Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/070212

An Improved Difference Type Estimator for Population Mean Under Two-Phase Sampling Design

Asra Nazir[*], *Rafia Jan and T. R. Jan*

Department Of Statistics, University Of Kashmir, India

Received: 18 Oct. 2018, Revised: 23 Mar. 2018, Accepted: 30 Mar. 2018.
Published online: 1 Jul. 2018.

Abstract: Two-phase sampling design offers a variety of possibilities for effective use of auxiliary information. A new class of regression-cum-ratio estimators has been proposed for two-phase sampling using information on two auxiliary variables. The Mean Square Error (MSE) of the proposed estimators has been obtained up to first order approximation. Efficiency comparison of the proposed estimators has been made with some traditional estimators. Numerical illustration has been carried out to examine the efficiency of the estimator.

Keywords: Auxiliary variable, Bias, Mean Square Error, Two phase sampling, Exponential chain-type estimator, Efficiency.

1 Introduction

In planning surveys, it is beneficial to take advantage from some auxiliary information either at stage of estimation or survey planning, in order to estimate a finite population, mean with higher degree of precision. For this purpose, many researchers suggested several ratios, product, and regression estimators by considering the relationship between the study and auxiliary variables, e.g. Hansen and Hurwitz (1943), Sukhatme (1962), Srivastava (1970), Chand (1975), Cochran (1977), Kiregyera (1980, 1984), Srivastava, Khare and Srivastava. (1990), Bahl and Tuteja (1991), Singh, Chauhan and Swan. (2006, 2007, 2011), Singh and Choudhury (2012), Khare, Srivastava and Kumar. (2013), proposed a generalized chain ratio in regression estimator for population mean using two auxiliary characters, Singh and Majhi (2014) using the information on two-auxiliary variables, three different exponential chain-type estimators of population mean of study variable have been proposed in two-phase (double) sampling and Khan, (2015, 2016) presents a ratio estimator for the estimation of finite population mean of the study variable under double sampling scheme when there is unusually low and unusually high values and analyzes their properties.

2. Symbols and Notations

Let us consider a finite population of size *N* of different units $U=\{U_1,U_2,....,U_N\}$. Let *y* and *x* be the study and the auxiliary variable with corresponding values y_i and x_i respectively for the i^{th} unit $i=\{1,2,3,...,N\}$ defined in a finite population *U* with

means, $\overline{Y}=\frac{1}{N}\sum_{i=1}^{N}y_i$ and $\overline{X}=\frac{1}{N}\sum_{i=1}^{N}x_i$ of the study as well as auxiliary variable respectively.

[*] Corresponding author e-mail: Asranazir44@gmail.com

Also let $S_y{}^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(y_i - \overline{Y})^2$ and $S_x{}^2 = \dfrac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{X})^2$ be the population variances of the study and the auxiliary variable respectively and let C_y and C_x be the coefficient of variation of the study as well as auxiliary variable respectively, and ρ_{yx} be the correlation coefficient between *x* and *y*. Let *y* and *x* be the study and the auxiliary variable in the sample with corresponding values y_i and x_i respectively for the i^{th} unit $i = \{1,2,3,...,N\}$ in the sample with unbiased means $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ and $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ respectively.

Also let $\hat{s}_y{}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(y_i - \overline{y})^2$ and $\hat{S}_x{}^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ be the corresponding sample variances of the study as well as auxiliary variable respectively. Let $S_{yx} = \dfrac{\sum_{i=1}^{N}(y_i - \overline{Y})(x_i - \overline{X})}{N-1}$, $S_{yz} = \dfrac{\sum_{i=1}^{N}(y_i - \overline{Y})(z_i - \overline{Z})}{N-1}$ and $S_{xz} = \dfrac{\sum_{i=1}^{N}(x_i - \overline{X})(z_i - \overline{Z})}{N-1}$ be the co variances between their respective subscripts respectively. Similarly $b_{yx} = \dfrac{\hat{S}_{xy}}{\hat{S}_x^2}$ is the corresponding sample regression coefficient of *y* on *x* based on a sample of size *n*.

Also $C_y = \dfrac{S_y}{\overline{Y}}$, $C_X = \dfrac{S_z}{\overline{X}}$ and $C_z = \dfrac{S_z}{\overline{Z}}$ are the coefficients of variations of the study and auxiliary variables respectively.

