Thailand Statistician
July 2024; 22(3): 610-617
http://statassoc.or.th
Contributed paper

Log-Product-Type Estimator for Estimation of Population Variance Using Auxiliary Information

Prabhakar Mishra [a], Ashish Sharma [b], Nitesh Kumar Adichwal* [c], Sakshi Rai [a] and Rajesh Singh [a]
[a] Department of Statistics, Banaras Hindu University, Varanasi, India.
[b] School of Business, UPES University, Dehradun, India.
[c] School of Management, IILM University, Greater Noida, Uttar Pradesh, India.
*Corresponding author; e-mail: nnitesh139@gmail.com

Received: 7 January 2021
Revised: 30 March 2022
Accepted: 17 April 2022

Abstract

This paper proposed a log product type estimator for estimating population variance under simple random sampling without replacement (SRSWOR) using auxiliary information. We have calculated the mean square error (MSE) and bias expressions up to the first order of approximation. To substantiate the result, an empirical study has been performed using three real population data sets. The properties of the estimators also verified through simulation study. The result shows that the performance of the proposed estimator is better than the existing estimators.

Keywords: Auxiliary variable, variance, SRSWOR, mean square error, bias.

1. Introduction

In sampling theory, the use of auxiliary variable increases the efficiency and precision as well. Auxiliary variables are correlated with study variable and accounted in sampling to reduce the mean square error, so as to get a highly précised estimate of the concerned population parameter. Mostly, auxiliary information is used at selection stage or at the stage of estimation. In sampling theory many authors have already proposed various estimators using auxiliary variables to improve the estimate of population parameter. Cochran (1940) was the first who suggested the use of auxiliary information in sampling to increase the efficiency of the estimators.

Several authors including Adichwal et al. (2018), Das and Tripathi (1978), Isaki (1983) and Upadhyaya and Singh (2001) studied the estimation of variance using auxiliary information. Shabbir (2006), Gupta and Shabbir (2008), Singh et al. (2011), Singh and Solanki (2013), Yadav and Kadilar (2013), Kumar and Saini (2020), Saini and Kumar (2020) suggested an estimator using auxiliary information to estimate population parameters under different sampling schemes. Sharma and Singh (2014) worked on the concept of improved estimation of population variance using the information of auxiliary attributes in simple random sampling. Subramani and Kumarpandiyan (2015) proposed a class of modified ratio estimators using information of auxiliary variable for estimating population variance. In this context, we have used auxiliary information and proposed log product type estimators for the estimation of population variance.

2. Notations and Terminology

Let us consider n sample observations for auxiliary variable and study variable y_i $(i = 1, 2, \ldots, n)$, respectively from a population of size N under simple random sampling without replacement (SRSWOR) scheme. S_y^2 and s_y^2 are the population and sample mean square for study variable Y, respectively. S_x^2 and s_x^2 are the population and sample mean square for auxiliary variable X. Let e_0 and e_1 be the sampling errors defined as

$$e_0 = \frac{s_y^2 - S_y^2}{S_y^2} x_i \text{ and } e_1 = \frac{s_x^2 - S_x^2}{S_x^2}. \tag{1}$$

If the expressions e_0 and e_1 are defined as (1),

$$E(e_0) = E(e_1) = 0, \tag{2}$$

$$E(e_0^2) = \xi(\lambda_{40} - 1), \tag{3}$$

$$E(e_1^2) = \xi(\lambda_{04} - 1), \tag{4}$$

$$E(e_0 e_1) = \xi(\lambda_{22} - 1), \tag{5}$$

where $\lambda_{mn} = \frac{\mu_{mn}}{\mu_{20}^{m/2} \mu_{02}^{n/2}}$, m and n be any non-negative integer, $\mu_{mn} = \frac{1}{N-1} \sum (Y_i - \bar{Y})^m (X_i - \bar{X})^n$, $\xi = \frac{1}{n} - \frac{1}{N}$.

3. Existing Estimators

Table 1 displays the existing estimators along with their variance/MSE.

