International Journal of Advanced Technology & Engineering Research (IJATER)
National Conference on Recent Trends in Science, Technology & Management (NCRTSTM-2018)

IMPROVED EXPONENTIAL PRODUCT TYPE ESTIMATORS OF FINITE POPULATION MEAN IN TWO-PHASE SAMPLING

Archana Panigrahi, Department of Statistics, Surajmal Saha College, Puri, Odisha, India;

Gopabandhu Mishra, Department of Statistics, Utkal University, Bhubaneswar, Odisha, India

Abstract

Some improved exponential product type estimators of finite population mean $\bar{\mathbf{Y}}$ 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 product estimator, the two-phase exponential product type estimator suggested by Singh and Vishwakarma [7] both theoretically and empirically.

Introduction

The use of auxiliary information at the estimation stage improves the efficiency of the estimators. In this context ratio estimator [1, 2], regression estimator [3] and product estimator [4, 5] are most commonly used in sampling literature. In the absence of the knowledge on the population mean of auxiliary variable we go for two-phase sampling or double sampling. In two-phase sampling a large preliminary sample is taken from the population to observe the auxiliary variable 'x' only to estimate $\bar{X}$ and $\bar{x}'$. Next a sub-sample from the large preliminary sample is considered to observe the study variable 'y' as well as the auxiliary variable 'x' to estimate $\bar{x}$ and $\bar{y}$.

Bahl and Tuteja [6] developed a exponential product type estimator to estimate finite population mean. Singh and Vishwakarma [7] suggested a exponential product type 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 exponential product type estimators in two-phase sampling. These estimators are compared theoretically and empirically with the mean per unit estimator ($\bar{y}$), conventional two-phase product estimator ($t_{TP} = \frac{\bar{y}}{\bar{x}'}\bar{x}$) and two-phase exponential product type 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 are negatively 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 exponential product estimator and two-phase exponential product type estimator suggested by Singh and Vishwakarma [7] are given by

$$t_{TP} = \frac{\bar{y}}{\bar{x}'}\bar{x} \quad (1.1)$$

$$t_{TEP1} = \bar{y}\exp\left(\frac{\bar{x}-\bar{x}'}{\bar{x}+\bar{x}'}\right) \quad (1.2)$$

Now the mean square errors (MSEs) of t_{TP} and t_{TEP1} to $O(\frac{1}{n})$ are given by

$$MSE(t_{TP}) = \bar{Y}^2(\theta_1 - \theta_1')(C_y^2 + C_x^2 + 2C_{yx}) + \theta_1'\bar{Y}^2C_y^2 \quad (1.3)$$

$$MSE\,(t_{TEP1}) = \bar{Y}^2[\theta_1\left(C_{02} + \frac{1}{4}C_{20} + C_{11}\right) - \theta_1'\left(C_{11} + \frac{1}{4}C_{20}\right)] \quad (1.4)$$

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 product estimator (t_{TP}) and MSE of two-phase exponential product type estimator (t_{TEP1}), we find t_{TEP1} performs better than the estimators ($\bar{y}$) and t_{TP} if

$$-\frac{3}{4}\frac{C_x}{C_y} < \rho < -\frac{1}{4}\frac{C_x}{C_y} \qquad (1.5)$$

where, ρ is the population correlation coefficient

Proposed Estimators

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

$$t_{TEP2} = \frac{\bar{y}}{1 + \theta_1 C_y^2} \exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) \qquad (2.1)$$

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

$$t_{TEP3} = \frac{\bar{y}}{1 + \theta_1 \hat{C}_y^2} \exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) \qquad (2.2)$$

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

$$t_{TEP4} = \bar{y}\left(1 + \theta_1 \hat{C}_y^2\right) \exp\left(\frac{\bar{x} - \bar{x}'}{\bar{x} + \bar{x}'}\right) \qquad (2.3)$$

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

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

$$B(t_{TEP1}) = E(t_{TEP1}) - \bar{Y}$$

$$= \bar{Y}(\theta_1 - \theta_1')\left(\frac{1}{2}C_{11} - \frac{1}{8}C_{20}\right) \qquad (3.1)$$

$$B(t_{TEP2}) = E(t_{TEP2}) - \bar{Y}$$

$$= \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{1}{2}C_{11} - \frac{1}{8}C_{20}\right) - \theta_1 C_{02}\right] \qquad (3.2)$$

$$B(t_{TEP3}) = E(t_{TEP3}) - \bar{Y}$$

$$= \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{1}{2}C_{11} - \frac{1}{8}C_{20}\right) - \theta_1 C_{02}\right] \qquad (3.3)$$

$$B(t_{TEP4}) = E(t_{TEP4}) - \bar{Y}$$

$$= \bar{Y}\left[(\theta_1 - \theta_1')\left(\frac{1}{2}C_{11} - \frac{1}{8}C_{20}\right) + \theta_1 C_{02}\right] \qquad (3.4)$$

