ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 42(2022) 105-122

On Estimation of Population Variance Using Ratio and Exponential Ratio Estimators

K. B. Panda and P. P. Mohanty
[Received on April, 2022 Accepted on November, 2022]

ABSTRACT

Invoking ratio and exponential ratio estimators for population variance, we have proposed three ratio-cum-exponential ratio estimators for estimating population variance. The estimators thus proposed are found to perform better than the competing estimators under conditions very likely to hold good in practice, as evidenced from the empirical investigations carried out in the paper. Analogous results are drawn when the study is extended to two-phase sampling.

1. Introduction

Use of auxiliary information in estimation stage has been a common phenomenon in survey sampling. A lot of research has so far been carried out in arriving at more precise estimators for population mean using ratio, product and regression methods of estimation where availability of auxiliary information is a precondition. Variance estimation is one of the major issues in survey sampling. It helps us to have a tentative idea about how the observations constituting a finite population vary among themselves. Besides this, variance estimates are used to build confidence intervals for estimating population mean. Two data sets with the same mean may vary considerably and variance is a means to quantify the heterogeneity.

There are several estimators constructed resorting to ratio, product and regression methods of estimation to address the problem of estimating population variance. Along the lines followed in ratio estimation for estimating population mean, Isaki (1983) used auxiliary information in the form of population variance of the auxiliary variate at estimation stage for the estimation of population variance. With a view to improving upon the estimator in terms of efficiency, Singh et.al. (2011) introduce a new concept called exponential method of estimation. It is of

: P. P. Mohanty
Email: mprajna96@gmail.com

Extended authors information available after reference list of the article.

interest to note that Bhusan *et al.* (2021 & 2022) make use of attributes in estimating population variance. Recently, Panda and Chattapadhyay (2022) have followed Singh (1967) in their work. Invoking the technique due to Singh (1967) and Kadilar (2016) into ratio and exponential ratio estimators, we have come up with the proposed ratio-cum-exponential ratio estimators in general forms introducing a preassigned constant.

Consider a finite population of size N, arbitrarily labelled 1, …, N. Let y_i and x_i be the value of the study variable y and the auxiliary variable x, respectively, for ith unit in the population. Now assume that the problem is to estimate the population variance S_y^2, where $S_y^2 = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^2$. With a view to estimating the variance of a finite population, Isaki (1983) proposed the following estimator:

$$t_1 = s_y^2 \frac{S_x^2}{s_x^2}, \qquad (1)$$

where the symbols have their usual meanings. Nevertheless, the details of the mathematical expressions, assumptions and the expected values of various error terms are discussed in the Appendix.

We derive the expressions for the Bias and Mean Square Error (MSE), up to $O(n^{-1})$, as

$$B(t_1) = \frac{S_y^2}{n}\{(\lambda_{04} - 1) - (\lambda_{22} - 1)\} \qquad (1.1)$$

and

$$M(t_1) = \frac{S_y^4}{n}\{(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)\}. \qquad (1.2)$$

It may be noted here that s_y^2 is an unbiased estimator of S_y^2 with

$$V\left(s_y^2\right) = \frac{S_y^4}{n}(\lambda_{40} - 1). \qquad (1.3)$$

We consider below the estimator proposed by Singh et.al. (2011) for estimating S_y^2 as

$$t_2 = s_y^2 exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right), \qquad (1.4)$$

whose Bias and MSE, up to $O(n^{-1})$, are given by

$$B(t_2) = \frac{S_y^2}{n}\left\{\frac{3}{8}(\lambda_{04} - 1) - \frac{1}{2}(\lambda_{22} - 1)\right\} \qquad (1.5)$$

and $$M(t_2) = \frac{S_y^4}{n}\left\{(\lambda_{40} - 1) + \frac{1}{4}(\lambda_{04} - 1) - (\lambda_{22} - 1)\right\}. \tag{1.6}$$

2. The Proposed Ratio-Cum-Exponential Ratio Estimators

Following Singh et. al. (2011), Kadilar (2016), Rather et. al. (2022) and Panda & Chattapadhyay (2022) we have, in this paper, come up with the following estimators for S_y^2 given by

$$t_3 = s_y^2\left(\frac{S_x^2}{s_x^2}\right)^{\alpha_1} exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right), \tag{2.1}$$

$$t_4 = s_y^2\frac{S_x^2}{s_x^2} exp\left(\frac{\alpha_2(S_x^2 - s_x^2)}{S_x^2 + s_x^2}\right) \tag{2.2}$$

and

$$t_5 = s_y^2\frac{S_x^2}{s_x^2} exp\left(\frac{S_x^2 - s_x^2}{\alpha_3(S_x^2 + s_x^2)}\right), \tag{2.3}$$

where α_1, α_2 and α_3 are any three constants.