Table 1 Estimators and variance/MSE of the estimators

No.	Estimators	Variance/MSE Expressions
1	$t_0 = s_y^2$ (Usual unbiased estimator)	$Var(t_0) = \xi(\lambda_{40} - 1) S_y^4$
2	$t_1 = s_y^2 \left[\frac{S_x^2}{s_x^2} \right]$ (Isaki 1983)	$MSE(t_1) = S_y^4 \xi \left[(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1) \right]$
3	$t_2 = s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right]$	$MSE(t_2) = S_y^4 \xi \left[(\lambda_{40} - 1) + \frac{(\lambda_{04} - 1)}{4} - (\lambda_{22} - 1) \right]$
	$t_3 = s_y^2 \exp\left[\frac{s_x^2 - S_x^2}{s_x^2 + S_x^2} \right]$ (Kumar et al. 2011)	$MSE(t_3) = S_y^4 \xi \left[(\lambda_{40} - 1) + \frac{(\lambda_{04} - 1)}{4} + (\lambda_{22} - 1) \right]$
4	$t_{reg} = s_y^2 + k_0 \left(S_x^2 - s_x^2 \right)$ (Isaki 1983)	$Var(t_{reg}) = S_y^4 \xi \left[(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} \right]$

Table 1 (Continued)

No.	Estimator	Variance/MSE Expressions
5	$t_{ss} = w_1 s_y^2 + w_2(S_x^2 - s_x^2)$ (Sharma and Singh 2014)	$\min.MSE\ (t_{ss}) = \dfrac{S_y^4}{D_1}\left[(D_1 - 1) - \dfrac{2\xi(\lambda_{22}-1)^2}{(\lambda_{04}-1)D_1}\right]$ where $D_1 = 1 + \xi(\lambda_{40}-1) + \xi\dfrac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}$
6	$t_4 = s_y^2 \exp\left[\dfrac{S_x^2 - s_x^2}{S_x^2 + (\alpha - 1)s_x^2}\right]$ (Yadav and Kadilar 2013)	$\min.MSE(t_4) = S_y^4 \xi\left[(\lambda_{40}-1) - \dfrac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]$

4. Proposed Estimators

Motivated by Mishra et al. (2017), we have proposed the following two estimators for estimating population variance which are given as

(i)
$$tl_1 = s_y^2 + \alpha \log\left(\frac{s_x^2}{S_x^2}\right). \tag{6}$$

Expressing (6) in the form of e's

$$tl_1 = s_y^2(1+e_0) + \alpha\left(e_1 - \frac{e_1^2}{2}\right),$$

$$tl_1 - S_y^2 = S_y^2 e_0 + \alpha\left(e_1 - \frac{e_1^2}{2}\right).$$

The expression of bias and MSE for the proposed estimator tl_1 up to the first order of approximation is given by

$$Bias(tl_1) = -\frac{\alpha}{2}\xi\left(\lambda_{04} - 1\right), \tag{7}$$

$$MSE(tl_1) = S_y^4 A + \alpha^2 B + 2S_y^2 \alpha C, \tag{8}$$

where $A = \xi(\lambda_{40}-1),\ B = \xi(\lambda_{04}-1),\ C = \xi(\lambda_{22}-1).$ The optimum value of α is given by

$$\alpha^* = -\frac{S_y^2\ C}{B}. \tag{9}$$

Then minimum MSE at optimum value of α is given by

$$\min.MSE(tl_1) = S_y^4\left(A - \frac{C^2}{B}\right). \tag{10}$$

(ii)
$$tl_2 = s_y^2(w_1 + 1) + w_2 \log\left(\frac{s_x^2}{S_x^2}\right). \tag{11}$$

Expressing (11) in the term's of e's

$$tl_2 = S_y^2\left(1+e_0\right)\left(w_1+1\right) + w_2\left(e_1 - \frac{e_1^2}{2}\right),$$

$$tl_2 - S_y^2 = S_y^2 e_0 + S_y^2\ w_1(1+e_0) + w_2\left(e_1 - \frac{e_1^2}{2}\right).$$