Also $\theta = \left(\dfrac{1}{n} - \dfrac{1}{N}\right)$, $\theta_1 = \left(\dfrac{1}{n'} - \dfrac{1}{N}\right)$ and $\theta_2 = \left(\dfrac{1}{n} - \dfrac{1}{n'}\right)$.

3. Some Existing Estimator

Let us consider a finite population *U={U₁,U₂,U₃ y ..U_N}* of size N units.To estimate the population mean $\overline{Y}$ of the variable of interest say y taking values y_i in the existence of two auxiliary variables say x and y taking values x_i and z_i for the i^{th} unit U_I.we assume that there is high correlation between y and x as compared to the correlation between y and z, $\left(i.e.\rho_{yx} > \rho_{yz} > 0\right)$. .When the population $\overline{X}$ of the auxiliary variable x is unknown ,but information on the other cheaply auxiliary variable say z closely related to x but compared to x remotely to y ,is available for all units in the population .In such situations we use a two –phase sampling .In two phase sampling scheme a large initial sample of size $n'(n' < N)$ is drawn from the population *U* by using (SRSWOR)scheme and measure x and z to estimate $\overline{X}$ .In the second phase ,we draw a sample (subsample)of size n from the first phase sample of size $n', i.e.(n < n')$ by using (SRSWOR) or directly from the population *U* and observed the study variable y .

The variance of the usual simple estimator $t_o = \overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$ up to first order of approximation is given by

$$V(t_o) = \theta S_y^2 \qquad (1)$$

The classical ratio and regression estimators in two –phase sampling and their mean square errors up to first order of approximation are, given by

$$t_1 = \frac{\overline{y}}{\overline{x}}\overline{x}' \qquad (2)$$

$$MSE(t_1) = \bar{Y}^2\left[\theta C_y^2 + \theta_2(C_x^2 - 2\rho_{yx}C_yC_x)\right] \tag{3}$$

$$t_2 = \bar{y} + b_{yx(n)}(\bar{x}' - \bar{x}) \tag{4}$$

$$MSE(t_2) = S_y^2\left[\theta(1-\rho^2{}_{yx}) + \theta_1\rho^2{}_{yx})\right] \tag{5}$$

Chand (1975) suggested the following chain ratio type estimator the suggested estimator is, given by

$$t_3 = \frac{\bar{y}}{\bar{x}}\frac{\bar{x}'}{\bar{z}'}\bar{Z} \tag{6}$$

The mean square error of the suggested estimator is,given as

$$MSE(t_3) = \bar{Y}^2\left[\theta C_y^2 + \theta_2(C_x^2 - 2\rho_{yx}C_yC_x) + \theta_1(C_z^2 - 2\rho_{yz}C_yC_z)\right] \tag{7}$$

Khare et al. (2013),proposed a generalized chain ratio in regression estimator for population mean, the recommended estimator is given by

$$t_4 = \bar{y} + b_{yx}\left\{\bar{x}'\left(\frac{\bar{Z}}{\bar{z}'}\right)^{\alpha} - \bar{x}\right\} \tag{8}$$

Where α is the unknown constant , and the minimum mean square error at the optimum value of

$\alpha = \dfrac{\rho_{yz}C_x}{\rho_{yz}C_x}$ is, given by

$$MSE(t_4) = \bar{Y}^2C_y^2\left[\theta + \theta_2(\rho^2{}_{yx}) - \theta_1\rho_{yz}{}^2)\right] \tag{9}$$

4 .The Proposed Estimator

On the lines of Khare et al. (2013),we propose a difference –type estimator for population mean under two- phase sampling scheme using two auxiliary variables;the suggested estimator is, given by

$$t_m = \bar{y} + k_1\left(\bar{x}'\left(\frac{\bar{Z}}{\bar{z}}\right)^{\alpha} - \bar{x}\right) + k_2\left(\bar{z}'\left(\frac{\bar{X}'}{\bar{x}}\right) - \bar{z}\right) \tag{10}$$

Where are k_1 and k_2 the unknown constants.

