Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/100118

An Efficient Log-Type Class Of Estimators Using Auxiliary Information Under Double Sampling

Chandni Kumari and Ratan Kumar Thakur*

Department of Statistics, Babasaheb Bhimrao Ambedkar University, (A Central University), Lucknow-226025, India

Received: 27 Jul. 2019, Revised: 3 Apr. 2020, Accepted: 1 May. 2020
Published online: 1 Mar. 2021

Abstract: In this paper, a class of log-type estimator using the auxiliary information in form of variable 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 used estimators available in literature.

Keywords: Mean squared error, efficiency, ratio method of estimation

1 Introduction

Double sampling proves to be a powerful technique when population mean or populataion total of auxiliary variable is unknown. Moreover in sampling theory, use of auxiliary information, is always beneficial in order to get more efficient estimates of population parameters. Various authors have made the use of auxiliary variable as a source of auxiliary information to increase the precision of the estimators, for the estimation of the population parameter under consideration. In recent years, many authors have also made use of various parameters associated with the auxiliary variable i.e. x for e.g., standard deviation S_x, coefficient of variation C_x, coefficient of kurtosis $b_2(x)$ and correlation coefficient r of the population in estimation of the population variance. [1–8, 10–13] etc are some of the authors in the list. In this paper, a family of estimators have been proposed by adapting the estimator of [15] and a class of log type estimators [14, 16–18] using the auxiliary information on a variable.

Consider a case of 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) under double sampling. Let $\bar{x}'$ is the larger sample mean which is used to estimate the unknown population mean of auxiliary variable and $s_x^{2'}$ be its sample variance.

Now, another sample of size n is taken from n' to estimate the sample mean of auxiliary variable only. Let y_i and x_i represent the values of the study and auxiliary variables for the ith unit $(i = 1, 2, ..., n)$ of the population. Further, let $\bar{y}$ and $\bar{x}$ be the sample means and $s_y^2 = \dfrac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{(n-1)}$ and $s_x^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$ be the sample variance of the study and auxiliary variables respectively.

2 Proposed estimators

In this paper, the following family of estimators has been proposed for the estimation of the population variance of the study variable y using auxiliary information on a variable.

$$T_c = \left[w_1 s_y^2 + w_2\left(\frac{s_x'^2}{s_x^2}\right)\right]\left[1 + a\,log\left(\frac{s_x'^{*2}}{s_x^{*2}}\right)\right] \tag{1}$$

* Corresponding author e-mail: chandnithakur897@gmail.com

where a is the characterizing scalar.
$s_x^{*2} = a s_x^2 + b, S_x^{*2} = a S_x^2 + b$
such that $a_1(\neq 0)$, b_1 and $a_2(\neq 0)$, b_2 are either real numbers or functions of the known parameters of the auxiliary variable x such as standard deviations S_x, coefficient of variation C_x, coefficient of kurtosis b_{2x}, coefficient of skewness b_{1xi} and correlation coefficient r of population.

3 Properties of proposed estimators

In order to obtain the bias and mean square error (MSE), let us consider
$E(\varepsilon_0) = 0 = E(\varepsilon_1) = E(\varepsilon_1^{'}), E(\varepsilon_0)^2 = Ib_{2y}^*, E(\varepsilon_1)^2 = Ib_{2x}^*, E(\varepsilon_1^{'})^2 = I'b_{2x}^*, E(\varepsilon_0\varepsilon_1) = II_{22yx}^*, E(\varepsilon_0\varepsilon_1^{'}) = I'I_{22yx}^*, E(\varepsilon_1\varepsilon_1^{'}) = I'b_{2x}^*$
where, $b_{2x}^* = b_{2x} - 1$, $b_{2y}^* = b_{2y} - 1$ and $I_{22yx}^* = I_{22yx} - 1$; $I_{pq} = m_{pq}/m_{20}^{\frac{p}{2}} m_{02}^{\frac{q}{2}}$, $\quad m_{pq} = \sum_{i=1}^{N}(Y_i - \bar{Y})^p (Y_i - \bar{Y})^q / N$, $I = 1/N$, $I' = 1/n'$, $\quad b_{2y} = m_{40}/m_{20}^2$, $\quad b_{2x} = m_{04}/m_{02}^2$ are the coefficient of kurtosis of study and auxiliary variable respectively.

