Estimation of Finite Population Variance using Auxiliary Information

K. B. Panda, N. Sahoo
Utkal University, Vani Vihar, Odisha, India
Utkal University, Vani Vihar, Odisha, India

Abstract: *This paper proposes a family of exponential estimators for estimating the finite population variance using auxiliary information in simple random sampling. Expressions for bias, mean squared error and its minimum values have been obtained. The comparisons have been made with the usual unbiased estimator, Isaki (J. Am. Stat. Assoc.78: 117-123, 1983), Kadilar and Cingi (Appl. Math. & Comput., 173, 1047-1059), Upadhyaya and Singh (Vikram Math. J. 19, 14-17, 1999a) and Lone and Tailor (Pak. J. Stat. Oper.res. Vol.XI, No.2, pp 213-220, 2015). An empirical study is carried out to judge the merits of proposed estimator over the traditional estimators.*

Keywords: *Study variable, Auxiliary variable, Mean squared error, Bias, Simple random sampling.*

1.Introduction

Consider a finite population $U = \{U_1, U_2, \ldots\ldots, U_i, \ldots, U_N\}$ consisting of N units. Let y and x be the study variable and auxiliary variables with population means $\bar{Y}$ and $\bar{X}$ respectively. Let there be a sample of size n drawn from this population using simple random sampling without replacement (SRSWOR). Let $s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$ and $s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$ be the sample variances and $S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$ the population variances of y and x respectively. Let $C_y = S_y/\bar{Y}$ and $C_x = S_x/\bar{X}$ be the coefficients of variation of y and x respectively, and ρ_{yx} the coefficient of correlation between y and x. We assume that all parameters of x are known. It is also assumed that the population size N is very large so that the finite population correction (FPC) term is ignored.

Let $s_y^2 = S_y^2(1+e_0)$, $s_x^2 = S_x^2(1+e_1)$, $\bar{x} = \bar{X}(1+e_2)$ such that $E(e_0) = E(e_1) = E(e_2) = 0$, $E(e_0^2) = \frac{1}{n}(\lambda_{40}-1)$, $E(e_1^2) = \frac{1}{n}(\lambda_{04}-1)$, $E(e_2^2) = \frac{1}{n}C_x^2$, $E(e_0e_1) = \frac{1}{n}(\lambda_{22}-1)$, $E(e_0e_2) = \frac{1}{n}\lambda_{21}C_x$ and $E(e_1e_2) = \frac{1}{n}\lambda_{03}C_x$.

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.

2. Existing Estimators

The variance of the usual unbiased estimator s_y^2 (t_0) is given by

$$V(t_0) = \frac{1}{n}S_y^4(\lambda_{40}-1) \tag{2.1}$$

Isaki (1983), suggested the following ratio estimator for estimating population variance S_y^2

$$t_1 = s_y^2\left(\frac{S_x^2}{s_x^2}\right) \tag{2.2}$$

Kadilar and Cingi (2006) considered the following ratio type estimators for S_y^2 as

$$t_2 = s_y^2\left(\frac{S_x^2 - C_x}{s_x^2 - C_x}\right),$$

(2.3) (2.3)

$$t_3 = s_y^2\left(\frac{S_x^2 - \beta_2\ \ x}{s_x^2 - \beta_2\ \ x}\right),$$

(2.4)

$$t_4 = s_y^2\left(\frac{S_x^2\beta_2 \; x \; - C_x}{s_x^2\beta_2 \; x \; - C_x}\right),$$
(2.5)

$$t_5 = s_y^2\left(\frac{C_x S_x^2 - \beta_2 \; x}{C_x s_x^2 - \beta_2 \; x}\right),$$
(2.6)

Upadhyaya and Singh (1999a) proposed ratio estimator for S_y^2 as

$$t_6 = s_y^2\left(\frac{S_x^2 + \beta_2 \; x}{s_x^2 + \beta_2 \; x}\right). \qquad (2.7)$$

The mean squared error of the estimators $t_i(i = 1, 2, 3, 4, 5, 6)$ up to the first degree of approximation are given as

$$MSE(t_i) = \frac{1}{n} S_y^4 \left[\; \lambda_{40} - 1 \; + \delta_i^2 \; \lambda_{04} - 1 \; - 2\delta_i \; \lambda_{22} - 1 \; \right]$$

