International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 13 (2018) pp. 11151-11155

A Class of Double Sampling Log Type Estimators for Population Variance Using Two Auxiliary Variable

Shashi Bhushan[a] and Chandni Kumari[b]

[a]*Department of Mathematics & Statistics, Dr. Shakuntala MishraNational Rehabilitaion University, Lucknow, 226017, India.*

[b]*Department of Applied Statistics, Bahasaheb BhimraoAmbedkar University, Lucknow, 226025, India.*

Abstract

In this paper, a class of log-type estimator using two auxiliary information is proposed. Double sampling technique has been considered as it is assumed that the auxiliary information about the auxiliary variable is unknown. Bias and mean squared error has been found up to the first order of approximation. The proposed classes are compared to some commonly used estimators both theoretically as well as empirically and they perform better than commonly known estimators available in the literature.

INTRODUCTION

The use of auxiliary information enhance the precision of an estimator. Here, two auxiliary information is used for estimating the population variance under double sampling. We select the sample in two phases for estimating the population variance. Various authors used multiple auxiliary information (Olkin (1958), Raj (1965), Singh (1967), Shukla(1966), etc.). Recently, Bhushan and Kumari (2018) had made the use of logarithmic relationship between the study variable and auxiliary variable for estimating the population variance.

Consider a finite population $U = U_1, U_2, ..., U_N$ of size N from which a large sample of size n' is drawn according to simple random sampling without replacement (SRSWOR) for estimating unknown auxiliary variables only. Then, we select a sample of size n from the remaining observation for estimating the sample mean of study and auxiliary variables.

Let y_i, x_{i_1} and x_{i_2} denotes the value of the study and two auxiliary variable for the ith unit $i = 1, 2, ..., N$ of the population. Let $\bar{Y}$, $\bar{X}_1$ and $\bar{X}_2$ be the population means of study variable and two auxiliary variables. Also, $S_y^2 = N^{-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^2$, $S_{x_1}^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{X}_1\right)^2$ and $S_{x_2}^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{X}_2\right)^2$ be the population variance of the study and two auxiliary variables respectively.

Further, $\bar{x}_1'$ and $\bar{x}_2'$ be the larger sample means of two auxiliary variables and $s_{x_1}^{2'}$, $s_{x_2}^{2'}$ be the sample variance of two auxiliary variables respectively of size n'.

Here, $\bar{y}$, $\bar{x}_1$ and $\bar{x}_2$ be the sample means of study variable and two auxiliary variables from the sample of size n. Also, $s_y^2 = N^{-1}\sum_{i=1}^{N}\left(y_i - \bar{y}\right)^2$, $s_{x_1}^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_1\right)^2$ and $s_{x_2}^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}_2\right)^2$ be the sample variance of the study and two auxiliary variables respectively.

THE SUGGESTED GENERALIZED CLASS OF LOG-TYPE DOUBLE SAMPLING ESTIMATORS

We propose the following new classes of log type estimators for the population variance S_y^2 as

$$T_1 = s_y^2\left[1+\log\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\right]^{a_1}\left[1+\log\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)\right]^{a_2} \tag{2.1}$$

$$T_2 = s_y^2\left[1+b_1\log\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\right]\left[1+b_2\log\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)\right] \tag{2.2}$$

$$T_3 = s_y^2\left[1+\log\left(\frac{s_{x_1}^{2*'}}{s_{x_1}^{2*}}\right)\right]^{c_1}\left[1+\log\left(\frac{s_{x_2}^{2*'}}{s_{x_2}^{2*}}\right)\right]^{c_2} \tag{2.3}$$

$$T_4 = s_y^2\left[1+d_1\log\left(\frac{s_{x_1}^{2*'}}{s_{x_1}^{2*}}\right)\right]\left[1+d_2\log\left(\frac{s_{x_2}^{2*'}}{s_{x_2}^{2*}}\right)\right] \tag{2.4}$$

where $s_{x_i}^{2*'} = a_i s_{x_i}^{2'} + b_i$ and $s_{x_i}^{2*} = a_i s_{x_i}^2 + b_i$ for i = 1, 2

such that a_i, b_i, c_i and d_i are optimizing scalars or functions of the known parameters of the auxiliary variable x_i's such as the standard deviations S_{x_i}, coefficient of variation C_{x_i}, coefficient of kurtosis b_{2x_i}, coefficient of skewness b_{1x_i} and correlation coefficient $r_{x_i x_j}$ of the population $(i \neq j = 0)$.