Theorem 1.*Bias and mean squared error of the proposed estimators are given by*

$$\begin{aligned} Bias(T) &= S_y^2 \left[w_1 \left\{ 1 + (I - I') \left(a\eta^2 b_{2x}^* - a\eta r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - \frac{a\eta^2}{2} b_{2x}^* \right) \right\} - 1 \right] + w_2 \left\{ 1 + (I - I') \right. \\ &\quad \left. \left(b_{2x}^* + a\eta I b_{2x}^* + \frac{a\eta^2}{2} I b_{2x}^* \right) \right\} \\ MSE(T) &= S_y^4 w_1^2 A + w_2^2 B + S_y^4 w_1 D + S_y^2 w_2 G + S_y^2 w_1 w_2 F + S_y^4 \\ &\quad where \qquad (2) \\ A &= 1 + (I - I') \left(b_{2y}^* + a^2\eta^2 b_{2x}^* - 4a\eta r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - a\eta^2 b_{2x}^* \right) \\ B &= 1 + (I - I') \left(b_{2x}^* + a^2\eta^2 b_{2x}^* + 4a\eta r_{yx} b_{2x}^* + a\eta^2 b_{2x}^* \right) \\ D &= \left[(I - I') \left(2a\eta r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - a\eta^2 b_{2x}^* \right) - 2 \right] \\ G &= \left[(I - I') \left(-a\eta^2 b_{2x}^* - 2a\eta b_{2x}^* \right) - 2 \right] \\ F &= 2 + 2(I - I') \left(2a\eta b_{2x}^* - 2a\eta r_{yx}\sqrt{b_{2y}^* b_{2x}^*} - r_{yx}\sqrt{b_{2y}^* b_{2x}^*} + a\eta^2 b_{2x}^* + a^2\eta^2 b_{2x}^* \right) \\ \eta &= \frac{aS_x^2}{aS_x^2 + b} \end{aligned}$$

where $\eta_1 = \frac{a_1 S_x^2}{a_1 S_x^2 + b_1}$, $\eta_2 = \frac{a_2 S_f^2}{a_2 S_f^2 + b_2}$, $r_{yx} = \frac{I_{22yx}^*}{\sqrt{b_{2y}^* b_{2x}^*}}$,

Corollary 1.*Mean square error of the proposed class of estimator* T_c *is minimum for the optimum value of the characterizing parameters, given*

$$\begin{aligned} w_{1opt} &= \left[\frac{GF - 2BD}{4AB - F^2} \right] \\ w_{2opt} &= S_y^2 \left[\frac{DF - 2GA}{4AB - F^2} \right] \end{aligned} \qquad (3)$$

and the minimum value of the mean square error within the proposed class of estimator is

$$M(T_c)_{opt} = S_y^4 \left[1 - \frac{BD^2 - DFG + AG^2}{4AB - F^2} \right] \qquad (4)$$

4 Some members of the proposed class of estimators

It can be easily seen that a proposed class T_c is a generalized form of class of estimators for constants $a_1\ (\neq 0)$, b_1 and $a_2 (\neq 0)$, b_2 is either real numbers or functions of the known parameters of the auxiliary variable x such as the standard deviations S_x, coefficient of variation C_x, coefficient of kurtosis $b_2(x)$, coefficient of skewness $b_1(x)$ and correlation coefficient r of the population. Therefore, a wide variety of estimators can be designed using the above known population parameters. Some of them are listed below.