Substituting the values of e_0, e_1 in the expression (2.1), we get

$$t_3 = S_y^2(1 + e_0)\left(\frac{S_x^2}{{S_x}^2(1 + e_1)}\right)^{\alpha_1} exp\left(\frac{S_x^2 - S_x^2(1 + e_1)}{{S_x}^2 + {S_x}^2(1 + e_1)}\right),$$

$$= S_y^2(1 + e_0)(1 + e_1)^{-\alpha_1} exp\left(\frac{-e_1}{2 + e_1}\right).$$

Retaining terms only up to 2nd degree, we find that

$$t_3 = S_y^2(1 + e_0)\left(1 - \alpha_1 e_1 + \frac{\alpha_1(\alpha_1 + 1)}{2}e_1^2\right) exp\left(\frac{-e_1}{2}\left(1 + \frac{e_1}{2}\right)^{-1}\right),$$

$$= S_y^2\left(1 + e_0 - \alpha_1 e_1 - \alpha_1 e_0 e_1 + \frac{\alpha_1(\alpha_1 + 1)}{2}e_1^2\right)\left(1 - \frac{e_1}{2} + \frac{3}{8}e_1^2\right),$$

$$= S_y^2\left[1 + e_0 - \left(\alpha_1 + \frac{1}{2}\right)e_1 - \left(\alpha_1 + \frac{1}{2}\right)e_0 e_1 + \left(\alpha_1^2 + 2\alpha_1 + \frac{3}{4}\right)\frac{e_1^2}{2}\right].$$

The Bias of t_3, to $O(n^{-1})$, is

$$B(t_3) = E(t_3) - S_y^2$$

$$= S_y^2 E\left[1 + e_0 - \left(\alpha_1 + \frac{1}{2}\right)e_1 - \left(\alpha_1 + \frac{1}{2}\right)e_0 e_1 + \left(\alpha_1^2 + 2\alpha_1 + \frac{3}{4}\right)\frac{e_1^2}{2}\right] - S_y^2,$$

$$= S_y^2 E\left[\left(\alpha_1^2 + 2\alpha_1 + \frac{3}{4}\right)\frac{e_1^2}{2} - \left(\alpha_1 + \frac{1}{2}\right)e_0 e_1\right],$$

$$\Rightarrow\ B(t_3) = \frac{S_y^2}{n}\left[\frac{1}{2}\left(\alpha_1^2 + 2\alpha_1 + \frac{3}{4}\right)(\lambda_{04} - 1) - \left(\alpha_1 + \frac{1}{2}\right)(\lambda_{22} - 1)\right]. \qquad (2.4)$$

The MSE of t_3, to $O(n^{-1})$, is found to be

$$M(t_3) = E\left(t_3 - S_y^2\right)^2$$

$$= E\left[S_y^2\left\{1 + e_0 - \left(\alpha_1 + \frac{1}{2}\right)e_1 - \left(\alpha_1 + \frac{1}{2}\right)e_0 e_1 + \left(\alpha_1^2 + 2\alpha_1 + \frac{3}{4}\right)\frac{e_1^2}{2}\right\} - S_y^2\right]^2,$$

$$= E\left[S_y^2\left\{1 + e_0 - \left(\alpha_1 + \frac{1}{2}\right)e_1\right\} - S_y^2\right]^2,$$

$$= S_y^4 E\left[e_0^2 + \left(\alpha_1 + \frac{1}{2}\right)^2 e_1^2 - 2\left(\alpha_1 + \frac{1}{2}\right)e_0 e_1\right]$$

$$\Rightarrow\ M(t_3) = \frac{S_y^4}{n}\left[(\lambda_{40} - 1) + \left(\alpha_1 + \frac{1}{2}\right)^2(\lambda_{04} - 1) - 2\left(\alpha_1 + \frac{1}{2}\right)(\lambda_{22} - 1)\right]. \qquad (2.5)$$

To find the optimum value of α_1, we proceed as follows:

$$\frac{\partial MSE(t_3)}{\partial \alpha_1} = 0$$

$$\Rightarrow \alpha_{1_{opt}} = \left[\frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)} - \frac{1}{2}\right].$$

Putting the optimum value of $\alpha_{1_{opt}}$ in (2.5), we get the minimum mean square error, as

$$MSE(t_3)_{min} = \frac{S_y^4}{n}\left[(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right]. \qquad (2.6)$$

Now, expressing (2.2) in terms of e's, we have

$$t_4 = S_y^2(1+e_0)\left(\frac{S_x^2}{{S_x}^2(1+e_1)}\right)exp\left[\frac{\alpha_2\{S_x^2 - S_x^2(1+e_1)\}}{\{{S_x}^2 + {S_x}^2(1+e_1)\}}\right]$$

$$= S_y^2(1+e_0)(1+e_1)^{-1}exp\left\{\alpha_2\left(\frac{-e_1}{2+e_1}\right)\right\}$$

$$= S_y^2(1+e_0)(1-e_1+e_1^2)exp\left\{\frac{-\alpha_2 e_1}{2}\left(1+\frac{e_1}{2}\right)^{-1}\right\}$$

$$= S_y^2\left(1+e_0-e_1+{e_1}^2-e_0e_1-\alpha_2\frac{e_1}{2}-\alpha_2\frac{e_0e_1}{2}+\alpha_2\frac{{e_1}^2}{2}+\alpha_2\frac{{e_1}^2}{4}+\alpha_2^2\frac{{e_1}^2}{8}\right).$$