PROPERTIES OF THE SUGGESTED CLASS OF ESTIMATORS

In order to obtain the bias and mean square error (MSE), let us consider

$$\varepsilon_0 = \frac{\left(s_y{}^2 - S_y{}^2\right)}{S_y{}^2},\ \varepsilon_1 = \frac{\left(s_{x_1}{}^2 - S_{x_1}{}^2\right)}{S_{x_1}{}^2},\ \varepsilon_1^{'} = \frac{\left(s_{x_1}{}^{2'} - S_{x_1}{}^2\right)}{S_{x_1}{}^2},$$

$$\varepsilon_2 = \frac{\left(s_{x_2}{}^2 - S_{x_2}{}^2\right)}{S_{x_2}{}^2},\ \varepsilon_2{}' = \frac{\left(s_{x_2}{}^{2'} - S_{x_2}{}^2\right)}{S_{x_2}{}^2}$$

$$E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0,\ E\left(\varepsilon_0{}^2\right) = Ib_{2y}^*,$$

$$E\left(\varepsilon_1{}^2\right) = Ib_{2x_1}^*,\ E\left(\varepsilon_1{}^{2'}\right) = I'b_{2x_1}^*,\ E\left(\varepsilon_2{}^2\right) = Ib_{2x_2}^*,$$

$$E\left(\varepsilon_2{}^{2'}\right) = I'b_{2x_2}^*,\ E(\varepsilon_0\varepsilon_1) = I\,I_{22yx_1}^*,$$

$$E(\varepsilon_0\varepsilon_1{}') = I'I_{22yx_1}^*,\ E(\varepsilon_0\varepsilon_2) = I\,I_{22yx_2}^*,$$

$E(\varepsilon_0\varepsilon_2{}') = I'I_{22yx_2}^*$, $E(\varepsilon_1\varepsilon_2) = I'I_{22x_1x_2}^*$ and $E(\varepsilon_1{}'\varepsilon_2{}') = I'I_{22x_1x_2}^*$ where $b_{2y}^* = b_{2y} - 1$, $b_{2x_1}^* = b_{2x_1} - 1$, $b_{2x_2}^* = b_{2x_2} - 1$ and $I_{22yx_1}^* = I_{22yx_1} - 1$, $I_{22yx_2}^* = I_{22yx_2} - 1$, $I_{22x_1x_2}^* = I_{22x_1x_2} - 1$;

$$I_{pq} = m_{pq} \big/ m_{20}^{p/2}\, m_{02}^{q/2},$$

$$m_{pq} = \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^p \left(X_i - \bar{X}\right)^q \Big/ N,\ I = 1/N,$$

$I' = 1/n'$, $b_{2y} = m_{40}/m_{20}^2$, $b_{2x} = m_{04}/m_{02}^2$ are the coefficient of kurtosis of y and x respectively.

Theorem 1

The bias and the mean squared error of the proposed estimator considered up to the terms of order n^{-1} are given by

$$Bias(T_1) = S_y^2 (I - I')\left[\frac{a_1^2}{2}b_{2x_1}^* + \frac{a_2^2}{2}b_{2x_2}^* - a_1 r_{yx_1}\sqrt{b_{2y}^* b_{2x_1}^*} - a_2 r_{yx_2}\sqrt{b_{2y}^* b_{2x_2}^*} + a_1 a_2 r_{x_1x_2}\sqrt{b_{2x_1}^* b_{2x_2}^*}\right]$$

$$MSE(T_1) = S_y^4\left[Ib_{2y}^* + (I - I')\left\{a_1^2 b_{2x_1}^* + a_2^2 b_{2x_2}^* - 2a_1 r_{yx_1}\sqrt{b_{2y}^* b_{2x_1}^*} - 2a_2 r_{yx_2}\sqrt{b_{2y}^* b_{2x_2}^*} + 2a_1 a_2 r_{x_1x_2}\sqrt{b_{2x_1}^* b_{2x_2}^*}\right\}\right]$$
$$-2w_1 S_y^4\left\{1 + (I - I')\left(\frac{a_1^2}{2}b_{2x_1}^* + \frac{a_2^2}{2}b_{2x_2}^* - a_1 r_{yx_1}\sqrt{b_{2y}^* b_{2x_1}^*} - a_2 r_{yx_2}\sqrt{b_{2y}^* b_{2x_2}^*} + a_1 a_2 r_{x_1x_2}\sqrt{b_{2x_1}^* b_{2x_2}^*}\right)\right\}$$

where $r_{yx_1} = \dfrac{I_{22yx_1}^*}{\sqrt{b_{2y}^* b_{2x_1}^*}}$, $r_{yx_2} = \dfrac{I_{22yx_2}^*}{\sqrt{b_{2y}^* b_{2x_2}^*}}$ and $r_{x_1x_2} = \dfrac{I_{22x_1x_2}^*}{\sqrt{b_{2x_1}^* b_{2x_2}^*}}$

