Modified Ratio Estimators Using Two Auxiliary Information for Estimating Population Variance in Two-Phase Sampling

Chandni Kumari[a], Shashi Bhushan[b] and Ratan Kumar Thakur[a]

[a]Department of Applied Statistics, Bahasaheb Bhimrao Ambedkar University, Lucknow, 226025, India.
[b]Department of Mathematics & Statistics, Dr. Shakuntala Mishra National Rehabilitaion University, Lucknow, 226017, India.

Abstract

A class of estimators are suggested for estimating the population variance using two auxiliary information in the form of variable and attribute under double sampling. It's bias and mean squared error are found to the first order of approximation. Further, an efficiency comparison and numerical study is also given.

1 Introduction

In sampling, auxiliary information proves to be fruitful way to increase the precision of the estimator. This auxiliary information may exist in both the forms: auxiliary variable and auxiliary attribute. Naik and Gupta (1996) defined ratio type estimator of population variance when the prior information of population proportion of attribute as auxiliary information is available. Bahl and Tuteja (1991) have proposed an exponential type estimator by using auxiliary variable. However, the fact that the known population proportion of an attribute also provides similar type of information has not drawn as much attention. Though some work has been done in this area like Koyuncu (2012), Singh and Solanki (2013), Bhushan et al. (2012), etc. But there are many situations, where the auxiliary information about the auxiliary attribute and auxiliary variable can be easily available, we can take the advantages of both the forms of auxiliary information to increase the efficiency of the estimators in various estimation problems. But in cases where such auxiliary information is lacking, we can make use of double sampling or two-phase sampling technique provided that such information may be easily and economically obtained. Neyman (1938) was the first person to introduce the double sampling technique in the history of sampling. However, in recent years, various authors (viz, Sisodia and Dwivedi (1981), Upadhayaya and Singh (1999), Kalidar and

Cingi (2006), Singh et al. (2004, 2008) etc. have made the use of prior knowledge about parameters of auxiliary variable like population variance along with coefficient of variation, correlation coefficient, coefficient of kurtosis, coefficient of skewness, standard deviation etc. for estimation of population variance of a variable of interest. In fact, such prior knowledge can also be very useful when a relation between the presence (or absence) of an attribute and the value of a variable is observed. Let Y be the study variable, X be the auxiliary variable and f be the auxiliary attribute.

Consider a finite population $U = U_1, U_2, ..., U_N$ of size N from which a sample is drawn according to simple random sampling without replacement. Let, y_i, x_i and f_i denotes the value of the study variable, auxiliary variable and auxiliary attribute for the ith unit $i = 1, 2, ..., N$ of the population. Let, $\bar{x}'$ and p' be the larger sample mean and $s_x^{2'} = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}'\right)^2$, $s_f^{2'} = n^{-1}\sum_{i=1}^{n}\left(f_i - p'\right)^2$ be the larger sample variance of the auxiliary variable and attribute respectively. Also, $\bar{y}$, $\bar{x}$ and p be the smaller sample means and $s_y^2 = n^{-1}\sum_{i=1}^{n}\left(y_i - \bar{y}\right)^2$, $s_x^2 = n^{-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2$ and $s_f^2 = n^{-1}\sum_{i=1}^{n}\left(f_i - p\right)^2$ be the smaller sample variance of the study variable and auxiliary attribute respectively.

2 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

$$T_1 = w_1 s_y^2 \left[1 + \log\left(\frac{s_x^{2'}}{s_x^2}\right)\right]^{a_1} \left[1 + \log\left(\frac{s_f^{2'}}{s_f^2}\right)\right]^{a_2} \tag{2.1}$$

$$T_2 = w_2 s_y^2 \left[1 + b_1 \log\left(\frac{s_x^{2'}}{s_x^2}\right)\right]\left[1 + b_2 \log\left(\frac{s_f^{2'}}{s_f^2}\right)\right] \tag{2.2}$$

$$T_3 = w_3 s_y^2 \left[1 + \log\left(\frac{s_x^{2*'}}{s_x^{2*}}\right)\right]^{c_1} \left[1 + \log\left(\frac{s_f^{2*'}}{s_f^{2*}}\right)\right]^{c_2} \tag{2.3}$$