The bias of t_4, to $\mathrm{O(n^{-1})}$, is

$$B(t_4) = E(t_4) - S_y^2,$$

$$= S_y^2E\left(1+e_0-e_1+{e_1}^2-e_0e_1-\alpha_2\frac{e_1}{2}-\alpha_2\frac{e_0e_1}{2}+\alpha_2\frac{{e_1}^2}{2}+\alpha_2\frac{{e_1}^2}{4}+\alpha_2^2\frac{{e_1}^2}{8}\right)-S_y^2$$

$$= S_y^2E\left({e_1}^2-e_0e_1-\alpha_2\frac{e_0e_1}{2}+\alpha_2\frac{3}{4}{e_1}^2+\alpha_2^2\frac{{e_1}^2}{8}\right)$$

$$= \frac{S_y^2}{n}\left[(\lambda_{04}-1)-(\lambda_{22}-1)-\frac{\alpha_2}{2}(\lambda_{22}-1)+\alpha_2\frac{3}{4}(\lambda_{04}-1)+\alpha_2^2\frac{1}{8}(\lambda_{04}-1)\right]$$

$$\Rightarrow B(t_4) = \frac{S_y^2}{n}\left[(\lambda_{04}-1)\left(1+\tfrac{3}{4}\alpha_2+\tfrac{1}{8}\alpha_2^2\right)-(\lambda_{22}-1)\left(1+\tfrac{\alpha_2}{2}\right)\right]. \quad (2.7)$$

The MSE of t_4, to $\mathrm{O(n^{-1})}$, is

$$M(t_4) = E\left(t_4 - S_y^2\right)^2.$$

$$= E\left[S_y^2\left(e_0-e_1-\alpha_2\frac{e_1}{2}\right)-S_y^2\right]^2$$

$$= S_y^4E\left(e_0^2+e_1^2+\frac{\alpha_2^2}{4}{e_1}^2-2e_0e_1+\alpha_2e_1^2-\alpha_2e_0e_1\right)$$

$$\Rightarrow M(t_4) = \frac{S_y^4}{n}\Big[(\lambda_{40}-1)+(\lambda_{04}-1)+\frac{\alpha_2^2}{4}(\lambda_{04}-1)-2(\lambda_{22}-1)+$$
$$\alpha_2(\lambda_{04}-1)-\alpha_2(\lambda_{22}-1)\Big] \quad (2.8)$$

The above MSE is minimized when:

$$\frac{\partial MSE(t_4)}{\partial \alpha_2} = 0$$

$$\Rightarrow \alpha_{2_{opt}} = 2\left[\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1\right].$$

The minimum MSE of the proposed estimator t_4 for the optimum value of α_2 is expressed as:

$$MSE(t_4)_{min} = \frac{S_y^4}{n}\left[(\lambda_{40}-1) - \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]. \quad (2.9)$$

Similarly, expressing (2.3) in terms of e's, we have

$$t_5 = S_y^2(1+e_0)\left(\frac{S_x^2}{{S_x}^2(1+e_1)}\right) exp\left[\frac{\{S_x^2 - S_x^2(1+e_1)\}}{\alpha_3\{S_x^2 - S_x^2(1+e_1)\}}\right]$$

$$= S_y^2(1+e_0)(1+e_1)^{-1} exp\left\{\frac{1}{\alpha_3}\left(\frac{-e_1}{2+e_1}\right)\right\}$$

$$= S_y^2\left(1 + e_0 - e_1 + {e_1}^2 - e_0e_1 - \frac{1}{\alpha_3}\frac{e_1}{2} - \frac{1}{\alpha_3}\frac{e_0e_1}{2} + \frac{1}{\alpha_3}\frac{{e_1}^2}{2} + \frac{1}{\alpha_3}\frac{{e_1}^2}{4} + \frac{1}{\alpha_3^2}\frac{{e_1}^2}{8}\right).$$

The bias of t_5, to $\mathrm{O(n^{-1})}$, is

$$B(t_5) = E(t_5) - S_y^2$$

$$= S_y^2 E\left(1 + e_0 - e_1 + {e_1}^2 - e_0e_1 - \frac{1}{\alpha_3}\frac{e_1}{2} - \frac{1}{\alpha_3}\frac{e_0e_1}{2} + \frac{1}{\alpha_3}\frac{{e_1}^2}{2} + \frac{1}{\alpha_3}\frac{{e_1}^2}{4} + \frac{1}{\alpha_3^2}\frac{{e_1}^2}{8}\right) - S_y^2$$

$$= S_y^2 E\left({e_1}^2 - e_0e_1 - \frac{1}{\alpha_3}\frac{e_0e_1}{2} + \frac{1}{\alpha_3}\frac{3}{4}{e_1}^2 + \frac{1}{\alpha_3^2}\frac{{e_1}^2}{8}\right)$$

$$= \frac{S_y^2}{n}\left[(\lambda_{04}-1) - (\lambda_{22}-1) - \frac{1}{\alpha_3}\frac{(\lambda_{22}-1)}{2} + \frac{1}{\alpha_3}\frac{3(\lambda_{04}-1)}{4} + \frac{1}{\alpha_3^2}\frac{(\lambda_{04}-1)}{8}\right]$$

$$\Rightarrow B(t_5) = \frac{S_y^2}{n}\left[(\lambda_{04}-1)\left(1 + \frac{3}{4\alpha_3} + \frac{1}{8\alpha_3^2}\right) - (\lambda_{22}-1)\left(1 + \frac{1}{2\alpha_3}\right)\right] \quad (2.10)$$

and its MSE, to $\mathrm{O(n^{-1})}$, is

$$M(t_5) = E\left(t_5 - S_y^2\right)^2$$

$$= E\left[S_y^2\left(e_0 - e_1 - \frac{1}{\alpha_3}\frac{e_1}{2}\right) - S_y^2\right]^2$$

$$= S_y^4 E\left(e_0^2 + e_1^2 + \frac{{e_1}^2}{4\alpha_3^2} - 2e_0e_1 - \frac{e_0e_1}{\alpha_3} + \frac{e_1^2}{\alpha_3}\right)$$

$$\Rightarrow M(t_5) = \frac{S_y^4}{n}\left[(\lambda_{40} - 1) + (\lambda_{04} - 1) + \frac{1}{4\alpha_3^2}(\lambda_{04} - 1) - 2(\lambda_{22} - 1) - \frac{(\lambda_{22}-1)}{\alpha_3} + \frac{(\lambda_{04}-1)}{\alpha_3}\right]. \quad (2.11)$$

