

Journal of Statistics, Computing and Interdisciplinary Research ISSN (P): 2707-7101

# Some Exponential Type Predictive Estimators of Finite Population Mean in Two-Phase Sampling

Asifa Kamal<sup>\*1</sup>, Nimra Amir<sup>1</sup>, and Huma Dastagir<sup>1</sup>

Department of Statistics, Lahore College Women University, Lahore, Pakistan Corresponding Author: asifa.kamal@lcwu.edu.pk

#### Abstract

This study is designed for predictive estimation of finite population mean in two phase sampling using two auxiliary variables. Two phase exponential ratio type estimator and exponential chain ratio type estimator are proposed under predictive approach suggested by Bahl & Tuteja (1991). The expressions of bias and mean square error of both suggested estimators have been carried out for theoretical comparison. The numerical study on real life data sets as well as simulation study has been conducted to examine the performance of suggested estimators. Finally, it is demonstrated that the suggested exponential ratio estimators are more efficient than competitive estimators in support of numerical and simulation study as well.

Key Words: Predictive, Auxiliary, Two phase Sampling

## Introduction

In survey sampling, sometimes some additional information is available on every unit in the population, this additional information is called an auxiliary information. In many situations, information on single auxiliary variable say X may be readily available on all the units of population which is useful for obtaining the reliable estimate of population parameter Y. Sometimes, such information is lacking as it is relatively cheap to take a large preliminary sample from the population, two phase sampling design is used then.

Consider a finite population  $U = U_1, U_2, ..., U_N$  of size N. In order to construct an efficient estimator of the population mean under two phase sampling, initially draw a first phase sample of size  $n_1(n_1 < N)$  by using simple random sampling denoted by  $s_1$ . Then second phase sample of size  $n_2(n_2 < n_1)$  is drawn from first phase sample which is denoted by  $s_2$ . Hence onward, the following notations will be used:

 $\overline{Y}$ ,  $\overline{X}$ ,  $\overline{Z}$ : Population means of the variable y, x and z respectively.

 $\bar{x}_1, \bar{z}_1$ : Sample means of respective variables based on sample size  $n_1$ .

 $\bar{x}_2$ ,  $\bar{y}_2$ : Sample mean of variable x and y respectively based on sample size $n_2$ .

Asifa et al. (2020)

By using auxiliary information, Chand (1975) and Sukhatme& Chand (1977) proposed a technique of chaining auxiliary variables with the study variable. Kiregyera (1980, 1984) also proposed some chain ratio and regression estimators based on two auxiliary variables. Further contribution are due to Isaki (1983), Singh &Kataria (1990), Prasad & Singh (1990, 1992), Ahmed et al. (2000), Singh & Singh (2001), Ahmed et al. (2003), Singh et al. (2006), Singh et al. (2009) and many others. Some popular upcoming ratio estimators are recorded based upon information of single and double auxiliary variable(s).

Singh & Choudhury (2012) propounded the following exponential chain ratio type estimator for estimating finite population mean of study variable Y using double sampling scheme.

$$\bar{Y}_{Re}^{dc} = \bar{y}_2 \exp\left(\frac{\bar{x}_1 \frac{Z}{\bar{z}_1} - \bar{x}_2}{\bar{x}_1 \frac{Z}{\bar{z}_1} + \bar{x}_2}\right) \tag{1}$$

Noor-ul-Amin (2012) proposed following improved form of exponential ratio estimator by using double sampling scheme.

$$t_{ND} = \bar{y}_2 \exp(\frac{\bar{x} - \bar{x}_1}{\bar{x} + \bar{x}_1} + \frac{\bar{z} - \bar{z}_2}{\bar{z} + \bar{z}_2}) \quad (2)$$

#### **Predictive Approach and Proposed Estimators**

Predictive approach is used to predict the values of single study variable for the unobserved units of the population by using observed units of the population. The concept of prediction theory was first given by Royal (1970). Royal (1970) suggested the use of model dependent optimal strategies and survey population as a sample from a super population. Basu (1971) gave the prediction criteria in the circus example and attempted to predict the mean of unobserved units of the population by using design based unbiased estimator. Supporting Basu's (1971) prediction approach, Srivastava (1983) used ratio and regression estimators in predictive approach for estimating population mean under single phase sampling. However, Sahoo&Sahoo (2001) used ratio-type estimators in the presence of two auxiliary variables under two-phase sampling for estimation of population mean.