(2.8)

where

$$\delta_i = \begin{cases} 1, & i = 1 \\ S_x^2/(S_x^2 - C_x), & i = 2 \\ S_x^2/(S_x^2 - \beta_2(x)), & i = 3 \\ S_x^2\beta_2(x)/(S_x^2\beta_2(x) - C_x), & i = 4 \\ S_x^2 C_x/(S_x^2 C_x - \beta_2(x)), & i = 5 \\ S_x^2/(S_x^2 + \beta_2(x)), & i = 6 \end{cases}$$

(2.9)

Lone and Tailor (2015) suggested the following class of estimators for population variance S_y^2 as

$$t_7 = \left[W_1 s_y^2 \left(\frac{aS_x^2 - b}{as_x^2 - b}\right) + W_2 s_y^2 \left(\frac{\theta \bar{x} - \varphi}{\theta \bar{X} - \varphi}\right) \right]$$
(2.10)

Where (W_1, W_2) are suitably chosen constant can be determined such that mean squared error of the estimator t_6 is minimum and (a, b, θ, φ) are either constants or functions of known parameters C_x, $\beta_2(x)$ and ρ_{yx} of the auxiliary variate x.

The MSE of the estimator t_7 is given by

$$MSE(t_7) = S_y^4[1 + CW_1^2 + DW_2^2 + 2EW_1W_2 - 2W_1F - 2W_2G]$$
(2.11)

where

$$C = 1 + 3M^2\frac{1}{n}(\lambda_{04} - 1) + \frac{1}{n}(\lambda_{40} - 1) - 4M\frac{1}{n}(\lambda_{22} - 1)$$

$$D = 1 + \frac{1}{n}(\lambda_{40} - 1) + S^2\frac{1}{n}C_x^2 + 4S\frac{1}{n}\lambda_{21}C_x$$

$$E = 1 + 2\frac{S}{n}\lambda_{21}C_x - 2M\frac{1}{n}(\lambda_{22} - 1) - MS\frac{1}{n}\lambda_{03}C_x + \frac{1}{n}(\lambda_{40} - 1) + M^2\frac{1}{n}(\lambda_{04} - 1)$$

$$F = 1 + M^2\frac{1}{n}(\lambda_{04} - 1) - M\frac{1}{n}(\lambda_{22} - 1)$$

$$G = 1 + S\frac{1}{n}\lambda_{21}C_x$$

and $W_1(opt.) = \frac{DF - EG}{CD - E^2}$, $W_2(opt.) = \frac{CG - EF}{CD - E^2}$

(2.12)

So Min. $MSE(t_7) = S_y^4\left[1 - \frac{(DF^2 + CG^2 - 2EFG)}{CD - E^2}\right]$

(2.13)

Min. $MSE(t_7) = S_y^4[1 - K^*]$ (2.14)

where $K^* = \frac{(DF^2 + CG^2 - 2EFG)}{CD - E^2}$

3. Proposed Estimator

Motivated by Singh, et al. (2009a.), we propose the following estimator for estimating the population variance S_y^2 as

$$t_{Re} = \left[W_1 s_y^2 + W_2\left(\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right)\right]\exp\left[\frac{\gamma\ S_x^2 - s_x^2}{\gamma\ S_x^2 + s_x^2\ + 2\beta}\right]$$

(3.1)

where γ and β are either real numbers or functions of the known parameters associated with an auxiliary attribute. (W_1, W_2) are suitably chosen scalars to be properly determined for minimum mean square error (MSE) of suggested estimators and $W_1 + W_2 \neq 1$(see Sharma & Singh 2013a).