To obtain the properties of the proposed estimator we define the following relative error terms and their expectations. Let

$$e_o = \frac{\bar{y}-\bar{Y}}{\bar{Y}},\ e_1 = \frac{\bar{x}-\bar{X}}{\bar{X}},\ e_1{}' = \frac{\bar{x}'-\bar{X}}{\bar{X}},\ e_2 = \frac{\bar{z}-\bar{Z}}{\bar{Z}},\ e_2{}' = \frac{\bar{z}'-\bar{Z}}{\bar{Z}}$$

$$E(e_o) = E(e_i) = E(e_i{}') = 0\ fori = 1{,}2.$$

$$E(e_o^2) = \theta C_y^2, E(e_1^2) = \theta C_x^2, E(e_1'^2) = \theta_1 C_x^2$$

$$E(e_1e_1^{'}) = \theta_1 C_x^2, E(e_2^{\,2}) = \theta C_x^2, E(e_o e_2^{'}) = \theta_1 C_{yz},$$

$$E(e_o e_1) = \theta C_{yx}, E(e_o e_1^{'}) = \theta_1 C_{yx}, E(e_o e_2) = \theta C_{yz},$$

$$E(e_1 e_2^{'}) = E(e_1 e_2^{'}) = \theta_1 C_{xz}, E(e_1 e_2) = \theta C_{xz}$$

$$E(e_2^{'2}) = E(e_2 e_2^{'}) = \theta_1 C_z^2$$

Rewriting equation (10) in terms of e’s we get,

$$t_m = \left[\bar{Y}(1+e_o) + k_1 \bar{X}\left((1+e_1^{'})\left(\frac{\bar{Z}}{\bar{Z}(1+e_1^{'})} \right)^{\alpha} - \bar{X}(1+e_1) \right) + k_2 \bar{Z}\left(\left(\frac{\bar{X}(1+e_1^{'})}{\bar{X}(1+e_1)} \right) - \bar{Z}(1+e_2) \right) \right]$$

$$t_m = \left[\bar{Y}(1+e_o) + k_1 \bar{X}\left((1+e_1^{'})(1+e_2^{'})^{-\alpha} - (1+e_1) \right) + k_2 \bar{Z}\left((1+e_1^{'})(1+e_1)^{-1} - (1+e_2) \right) \right]$$

Expanding the R.H.S of the above equation, and neglecting terms of e’s having power greater than two, we have

$$t_m - \bar{Y} = \begin{bmatrix} \bar{Y}(1+e_o) + k_1 \bar{X}\left((e_1 - e_1^{'} + \alpha e_2^{'} - \frac{\alpha(\alpha-1)}{2} e_2^{'2} + \alpha e_1^{'} e_2^{'} \right) - \\ k_2 \bar{Z}\left(e_1 - e_1^{\,2} - e_1^{'} - e_2 + e_1 e_1^{'} \right) \end{bmatrix} \qquad (11)$$

On squaring and taking expectation on both sides of equation (11), and keeping terms up to second order, we have

$$MSE(t_m) = E\begin{bmatrix} \bar{Y}^2 e_o^2 + k_1^2 \bar{X}^2 (e_1 - e_1^2 + \alpha e_2' - \frac{\alpha(\alpha-1)}{2} e_2^{'2} + \alpha e_1^{'} e_2^{'})^2 \\ + k_2^2 \bar{Z}\left(e_1 - e_1^{\,2} - e_1^{'} - e_2 + e_1 e_1^{'} \right)^2 \\ + 2k_1 k_2 \overline{XZ}(e_1 - e_1^2 + \alpha e_2' - \frac{\alpha(\alpha-1)}{2} e_2^{'2} + \alpha e_1^{'} e_2^{'}) \\ - 2k_1 \overline{YX}(e_o e_1 - e_o e_1^{'} + \alpha e_2^{'} e_o) \\ - 2k_2 \overline{YZ}(e_o e_1 - e_o e_1^{'} + e_o e_2) \\ \end{bmatrix}$$