The expression of bias and MSE for the proposed estimator tl_2 up to the first order of approximation is given by

$$Bias(tl_2) = S_y^2 w_1 - \frac{w_2}{2}\xi(\lambda_{04}-1), \tag{12}$$

$$MSE(tl_2) = C_1 + w_1^2 A_1 + w_2^2 B_1 + 2w_1 C_1 + 2w_2 D + 2w_1 w_2 E, \tag{13}$$

where $A_1 = S_y^4(1+\xi(\lambda_{40}-1)$, $B_1 = \xi(\lambda_{04}-1)$, $C_1 = S_y^4\xi(\lambda_{40}-1)$, $D = S_y^2\xi(\lambda_{22}-1)$, $E = S_y^2\xi\left((\lambda_{22}-1) - \frac{(\lambda_{04}-1)}{2}\right)$. At optimum value of w_1 and w_2, i.e. w_1^* and w_2^*, the minimum MSE expression is given by

$$w_1^* = \frac{C_1 B_1 - DE}{E^2 - A_1 B_1}, \tag{14}$$

$$w_2^* = \frac{A_1 D - C_1 E}{E^2 - A_1 B_1}, \tag{15}$$

$$\min.MSE(tl_2) = C_1 + \frac{(B_1 C_1^2 + A_1 D^2 - 2C_1 DE)}{(E^2 - A_1 B_1)}. \tag{16}$$

5. Efficiency Comparison

This section compare the MSE of the proposed estimators tl_1 and tl_2 with the existing estimators t_0, t_1, t_2, t_3, t_{reg}, t_{ss}, and t_4.

$$Var(t_0) - \min.MSE(tl_1) \ \geq\ 0, \ \Rightarrow\ \frac{\xi S_y^2(\lambda_{22}-1)^2}{(\lambda_{04}-1)} \ \geq\ 0, \tag{17}$$

$$MSE(t_1) - \min.MSE(tl_1) \ \geq\ 0, \ \Rightarrow\ S_y^4\xi\left[\frac{(\lambda_{04}-\lambda_{22})^2}{(\lambda_{04}-1)}\right] \ \geq\ 0, \tag{18}$$

$$MSE(t_2) - \min.MSE(tl_1) \ \geq\ 0, \ \Rightarrow\ \frac{S_y^4\xi}{(\lambda_{04}-1)}\left[\frac{(\lambda_{04}-2\lambda_{22}+1)}{2}\right]^2 \ \geq\ 0 \tag{19}$$

$$MSE(t_3) - \min.MSE(tl_1) \ \geq\ 0, \ \Rightarrow\ \frac{S_y^4\xi}{(\lambda_{04}-1)}\left[\frac{(\lambda_{04}+2\lambda_{22}-3)}{2}\right]^2 \ \geq\ 0, \tag{20}$$

$$\min.MSE(t_{ss}) - \min.\ MSE(tl_1) \ \geq\ 0, \ \Rightarrow\ \frac{2\ S_y^2\xi(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\left[\frac{1}{D_1} - \frac{1}{D_1^2}\right] \ \geq\ 0, \tag{21}$$

$$Var(t_0) - \min.MSE(tl_2) \ \geq\ 0, \ \Rightarrow\ \frac{B_1 C_1^2 + A_1 D^2 - 2C_1 DE}{E^2 - A_1 B_1} \ \leq\ 0, \tag{22}$$

$$MSE(t_1) - \min.\ MSE(tl_2) \ \geq\ 0, \Rightarrow\ S_y^4\xi\left[(\lambda_{04}-1) - 2(\lambda_{22}-1)\right] - \left[\frac{B_1 C_1^2 + A_1 D^2 - 2C_1 DE}{E^2 - A_1 B_1}\right] \ \geq\ 0, \tag{23}$$