The minimum MSE can be obtained when:

$$\frac{\partial MSE(t_5)}{\partial \alpha_3} = 0$$

$$\Rightarrow \alpha_{3_{opt}} = \frac{1}{2\left[\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1\right]}.$$

The minimum MSE of the proposed estimator t_5 for the optimum value of α_3 can be expressed as:

$$MSE(t_5)_{min} = \frac{S_y^4}{n}\left[(\lambda_{40} - 1) - \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]. \quad (2.12)$$

Surprisingly, (2.6), (2.9), and (2.12) show that all the three proposed estimators are, under optimality of $\alpha's$, found to be equally efficient. Moreover, the minimum MSE coincides with the minimum MSE of the regression estimator for population variance, i.e., when the regression coefficient coincides with $\frac{S_y^2}{S_x^2}\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)}$.

Given the above findings, when conditions are such that either the ratio estimator or the regression estimator can be used for estimating the population variance, one can choose any one of the three proposed exponential type estimators or the customary regression estimator subject to one's convenience. It is also of interest to note that the minimum absolute bias of all the four estimators, three exponential type and the other regression, remains the same and is given by

$$|B_{min}| = \left|\frac{S_y^2}{n}\frac{(\lambda_{22} - 1)}{2}\left\{1 - \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)}\right\}\right|.$$

3. Efficiency Comparison

I. Between the customary ratio-type estimator t_1 due to Isaki(1983) and s_y^2, the former is more efficient than the latter iff

$$M(t_1) - V(s_y^2) < 0$$

$$\Rightarrow \frac{S_y^4}{n}\{(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)\} - \frac{S_y^4}{n}(\lambda_{40} - 1) < 0$$

$$\Rightarrow \frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} > \frac{1}{2}. \tag{3.1}$$

II. The ratio-type exponential estimator t_2 performs better than the simple variance estimator iff

$$M(t_2) - V(s_y^2) < 0$$

$$\Rightarrow \frac{S_y^4}{n}\left\{(\lambda_{40} - 1) + \frac{1}{4}(\lambda_{04} - 1) - (\lambda_{22} - 1)\right\} - \frac{S_y^4}{n}(\lambda_{40} - 1) < 0$$

$$\Rightarrow \frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} > \frac{1}{4}. \tag{3.2}$$

III. The ratio-type exponential estimator t_2 is better than t_1 iff

$$M(t_2) - M(t_1) < 0$$

$$\Rightarrow \frac{S_y^4}{n}\left\{(\lambda_{40} - 1) + \frac{1}{4}(\lambda_{04} - 1) - (\lambda_{22} - 1)\right\} - \frac{S_y^4}{n}\{(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)\} < 0$$

$$\Rightarrow \frac{S_y^4}{n}\left\{\frac{1}{4}(\lambda_{04} - 1) - (\lambda_{22} - 1)\right\} - \frac{S_y^4}{n}\{(\lambda_{04} - 1) - 2(\lambda_{22} - 1)\} < 0$$

$$\Rightarrow \frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} < \frac{3}{4}. \tag{3.3}$$

The proposed estimators t_3, t_4 and t_5 , which are of same optimum MSE, are better than the existing estimators if the following conditions hold good:

IV. The estimator t_3 performs better than s_y^2 iff

$$M(t_3)_{min} - V(s_y^2) < 0$$

$$\Rightarrow \frac{S_y^4}{n}\left[(\lambda_{40}-1) - \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right] - \frac{S_y^4}{n}(\lambda_{40}-1) < 0$$

$$\Rightarrow \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} > 0$$

$$\Rightarrow (\lambda_{22}-1) > 0. \tag{3.4}$$

V. Between the estimator t_3 and t_1, the former is found to be more efficient iff

$$M(t_3)_{min} - M(t_1) < 0$$

$$\Rightarrow \frac{S_y^4}{n}\left[(\lambda_{40}-1) - \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right] - \frac{S_y^4}{n}\{(\lambda_{40}-1) + (\lambda_{04}-1) - 2(\lambda_{22}-1)\} < 0$$

$$\Rightarrow -\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} - (\lambda_{04}-1) + 2(\lambda_{22}-1) < 0$$

$$\Rightarrow \frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} > 1. \tag{3.5}$$

VI. When compared with the ratio-type exponential estimator t_2, the proposed estimator can fare better iff

$$M(t_3)_{min} - M(t_2) < 0$$

$$\Rightarrow \frac{S_y^4}{n}\left[(\lambda_{40}-1) - \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right] - \frac{S_y^4}{n}\left\{(\lambda_{40}-1) + \frac{1}{4}(\lambda_{04}-1) - (\lambda_{22}-1)\right\} < 0$$

$$\Rightarrow -\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)} - \frac{1}{4}(\lambda_{04}-1) + (\lambda_{22}-1) < 0$$

$$\Rightarrow [2(\lambda_{22}-1) - (\lambda_{04}-1)]^2 > 0$$

$$\Rightarrow \frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} > \frac{1}{2}. \tag{3.6}$$

4. Empirical Investigation

To have tangible idea about the percent relative efficiency of different estimators over simple variance estimator, we consider as many as ten natural population data sets, the description of which is given below in Table 4.1:

Table 4.1 : Population data sets.