By using two phase predictive approach under simple random sampling, the population mean can be written as:

$$\overline{Y} = \frac{1}{N} \sum_{i \in S} y_i + \frac{1}{N} \sum_{i \in r_2} y_i \frac{1}{N} \sum_{i \in r_1} y_i \quad (3)$$

The total population is divided into three mutually exclusive domains as s,  $r_2 = \bar{s} \cap s'$  and  $r_1 = U - s'$  having n, (n' - n) and (N-n') units respective. By writing  $(n' - n)\bar{Y}_2 = \sum_{i \in r_2} y_i$  and  $(N-n')\bar{Y}_1 = \sum_{i \in r_1} y_i$ , we have

$$\bar{Y} = f\bar{y}_2 + (f' - f)T_2 + (1 - f')T_1$$
(4)

where  $f = \frac{n}{N}$  and  $\frac{n'}{n}$ .

Using  $T_1$  and  $T_2$  as predictors to predict quantities  $\overline{Y}_1$  and  $\overline{Y}_2$  from sample data, the estimator of  $\overline{Y}$  under predictive approach is expressed as:

$$\overline{Y} = f\overline{y}_2 + (f' - f)T_2 + (1 - f')T_1$$
(5)

#### Some Exponential Type Predictive Estimators of Finite Population Mean in Two-Phase

#### Sampling

Using Noor-ul-Amin (2012) two phase improved form of exponential ratio estimator as  $T_1$  by gathering first phase sample information of x and z along the second phase sample information of x in predictive equation. Also taking Sisodia & Dwivedi (1981) ratio estimator as  $T_2$  in predictive equation, the proposed estimator  $\overline{y}_{1ND}$  is as follows:

 $\bar{y}_{1ND} = f\bar{y}_2 + (f'-f)\bar{y}_2\frac{\bar{x}_2 + C_x}{\bar{x}_2 + C_x} + (1-f')\bar{y}_2\exp\{\frac{\bar{x}_1 - \bar{x}_2}{\bar{x}_1 + \bar{x}_2} + \frac{\bar{z} - \bar{z}_1}{\bar{z} + \bar{z}_1}\} \quad (6)$  $(n'-n)\bar{X}_2 = \sum_{i \in r_2} x_i$ 

Where,

$$\bar{X}_2 = \frac{n'\bar{x}_1 - n\bar{x}_2}{(n'-n)} = \frac{f'\bar{x}_1 - f\bar{x}_2}{(f'-f)}$$

By using Singh & Choudhury (2012) exponential chain ratio type estimator as  $T_1$  and two phase ratio estimator as  $T_2$  in predictive equation, the proposed estimator  $\bar{y}_{2ND}$  is as follows:

$$\bar{y}_{2ND} = f\bar{y}_2 + (f'-f)\bar{y}_2\frac{\bar{x}_2}{\bar{x}_2} + (1-f')\bar{y}_2\exp\left(\frac{\bar{x}_1\frac{\bar{z}}{\bar{z}_1} - \bar{x}_2}{\bar{x}_1\frac{\bar{z}}{\bar{z}_1} + \bar{x}_2}\right)$$
(6)

Substituting the expanded expression of population mean of auxiliary variable X with domain  $r_2$  as  $\overline{X}_2$ , the above equation becomes:

7

$$\bar{y}_{2ND} = \frac{n}{N} + \frac{(n'-n)}{N} \bar{y}_2 \frac{(n'\bar{x}_1 - n\bar{x}_2)}{(n'-n)\bar{x}_2} + (1 - \frac{n'}{N}) \bar{y}_2 \exp\left(\frac{\bar{x}_1 \frac{z}{\bar{z}_1} - \bar{x}_2}{\bar{x}_1 \frac{\bar{z}}{\bar{z}_1} + \bar{x}_2}\right)$$