Proof. Consider the estimator

$$T_1 = s_y^2\left[1 + \log\left(\frac{s_{x_1}^{2'}}{s_{x_1}^2}\right)\right]^{a_1}\left[1 + \log\left(\frac{s_{x_2}^{2'}}{s_{x_2}^2}\right)\right]^{a_2} = S_y^2(1 + \varepsilon_0)\left[1 + \log\left(1 + \varepsilon_1^{'}\right)(1 + \varepsilon_1)^{-1}\right]^{a_1}\left[1 + \log\left(1 + \varepsilon_2^{'}\right)(1 + \varepsilon_2)^{-1}\right]^{a_2}$$

$$= S_y^2(1 + \varepsilon_0)\left[1 + a_1\left(\varepsilon_1^{'} - \varepsilon_1 - \varepsilon_1\varepsilon_1^{'} + \varepsilon_1^2\right) + \frac{a_1^2}{2}\left(\varepsilon_1^{'} - \varepsilon_1\right)^2 - a_1\left(\varepsilon_1^{'} - \varepsilon_1\right)^2\right]$$
$$\left[1 + a_2\left(\varepsilon_2^{'} - \varepsilon_2 - \varepsilon_2\varepsilon_2^{'} + \varepsilon_2^2\right) + \frac{a_2^2}{2}\left(\varepsilon_2^{'} - \varepsilon_2\right)^2 - a_2\left(\varepsilon_2^{'} - \varepsilon_2\right)^2\right] \quad (3.1)$$

On solving and then taking expectation on both the sides, we get

$$Bias(T_1) = S_y^2 (I - I')\left[\frac{a_1^2}{2}b_{2x_1}^* + \frac{a_2^2}{2}b_{2x_2}^* - a_1 r_{yx_1}\sqrt{b_{2y}^* b_{2x_1}^*} - a_2 r_{yx_2}\sqrt{b_{2y}^* b_{2x_2}^*} + a_1 a_2 r_{x_1x_2}\sqrt{b_{2x_1}^* b_{2x_2}^*}\right]$$

Squaring and then taking

expectation on both the sides of equation (3.1), we get required expression for MSE.

Corollary 1. *The optimum values of constant are obtained as*

$$a_{1opt} = \left[\frac{r_{yx_1} - r_{yx_2}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}} \right] \frac{\sqrt{b^*_{2y}}}{\sqrt{b^*_{2x_1}}}$$

$$a_{2opt} = \left[\frac{r_{yx_2} - r_{yx_1}\, r_{x_1x_2}}{1 - r^2_{x_1x_2}} \right] \frac{\sqrt{b^*_{2y}}}{\sqrt{b^*_{2x_2}}}$$

The optimum mean squared error is given by

$$M\left(T_1\right)_{opt} = S_y^4 \left[I - \left(I - I'\right) R^{2*}_{y.x_1x_2} \right] \tag{3.2}$$

where $R^{2*}_{y.x_1x_2}$ *is the transformed multiple correlation coefficient between* y *and* x_1, x_2

MULTIVARIATE EXTENSION OF PROPOSED CLASS OF ESTIMATORS

Let there are k auxiliary variables then we can use the variables by taking a linear combination of these k estimators of the form given in section 2, calculated for every auxiliary variable separately, for estimating the population variance. Then the estimators for population variance will be defined as

$$T_1^* = s_y^2 \ \pi_{i=1}^{k} \left[1 + \log\left(\frac{s^{2'}_{x_i}}{s^2_{x_i}} \right) \right]^{a_i}$$

$$T_2^* = s_y^2 \ \pi_{i=1}^{k} \left[1 + b_i \log\left(\frac{s^{2'}_{x_i}}{s^2_{x_i}} \right) \right]$$

$$T_3^* = s_y^2 \ \pi_{i=1}^{k} \left[1 + \log\left(\frac{s^{2*'}_{x_i}}{s^{2*}_{x_i}} \right) \right]^{c_i}$$

$$T_4^* = s_y^2 \ \pi_{i=1}^{k} \left[1 + d_i \log\left(\frac{s^{2*'}_{x_i}}{s^{2*}_{x_i}} \right) \right]$$

where a_i, b_i, c_i and d_i are the optimizing scalars i = 1, 2, ..., k.