Population	**N**	**n**	ρ_{yx}	λ_{40}	λ_{04}	λ_{22}
I: Gujarati [5] y: Telephone Ownership in Singapore x: Per capita GDP in Singapore	22	8	0.97	2.78	1.97	2.27
II: DNase [12] y: concentration of the protein x: optical density	176	20	0.93	3.78	1.92	2.42
III: Gujarati [5] y: U.S. Defense budget-outlay x: GNP	20	5	0.95	4.03	2.48	3.02
IV: SWISS [13] y: Education x: Examination	47	9	0.69	9.54	2.98	8.84
V: Gujarati [5] y: Real gross product x: Labor input	15	4	0.96	2.65	1.64	1.92
VI: Cochran [4] y: Number of inhabitants in 1930 x: Number of inhabitants in 1920	49	11	0.98	8.71	7.51	8.04
VII: Murthy [9] y: Cultivated area in 1964 x: Area under wheat 1963	34	6	0.98	3.72	2.90	3.10
VIII: Singh et al. [14] y: No of agriculture labourers for 1971 x: No of agriculture labourers for 1961	278	32	0.73	24.89	37.88	25.81
IX: Sukhatme and Sukhatme [17] y: Wheat acreage in 1937 x: Wheat acreage in 1936	170	25	0.97	3.18	2.20	2.56
X: Johnston [7] y: %age of hives affected by disease x: mean January temperature, ^{0}F	10	3	0.79	1.89	1.85	1.29

Table 4.2: $\left|\frac{\text{Bias}}{s_y^2/n}\right|$ of the competing estimators for optimum values of α_i's.

Population	*Estimators*			
	s_y^2	t_1	t_2	t_3, t_4, t_5
I	0	0.3014	0.2721	0.1975
II	0	0.5013	0.3662	0.3865
III	0	0.5384	0.4544	0.3670
IV	0	0.8513	0.6743	0.6079
V	0	0.2874	0.2235	0.2084
VI	0	0.5289	1.0784	0.2859
VII	0	0.1974	0.3369	0.1089
VIII	0	0.9173	3.4457	0.4763
IX	0	0.6245	0.0392	0.2229
X	0	0.5705	0.1778	0.0958

Table 4.3: $\left(\frac{MSE}{S_y^4/n}\right)$ of the competing estimators for optimum values of α_i's.

Population	*Estimators*			
	s_y^2	t_1	t_2	t_3, t_4, t_5
I	1.778	0.204	0.748	0.110
II	2.784	0.857	1.589	0.585
III	3.031	0.473	1.381	0.277
IV	8.541	4.849	6.198	4.485
V	1.653	0.439	0.887	0.310

VI	7.714	0.145	2.302	0.102
VII	2.726	0.425	1.099	0.404
VIII	36.889	11.158	18.049	11.123
IX	1.203	0.268	0.189	0.089
X	0.886	1.167	0.812	0.789

Table 4.4 : PRE of the competing estimators over s_y^2.

Estimator	**Population**			
	s_y^2	t_1	t_2	t_3, t_4, t_5
I	100	872.76	237.69	1614.215
II	100	324.73	175.14	475.13
III	100	641.48	219.43	1094.66
IV	100	176.12	137.80	190.42
V	100	375.73	186.37	532.46
VI	100	5318.24	335.13	7556.97
VII	100	641.35	248.04	673.77
VIII	100	330.60	204.38	331.65
IX	100	449.21	635.33	1374.97
X	100	75.865	109.108	112.30

From the above investigation, it is evident that relative efficiency remains invariant of sample size for a given population size.
Furthermore, we have computed the percent relative efficiency (PRE) of different estimators with respect to s_y^2 using the following formula:

$$PRE\left(t_i, s_y^2\right) = \frac{V\left(s_y^2\right)}{MSE(t_i)} \times 100, (i = 1,2,3,4,5).$$

5. Performance of Ratio-Type Estimators in Two-Phase Sampling

In certain practical situations, the population variance of the auxiliary variable S_x^2 is not known in advance. The usual procedure in such a case is to use the technique of two-phase or double-sampling. This technique consists of taking a larger sample of size n' by simple random sampling without replacement to estimate S_x^2 while a sub-sample of n out of n' units is drawn by Simple random sampling without replacement (SRSWOR) to observe the characteristic S_y^2 under study. Ratio-type estimators analogous to those considered in the previous section can be obtained by replacing $S_x^2\ by\ {s_x'}^2$, the sample variance based on n'units.