$$T_4 = w_4 s_y^2 \left[1 + d_1 \log\left(\frac{s_x^{2*'}}{s_x^{2*}}\right)\right]\left[1 + d_2 \log\left(\frac{s_f^{2*'}}{s_f^{2*}}\right)\right] \tag{2.4}$$

where $S_x^{2*} = a_i S_x^2 + b_i$ and $s_x^{2*} = a_i s_x^2 + b_i$ for i = 1, 2

$S_f^{2*} = a_i S_f^2 + b_i$ and $s_f^{2*} = a_i s_f^2 + b_i$ for i = 1, 2

such that a_i, b_i, c_i and d_i are either real numbers or functions of the known parameters of the auxiliary variable x and auxiliary attribute f such as the standard deviations S_x , S_f, coefficient of variation C_x, C_f, coefficient of kurtosis b_{2x}, b_{2f}, coefficient of skewness b_{1x}, b_{1f} and correlation coefficient ρ of the population.

3 Properties of the suggested class of estimators

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

$$\varepsilon_0 = \frac{\left(s_y{}^2 - S_y{}^2\right)}{S_y{}^2},\ \varepsilon_1 = \frac{\left(s_x{}^2 - S_x{}^2\right)}{S_x{}^2},\ \varepsilon_1' = \frac{\left(s_x{}^{2'} - S_x{}^2\right)}{S_x{}^2},\ \varepsilon_2 = \frac{\left(s_f{}^2 - S_f{}^2\right)}{S_f{}^2} \text{ and } \varepsilon_2' = \frac{\left(s_f{}^{2'} - S_f{}^2\right)}{S_f{}^2}$$

$E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = 0$, $E\left(\varepsilon_0{}^2\right) = Ib_{2y}^*$, $E\left(\varepsilon_1{}^2\right) = Ib_{2x}^*$, $E\left(\varepsilon_1{}^{2'}\right) = I'b_{2x}^*$, $E\left(\varepsilon_2{}^2\right) = Ib_{2f}^*$, $E\left(\varepsilon_2{}^{2'}\right) = I'b_{2f}^*$, $E(\varepsilon_0\varepsilon_1) = I\,I_{22yx}^*$, $E\left(\varepsilon_0\varepsilon_1'\right) = I'I_{22yx}^*$, $E(\varepsilon_0\varepsilon_2) = I\,I_{22yf}^*$, $E\left(\varepsilon_0\varepsilon_2'\right) = I'I_{22yf}^*$ and $E(\varepsilon_1\varepsilon_2) = I\,I_{22xf}^*$, $E\left(\varepsilon_1'\varepsilon_2'\right) = I'I_{22xf}^*$ where $b_{2y}^* = b_{2y} - 1$, $b_{2x}^* = b_{2x} - 1$, $b_{2f}^* = b_{2f} - 1$ and $I_{22yx}^* = I_{22yx} - 1$, $I_{22yf}^* = I_{22yf} - 1$, $I_{22xf}^* = I_{22xf} - 1$; $I_{pq} = m_{pq} \big/ m_{20}^{p/2}\, m_{02}^{q/2}$, $m_{pq} = \sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^p \left(X_i - \bar{X}\right)^q \Big/ N$, $I = 1/N$, $I = 1/n'$, $b_{2y} = m_{40}/m_{20}^2$, $b_{2x} = m_{04}/m_{02}^2$ are the coefficient of kurtosis of y and x respectively.

Theorem 1 *The bias and the mean squared error of the proposed estimator considered upto the terms of order n⁻¹ are given by*

$$Bias(T_1) = S_y^2\left[w_1\left\{1 + (I - I')\left(\frac{a_1^2}{2}b_{2x}^* + \frac{a_2^2}{2}b_{2f}^* - a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*} + a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right)\right\} - 1\right]$$

$$MSE(T_1) = S_y^4 + w_1^4 S_y^4\left[1 + (I - I')\left\{b_{2y}^* + 2a_1^2 b_{2x}^* + 2a_2^2 b_{2f}^* - 4a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - 4a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*} + 4a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right\}\right]$$
$$-2w_1 S_y^4\left\{1 + (I - I')\left(\frac{a_1^2}{2}b_{2x}^* + \frac{a_2^2}{2}b_{2f}^* - a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*} + a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right)\right\}$$

where $r_{yx} = \frac{I_{22yx}^*}{\sqrt{b_{2y}^* b_{2x}^*}}$, $r_{yf} = \frac{I_{22yf}^*}{\sqrt{b_{2y}^* b_{2f}^*}}$ and $r_{xf} = \frac{I_{22xf}^*}{\sqrt{b_{2x}^* b_{2f}^*}}$