Theorem 2

The bias and the mean squared error of the proposed estimator considered upto the terms of order n^{-1} are given by

$$Bias\left(T_1^*\right) = S_y^2 \left(I - I'\right) \left[\sum_{i=1}^{k} \frac{a_i^2}{2} b^*_{2x_i} - \sum_{i=1}^{k} a_i\, r_{yx_i} \sqrt{b^*_{2y}\, b^*_{2x_i}} + \sum_{i \neq j=1}^{k} a_i\, a_j r_{x_ix_j} \sqrt{b^*_{2x_i} b^*_{2xj}} \right]$$

$$MSE\left(T_1^*\right) = S_y^4 \left[I b^*_{2y} + \left(I - I'\right) \left\{ \sum_{i=1}^{k} a_i^2 b^*_{2x_i} - 2 \sum_{i=1}^{k} a_i r_{yx_i} \sqrt{b^*_{2y} b^*_{2x_i}} + 2 \sum_{i \neq j=1}^{k} a_i a_j r_{x_ix_j} \sqrt{b^*_{2x_i} b^*_{2x_j}} \right\} \right]$$

$$-2 w_1 S_y^4 \left\{ 1 + \left(I - I'\right) \left(\sum_{i=1}^{k} \frac{a_i^2}{2} b^*_{2x_i} - \sum_{i=1}^{k} a_i\, r_{yx_i} \sqrt{b^*_{2y}\, b^*_{2x_i}} + \sum_{i \neq j=1}^{k} a_i\, a_j\, r_{x_ix_j} \sqrt{b^*_{2x_i} b^*_{2x_j}} \right) \right\}$$

EFFICIENCY COMPARISON

In this section, we compare the proposed classes of estimators with some important estimators. The comparison will be in terms of their MSE up to the order of n^{-1}. The optimum mean squared error of proposed estimator is given by

$$M\left(T_1\right)_{opt} = S_y^4 \left[I - \left(I - I'\right) R^{2*}_{y.x_1x_2} \right]$$

General variance estimator

$$\hat{S}_y^2 = s_y^2$$

It's mean squared error is given by

$$MSE(\hat{S}_y^2) = S_y^4\, I\, b^*_{2y}$$

The usual ratio type variance estimator

$$\hat{S}_r^{2'} = s_y^2 \left(\frac{s^{2'}_{x_1}}{s^2_{x_1}} \right) \left(\frac{s^{2'}_{x_2}}{s^2_{x_2}} \right)$$

It's mean squared error is given by

$$MSE(\hat{S}_r^{2'}) = S_y^4 \begin{bmatrix} Ib_{2y}^* + (I - I')b_{2x_1}^* + b_{2x2}^* \\ -2I_{22yx_1}^* - 2I_{22yx_2}^* + 2I_{22x_1x_2}^* \end{bmatrix} > MSE(T_1)_{opt}$$

The product type variance estimator

$$\hat{S}_p^{2'} = s_y^2 \left(\frac{s_{x_1}^2}{s_{x_1}^{2'}} \right) \left(\frac{s_{x_2}^2}{s_{x_2}^{2'}} \right)$$

Its mean squared error is given by

$$MSE(\hat{S}_p^{2'}) = S_y^4 \begin{bmatrix} Ib_{2y}^* + (I - I')b_{2x_1}^* + b_{2x2}^* + 2I_{22yx_1}^* \\ +2I_{22yx_2}^* + 2I_{22x_1x_2}^* \end{bmatrix} > MSE(T_1)_{opt}$$

Singh, Chauhan, Sawan and Smarandache (2011) type variance estimator

$$\hat{S}_s^{2'} = s_y^2 \exp\left(\frac{s_{x_1}^{2'} - s_{x_1}^2}{s_{x_1}^{2'} + s_{x_1}^2} \right) \left(\frac{s_{x_2}^{2'} - s_{x_2}^2}{s_{x_2}^{2'} + s_{x_2}^2} \right)$$

It's mean squared error is given by

$$MSE(\hat{S}_s^{2'}) = S_y^4 \left[Ib_{2y}^* + (I - I') \left\{ \begin{matrix} \frac{b_{2x_1}^*}{4} + \frac{b_{2x2}^*}{4} - I_{22yx_1}^* \\ -I_{22yx_2}^* + \frac{I_{22x_1x_2}^*}{4} \end{matrix} \right\} \right] > MSE(T_1)_{opt}$$

Olufadi and Kadilar (2014) variance estimator

$$\hat{S}_K^{2'} = s_y^2 \left(\frac{s_{x_1}^{2'}}{s_{x_1}^2} \right)^{a_1} \left(\frac{s_{x_2}^{2'}}{s_{x_2}^2} \right)^{a_2}$$

It's mean squared error is given by

$$MSE(\hat{S}_K^{2'}) = MSE(T_1)_{opt}$$

Das and Tripathi (1978) type variance estimator

$$\hat{S}_D^{2'} = s_y^2 \left(\frac{s_{x_1}^{2'}}{s_{x_1}^{2'} + a_1 \left(s_{x_1}^2 - s_{x_1}^{2'} \right)} \right) \left(\frac{s_{x_2}^{2'}}{s_{x_2}^{2'} + a_2 \left(s_{x_2}^2 - s_{x_2}^{2'} \right)} \right)$$

It's mean squared error is given by

$$MSE(\hat{S}_D^{2'}) = MSE(T_1)_{opt}$$

EMPIRICAL STUDY

The data on which we performed the numerical calculation is taken from some natural populations. The source of the data is given as follows.

