of (x, y) the inequality (4.5) reduces to

$$\rho > \frac{1}{4(\theta_1 - \theta_1')Z}[5(\theta_1 - \theta_1')Z^2 - 4\theta_1] \quad (4.6)$$

iv. The estimator t_{TER3} is more efficient than t_{TER2} if

$$C_{11} > \frac{1}{2(\theta_1 - \theta_1')C_{02}}[\theta_1(4C_{02}^2 - 2C_{03}) + (\theta_1 - \theta_1')C_{12}] \quad (4.7)$$

Assuming symmetrical bivariate distribution of (x, y) the inequality (4.7) reduces to

$$\rho > \frac{2\theta_1}{(\theta_1 - \theta_1')Z} \quad (4.8)$$

v. The estimator t_{TER4} is more efficient than t_{TER2} if

$$C_{11} > \frac{1}{8(\theta_1 - \theta_1')C_{02}}[4\theta_1 C_{03} + (\theta_1 - \theta_1')(5C_{20}C_{02} - 2C_{12})] \quad (4.9)$$

Assuming symmetrical bivariate distribution of (x, y) the inequality (4.9) reduces to

$$\rho > \frac{5}{8}Z \quad (4.10)$$

vi. The estimator t_{TER4} is more efficient than t_{TER3} if

$$C_{11} > \frac{1}{4(\theta_1 - \theta_1')C_{02}}[2\theta_1(4C_{03} - 4C_{02}^2) + (\theta_1 - \theta_1')(5C_{20}C_{02} - 4C_{12})] \quad (4.11)$$

Assuming symmetrical bivariate distribution of (x, y) the inequality (4.11) reduces to

$$\rho > \frac{1}{4(\theta_1 - \theta_1')Z}[5(\theta_1 - \theta_1')Z^2 - 8\theta_1] \quad (4.12)$$

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. The exact mean square errors are calculated from given populations. Table 1 gives the descriptions of population with Correlation Coefficient ρ and the Coefficient of Variation C_x and C_y. Table 2 gives the exact MSE of different estimators i.e. mean per unit estimator $t_0(=\bar{y})$, $t_{TR} = \frac{\bar{y}}{\bar{x}}\bar{x}'$, t_{TER1}, t_{TER2}, t_{TER3} and t_{TER4}.

TABLE 1: DESCRIPTION OF POPULATIONS

Population No.	Description	N	n'	n	Y	X	ρ	C_x	C_y
1	Black (2009) p.541 [12]	12	8	4	Price of Gold (in dollars/ounce)	Price of Aluminum (in cents/pound)	0.4812	0.2145	0.2771
2	Brownlee (1965) p.463 [13]	22	13	7	Deaths from Heart Disease (in hundreds in log scale)	Percentage of Fat Calories	0.4462	0.2865	0.3403
3	Cochran (1999) p.325 [14]	10	7	4	Number of Persons	Number of Rooms	0.6515	0.1350	0.1527
4	Draper&Smith (1966) p.352 [15]	25	14	7	Pounds of Steam Used per Month	Pounds of Crude Gasoline Made	0.3055	0.1810	0.1730
5	Draper&Smith (1966) p.366 [15]	13	8	4	Heat Evolved (in cal/g)	Weight Percent of Tricalcium Silicate	0.8162	0.3231	0.1576
6	Härdle&Hlávka	10	7	4	Number of	Advertisement	0.8672	0.5191	0.1965

	(2007) p.336 [16]				Classic Blue Pullovers sold	Costs (in EUR)			
7	Rao (2000) p.173 [17]	15	9	5	Diastolic Pressure	Weight	0.6879	0.0879	0.0765
8	Stopher (2012) p.276 [18]	20	12	6	Number of Vehicles	Number of Persons	0.5548	0.5105	0.3554

TABLE 2: MSE OF DIFFERENT ESTIMATORS

Population No.	$t_0 = \bar{y}$	t_{TR}	t_{TER1}	t_{TER2}	t_{TER3}	t_{TER4}
1	1928.6443	1602.4760	1491.1310	1476.1181	1504.9323	1518.8229
2	36.3096	34.5266	30.1801	29.8678	30.7076	30.3993
3	35.7816	23.9416	23.5010	23.4539	23.5221	23.6828
4	0.2734	0.3477	0.2542	0.2532	0.2526	0.2574
5	39.1696	65.7078	20.2614	20.0938	20.2136	20.5965
6	172.8683	576.6292	106.9808	105.0976	107.8486	107.7362
7	6.2069	4.8008	4.01895	4.01676	4.01331	4.03303
8	0.0230	0.0327	0.0177	0.0173	0.0167	0.0192

VI. CONCLUSION

a. For all populations, the estimators, t_{TER1}, t_{TER2}, t_{TER3} and t_{TER4} are more efficient than the mean per unit estimator t_0 and t_{TR} .
b. For populations 1, 2, 3, 5 and 6, the estimator t_{TER2} is most efficient.
c. For population 4, 7 and 8 the estimator t_{TER3} is most efficient.

As the estimator t_{TER2} perform better than other estimators in most of populations considered here, so it may be used as an alternative estimator of t_{TER1}.

REFERENCES

[1] P.S. Laplace, "A philosophical essay on probabilities", 1820 (English translation, Dover, 1951)
[2] 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.
[3] D. J. Watson, "The estimation of leaf area in field crops." *Journal of the Agricultural Science*, vol. 27, pp. 474-483, 1937.
[4] 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.
[5] M. N. Murthy, "Product method of estimation", *Sankhya A*, vol. 26, pp. 69-74, 1964.
[6] S. Bahl and R. K. Tuteja., "Ratio and product type exponential estimator", *Information and Optimization Sciences*, vol. 12, pp. 159-163, 1991.
[7] H. P. Singh and G. K. Vishwakarma, "Modified Exponential Ratio and Product Estimators for Finite Population Mean in Double Sampling", *Austrailian Journal of Statistics*, vol. 36, pp. 217-225, 2007.
[8] 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.
[9] 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.
[10] 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.
[11] 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.
[12] K. Black, *Business Statistics for Contemporary Decision Making*, 6th Edition, New York, US: John Wiley & Sons, Inc., 2009
[13] K. A. Brownlee, *Statistical Theory and Methodology*, 2nd Edition, New York, US: John Wiley & Sons, Inc., 1965
[14] W. G. Cochran, *Sampling Techniques*, 3rd ed., New York, US: Wiley Eastern Limited, 1999.
[15] N. R. Draper and H. Smith, *Applied Regression Analysis*, New York, US: John Wiley & Sons, 1966.

ISSN (Online) : 2348 - 2001

International Refereed Journal of Reviews and Research

Volume 5 Issue 6 November - December 2017

International Manuscript ID : 23482001V5I6112017-05

(Approved and Registered with Govt. of India)

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

[17] P. S. R. S. Rao, *Sampling Methodologies with Applications*, Boca Raton, US: Chapman & Hall, 2000

[18] P. Stopher, *Collecting, Managing, and Assessing Data using Sample Surveys*, Cambridge, UK : Cambridge University Press , 2012