A New Product Type Estimator of Population Mean and Its Efficiency

K.B.Panda[#1], M. Samantaray[#2]

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

ABSTRACT:

In this paper, a new product type estimator for the population mean using population correlation coefficient between the study variable and auxiliary variable has been proposed. The suggested estimator, both in first and second phase sampling,fares better than its competitors when compared with respect to bias and mean square error up to first degree of approximation. Empirical investigations have been carried out to support the theoretical findings.

KEYWORDS*: Auxiliary variable, correlation coefficient, first and second phase sampling, product type estimator, mean square error.*

I. INTRODUCTION

The use of auxiliary information has been studied by various authors in various form to improve the efficiency of their suggested estimators. In the use of auxiliary variable, ratio, product and regression estimators are corner-stones in the estimation of population characteristics.

In order to improve the efficiency of the estimators, auxiliary information is used at both selection as well as estimation stage. While Cochran, (1940) used auxiliary information at estimation stage and proposed ratio estimator, Murthy(1964) envisaged product estimator and Searl(1964), Sisodia and Dwibedi(1981) utilised co-efficient of variation of auxiliary variable in their respective ratio and product method of estimation. Srivenkataraman(1980) first proposed dual to ratio estimator, Singh and Tailor(2005) and Tailor and Sharma(2009) worked on ratio cum product estimator. Deriving inspiration from the above works completed with the estimator due to Mallick and Tailor(2013), we have proposed a new product -cum-dual to product estimator of finite population mean.

Consider a finite population $U : U_1, U_2, \ldots, U_N$ of N units. Let (y_i, x_i), i = 1, 2...n denote the values of the units included in a sample of size n drawn by simple random sampling without replacement (SRSWOR). In order to have a survey estimate of the population mean $\overline{Y}$ of the study variable Y, assuming the knowledge of the population mean $\overline{X}$ of the auxiliary variable X, Murthy(1964) proposed the classical product estimator defined as

$$\overline{y}_p = \overline{y}\frac{\overline{x}}{\overline{X}}$$

(1.1)

where, $\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ are unbiased estimator of population mean $\overline{Y} = \frac{1}{N}\sum_{i=1}^{n} Y_i$ and

$\overline{X} = \frac{1}{N}\sum_{i=1}^{n} X_i$ respectively.

Then, the bias and MSE of $\overline{y}_p$ up to first degree of approximation are obtained as

$$Bias(\overline{y}_p) = \overline{Y}f_1\rho C_y C_x = \overline{Y}f_1 C_{yx}$$

(1.2)

$$MSE(\overline{y}_p) = \overline{Y}^2 f_1(C_y^2 + C_x^2 + 2C_{yx})$$

(1.3)

Utilizing information on correlation co-efficient ρ between the study variable Y and auxiliary variable X Singh and Tailor(2003) the suggested ratio type estimator

$$\bar{y}_r^* = \bar{y}\frac{(\bar{X}+\rho)}{(\bar{x}+\rho)}$$

(1.4)

and in double sampling, the corresponding ratio estimator is expressed as

$$\bar{y}'^*_r = \bar{y}\frac{(\bar{x}'+\rho)}{(\bar{x}+\rho)}$$

(1.5)

where, $f_1 = \dfrac{1}{n} - \dfrac{1}{N}$,

$$\sigma_y = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^2}\ ,$$

$$\sigma_X = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})^2}\ ,$$

$$\sigma_{XY} = \frac{1}{N-1}\sum_{i=1}^{N}(X_i-\bar{X})(Y_i-\bar{Y})\ \ ,$$

$\rho = \dfrac{\sigma_{XY}}{\sigma_X\sigma_Y} = $ population correlation coefficient,

$C_{yx} = \rho C_Y C_X$,

$C_X^2 = \dfrac{\sigma_X^2}{\bar{X}^2}$ and $C_Y^2 = \dfrac{\sigma_Y^2}{\bar{Y}^2}$ are co-efficient of variation of X and Y respectively.