$$MSE(t_2) - \min.MSE(tl_2) \ \geq 0, \ \Rightarrow\ S_y^4\xi\left[\frac{(\lambda_{04}-1)}{4} - (\lambda_{22}-1)\right] - \left[\frac{B_1 C_1^2 + A_1 D^2 - 2\ C_1 DE}{E^2 - A_1 B_1}\right] \ \geq\ 0, \tag{24}$$

$$MSE(t_3)-\min.MSE(tl_2)\ \geq\ 0,\Rightarrow\ S_y^4\xi\left[\frac{(\lambda_{04}-1)}{4}+(\lambda_{22}-1)\right]-\left[\frac{B_1C_1^2+A_1D^2-2C_1DE}{E^2-A_1B_1}\right]\ \geq\ 0,\quad (25)$$

$$V(t_{reg})-\min.MSE(tl_2)\ \geq\ 0,\quad\Rightarrow\ S_y^4\xi\left[\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]+\left[\frac{B_1C_1^2+A_1D^2-2C_1DE}{E^2-A_1B_1}\right]\ \leq\ 0,\quad (26)$$

$$\min.MSE\ (t_{ss})-\min.MSE(tl_2)\ \geq\ 0,$$

$$\Rightarrow\ \frac{S_y^4}{D_1}\left[(D_1-1)-\frac{2\xi(\lambda_{22}-1)^2}{(\lambda_{04}-1)D_1}\right]-C_1-\left[\frac{B_1C_1^2+A_1D^2-2C_1DE}{E^2-A_1B_1}\right]\ \geq\ 0,\quad (27)$$

$$\min.MSE(t_4)-\min.MSE(tl_2)\ \geq\ 0,\quad\Rightarrow\ S_y^4\xi\left[\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]+\left[\frac{B_1C_1^2+A_1D^2-2C_1DE}{E^2-A_1B_1}\right]\ \leq\ 0.\quad (28)$$

6. Empirical Study

Here, we have considered three data sets to examine the performance of the proposed estimator for the population mean square of the study variable.

Population 1 (Cochran 1977, p.325)
X : number of rooms per block, Y : number of persons per block,
$N=100,\ n=10,\ S_y^2=214.69,\ S_x^2=56.76,\ \lambda_{40}=2.2387,\ \lambda_{04}=2.2523,\ \lambda_{22}=1.5432,\ \rho=0.6515.$

Population 2 (Cochran 1977, p.203)
X : eye estimate of weight of peaches on each tree, Y : actual weight of peaches on each tree,
$N=200,\ n=10,\ S_y^2=99.18,\ S_x^2=85.09,\ \lambda_{40}=1.9249,\ \lambda_{04}=2.5932,\ \lambda_{22}=2.1149,\ \rho=0.9937.$

Population 3 (Sukhatme and Sukhatme 1970, p.185)
X : wheat acreage in 1936, Y : wheat acreage in 1937.
$N=170,\ n=10,\ S_y^2=26456.89,\ S_x^2=22355.76, \lambda_{40}=3.1842,\ \lambda_{04}=2.2030,\ \lambda_{22}=2.5597,$
$\rho=0.977.$

Table 2 PRE's of the estimators

Estimators	Population 1	Population 2	Population 3
t_0	100.000	100.000	100.000
t_1	88.164	320.812	815.609
t_2	122.793	444.023	236.066
t_3	59.133	37.935	54.002
t_4	123.468	639.150	1347.978
t_{reg}	123.468	639.150	1347.978
t_{ss}	132.561	647.931	1368.535
tl_1	123.468	639.150	1347.978
tl_2	142.116	912.013	2471.134

Table 2 clearly shows that the estimator tl_1 performs better than usual unbiased estimator t_0, Isaki's estimator t_1 and Singh et al.'s estimators t_2 and t_3 and equally efficient to the Yadav and

Kadilar's estimator t_4 and Isaki's regression type estimator t_{reg}. On the other hand the proposed estimator tl_2 is more efficient than the existing estimators t_0, t_1, t_2, t_3, t_{reg}, t_{ss}, and t_4 as well as the proposed estimator tl_1.