The estimators $t_1,\ t_2,\ t_3,\ t_4\ and\ t_5$ in two-phase sampling will take the following forms, respectively

$$t_{1d} = s_y^2\frac{{s_x'}^2}{s_x^2}, \tag{5.1}$$

$$t_{2d} = s_y^2 exp\left(\frac{{s_x'}^2 - s_x^2}{{s_x'}^2 + s_x^2}\right), \tag{5.2}$$

$$t_{3d} = s_y^2\left(\frac{{s_x'}^2}{s_x^2}\right)^{\alpha_1} exp\left(\frac{{s_x'}^2 - s_x^2}{{s_x'}^2 + s_x^2}\right), \tag{5.3}$$

$$t_{4d} = s_y^2\frac{{s_x'}^2}{s_x^2} exp\left(\frac{\alpha_2\left({s_x'}^2 - s_x^2\right)}{{s_x'}^2 + s_x^2}\right) \tag{5.4}$$

and

$$t_{5d} = s_y^2\frac{{s_x'}^2}{s_x^2} exp\left(\frac{{s_x'}^2 - s_x^2}{\alpha_3\left({s_x'}^2 + s_x^2\right)}\right), \tag{5.5}$$

The expressions for Bias and MSE of t_{1d} and t_{2d}, to $O(n^{-1})$, are found to be

$$B(t_{1d}) = S_y^2\left(\frac{1}{n} - \frac{1}{n'}\right)\{(\lambda_{04} - 1) - (\lambda_{22} - 1)\}, \tag{5.6}$$

$$M(t_{1d}) = S_y^4\left\{\frac{1}{n}(\lambda_{40} - 1) + \left(\frac{1}{n} - \frac{1}{n'}\right)(\lambda_{04} - 1) - 2\left(\frac{1}{n} - \frac{1}{n'}\right)(\lambda_{22} - 1)\right\}, \tag{5.7}$$

$$B(t_{2d}) = S_y^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left\{\frac{3}{8}(\lambda_{04} - 1) - \frac{1}{2}(\lambda_{22} - 1)\right\}, \tag{5.8}$$

$$M(t_{2d}) = S_y^4\left\{\frac{1}{n}(\lambda_{40}-1)+\frac{1}{4}\left(\frac{1}{n}-\frac{1}{n'}\right)(\lambda_{04}-1)-\left(\frac{1}{n}-\frac{1}{n'}\right)(\lambda_{22}-1)\right\}. \qquad (5.9)$$

The proposed estimator t_{3d} can be expressed in the form of e's as:

$$t_{3d} = S_y^2(1+e_0)\left(\frac{S_x^2(1+e_1')}{{S_x}^2(1+e_1)}\right)^{\alpha_1} exp\left(\frac{S_x^2(1+e_1')-S_x^2(1+e_1)}{S_x^2(1+e_1')+{S_x}^2(1+e_1)}\right)$$

$$= S_y^2(1+e_0)\left((1+e_1')(1+e_1)^{-1}\right)^{-\alpha_1} exp\left[\frac{(e_1'-e_1)}{2+(e_1'+e_1)}\right].$$

Retaining terms only up to 2nd degree, we find that

$$t_{3d} = S_y^2\left[1+e_0-\left(\alpha_1+\frac{1}{2}\right)e_1+\left(\alpha_1+\frac{1}{2}\right)e_1'-\left(\alpha_1+\frac{1}{2}\right)e_0e_1 + \left(\alpha_1+\frac{1}{2}\right)e_0e_1'-\left(\alpha_1^2+\alpha_1+\frac{1}{4}\right)e_1e_1'+\left(\alpha_1^2+2\alpha_1+\frac{3}{4}\right)\frac{e_1^2}{2} + \left(\alpha_1^2-\frac{1}{4}\right)\frac{{e_1'}^2}{2}\right].$$

The Bias of t_{3d}, to $O(n^{-1})$, is

$$B(t_{3d}) = E(t_{3d})-S_y^2$$

$$= S_y^2E\left[-\left(\alpha_1+\frac{1}{2}\right)e_0e_1+\left(\alpha_1+\frac{1}{2}\right)e_0e_1'-\left(\alpha_1^2+\alpha_1+\frac{1}{4}\right)e_1e_1' + \left(\alpha_1^2+2\alpha_1+\frac{3}{4}\right)\frac{e_1^2}{2}+\left(\alpha_1^2-\frac{1}{4}\right)\frac{{e_1'}^2}{2}\right]$$

$$= S_y^2\left[-\left(\alpha_1+\frac{1}{2}\right)\frac{(\lambda_{22}-1)}{n}+\left(\alpha_1+\frac{1}{2}\right)\frac{(\lambda_{22}-1)}{n'}-\left(\alpha_1^2+\alpha_1+\frac{1}{4}\right)\frac{(\lambda_{04}-1)}{n'} + \frac{1}{2}\left(\alpha_1^2+2\alpha_1+\frac{3}{4}\right)\frac{(\lambda_{04}-1)}{n}+\frac{1}{2}\left(\alpha_1^2-\frac{1}{4}\right)\frac{(\lambda_{04}-1)}{n'}\right]$$

$$= S_y^2\left(\frac{1}{n}-\frac{1}{n'}\right)\left[(\lambda_{04}-1)\frac{1}{2}\left(\alpha_1^2+2\alpha_1+\frac{3}{4}\right)-(\lambda_{22}-1)\left(\alpha_1+\frac{1}{2}\right)\right]. \qquad (5.10)$$