II. THE SUGGESTED ESTIMATOR

We propose a new product type estimator as

$$\bar{y}_p^* = \bar{y}\frac{(\bar{x}+\rho)}{(\bar{X}+\rho)}$$

(2.1)

To obtain the bias and MSE of the suggested estimators y_p^* and $y_p^{*'}$, we let

$$\bar{y} = \bar{Y}(1+e_0),\ \ \bar{x}=\bar{X}(1+e_1),\ \ \bar{x}'=\bar{X}(1+e_2)\ ,$$

$$\bar{x}+\rho = \bar{X}(1+\theta e_1)\ ,\ \bar{x}'+\rho=\bar{X}(1+\theta e_2)\ \text{ where }\ \theta = \frac{\bar{X}}{\bar{X}+\rho}$$

Now, $E(e_0)=E(e_1)=E(e_2)=0$,

$E(e_0^2)=f_1 C_Y^2,\ E(e_1^2)=f_1 C_X^2,\ E(e_2^2)=f_2 C_X^2$,

$E(e_0 e_1)=f_1 C_{YX},\ E(e_0 e_2)=f_2 C_{YX},\ E(e_1 e_2)=f_2 C_X^2$,

Substituting the above values in equation (2.1) we have (up to first degree of approximation)

$$Bias(\bar{y}_p^*) = E(\bar{y}_p^* - \bar{Y}) = \theta.\bar{Y}.f_1.\rho.C_Y.C_X = \theta.\bar{Y}.f_1.C_{YX}$$

(2.2)

$$MSE(\bar{y}_p^*) = E(\bar{y}_p^* - \bar{Y})^2 = \bar{Y}.f_1[C_Y^2 + \theta C_X^2(\theta+2\rho.C_Y/C_X) = \bar{Y}.f_1[C_Y^2 + \theta C_X^2(\theta+2k)]$$

(2.3)

III. EFFICIENCY COMPARISON:

Using simple random sampling without replacement (SRSWOR) we know that sample mean is unbiased estimator of population mean, having

$$V(\bar{y}) = f_1\bar{Y}^2 C_Y^2$$

(3.1)

Comparing equations (2.4) and (3.1), we have find that suggested estimator $\bar{y}_p^*$ would be more efficient than that of sample mean estimator $\bar{y}$, if

$$\left.\begin{array}{l}\text{i.e. } MSE(\bar{y}_p^*) - V(\bar{y}) = \theta C_X^2(\theta + 2k) < 0 \\ \text{provided, } \theta > 0 \, and \, (\theta + 2k) < 0 \Rightarrow 0 < \theta < (-2k) \\ \text{or,} \qquad \theta < 0 \, and \, (\theta + 2k) > 0 \Rightarrow -2k < \theta < 0 \end{array}\right\} \tag{3.2}$$

Comparing equations (2.4) and (1.3) , we find that, $\bar{y}_p^*$ would be more efficient than that of sample mean estimator $\bar{y}_P$. if

$$\left.\begin{array}{l} MSE(\bar{y}_p^*) - MSE(\bar{y}_p) = (\theta - 1)C_X^2(\theta + 1 + 2k) < 0 \\ \text{Provided, } (\theta - 1) > 0 \, and \, (\theta + 1 + 2k) < 0 \Rightarrow 1 < \theta < (-2k - 1) \\ \text{or, } (\theta - 1) < 0 \, and \, (\theta + 1 + 2k) > 0 \Rightarrow -(2k + 1) < \theta < 1 \end{array}\right\} \tag{3.3}$$

Combining equations (3.2) and (3.3), we find that $\bar{y}_p^*$ is more efficient than $\bar{y}$ and $\bar{y}_p$,under the following conditions

$$\left.\begin{array}{l} \text{either, } 1 < \theta < (-2k - 1) \\ \text{or} \qquad -2k < \theta < 0. \end{array}\right\} \tag{3.4}$$

IV. PERFORMANCE OF THE PROPOSEDESTIMATOR IN TWO-PHASE SAMPLING

There exit cases when $\bar{X}$ is unknown. T get rid of such cases, two phase sampling or double sampling procedure come into play, wherein we replace $\bar{X}$ by $\bar{x}'$, the sample mean based on large preliminary sample of size n' drawn with SRSWOR from the population of size N, corresponding to ith auxiliary variable. Thus, the product estimator due to classical estimator and the proposed estimator can be expressed respectively, as

When population mean of auxiliary variable X is not known, then double sampling product estimator is defined as,

$$\bar{y}_p' = \bar{y}\frac{\bar{x}}{\bar{x}'} \tag{4.1}$$

$$\text{and} \qquad \bar{y}_p'^* = \bar{y}\frac{(\bar{x} + \rho)}{(\bar{x}' + \rho)} \tag{4.2}$$

where, $\bar{x}' = \dfrac{1}{n'}\sum_{i=1}^{n'} x_i$ is unbiased estimator of population mean $\bar{X}$ based on sample size n'.