Expanding equation (3.1) in terms of e's up to the first order of approximation, we have,

$$t_{Re} - S_y^2 = W_1 S_y^2 + W_1 S_y^2 e_0 + W_2 \frac{e_2}{2} + W_2 \frac{e_2^2}{4} - W_1 S_y^2 \frac{\theta e_1}{2} - W_1 S_y^2 \frac{\theta e_0 e_1}{2} + W_2 \frac{\theta e_1 e_2}{4} + W_1 S_y^2 \frac{3}{8}\theta^2 e_1^2 - S_y^2$$

(3.2)

where , $e_0$and e_1 are defined earlier and

$$\theta = \frac{\gamma S_x^2}{\gamma S_x^2 + \beta}.$$

Now $B(t_{Re}) = E\left(t_{Re} - S_y^2\right) = (W_1 - 1)S_y^2 + W_2 \frac{C_x^2}{4n} - W_1 S_y^2 \frac{\theta}{2n}(\lambda_{22} - 1) + W_2 \frac{\theta}{4n}\lambda_{03} C_x + \frac{3}{8} W_1 S_y^2 \frac{\theta^2}{n}(\lambda_{04} - 1)$

(3.3)

Then $MSE(t_{Re}) = S_y^4 + W_1^2 S_y^4 A_1 + W_2^2 A_2 + 2W_1 S_y^4 A_3 + 2W_2 S_y^2 A_4 + 2W_1 W_2 S_y^2 A_5$

(3.4)

where ,

$$A_1 = 1 + \frac{1}{n}(\lambda_{40} - 1) + \frac{\theta^2}{n}(\lambda_{04} - 1) - \frac{2\theta}{n}(\lambda_{22} - 1)$$

$$A_2 = \frac{C_x^2}{4n}$$

$$A_3 = -1 + \frac{\theta}{2n}(\lambda_{22} - 1) - \frac{3\theta^2}{8n}(\lambda_{04} - 1)$$

$$A_4 = -\frac{C_x^2}{4n} - \frac{\theta}{4n}\lambda_{03} C_x$$

$$A_5 = \frac{C_x^2}{4n} - \frac{1}{2n}\lambda_{21} C_x + \frac{\theta}{2n}\lambda_{03} C_x$$

Partially differentiating eq[n] (3.4) with respect to $W_1$and W_2 and equating to zero, we get the optimum value of W_1 and W_2 as

$$W_1(opt.) = \frac{A_4 A_5 - A_2 A_3}{A_1 A_2 - A_5^2} \quad \text{and}$$

$$W_2(opt.) = \frac{S_y^2\ \ A_3 A_5 - A_1 A_4}{A_1 A_2 - A_5^2}$$

(3.5) Putting (3.5) in (3.4), we get the optimum mean squared error of the estimator t_{Re} as

$$\text{Min. } MSE(t_{Re}) = S_y^4\left[1 + \frac{2A_3 A_4 A_5 - A_2 A_3^2 - A_1 A_4^2}{A_1 A_2 - A_5^2}\right]$$

(3.6)

$$\text{Min. } MSE(t_{Re}) = S_y^4\,[1 + R]$$

(3.7)

where $R = \frac{2A_3 A_4 A_5 - A_2 A_3^2 - A_1 A_4^2}{A_1 A_2 - A_5^2}$, and A_1, A_2, A_3, A_4 and A_5 defined earlier.

Table-1: Members of class of estimators t_{Re}

Estimators	Constants	
	γ	β
$t_{Re1} = \left[W_1 s_y^2 + W_2\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)\right] exp\left[\frac{(S_x^2 - s_x^2)}{(S_x^2 + s_x^2)+2}\right]$	1	1
$t_{Re2} = \left[W_1 s_y^2 + W_2\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)\right] exp\left[\frac{(S_x^2 - s_x^2)}{(S_x^2 + s_x^2)+2C_x}\right]$	1	C_x
$t_{Re3} = \left[W_1 s_y^2 + W_2\left(\frac{\bar{X}-\bar{x}}{\bar{X}+\bar{x}}\right)\right] exp\left[\frac{(S_x^2 - s_x^2)}{(S_x^2 + s_x^2)+2S_x}\right]$	1	S_x

4. Efficiency comparison of the estimator t_{Re} with the estimators t_i $(i = 0, 1, 2, 3, 4, 5, 6, 7)$

From (2.1), (2.8), (2.14) and (3.7), it is observed that the proposed estimator would be more efficient than