The MSE of t_{3d}, to $O(n^{-1})$, is found to be

$$M(t_{3d}) = E\left(t_{3d}-S_y^2\right)^2,$$

$$= S_y^4 E\left[e_0 - \left(\alpha_1 + \frac{1}{2}\right)e_1 + \left(\alpha_1 + \frac{1}{2}\right)e_1'\right]^2$$

$$= S_y^4\left[\frac{(\lambda_{40}-1)}{n} + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{(\lambda_{04}-1)\left(\alpha_1 + \frac{1}{2}\right)^2 - 2(\lambda_{22}-1)\left(\alpha_1 + \frac{1}{2}\right)\right\}\right]. \quad (5.11)$$

Minimizing equation (5.11) with respect to α_1, we can get $\alpha_{1_{opt}}$, i.e.,

$$\Rightarrow \alpha_{1_{opt}} = \left[\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} - \frac{1}{2}\right].$$

Similarly, the expressions for Bias and MSE of t_{4d} and t_{5d} to $O(n^{-1})$, are found to be

$$B(t_{4d}) = S_y^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left[(\lambda_{04}-1)\left(1 + \frac{3}{4}\alpha_2 + \frac{1}{8}\alpha_2^2\right) - (\lambda_{22}-1)\left(1 + \frac{\alpha_2}{2}\right)\right], \quad (5.12)$$

$$M(t_{4d}) = S_y^4\left[\frac{(\lambda_{40}-1)}{n} + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{(\lambda_{04}-1)\left(1 + \frac{\alpha_2}{2}\right)^2 - 2(\lambda_{22}-1)\left(1 + \frac{\alpha_2}{2}\right)\right\}\right]. \quad (5.13)$$

The above MSE can be minimized when:

$$\Rightarrow \alpha_{2_{opt}} = 2\left[\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1\right].$$

$$B(t_{5d}) = S_y^2\left(\frac{1}{n} - \frac{1}{n'}\right)\left[(\lambda_{04}-1)\left(1 + \frac{3}{4\alpha_3} + \frac{1}{8\alpha_3^2}\right) - (\lambda_{22}-1)\left(1 + \frac{1}{2\alpha_3}\right)\right], \quad (5.14)$$

$$M(t_{5d}) = S_y^4\left[\frac{(\lambda_{40}-1)}{n} + \left(\frac{1}{n} - \frac{1}{n'}\right)\left\{(\lambda_{04}-1)\left(1 + \frac{1}{2\alpha_3}\right)^2 - 2(\lambda_{22}-1)\left(1 + \frac{1}{2\alpha_3}\right)\right\}\right]. \quad (5.15)$$

The optimum value of α_3 is given by:

$$\Rightarrow \alpha_{3_{opt}} = \frac{1}{2\left[\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} - 1\right]}$$

Putting the optimum value of $\alpha_{i_{opt}}, (i = 1, 2, 3)$ respectively in equation (5.11), (5.13) and (5.15) we get the minimum mean square error as

$$M(t_{3d})_{min} = M(t_{4d})_{min} = M(t_{5d})_{\min} = S_y^4\left[\frac{(\lambda_{40}-1)}{n} - \left(\frac{1}{n} - \frac{1}{n'}\right)\frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right]. \quad (5.16)$$

From the above expressions, it is evident that the performance of the exponential type estimators in two-phase sampling remains the same as in one-phase sampling dealt with in the preceding section. The efficiency conditions are also exactly the same as in one-phase sampling.

6. Empirical Study for Two-Phase Sampling

For the purpose of practical application, the following data from a real population are considered:

Population 11: (Source: DNase: R-dataset package) [12]

y: a numeric vector giving the known concentration of the protein

x: a numeric vector giving the measured optical density (dimensionless) in the assay

$N = 176,\ n' = 65,\ \ n = 10\ , \bar{Y} = 3.106,\ \ \bar{X} = 0.719,\ \ \lambda_{40} = 3.784,\ \ \lambda_{04} = 1.924,\ \lambda_{22} = 2{,}425,\ \rho_{yx} = 0.9309.$

Table 6.1: Bias, MSE and PRE of the Competing Estimators for optimum values of α_i, (i = 1,2,3).

Estimator	$\left\lvert\frac{BIAS}{S_y^2}\right\rvert$	$\frac{MSE}{S_y^4}$	PRE
s_y^2	0.0000	0.2784	100
t_{1d}	0.0424	0.1153	241.3022
t_{2d}	0.0309	0.1773	156.9906
t_{3d}, t_{4d}, t_{5d}	0.0327	0.0923	301.3778

It is evident from the above Table that the estimators proposed under two-phase sampling outperform their competing estimators in terms of efficiency, bias being ignored.

7. Conclusion

In survey sampling, use of auxiliary information in the improvement of estimation of parameters is a regular phenomenon. We have, in this paper, come up with three structurally different ratio-cum-exponential ratio estimators for population variance. Coincidentally, the three proposed estimators are found to have the same minimum mean square error both under one-phase and two-phase sampling. Numerical investigation based on as many as ten real populations is in agreement with theoretical observations so far as one-phase sampling is concerned. Another interesting feature is the minimum MSE of the proposed estimators is no different from the minimum MSE due to the already existing regression estimator. When conditions permit use of either ratio type or regression estimator, one finds a choice among the four estimators, three exponential ratio type and the regression, for practical purpose. It may be noted here that the minimum absolute biases of these three proposed estimators coincide and is also found to be equal to that of the regression estimator.