Thus, the bias and MSE of $\bar{y}_p'$ up to first degree of approximation are obtained as,

$$Bias(\bar{y}_p') = \bar{Y}f_3\rho C_Y C_X = \bar{Y}f_3 C_{YX} \, , \tag{4.3}$$

$$MSE(\bar{y}_p') = \bar{Y}[f_1 C_Y^2 + C_X^2 f_3(1 + 2\rho \, C_y/C_X)] = \bar{Y}[f_1 C_Y^2 + C_X^2 f_3(1 + 2k)], $$

$$\text{where } f_2 = \frac{1}{n'} - \frac{1}{N}, \, f_3 = f_1 - f_2 = \frac{1}{n} - \frac{1}{n'} \, k = \rho \, C_Y/C_X \, , \tag{4.4}$$

The bias and MSE of $\bar{y}_p^{*'}$ up to first degree of approximation are defined as,

$$Bias(\bar{y}_p'^*) = E(\bar{y}_p'^* - \bar{Y}) = \theta.\bar{Y}.f_3.\rho.C_Y.C_X = \theta.\bar{Y}.f_3.C_{YX} \tag{4.5}$$

$$MSE(\bar{y}_p'^*) = E(\bar{y}_p'^* - \bar{Y})^2 = \bar{Y}.[f_1 C_Y^2 + \theta C_X^2 f_3(\theta + 2\rho.C_Y/C_X) = \bar{Y}[f_1 C_Y^2 + \theta C_X^2 f_3(\theta + 2k)]$$

$$(4.6)$$

It can be easily be seen that, in two phase sampling , the performance of the propose estimator as measured in terms of bias and mean square error is better than its competing estimator.

V. EFFICIENCY COMPARISON IN DOUBLE SAMPLING

With a view to comparing the efficiency of the proposal estimator, we proceed as follows:

From equations (4.6)with (3.1) , we have $\bar{y}_p'^*$ would be more efficient than that of sample mean estimator $\bar{y}$. if

$$MSE(\bar{y}_p'^*) - V(\bar{y}) = \theta C_X^2 f_3(\theta + 2k) < 0$$
provided, $\theta > 0 \; and \; (\theta + 2k) < 0 \Rightarrow 0 < \theta < (-2k)$
or, $\theta < 0 \; and \; (\theta + 2k) > 0 \Rightarrow -2k < \theta < 0$

$$(5.1)$$

Again, comparing equations (4.6)and (4.4) , $\bar{y}_p'^*$ would be more efficient than that of sample mean estimator $\bar{y}_P'$,

$$MSE(\bar{y}_p'^*) - MSE(\bar{y}_p') = (\theta + 1)(\theta - 1 - 2k) < 0$$
provided, $(\theta + 1) > 0 \; and \; (\theta - 1 - 2k) < 0 \Rightarrow -1 < \theta < (2k + 1)$
or, $(\theta + 1) < 0 \; and \; (\theta - 1 - 2k) > 0 \Rightarrow (2k + 1) < \theta < -1$

$$(5.2)$$

Combining equations (5.1) and (5.2), we found that $\bar{y}_p'^*$ is more efficient than $\bar{y}$ and $\bar{y}_p'$,under the following conditions , if

provided, $0 < \theta < -2k$
or $(2k + 1) < \theta < -1.$

$$(5.3)$$

VI. EMPIRICAL STUDY

To check, the performance of suggested estimators $\bar{y}_p^*$ and $\bar{y}_p^{*'}$ over their competitors two natural population data set are being considered. Descriptions of the population are given below:

Population I:

we have considered a real population data taken from Adewaraet. al.(2012) wherein the variables of interest are as follows;

X: The number of rooms per block

Y: The number of person per block

$\bar{X}$ =75.6375, $\bar{Y}$ =7.6375, C_Y =0.2278, C_X =0.098, ρ =-0.6823, n= 3, n' =20

The bias and MSE of the above comparing estimators have been computed and presented in the following table:

Table I: Bias and MSE of the competing estimators

Sl. No.	Estimators	bias	Mean square error(MSE)
1	$\bar{y}$	0.0000	3.0268
2	$\bar{y}_p$	0.0349	1.8088
3	$\bar{y}_p^*$	0.03521	0.9204

Table II: Bias and MSE of the competing estimators in case of double sampling

Sl. No.	Estimators	bias	Mean square error(MSE)
1	$\bar{y}$	0.0000	3.0268
2	$\bar{y}'_p$	0.0329	1.3118
3	$\bar{y}'^*_p$	0.03322	0.5639

Population II:

We refer to Example-8.1(Highway data) given in weiberg (1980, p. 1979). the sample quantities given therein have been taken as the corresponding population quantities whch are as follows:

X: LEN = Length of the segment in miles

Y: RATE= 1973 accident rate per million vehicle miles.

$\bar{X}$ =49.23, $\bar{Y}$ =9.81, C_Y =0.7152, C_X =0.1499, ρ = -   0.00042, N=69, n= 3, n' =20

The bias and MSE of the above comparing estimators have been computed and presented in the following table:

Table III: Bias and MSE of the competing estimators

Sl. No.	Estimators	Bias	Mean square error(MSE)
1	$\bar{y}$	0.0000	132.774
2	$\bar{y}_p$	0.0001472	63.756
3	$\bar{y}_p^*$	0.0001480	46.241

Table IV: Bias and MSE of the competing estimators in case of double sampling

Sl. No.	Estimators	Bias	Mean square error(MSE)
1	$\bar{y}$	0.0000	132.774
2	$\bar{y}_p'$	0.000372	50.4055
3	$\bar{y}_p^{*'}$	0.00037	44.8678

The above tables clearly points to the fact that, the proposed estimators better than its competing estimators with respect to bias and MSE. The estimator is found to be unbiased.

VII.CONCLUSION:

A new product type estimator which is vindicated to be more efficient than its competitors both under single and double sampling has been proposed. Theoretical findings are numerically supported.

REFERENCES:

[1] Cochran W.G., 1977: Sampling Techniques.Third U.S. Edition.Wiley Eastern Limited. A.; 325

[2] Cochran W. G., Sampling Theory when sampling Units are of Unequal sizes.Journal American Statist.Association. 1942. 37; 191-212

[3] Das A. K., 1988: Contribution to the Theory of Sampling Strategies based on Auxiliary Information. Ph. D. Thesis, BCKY, West Bengal, India.

[4] Housila P. Singh, Ritesh Tailor and Tailor Rajesh, 2012: Estimation of Finite Population Mean In Two Phase Sampling with Known Coefficient of Variation of Auxiliary Character. Statistica. 72(1) ;1-22

[5] Kalidar C. and Cingi H., 2004: Ratio Estimators in Simple Random sampling. Applied Mathematics and Computation. 151; 893-902

[6] Malik K. A. and Tailor R., 2013: Ratio Type Estimator Of Population Mean in double sampling. International Jour. Of Advanced Math.and Statist. Volume 1.Issue 1; 34-39

[7] Murthy M. N., 1964: Product Method Of Estimation. Sankhya. A. 26; 294-307

[8] Searls D.T., 1964: The Utilization of Known Coefficient of Variation in the estimation procedure. Journal of American Statistical Association. 59; 1125-1126

[9] Singh H. P. and Tailor R. , 2005: Estimation of Population Mean Using Known Correlation Coefficient between Auxiliary Characters. Statistica. Anno LXV. 4; 407-418

[10] Sisodia B. V. S. and Dwibedi V. K., 1981: A Modified Ratio Estimator Using CCo-efficient of Variation Of Auxiliary variable. Journal.Ind. Soc. Agril.Statist. 33(2); 13-18

[11] Sodipo A. A. and Obisesan K. O., 2007: Estimation of The Population Mean Using Difference Cum Ratio Estimator with Full Response on the Auxiliary Character. Research Journal Of Applied Sciences. 2(6); 769-772

[12] Srivenkataramana T., 1980: A Dual Of Ratio Estimator in sample surveys. Biometrika.67(1). 199-204

[13] Tailor R. and Sharma B. K., 2009: A Modified Ratio-Cum- Product Estimator Of Finite Population Mean Using Known Co-efficient of Variation and Co-efficient Of Kurtosis. Statistics in Transition-New Series.10(1). 15-24