7. Simulation Study

This section shows the study based on simulated data to check and compare the performance of the proposed estimators with the existing estimators. For simulation study, we have use the algorithm proposed by Reddy et al. (2010). The steps of the algorithm are as follows.

Step 1: Generate $X \sim N(5,3)$ and $X' \sim N(5,3)$ for Population-I and $X \sim N(3,2)$ and $X' \sim N(5,3)$ for Population-II using the method of Box-Muller.

Step 2: Let $Y = rX + \sqrt{(1+r^2)}X'$ such that $0 < r = 0.5, 0.6, 0.8 < 1$.

Step 3: Gives the pair (Y, X).

Step 4: For both the data sets I and II repeat Steps 2 to 3 for 2,000 times. Variable Y and X will have the same variances in the population-I and different variances in the population-II.

Step 5: Use simple random sampling (SRS) to draw 1,000 samples (y_i, x_i), for i = 1, 2,…, n from the population of size N = 2000, without replacement (WOR) of size n = 40, 50 and 60.

Step 6: Calculate $AMSE(t) = \frac{1}{1,000}\sum_{k=1}^{1,000} E(t_k - \bar{y})^2$ and PRE of an estimator t with respect to the usual estimator $\bar{y}$ is $PRE(t) = \frac{Var(\bar{y}) \times 100}{MSE(t)}$.

The obtained results are as follows in Table 3.

Table 3 Comparison of proposed estimator tl_1 and tl_2 with other existing estimators for Population-I

r	n	Average percent relative efficiency (PRE)								
		s_y^2	t_1	t_2	t_3	t_{reg}	t_{ss}	t_4	tl_1	tl_2
0.5	40	100.00	65.12	101.20	63.65	108.26	112.88	108.26	108.26	114.55
	50	100.00	65.09	101.21	63.61	108.26	111.96	108.26	108.26	113.31
	60	100.00	65.10	101.23	63.60	108.27	111.35	108.27	108.27	112.48
0.6	40	100.00	77.73	115.25	59.40	117.82	122.06	117.82	117.82	125.16
	50	100.00	77.72	115.29	59.36	117.85	121.23	117.85	117.85	123.74
	60	100.00	77.72	115.29	59.36	117.84	120.66	117.84	117.84	122.75
0.8	40	100.00	143.95	172.37	51.17	179.17	173.60	179.17	179.17	192.12
	50	100.00	144.06	172.54	51.14	179.35	174.59	179.35	179.35	189.74
	60	100.00	143.96	172.46	51.15	179.23	175.12	179.23	179.23	187.88

Table 4 Comparison of proposed estimator tl_1 and tl_2 with other existing estimators for Population-II

r	n	Average percent relative efficiency (PRE)								
		s_y^2	t_1	t_2	t_3	t_{reg}	t_{ss}	t_4	tl_1	tl_2
0.5	40	100.00	54.75	88.66	69.31	102.25	106.89	102.25	102.25	107.61
	50	100.00	54.71	88.64	69.28	102.24	105.95	102.24	102.24	106.54
	60	100.00	54.72	88.66	69.26	102.24	105.34	102.24	102.24	105.84
0.6	40	100.00	60.49	95.72	65.84	105.28	109.93	105.28	105.28	111.17
	50	100.00	60.45	95.71	65.80	105.28	109.00	105.28	105.28	110.00
	60	100.00	60.47	95.73	65.79	105.28	108.38	105.28	105.28	109.22
0.8	40	100.00	90.15	127.94	56.68	128.44	131.85	128.44	128.44	136.78
	50	100.00	90.16	128.01	56.64	128.49	131.20	128.49	128.49	135.19
	60	100.00	90.15	128.00	56.64	128.47	130.71	128.47	128.47	134.05