Proof. Consider the estimator

$$T_1 = w_1 s_y^2\left[1 + \log\left(\frac{s_x^{2'}}{s_x^2}\right)\right]^{a_1}\left[1 + \log\left(\frac{s_f^{2'}}{s_f^2}\right)\right]^{a_2}$$

$$= w_1 S_y^2 (1+\varepsilon_0)\left[1+\log(1+\varepsilon_1^{'})(1+\varepsilon_1)^{-1}\right]^{a_1}\left[1+\log(1+\varepsilon_2^{'})(1+\varepsilon_2)^{-1}\right]^{a_2}$$

$$= w_1 S_y^2 (1+\varepsilon_0)\left[1+a_1\left(\varepsilon_1^{'}-\varepsilon_1-\varepsilon_1\varepsilon_1^{'}+\varepsilon_1^2\right)+\frac{a_1^2}{2}\left(\varepsilon_1^{'}-\varepsilon_1\right)^2-a_1\left(\varepsilon_1^{'}-\varepsilon_1\right)^2\right]$$

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

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

$$Bias(T_1)=S_y^2\left[w_1\left\{1+(I-I')\left(\frac{a_1^2}{2}b_{2x}^*+\frac{a_2^2}{2}b_{2f}^*-a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*}-a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*}+a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right)\right\}-1\right]$$

Squaring and by considering expectation on both the sides of equation (3.1), we get

$$MSE(T_1)=S_y^4+w_1^4 S_y^4\left[1+(I-I')\left\{b_{2y}^*+2a_1^2 b_{2x}^*+2a_2^2 b_{2f}^*-4a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*}-4a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*}+4a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right\}\right]$$

$$-2w_1 S_y^4\left\{1+(I-I')\left(\frac{a_1^2}{2}b_{2x}^*+\frac{a_2^2}{2}b_{2f}^*-a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*}-a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*}+a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right)\right\}$$

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

$$w_{1opt}=\frac{B}{A}$$

where

$$A=\left[1+Ib_{2y}^*+(I-I')\left\{2a_1^2 b_{2x}^*+2a_2^2 b_{2f}^*-4a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*}-4a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*}+4a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right\}\right]$$

$$B=\left\{1+(I-I')\left(\frac{a_1^2}{2}b_{2x}^*+\frac{a_2^2}{2}b_{2f}^*-a_1 r_{yx}\sqrt{b_{2y}^* b_{2x}^*}-a_2 r_{yf}\sqrt{b_{2y}^* b_{2f}^*}+a_1 a_2 r_{xf}\sqrt{b_{2x}^* b_{2f}^*}\right)\right\}$$

The optimum mean squared error is given by

$$M(T_1)_{opt}=S_y^4\left(1-\frac{B^2}{A}\right) \quad (3.4)$$

5 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

$$M\left(T_1\right)_{opt}=S_y^4\left(1-\frac{B^2}{A}\right)$$

5.1 General variance estimator

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

It’s mean squared error is given by

$$MSE(\hat{S}_y^2)=S_y^4\,I\,b_{2y}^*>MSE(T_1)_{opt}$$

5.2 The usual ratio type variance estimator

$$\hat{S}_r^{2'}=s_y^2\left(\frac{s_x^{2'}}{s_x^2}\right)\left(\frac{s_f^{2'}}{s_f^2}\right)$$

It’s mean squared error is given by

$$MSE(\hat{S}_r^2)=S_y^4\left[I\,b_{2y}^*+\left(I-I'\right)\left\{b_{2x}^*+b_{2f}^*-2I_{22yx}^*-2I_{22yf}^*+2I_{22xf}^*\right\}\right]>MSE(T_1)_{opt}$$