Thus, the simplified form of proposed estimator is as:

$$\bar{y}_{2ND} = f'\bar{y}_2 \frac{\bar{x}_1}{\bar{x}_2} + (1 - f')\bar{y}_2 \exp\left(\frac{\bar{x}_1 \frac{\bar{z}}{\bar{z}_1} - \bar{x}_2}{\bar{x}_1 \frac{\bar{z}}{\bar{z}_1} + \bar{x}_2}\right)$$
(7)

#### **Bias and Mean Square Error of Proposed Estimator**

The bias and mean square errors of the proposed estimators under large sample assumptions are derived up to the first order of approximations and using the following transformations:

$$e_0 = (\bar{y}_2 - \bar{Y}), e_1 = (\bar{x}_2 - \bar{X}), e_2 = (\bar{z}_2 - \bar{Z}), e_1' = (\bar{x}_1 - \bar{X}), e_2' = (\bar{z}_1 - \bar{Z})$$

By excluding the details of calculations, the expressions of bias and mean square error of both proposed estimators are obtained as stated below.

$$Bias(\bar{y}_{1ND}) = \bar{Y}[\theta_1 C_x^2 f' k^2 + \frac{1}{8} C_x^2 \{(1 - f')(3\theta_1 - \theta_2) - 4\theta_3 C_{yx}(1 - f' + 2f' + 2f'k)\}]$$
(8)

$$Bias(\bar{y}_{2ND}) = \bar{Y}[\theta_1 C_x^2 f' - \frac{1}{2} \{ \theta_3 (1+f') C_{yx} C_x^2 + (1-f') \theta_2 C_{yz} C_z^2 \} + \frac{(1-f')}{8} \{ (3\theta_1 - \theta_2) C_x^2 + 3\theta_2 C_z^2 \} ](9)$$

$$MSE(\bar{y}_{1ND}) = \bar{Y}^{2} \left[ \theta_{1} C_{y}^{2} + \frac{1}{4} (1 - f' + 2f'k) \theta_{3} C_{x}^{2} \left\{ (1 - f' + 2f'k) - 4C_{yx} \right\} + \frac{1}{4} (1 - f') \\ \theta_{2} C_{z}^{2} (1 - f' - 4C_{yz}) \right]$$
(10)

$$MSE(\bar{y}_{2ND}) = \bar{Y}^{2}[\theta_{1}C_{y}^{2} + \frac{1}{4}(1+f')^{2}(\theta_{1}+\theta_{2})C_{x}^{2} - \theta_{3}(1+f')\rho_{yx}C_{y}C_{x} + \frac{\theta_{2}}{4}((1-f')^{2}C_{z}^{2}) - 4(1-f')\rho_{yz}C_{y}C_{z} - 2(1+f')^{2}C_{x}^{2}]$$
(11)

Where  $\theta_1 = \left(\frac{1}{n} - \frac{1}{N}\right), \theta_2 = \left(\frac{1}{n'} - \frac{1}{N}\right), \theta_3 = (\theta_1 - \theta_2), k = \frac{\bar{X}}{(\bar{X} + C_3)}, C_y, C_x, C_z$  are noted for the coefficient of variations and  $\rho_{yx}$ ,  $\rho_{yz}$  are noted for the correlation coefficients.

#### **EFFICIENCY COMPARISONS**

The proposed estimator  $\bar{y}_{1ND}$  is more efficient than Chand (1975) chain ratio estimator under these conditions.

$$4(1 - 2C_{yx}) < (1 - f' + 2f'k)(1 - f' + 2f'k - 4C_{yx}) and 4(1 - 2C_{yz}) < (1 - f')(1 - f' - 4C_{yz})$$
(12)

The proposed estimator  $\bar{y}_{1ND}$  is more efficient than Sahoo & Sahoo (2001) ratio estimator under these conditions.

$$[(1 - f' + 2f'k)^{2} + 4(2C_{yx} - 1)] < 4C_{yx}(1 - f' + 2f'k)$$
  
and  $-C_{z}^{2}[3(1 - f') - 4C_{yz}]$   
 $< C_{x}^{2}\left[\left(1 + \frac{2f'k}{(1 - f')}\right)\{(1 - f' + 2f'k) - 4C_{yx}\} - \frac{4(C_{yx} - 1)}{(1 - f')}\right](13)$ 