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

$$\begin{aligned} MSE(t_{TEP1}) = \bar{Y}^2 [\{ (\theta_1 - \theta_1') \left(\frac{1}{4}C_{20} + C_{11}\right) \\ + \theta_1 C_{02} \} + \{ \left(\theta_2 - \frac{3\theta_1}{N}\right) - \left(\theta_2' - \frac{3}{N}\theta_1'\right) \} \\ \left(\frac{1}{4}C_{21} - \frac{1}{8}C_{30} + C_{12}\right) + \{ (\theta_1^2 - \theta_1'^2) \\ \left(\frac{7}{64}C_{20}^2 - \frac{5}{8}C_{11}C_{20}\right) \\ + \theta_1'^2\left(\frac{1}{4}C_{20}C_{02} + \frac{1}{2}C_{11}^2\right)\}] \qquad (3.5) \end{aligned}$$

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

$$\begin{aligned} MSE(t_{TEP2}) = MSE(t_{TEP1}) - \bar{Y}^2 [\theta_1(\theta_1 - \theta_1') \\ \left(3C_{11}C_{02} + \frac{1}{4}C_{02}C_{20}\right) + \theta_1^2 C_{02}^2] \qquad (3.6) \end{aligned}$$

$$\begin{aligned} MSE(t_{TEP3}) = MSE(t_{TEP1}) - \bar{Y}^2 [\theta_1(\theta_1 - \theta_1') \\ \left(C_{11}C_{02} + \frac{1}{4}C_{02}C_{20} + C_{12}\right) \\ + \theta_1^2 (2C_{03} - 3C_{02}^2)] \qquad (3.7) \end{aligned}$$

$$\begin{aligned} MSE(t_{TEP4}) = MSE(t_{TEP1}) - \bar{Y}^2 [\theta_1^2 (C_{02}^2 \\ - 2C_{03}) - \theta_1(\theta_1 - \theta_1') \\ \left(C_{11}C_{02} + \frac{1}{4}C_{02}C_{20} + C_{12}\right)] \qquad (3.8) \end{aligned}$$

Comparison of Biases and Mean Square Errors

The biases of t_{TEP1}, t_{TEP2}, t_{TEP3} and t_{TEP4} are of order $O(\frac{1}{n})$ and hence, are negligible when sample size is large. The mean square errors of t_{TEP1}, t_{TEP2}, t_{TEP3} and t_{TEP4} to $O(\frac{1}{n})$ are same. Thus for the purpose of comparison, the mean square error of estimators are considered to $O(\frac{1}{n^2})$.

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

i. t_{TEP2} is more efficient than t_{TEP1} if
Case (a) $C_{11} > -\frac{1}{12(\theta_1 - \theta_1')}[(\theta_1 - \theta_1') C_{20} + 4\theta_1 C_{02}] \qquad (4.1)$

$$\text{i.e. } \rho > -\frac{1}{12Z(\theta_1 - \theta_1')}[(\theta_1 - \theta_1')Z^2 + 4\theta_1] \qquad (4.2)$$

Case (b) same as above condition

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

ii. t_{TEP3} is more efficient than t_{TEP1} if

Case (a) $C_{11} > \frac{-1}{4C_{02}(\theta_1 - \theta_1')}[\theta_1(2C_{03} - 3C_{02}^2) + (\theta_1 - \theta_1')(C_{20}C_{02} + 4C_{12})] \qquad (4.3)$

Case (b) $\rho > \frac{-1}{4Z(\theta_1 - \theta_1')}[(\theta_1 - \theta_1')Z^2 - 3\theta_1) \qquad (4.4)$

iii. t_{TEP4} is more efficient than t_{TEP1} if

Case (a) $C_{11} < \frac{1}{4C_{02}(\theta_1 - \theta_1')}[4\theta_1(C_{02}^2 - 2C_{03}) - (\theta_1 - \theta_1')(C_{20}C_{02} + 4C_{12}] \qquad (4.5)$