Further, simplifying we get,

$$MSE(t_m) = \begin{bmatrix} \bar{Y}^2 \theta C_y^2 + k_1^2 \bar{X}^2 \left\{ \theta_2 C_x^2 + \alpha \theta_1 C_z^2 \right\} + k_2^2 \bar{Z}^2 \left\{ \theta_2 C_x^2 + \theta C_z^2 + 2\theta_2 C_{xz} \right\} + \\ 2k_1 k_2 \overline{XZ} \left\{ \theta_2 C_x^2 + \theta_2 C_{xz} + \alpha \theta_1 C_z^2 \right\} \\ - 2k_1 \overline{XY}(\theta_2 C_{yx} + \alpha \theta_1 C_{Zyz}) - 2k_2 \overline{YZ}(\theta_2 C_{yx} + \theta C_{yz}) \end{bmatrix} (12)$$

Now to find the minimum mean squared error of *t*t_m, we differentiate equation (12) with respect to k_1 and k_2 respectively and putting it equal to zero, that is

$$\frac{\partial MSE(t_m)}{\partial k_1} = 0 \text{ and } \frac{\partial MSE(t_m)}{\partial k_2} = 0$$

$$k_1 opt = \frac{\bar{Y}(BC - DE)}{\bar{X}(AB - E^2)} \text{ and } k_2 opt = \frac{\bar{Y}(AD - CE)}{\bar{Z}(AB - E^2)}$$

Where
$$A=\alpha\theta_1C_z^2+\theta_2C_x^2, B=\theta C_z^2+\theta_2C_x^2+2\theta_2C_{xz}, C=\theta_2C_{yx}+\alpha\theta_1C_{yz},$$
$$D=\theta_2C_{yx}+\theta C_{yz}, E=\alpha\theta_1C_z^2+\theta_2C_x^2+\theta_2C_{xz}$$

Putting $\alpha=1$ and substituting the optimum values of k_1 and k_2 in equation(12) we get the minimum mean square error (MSE) of the proposed estimator tm up to order one is, given as

$$MSE(t_m)=\bar{Y}^2\left[\theta C_y^2-\frac{(AD^2+BC^2-2CDE)}{(AB-E^2)}\right] \qquad (13)$$

5.Efficiency Comparisons

In this section, we have compare the propose estimator with the other existing estimators.

a. By equations(1) and (13),

$$MSE(t_m)_{\min}<MSE(t_0)\text{ if}\left[\frac{(AD^2+BC^2-2CDE)}{(AB-E^2)}\right]>0.$$

b. By equations (3) and (13),

$$MSE(t_m)_{\min}<MSE(t_1)\text{ if}\left[\frac{(AD^2+BC^2-2CDE)}{(AB-E^2)}+\theta_2(C_x^2-2\rho_{yx}C_yC_x)\right]>0.$$

c. By equations(5) and (13),

$$MSE(t_m)_{\min}<MSE(t_3)\text{ if}\left[\frac{(AD^2+BC^2-2CDE)}{(AB-E^2)}+\theta_2C_y^2\rho_{yx}^2\right]>0.$$

d. By equations (7) and (13),

$$MSE(t_m)_{\min}<MSE(t_5)\text{ if}$$

$$\left[\theta_2C_x(C_x-2\rho_{yx}C_y)+\theta_1C_z(C_z-2\rho_{yz}C_y)+\frac{(AD^2+BC^2-2CDE)}{(AB-E^2)}\right]>0.$$

e. By equations (9) and (13),

$$MSE(t_m)_{\min}<MSE(t_7)\text{ if}$$

$$\left[\theta_2C_x(C_x-2\rho_{yx}C_y)+\theta_1C_z(C_z-2\rho_{yz}C_y)+\frac{(AD^2+BC^2-2CDE)}{(AB-E^2)}-(\theta_2\rho_{yx}^2+\theta_1\rho_{yz}^2)C_y^2\right]>0.$$

6 Numerical Comparisons

To illustrate the performance of various estimators of $\bar{Y}$, we consider the data used by Anderson (1958). The variates are

y: Head length of second son,

x: Head length of first son,

z: Head breadth of first son,

$$N=25, \bar{Y}=183.84, \bar{X}=185.72, \bar{Z}=151.12, \rho_z=7.2, \rho_z=7.2, C_y=0.05, C_x=0.05, C_z=0.04$$
$$\rho_{yx}=0.71, \rho_{yz}=0.69, \rho_{xz}=0.73, n'=10, n=7$$

We have computed the percent relative efficiency (PRE) of different estimators of *Y* with respect to usual estimator $\bar{y}$ and compiled in the Table 1.1