The proposed estimator  $\bar{y}_{2ND}$  is more efficient than Singh & Choudhry (2012) ratio estimator under these conditions.

$$C_z^2(f'-2+4C_{yz}) < C_x^2(f'+2-4C_{yx}) \text{ and } f' < (4C_{yx}-2)$$
 (14)

The proposed estimator  $\bar{y}_{2ND}$  is more efficient than Sahoo & Sahoo (2001) ratio estimator under these conditions.

$$C_z^2 \left( C_{yz} - \frac{3}{4} (1 - f') \right) + \frac{C_x^2}{4} \left( \frac{3 - (f' + 2f')}{(1 - f')} \right) < C_x^2 C_{yx} \text{ and } C_{yx} = \frac{C_x^2}{4} (3 - (f'^2 + 2f'))$$
(15)

#### **Empirical Study**

The real life data sets of Murthy (1977), Fisher (1936), Cochran (1977) and Srivastava (1989) used to examine the performance of proposed estimators over their competitors numerically. To see the relative performances of all estimators of the population mean, the percent relative efficiencies (PREs) of the estimators are computed with respect to the usual unbiased estimator by following formula Where  $*= \bar{y}_{12}$ ,  $\bar{y}_{R}^{dc}$ ,  $\bar{y}_{1ND}$ ,  $\bar{Y}_{2Re}^{dc}$ ,  $\bar{y}_{2ND}$  respectively.

Some Exponential Type Predictive Estimators of Finite Population Mean in Two-Phase

#### Sampling

$$PRE(*, \bar{y}) = \frac{var(\bar{y})}{MSE(*)} \times 100$$

Estimators	Pop 1	Pop 2	Pop 3	Pop 4	Pop 5
Usual Unbiased Estimator $\overline{Y}$	100	100	100	100	100
Sahoo&Sahoo (2001) $\bar{y}_{12}$	*	114.96	156.66	*	*
Chand (1975) $\bar{y}_R^{dc}$	244.35	111.438	130.320	*	*
Proposed Estimator $\bar{y}_{1ND}$	349.47	124.51	172.49	*	*
Singh & Choudhury $\bar{Y}^{dc}_{2Re}$	*	123.63	*	130.16	120.57
Proposed Estimator $\bar{y}_{2ND}$	349.55	124.43	*	131.45	122.61

## **Table 1: The PREs of different estimators**

\*indicates that data is not applicable for these estimators.

The results in above table show that  $\bar{y}_{1ND}$  is better than  $\bar{y}_{R}^{dc}$  for first three populations and  $\bar{y}_{2ND}$  is performing better than  $\bar{Y}_{2Re}^{dc}$  for population 2, 4 and 5. However both proposed estimators are more efficient than  $\bar{y}_{12}$  for 2 and 3 populations.

## **Simulation Study**

The simulation study has been conducted in this section where 5000 number of simulations are generated by using R-code. The multivariate normal population is considered with three variables. By considering the population of size 50 where first phase sample size 18 is selected from it and second phase sample size 8 is selected. The theoretical mean vector and co-variance matrices of the simulated multivariate population are defined. For simulation study data has been taken from Fisher (1936) Iris flowers population which follows multivariate normal distribution. The obtained simulated mean, bias and mean square error are as:

Estimators	Simulation No.	Simulation No.	Bias	MSE
$\bar{y}_{12}$	5000	1.954301	-0.03648206	0.007722487
$ar{y}_R^{dc}$	5000	2.08158	.008459	0.01286466
$\overline{\mathcal{Y}}_{1ND}$	5000	1.988833	-0.0019504	0.005340948
$\overline{Y}^{dc}_{2Re}$	5000	2.006141	0.01535801	0.005501707
$\overline{\mathcal{Y}}_{2ND}$	5000	1.987628	-0.003154992	0.005366724

## Conclusion

For two phase sampling, the better performance of proposed estimators  $\bar{y}_{1ND}$  and  $\bar{y}_{2ND}$  over usual unbiased estimator, Chand (1975), Sahoo&Sahoo (2001) and Singh &Choudury(2012) have been observed theoretically, empirically and with simulation study too. It is therefore found that these two exponential proposed estimators perform better than its competitive estimators in the presence of two auxiliary variables.