From the above Tables 3 and 4, we observe that the average PRE's of the estimators increasing with an increase of sample size n and the value of correlation coefficient r for both the simulated data sets I and II. The proposed estimator tl_1 is equally efficient to the Isaki regression estimator t_{reg} and Yadav and Kadilar estimator t_4. The Proposed estimator tl_1 performs better than the existing estimators t_0, t_1, t_2 and t_3, as the average PRE of the proposed estimator tl_1 is higher comparatively. In addition, the average PRE of the estimator 'tl_2' is higher as compared to the existing estimators t_0, t_1, t_2, t_3, t_{reg}, t_{ss}, and t_4 as well as the proposed estimator tl_1. Which reflects that the estimator tl_2 will perform better compared to the proposed estimator tl_1 as well as the existing estimators t_0, t_0, t_1, t_2, t_3, t_{reg}, t_{ss}, and t_4.

8. Conclusions

This paper proposed log product type estimators for estimating population variance under SRSWOR using auxiliary information. We have check the pperformance of the proposed estimators by using three real population data sets as well as simulated data sets. From Tables 2-4, it can be seen that the proposed estimator tl_2 performs better as compared to the existing estimators t_0, t_1, t_2, t_3, t_{reg}, t_{ss}, and t_4 as well as the proposed estimator tl_1 for both real and simulated data sets. Hence it can be recommended for use in practice.

References

Adichwal NK, Kumar J, Singh R. An improved generalized class of estimators for population variance using auxiliary variables. Cogent Math Stat. 2018; 5(1): 1-8.

Cochran WG. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. J Agric Sci. 1940; 30(2): 262-275.

Cochran WG. Sampling Techniques. New York: John Wiley & Sons; 1977.

Das AK, Tripathi TP. Use of auxiliary information in estimating the finite population variance. Sankhya. 1978; 40(2): 139-148.

Gupta S, Shabbir J. Variance estimation in simple random sampling using auxiliary information. Hacet J Math Stat. 2008; 37(1): 57-67.

Isaki CT. Variance estimation using auxiliary information. J Am Stat Assoc. 1983; 78(381): 117-123.

Kumar A, Saini M. Exponential estimators for finite population mean using auxiliary attribute: A predictive approach. Invest Oper. 2020; 41(7): 999-1009.

Kumar M, Singh R, Sawan N, Chauhan P. Exponential ratio method of estimation in the presence of measurement errors. Int J Agric Stat Sci. 2011; 7(2): 457-461.

Mishra P, Adichwal NK, Singh R. A new log-product-type estimator using auxiliary information. J Sci Res. 2017; 61(1, 2): 179-183.

Reddy MK, Rao KR, Boiroju NK. Comparison of ratio estimators using Monte Carlo simulation. Int J Agric Stat Sci. 2010; 6(2): 517-527.

Saini M, Kumar A. Estimators of population mean using information on auxiliary attribute. J Appl Probab Stat. 2020; 15(3): 17-26.

Shabbir J. A dual to variance ratio-type estimator in simple random sampling. Proc Pakistan Acad Sci. 2006; 43(4): 279-283.

Sharma P, Singh R. Improved dual to variance ratio type estimators for population variance. Chil J Stat. 2014; 5(2): 45-54.

Singh HP, Solanki RS. A new procedure for variance estimation in simple random sampling using auxiliary information. Stat Pap. 2013; 54(2): 479-497.

Singh R, Chauhan P, Sawan N, Smarandache F. Improved exponential estimator for population variance using two auxiliary variables. Ital J Pure Appl Math. 2011; 28: 101-108.

Subramani J, Kumarapandiyan G. A class of modified ratio estimators for estimation of population variance. J Appl Math Stat Inform. 2015; 11(1): 91-114.

Sukhatme PV, Sukhatme BV. Sampling theory of surveys with applications. Iowa: The Iowa State College Press; 1970.

Upadhyaya LN, Singh HP. Estimation of the population standard deviation using auxiliary information. AM J Math-S. 2001; 21(3-4): 345-358.

Yadav SK, Kadilar C. Improved exponential type ratio estimator of population variance. Rev Colomb Estad. 2013; 36(1): 145-152.