Table 1: *Some members of the proposed class of estimators i.e. T_c*

Log-type estimators T_c	a_1	b_1
$T_{c_1} = \left[w_1 s_y^2 + w_2 \left(\frac{s_x^{2'}}{s_x^2}\right)\right]^a \left[1 + log\left(\frac{s_x^{*2'}}{s_x^{*2}}\right)\right]^b$	1	0
$T_{c_2} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{s_x^{2'} + C_x}{s_x^2 + C_x}\right)\right]^a \left[1 + log\left(\frac{s_x^{*2'} + C_x}{s_x^{*2} + C_x}\right)\right]^b$	1	C_x
$T_{c_3} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{b_2(x) s_x^{2'} + C_x}{b_2(x) s_x^2 + C_x}\right)\right]^a \left[1 + log\left(\frac{b_2(x) s_x^{*2'} + C_x}{b_2(x) s_x^{*2} + C_x}\right)\right]^b$	b_{2x}	C_x
$T_{c_4} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{C_x s_x^{2'} + b_2(x)}{C_x s_x^2 + b_2(x)}\right)\right]^a \left[1 + log\left(\frac{C_x s_x^{*2'} + b_2(x)}{C_x s_x^{*2} + b_2(x)}\right)\right]^b$	C_x	b_{2x}
$T_{c_5} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{s_x^{2'} + S_x}{s_x^2 + S_x}\right)\right]^a \left[1 + log\left(\frac{s_x^{*2'} + S_x}{s_x^{*2} + S_x}\right)\right]^b$	1	S_x
$T_6 = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{b_1(x) s_x^{2'} + S_x}{b_1(x) s_x^2 + S_x}\right)\right]^a \left[1 + log\left(\frac{b_1(x) s_x^{*2'} + S_x}{b_1(x) s_x^{*2} + S_x}\right)\right]^b$	b_{1x}	S_x
$T_{c_7} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{b_2(x) s_x^{2'} + S_x}{b_2(x) s_x^2 + S_x}\right)\right]^a \left[1 + log\left(\frac{b_2(x) s_x^{*2'} + S_x}{b_2(x) s_x^{*2} + S_f}\right)\right]^b$	b_{2x}	S_x
$T_{c_8} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{s_x^{2'} + r}{s_x^2 + r}\right)\right]^a \left[1 + log\left(\frac{s_x^{*2'} + r}{s_x^{*2} + r}\right)\right]^b$	1	r
$T_{c_9} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{s_x^{2'} + b_2(x)}{s_x^2 + b_2(x)}\right)\right]^a \left[1 + log\left(\frac{s_x^{*2'} + b_2(x)}{s_x^{*2} + b_2(x)}\right)\right]^b$	1	$b_2(x)$
$T_{c_{10}} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{C_x s_x^{2'} + r}{C_x s_x^2 + r}\right)\right]^a \left[1 + log\left(\frac{C_x s_x^{*2'} + r}{C_x s_x^{*2} + r}\right)\right]^b$	C_x	r
$T_{c_{11}} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{r s_x^{2'} + C_x}{r s_x^2 + C_x}\right)\right]^a \left[1 + log\left(\frac{r s_x^{*2'} + C_x}{r s_x^{*2} + C_x}\right)\right]^b$	r	C_x
$T_{c_{12}} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{b_2(x) s_x^{2'} + r}{b_2(x) s_x^2 + r}\right)\right]^a \left[1 + log\left(\frac{b_2(x) s_x^{*2'} + r}{b_2(x) s_x^{*2} + r}\right)\right]^b$	b_{2x}	r
$T_{c_{13}} = s_y^2 \left[w_1 s_y^2 + w_2 \left(\frac{r s_x^{2'} + b_2(x)}{r s_x^2 + b_2(x)}\right)\right]^a \left[1 + log\left(\frac{r s_x^{*2'} + b_2(x)}{r s_x^{*2} + b_2(x)}\right)\right]^b$	r	$b_2(x)$

5 Dominance Condition

Let us now compare the proposed classes of estimators with the conventional estimators.

5.1 Variance estimator in case of SRSWOR

$$MSE(t_1) > MSE(T_c)_{opt}$$

5.2 Ratio type variance estimator

$$MSE(t_2) > MSE(T_c)_{opt}$$

5.3 Product type variance estimator

$$MSE(t_3) > MSE(T_c)_{opt}$$

5.4 Singh et al. (2011) estimator

$$MSE(t_4) > MSE(T_c)_{opt}$$

5.5 Isaki (1983) estimator

$$MSE(t_5) > MSE(T_c)_{opt}$$

5.6 Das and Tripathi (1978) estimator

$$MSE(t_6) > MSE(T_c)_{opt}$$

5.7 Kadilar and Cingi (2006) estimator

$$MSE(t_7) > MSE(T_c)_{opt}$$

Thus, it can be easily seen and verified that the proposed class of log-type estimators is far better than these above mentioned estimators which is available in sampling literature.