5.3 The product type variance estimator

$$\hat{S}_p^{2'}=s_y^2\left(\frac{s_x^2}{s_x^{2'}}\right)\left(\frac{s_f^2}{s_f^{2'}}\right)$$

Its mean squared error is given by

$$MSE(\hat{S}_p^2)=S_y^4\left[I\,b_{2y}^*+\left(I-I'\right)\left\{b_{2x}^*+b_{2f}^*+2I_{22yx}^*+2I_{22yf}^*+2I_{22xf}^*\right\}\right]>MSE(T_1)_{opt}$$

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

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

It’s mean squared error is given by

$$MSE(\hat{S}_s^{2'})=S_y^4\left[Ib_{2y}^*+\left(I-I'\right)\left\{\frac{b_{2x}}{4}+\frac{b_{2f}^*}{4}-I_{22yx}^*-I_{22yf}^*+\frac{I_{22xf}^*}{4}\right\}\right]>MSE(T_1)_{opt}$$

5.5 Olufadi and Kadilar (2014) variance estimator

$$\hat{S}_K^{2'}=s_y^2\left(\frac{s_x^{2'}}{s_x^2}\right)^{a_1}\left(\frac{s_f^{2'}}{s_f^2}\right)^{a_2}$$

It's mean squared error is given by

$$MSE(\hat{S}_K^{2'}) = S_y^4 \left[Ib_{2y}^* + \left(I - I'\right)\left\{a_1^2 b_{2x}^* + a_2^2 b_{2f}^* - 2a_1 I_{22yx}^* - 2a_2 I_{22yf}^* + 2a_1 a_2 I_{22xf}^*\right\}\right] > MSE(T_1)_{opt}$$

5.6 Das and Tripathi (1978) type variance estimator

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

It's mean squared error is given by

$$MSE(\hat{S}_D^{2'}) = S_y^4 \left[Ib_{2y}^* + \left(I - I'\right)\left\{a_1^2 b_{2x}^* + a_2^2 b_{2f}^* - 2a_1 I_{22yx}^* - 2a_2 I_{22yf}^* + 2a_1 a_2 I_{22xf}^*\right\}\right] > MSE(T_1)_{opt}$$

6 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 : number of persons

f : the weekly family income

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

y : area under wheat (in acres) in 1974

x : area under wheat (in acres) in 1971

f : 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}^*	1.388	12.366
b_{2f}^*	1.143	1.912
I_{22yx}^*	0.305	0.224
I_{22yf}^*	1.155	2.104
I_{22xf}^*	0.492	0.152

Table 3: **PRE of the estimators**

Estimator	Pop. 1	Pop. 2
$\hat{S}_y^2$	100	100
$\hat{S}_r^{2'}$	119.692	38.956
$\hat{S}_p^{2'}$	65.714	20.984
$\hat{S}_s^{2'}$	114.495	90.921
$\hat{S}_D^{2'}$	122.486	231.107
$\hat{S}_K^{2'}$	122.486	231.107
$T_{1_{opt}}$	151.509	242.192

7 Conclusion

This work utilizes the auxiliary information in form of auxiliary variable and auxiliary attribute to the study variable under double sampling. It is clear from the comparative study and numerical study that the proposed estimators perform better than conventional estimators viz. variance estimator, ratio type estimator, product type estimator, Das & Tripathi type (1978) estimator, Singh et al. type (2011) estimator, Olufadi and Kadilar type (2014) estimator, Bhushan et al. type (2018) estimator. 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] Bhushan, S. and Kumari, C. (2018). “A new log type estimators for estimating the population variance”, International Journal of Computational and Applied Mathematics (ISSN 1819-4966), 13 (1), 43-54.

[3] Bhushan, S. and Kumari, C. “A Class of Double Sampling Log-Type Estimators for Population Variance Using Two Auxiliary Variable”, International Journal of Applied Engineering Research, Volume 13, Number (13) 2018, pg 11151-11155.

[4] Bhushan, S. and Kumari, C. “A Class of Double Sampling Log-Type Estimators for Population Variance Using Two Auxiliary Variable”, International Journal of Applied Engineering Research, Volume 13, Number (13) 2018, pg 11151-11155.

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

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

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

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

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

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

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

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

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

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

[15] 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.

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

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