Case (b) $\rho < \frac{-1}{4(\theta_1 - \theta_1')Z}[Z^2(\theta_1 - \theta_1') - 4\theta_1] (4.6)$

iv. t_{TEP3} is more efficient than t_{TEP2} if

Case (a) $C_{11} < \frac{1}{2C_{02}(\theta_1 - \theta_1')}[(\theta_1 - \theta_1') C_{12} - 2\theta_1(2C_{02}^2 - C_{03})] \qquad (4.7)$

Case (b) $\rho < -\frac{2\theta_1}{Z(\theta_1 - \theta_1')} \qquad (4.8)$

v. t_{TEP4} is more efficient than t_{TEP2} if

Case (a) $C_{11} < \frac{-1}{8C_{02}(\theta_1 - \theta_1')}[4\theta_1 C_{03} + (\theta_1 - \theta_1')(C_{20}C_{02} + 2C_{12})] \qquad (4.9)$

Case (b) $\rho < -\frac{Z}{8} \qquad (4.10)$

vi. t_{TEP4} is more efficient than t_{TEP3} if

Case (a) $C_{11} < \frac{-1}{4C_{02}(\theta_1 - \theta_1')}[4\theta_1(C_{03} - C_{02}^2) + (\theta_1 - \theta_1')(C_{20}C_{02} - 4C_{12})] \qquad (4.11)$

Case (b) $\rho < \frac{-1}{4(\theta_1 - \theta_1')}[(\theta_1 - \theta_1')Z^2 + 4\theta_1] \qquad (4.12)$

Empirical Study

To study the efficiency of different estimators we consider six natural populations collected from different textbooks. Table 1 gives the description of the populations and the size of the population (N). Table 2 gives Correlation Coefficient (ρ) and the Coefficient of Variations of x and y (C_x and C_y). Table 3 gives the choice of sample size. Table 4 gives the exact MSE of different estimators i.e. mean per unit estimator ($\bar{y}$), the conventional product estimator t_{TP}, t_{TEP1}, t_{TEP2}, t_{TEP3} and t_{TEP4}.

Table 1. Description of Populations

Popl. No.	N	X	Y	Ref.
1	25	Number of Startups	Average Atmospheric Temperature	[13] p.352
2.	12	Hundreds of Fruits on a Tree	Percentage of Fruits Wormy	[15] p.252
3	70	Loads of Garbage	Cost of Garbage Disposal	[16] p.112
4	24	Marks for Car Safety	Marks for Car Price	[14] p.335
5	25	Average Atmospheric Temperature	Amount of Steam Used per Month	[13] p.352
6	12	Year	Number of Farms (in millions)	[12] p.476

Table 2. Population Parameters

Popl. No.	ρ	C_x	C_y	Ref.
1	-0.24	0.197	0.328	[13] p.352
2.	-0.54	0.429	0.356	[15] p.252
3	-0.55	0.204	0.151	[16] p.112
4	-0.7	0.335	0.361	[14] p.335
5	-0.85	0.328	0.173	[13] p.352
6	-0.9	0.009	0.385	[12] p.476

Table 3. Choice of Sample Size

Popl. No.	N	n′	n	Ref.
1	25	14	7	[13] p.352
2.	12	8	4	[15] p.252
3	70	34	17	[16] p.112
4	24	14	7	[14] p.335
5	25	14	7	[13] p.352
6	12	8	4	[12] p.476

Table 4. MSE of Estimators

Popl. No.	$t_0 = \bar{y}$	t_{TP}	t_{TEP1}	t_{TEP2}	t_{TEP3}	t_{TEP4}
1	30.66	31.25	28.37	28.13	29.59	27.78
2	21.63	22.45	15.69	15.77	16.73	15.43
3	0.619	0.743	0.4901	0.4902	0.487	0.495
4	0.139	0.093	0.092	0.093	0.091	0.097
5	0.273	0.342	0.137	0.139	0.14	0.136
6	0.226	0.201	0.204	0.199	0.188	0.229

Conclusion

a. For all populations, the estimators, t_{TEP1}, t_{TEP2}, t_{TEP3} and t_{TEP4} are more efficient than the mean per unit estimator t_0 and t_{TP} .

b. For population 3, 4 and 6 the estimator t_{TEP3} is most efficient.

c. For populations 1, 2 and 5, the estimator t_{TEP4} is most efficient.

As the estimator t_{TEP3} and t_{TEP4} perform better than other estimators in most of populations considered here, so anyone of them may be used as an alternative estimator of t_{TEP1}.

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

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

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

[16] J. D. Jobson (1991) *Applied Multivariate Data Analysis Volume I: Regression and Experimental Design*, New York, USA : Springer Science+Business Media, LLC, , 1991