6 Empirical Performance

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

Table 2: *Parameters of Data*

Parameters	Population 1	Population 2	Population 3	Population 4
n'	33	36	22	34
n	11	11	8	10
b^*_{2y}	4.032	2.632	9.433	2.725
b^*_{2x}	1.388	2.402	7.105	12.366
I^*_{22xy}	0.305	1.835	8.140	0.224

Table 3: *PRE of the estimators*

Estimator	Population 1	Population 2	Population 3	Population 4
t_1	100	100	100	100
t_2	87.167	65.093	162.742	21.544
t_3	38.525	13.593	13.009	12.407
t_4	116.860	248.212	5107.79	67.4228
t_5	141.940	277.598	574168.1	637.142
t_6	142.235	293.078	605514.2	666.034
t_7	142.235	293.078	605514.2	666.034
T_{iopt}	98.67	185.87	322345.2	456.786

7 Conclusion

The present study extends the idea regarding the effective use of auxiliary information if the relationship between the study variable and the auxiliary variable is of logarithmic type. The present study provides some novel estimators that are very efficient to some known estimators when such auxiliary information is available. This study is also supported through a numerical study and the results of the study are quite encouraging.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

[1] Kumari, C. and Thakur. R. K. Optimal two parameter logarithmic estimators for estimating the population variance. *Global J. Pure App. Math.*, 15 (5), 527-536, 2019.
[2] Bhushan S. and Kumari C. Double sampling log type estimators using auxiliary attribute for population variance, *J. Stat. Appl. Pro.*, 6 (3), 1-6, 2019.
[3] Bhushan S. and Kumari C. A new log type estimators for estimating the population variance, *Int. J. Comp. App. Math.*, 13 (1), 43-54, 2018.
[4] Bhushan S. and Kumari C. A class of double sampling log type estimators for population variance using two auxiliary variable, *Int. J. Appl. Eng. Res.*, 13 (13) ,11151-11155, 2018.
[5] Bhushan S. and Kumari C. Estimation of variance of finite population using double sampling scheme, *Int. J. Sci. Eng. Res.*, 9 (8), 1893-1901, 2018.
[6] Bhushan S. and Kumari C. Modified ratio estimators using two auxiliary information for estimating population variance in two-phase sampling, *Int. J. Sci. Eng. Res.*, 9 (8), 1884-1892, 2018.
[7] Bhushan S. and Kumari C. Some classes of log type estimators using auxiliary attribute for population variance, *Int. J. Sci. Eng. Res.*, 9 (7), 1823-1832, 2018.
[8] Singh, R., Chauhan, P., Sawan, N. & Smarandache, F. Improved exponential estimator for population variance using two auxiliary variables, *Ital. J. Pure Appl. Maths.* 28, 101–108, 2011.
[9] Gupta S. and Shabbir J. Variance estimation in simple random sampling using auxiliary information, *Hacettepe Journal of Mathematics and Statistics*, 37, 57-67, 2008.
[10] Isaki C. T. Variance estimation using auxiliary information , *Jour. Amer. Statist. Asssoct*, 78, 117-123, 1983.
[11] Kadilar C. and Cingi H. Improvement in variance estimation using auxiliary information, *Hacettepe Journal of Mathematics and Statistics*, 1 (35), 111-115, 2006a
[12] Kadilar C. and Cingi H. Ratio estimators for population variance in simple and stratified sampling, *Applied Mathematics and Computation*, 1 (73), 1047-1058, 2006b.

[13] Bahl S. and Tuteja R. K. Ratio and product type exponential estimator, *Info. Optim. Sci.*, Vol. XII(I), 159-163, 1991.
[14] Hidiroglou M. A. and Sarndal C. E. Use of auxiliary information for two-phase sampling, *Survey Methodology*, 24 (1), 11-20, 1998.
[15] Neyman J. Contribution to the theory of sampling human populations, *J. Amer. Stat. Asso.*, 33, 101-116, 1938.
[16] Cochran W. G. Sampling techniques, *Wiley Eastern Private Limited*, New Delhi, 307-310, 1963.
[17] Chaudhury A. On estimating the variance of a finite population. *Metrika*, 25, 66-67, 1978.
[18] Das A. K. and Tripathi T. P. Use of auxiliary information in estimating the finite population variance. *Sankhya*, C(4), 139-148, 1978.