Appendix

We consider $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ and $s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ as the sample variances of y and x, respectively, and $S_x^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{X})^2$, the population variance of x.

The following assumptions on error terms and expected values are due to Singh et.al., (2011):

In one-phase sampling,

$$e_0 = \frac{s_y^2 - S_y^2}{S_y^2}, i.e., s_y^2 = S_y^2(1 + e_0)$$

and
$$e_1 = \frac{s_x^2 - S_x^2}{S_x^2}, i.e., s_x^2 = S_x^2(1 + e_1),$$

such that
$$E(e_0) = E(e_0) = 0, E({e_0}^2) = \frac{1}{n}(\lambda_{40} - 1),$$
$$E({e_1}^2) = \frac{1}{n}(\lambda_{04} - 1), E(e_0 e_1) = \frac{1}{n}(\lambda_{22} - 1),$$

where $\lambda_{pq} = \frac{\mu_{pq}}{\mu_{20}^{p/2}\mu_{02}^{q/2}}$ and $\mu_{pq} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^p (x_i - \bar{X})^q$, (p, q) being non-negative integers.

In two-phase sampling,

$$e_1' = \frac{{s_x'}^2 - S_x^2}{S_x^2}, i.e., {s_x'}^2 = S_x^2(1 + e_1'),$$

where
$${s_x'}^2 = \frac{1}{n'-1}\sum_{i=1}^{n'}(x_i - \bar{x}')^2 \text{ and } \bar{x}' = \frac{1}{n'}\sum_{i=1}^{n'} x_i.$$

Also
$$E(e_i') = 0, \qquad E\left({e_1'}^2\right) = \frac{1}{n'}(\lambda_{40} - 1),$$
$$E(e_0 e_1') = \frac{1}{n'}(\lambda_{22} - 1), \qquad E(e_1 e_1') = \frac{1}{n'}(\lambda_{22} - 1),$$

where λ_{pq} and μ_{pq} remain the same as in one-phase sampling.

Acknowledgement

The authors are highly grateful to the reviewers for their valuable suggestions leading to an improvement in the overall quality of the paper.

References

Bahl, S. and Tuteja, R. K. (1991): Ratio and product type exponential estimator, *Journal of Information and Optimization Sciences*, **12**, 159–163.

Bhushan, S., Kumar, A. and Kumar, S. (2021): Efficient classes of estimators for population variance using attribute. *International Journal of Mathematics in Operation Research*. Vol.**22, No.1**. DOI:10.1504/IJMOR.2021. 10040442.

Bhushan, S., Kumar, A., Kumar, S. and Singh, S. (2022): Some modified classes of estimators for population variance using auxiliary attribute. *Pakistan Journal of Statistics*, **38(2),** 235-252.

Cochran, W.G. (1977): Sampling Techniques (3rd ed.), New York: John Wiley, pp−152.

Gujarati, D. (1995): Basic Econometrics, Mc-Graw Hill, India, International Editions, pp-92,227.
Isaki, C.T. (1983): Variance estimation using auxiliary information, *Journal of the American Statistical Association* **78,** pp-117−123.
Johnston, J. (1960): Econometric Methods, Mc-Graw Hill, New York, International Student Edition, pp-399−400.
Kadilar, G. O. (2016): A new exponential type estimator for the population mean in simple random sampling. *Journal of Modern Applied Statistical Methods*., **15**, 207-214.
Murthy, M. N. (1977): Sampling Theory and Methods, Tokyo, Japan, pp-399.
Panda, K. B. and Chattopadhyay, G. (2022): On Estimation of Finite Population Mean Using Variants of Ratio and Product Estimators, Dogo Rangsang Research Journal, Vol-**12**, Issue-**06** No**. 03** June, ISSN: 2347-7180.
Rather, K. U. I., Jeelani, M. I., Rizvi, S. E. H. and Sharma, M. (2022): A New Exponential Ratio Type Estimator for the Population Means Using Information on Auxiliary Attribute. *Journal of Statistics Applications & Probability Letters 9*, No. **1**, 31-42.
R-dataset Package. DNase: The DNase data frame has 176 rows and 3 columns of data obtained during development of an ELISA assay for the recombinant protein DNase in rat serum.
R-dataset Package. Swiss (1888): Standardized fertility measure and socio-economic indicators for each of 47 French-speaking provinces of Switzerland at about.
Singh, H.P. Upadhyaya, L.N. and Namjoshi, U.D. (1988): Estimation of finite population variance. *Curr. Sc*., **57**, 1331−1334.
Singh, M. P. (1967): Ratio-Cum-Product Method of Estimation, *Metrika*, **12**, 34−42.
Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. (2011): Improved Exponential Estimator for Population Variance Using Two Auxiliary Variables. *Italian journal of pure and applied mathematics* – n. **28** (101−108).
Sukhatme, P.V and Sukhatme, B.V. (1970). Sampling Theory of Surveys with Applications, Asia Publishing House, New Delhi, pp-184−235.
Singh, R., Kumar, M., Singh, A.K. and Smarandache, F. (2011): A family of estimators of population variance using information on auxiliary attribute. Zip Publishing, Columbus, USA, pp.63-70.

Authors and Affiliations

K. B. Panda[1] and P. P. Mohanty[2]

K. B. Panda
Email: kunja.st@utkaluniversity.ac.in

[1,2]Department of Statistics, Utkal University, Bhubaneswar, Odisha, India