## References

- Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2), 179-188.
- Shukla, G. K. (1966). An alternative multivariate ratio estimate for finite population. *Calcutta Statistical Association Bulletin*, 15(2-3), 127-134.
- Royall, R. M. (1970). On finite population sampling theory under certain linear regression models. *Biometrika*, 57(2), 377-387.
- Basu, D. (1971). An Essay on the Logical Foundations of Survey Sampling, Part One: The University of New Mexico and Indian Statistical Institute. In A. Das Gupta(Ed), Selected Works of DebabrataBasu, 167-206. Springer New York.doi:10.1007/978-1-4419-5825-9\_24
- Chand, L. (1975). Some Ratio-type Estimators based on two or more Auxiliary Variables.Unpublished Ph.D. Dissertation. Iowa State University, USA.https://doi.org/10.31274/rtd-180813-485.
- Cochran, W. G. (1977). Sampling Techniques. John Wiley & Sons, New York.
- Murthy, M. N. (1977). Sampling: Theory and Methods, Statistical Pub. Society, Calcutta.
- Sukhatme, B. V., & Chand, L. (1977). Multivariate ratio-type estimators. *Proceedings of American Statistical Association, Social Statistics Section*, 927–931
- Kiregyera, B. (1980). Chain ratio-type estimators in finite population double sampling using two auxiliary variables. *Metrika*, 27(1), 217-223.
- Sisodia, B. V. S., &Dwivedi, V. K. (1981).Modified ratio estimator using coefficient of variation of auxiliary variable.*Journal-Indian Society of Agricultural Statistics*, 33, 13-18
- Srivastava, S. K. (1983). Predictive estimation of finite population mean using product estimator. *Metrika*, *30*(1), 93-99.
- Kiregyera, B. (1984). Regression-type estimators using two auxiliary variables and the model of double sampling from finite populations.*Metrika*, 31(1), 215-226.

#### Sampling

- Srivnstava, R. S., Srivastava, S. P., &Khare, B. B. (1989). Chain ratio type estimator for ratio of two population means using auxiliary characters. *Communications in Statistics-Theory* and Methods, 18(10), 3917-3926.
- Singh, S., &Kataria, P. (1990). An estimator of finite population variance *Journal of Industrial Society Agricultural Statistics*, 42(2), 186–188.
- Prasad, B., & Singh, H. P. (1990). Some improved ratio type estimators of finite population variance in sample surveys. *Communication in Statistics –Theory and Methods*, 19(3), 1127–1139.
- Bahl, S., &Tuteja, R. K. (1991). Ratio and product type exponential estimators. *Journal of information and optimization sciences*, 12(1), 159-164.
- Prasad, B., & Singh, H. P. (1992). Unbiased estimators of finite population variance using auxiliary information in sample surveys. *Communication in Statistics -Theory and Methods*, 21(5), 1367–1376.
- Singh, A. K., & Singh, H. P. (2001).Dual to chain ratio type estimator in double sampling using two auxiliary variables. *Journal of Ravishankar University B*, 14, 99-106.
- Ahmed, M. S., Walid, A. D. & Ahmed, A. O. H. (2003). Some estimators for finite population variance under two-phase sampling. *Statistics in Transition*, 6(1), 143-150.
- Singh, H.P., Singh, S., & Kim, J.M. (2006). General families of chain ratio type estimators of the population mean with known coefficient of variation of the second auxiliary variable in two phase sampling. *Journal of the Korean Statistical Society*, 35(4): 377-395.
- Singh, R., Chauhan, P., Sawan, N., & Smarandache, F. (2009). Improvement in Estimating the Population Mean Using Exponential Estimator in Simple Random Sampling. *International Journal of Statistics & Economics*, 3(A09), 13-18.
- Singh, B. K., &Choudhury, S. (2012). Exponential chain ratio and product type estimators for finite population mean under double sampling scheme. *Journal of Science Frontier Research in Mathematics and Design Sciences*, 12(6), 0975-5896.
- Noor-ul-Amin, M. and Hanif, M. (2012) Some exponential estimators in survey sampling. *Pak. J. Statist.* 28(3), 367-374.