Population 1. (Chochran, Pg. no. 155). The data concerns about weekly expenditure on food per family.

y : weekly expenditure on food

x_1 : number of persons

x_2 : the weekly family income

Population 2. (Choudhary F. S., Pg. no. 117).

y : area under wheat (in acres) in 1974

x_1 : area under wheat (in acres) in 1971

x_2 : area under wheat (in acres) in 1973

The summary and the percent relative efficiency of the following estimators are as follows:

Table 2: Parameters of the data

Parameter	Population 1	Population 2
N	33	34
n'	29	30
n	11	10
b_{2y}^*	4.032	2.725
$b_{2x_1}^*$	1.388	12.366
$b_{2x_1}^*$	1.143	1.912
$I_{22yx_1}^*$	0.305	0.224
$I_{22yx_2}^*$	1.155	2.104
$I_{22x_1x_2}^*$	0.492	0.152

Table 3: PRE of the estimators

Estimator	Pop. 1	Pop. 2
$\hat{S}_y^2$	100	100
$\hat{S}_r^{2'}$	91.627	29.173
$\hat{S}_p^{2'}$	50.236	17.524
$\hat{S}_s^{2'}$	109.836	75.636
$\hat{S}_D^{2'}$	122.595	230.718
$\hat{S}_K^{2'}$	122.595	230.718
$T_{1_{opt}}$	122.595	230.718

CONCLUSION

This work utilizes the two auxiliary information for estimating the study variable under double sampling. It is clear from the comparative study and numerical study that the proposed estimator perform better than conventional estimators viz. variance estimator, ratio type estimator, product type estimator and found equally efficient to Das & Tripathi type (1978) estimator, Olufadi and Kadilar type (2014) estimator etc. Hence, the proposed estimators have much more practical utility than the conventional estimators.

REFERENCES

[1] Bhushan S. and Kumari C. (2018). A new log type estimators for estimating the population variance, Int. J. Comp. App. Math., 13 (1), 43-54.

[2] Bahl S. and Tuteja R. K. (1991). Ratio and Product type exponential estimator, Info. Optim. Sci., Vol. XII(I), 159-163.

[3] Bhushan S., Gupta R. and Pandey S. K. (2015). Some log-type classes of estimators using auxiliary information.

[4] Hidiroglou M. A. and Sarndal C. E. (1998). Use of auxiliary information for two-phase sampling, Survey Methodology, 24(1), 11-20.

[5] Neyman J. (1938). Contribution to the theory of sampling human populations, J. Amer. Stat. Asso., 33, 101-116.

[6] Cochran W. G. (1963). Sampling Techniques, Wiley Eastern Private Limited, New Delhi, 307-310.

[7] Chaudhury A. (1978). On estimating the variance of a finite population. Metrika, 25, 66-67.

[8] Das A. K. and Tripathi T. P. (1978). Use of auxiliary information in estimating the nite population variance. Sankhya, C(4), 139-148.

[9] Gupta S. and Shabbir J. (2008). Variance estimation in simple random sampling using auxiliary information, Hacettepe Journal of Mathematics and Statistics, 37, 57-67.

[10] Isaki C. T. (1983). Variance estimation using Auxiliary Information, Jour. Amer. Statist. Asssoc., 78, 117-123.

[11] Kadilar C. and Cingi H. (2006)a . Improvement in variance estimation using auxiliary information, Hacettepe Journal of Mathematics and Statistics, 1(35), 111-115.

[12] Kadilar C. and Cingi H. (2006)b . Ratio estimators for population variance in simple and stratied sampling, Applied Mathematics and Computation, 1(73), 1047-1058.

[13] Sukhatme P. V., Sukhatme B. V., Sukhatme S. and Ashok C. (1984). Sampling Theory of Surveys with Applications, Iowa State University Press, Ams.

[14] Swain A. K. P. C. and Mishra G. (1994). Estimation of population variance under unequal probability sampling, Sankhya, B (56), 374-384