We have use the following expression for Percentage Relative Efficiency(PRE)

$$PRE = \left[\frac{Var(t_0)}{MSE(t_j) or Var(t_j)} \right] *100 \textit{ for } j=0,1,2,3,4 \textit{ and } m.$$

Table 1: The mean square error (MSE's) and the percent relative efficiencies (PRE's) of the estimators with respect to t_o

Estimator	MSE's	PRE(t_0, t_1)
t_0	8.44	100.0
t_1	7.09	118.92
t_2	6.69	126.15
t_3	5.47	154.96
t_4	4.69	179.95
t_m	4.39	192.20

7. Conclusion

In this article we have proposed a difference –type estimator for population mean under two- phase sampling scheme using two auxiliary variables for the population mean of a study variable when information is available on an auxiliary variable in simple random sampling without replacement (SRSWOR).From the above table, we have observed that the proposed estimator has smaller mean square error and has higher percent relative efficiency than the other existing estimators.

References

[1] Bahl, S., Tuteja, R.K Ratio and product type exponential estimator. Inf Optim Sci.,**12**,159–163,1991.

[2] Chand, L. Some ratio type estimator based on two or more auxiliary variables. Unpublished Ph.D. dissertation, Lowa State University, Ames, Lowa,1975.

[3] Cochran, W.G Sampling techniques. Wiley, New-York Hansen MH, Hurwitz WN (1943) On the theory of sampling from finite populations. Ann Math Stat **14**, 333–362,1977.

[4] Khan, M. A ratio chain-type exponential estimator for finite population mean using double sampling. SpringerPlus.,**5**, 1–9, 2016.

[5] Khan, M. Improvement in estimating the finite population mean under maximum and minimum values in double sampling scheme. J Stat Appl Probab Lett,**2(2)**,1–7, 2015.

[6] Khare, B, B., Srivastava, U., Kumar, K(2013). A generalized chain ratio in regression estimator for population mean using two auxiliary characters in sample survey. J Sci Res Banaras Hindu Univ Varanasi,**57**, 147–153, 2013.

[7] Kiregyera, B. A chain ratio-type estimator in finite population mean in double sampling using two auxiliary variables. Metrika **27**, 217–223, 1980.

[8] Kiregyera, B. Regression-type estimator using two auxiliary variables and model of double sampling from finite populations. Metrika.,**31**, 215–223, 1984.

[9] Singh, B.K., Choudhury, S. Exponential chain ratio and product-type estimators for finite population mean under double sampling scheme. Glob J Sci Front Res Math Decis Sci.,**12(6)**, 2249–4626, 2012.

[10] Singh, G.N., Majhi, D. Some chain-type exponential estimators of population mean in two-phase sampling. Stat Trans ., **15(2)**, 221–230, 2014.

[11] Singh, H.P., Singh, S. and Kim, J.M. 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. J Korean Stat Soc.,**35(4)**, 377–395, 2006.

[12] Singh, R., Chuhan, P.,and Swan, N. Families of estimators for estimating population mean using known correlation coef- ficient in two phase sampling. Stat Trans.,**8(1)**, 89–96, 2007.

[13] Singh,R., Chuhan,P., Swan,N.,and Smarandache, F. Improved exponential estimator for population variance using two auxiliary variables. Ital J Pure Appl Math.,**28**,101–108, 2011.

[14] Srivastava, S.R., Khare, B.B., Srivastava, S.R A generalized chain ratio estimator for mean of finite population. J Indian Soc Agric Stat.,**42(1)**, 108–117,1990.

[15] Srivastava, S.K.A two phase estimator in sampling surveys. Austr J Stat.,**12**, 23–27, 1970.

[16] Sukhatme, B. V. Some ratio type estimators in two-phase sampling. J Am Stat Assoc.,**57**, 628–632,1962.

AsraNazir Research Scholar, Pursuing Ph.D in the Department of Statistics, University of Kashmir.

Rafia Jan Research Scholar, Pursuing Ph.D in the Department of Statistics University of Kashmir.

Tariq Rashid Jan. is faculty member in the Department of Statistics, University of Kashmir. He Obtained his Doctorate from the University of Kashmir.His field of interest are mainly in the area Sampling Theory, Generalized Probability Distribution.