(i) Usual unbiased estimator s_y^2 (t_0) if

$$R < \frac{\lambda_{40} - 1}{n} - 1$$

(4.1)

(ii) Isaki (1983) estimator t_1 if

$$R < \left(\frac{\lambda_{40} + \lambda_{04} - 2\lambda_{22}}{n}\right) - 1$$

(4.2)

(iii) Kadilar and Cingi (2006) estimator t_2 if

$$R < \frac{1}{n}\left[(\lambda_{40} - 1) + \left(\frac{S_x^2}{S_x^2 - C_x}\right)^2 (\lambda_{04} - 1) - 2\left(\frac{S_x^2}{S_x^2 - C_x}\right)(\lambda_{22} - 1)\right] - 1$$

(4.3)

(iv) Kadilar and Cingi (2006) estimator t_3 if

$$R < \frac{1}{n}\left[(\lambda_{40} - 1) + \left(\frac{S_x^2}{S_x^2 - \beta_2\ x}\right)^2 (\lambda_{04} - 1) - 2\left(\frac{S_x^2}{S_x^2 - \beta_2\ x}\right)(\lambda_{22} - 1)\right] - 1$$

(4.4)

(v) Kadilar and Cingi (2006) estimator t_4 if

$$R < \frac{1}{n}\left[\left(\lambda_{40} - 1\right) + \left(\frac{S_x^2 \beta_2\ x}{S_x^2 \beta_2\ x - C_x}\right)^2 (\lambda_{04} - 1) - 2\left(\frac{S_x^2 \beta_2\ x}{S_x^2 \beta_2\ x - C_x}\right)(\lambda_{22} - 1)\right] - 1$$

(4.5)

(vi) Kadilar and Cingi (2006) estimator t_5 if

$$R < \frac{1}{n}\left[(\lambda_{40} - 1) + \left(\frac{S_x^2 C_x}{S_x^2 C_x - \beta_2\ x}\right)^2 (\lambda_{04} - 1) - 2\left(\frac{S_x^2 C_x}{S_x^2 C_x - \beta_2\ x}\right)(\lambda_{22} - 1)\right] - 1$$

(4.6)

(vii) Updhyaya and Singh (199a) estimator t_6 if

$$R < \frac{1}{n}\left[(\lambda_{40}-1)+\left(\frac{S_x^2}{S_x^2+\beta_2\ \ x}\right)^2(\lambda_{04}-1)-2\left(\frac{S_x^2}{S_x^2+\beta_2\ \ x}\right)(\lambda_{22}-1)\right]-1$$

(4.7)

(viii) Lone and Tailor (2015) estimator t_7 if

$$R < -K^*$$

(4.8)

5. Empirical study

To illustrate the performance of estimators t_{Rei} and min.(t_{Re}) over the existing estimators, we consider a natural population from [Singh(2003), p. 1111-1112]. The description of population is given below.

y: Amount (in $000) of real estate farm loans in different state during 1997,

x: Amount (in $000) of non-real estate farm loans in different state during 1997.

Table 5.1

$\lambda_{40} = 3.5822, \lambda_{04} = 4.5247, \lambda_{22} = 2.8411,$
$\lambda_{21} = 0.9387, \lambda_{03} = 1.5936, \bar{Y} = 555.43,$

$\bar{X} = 878.16, C_x = 1.2351, C_y = 1.0529, n = 10$

Table 5.2

Percent Relatives Efficiencies of S_y^2, $t_i(i = 0,1,2,3,4,5,6)$ and $T_i(i = 1,2,3)$ with respect to S_y^2

Estimators	PRE
t_0	100
t_1	156.0173
t_2	156.0157
t_3	156.0168
t_4	156.0172
t_5	156.0176
t_6	156.0179
t_7	163.8827
$\boldsymbol{t_{Re1}}$	**230.7596**
$\boldsymbol{t_{Re2}}$	**230.7596**
t_{Re3}	226.3102

6. Conclusion

In table 5.2, it is observed that the proposed estimator is more efficient than the usual unbiased estimator, Isaki (J. Am. Stat. Assoc.78: 117-123, 1983), Kadilar and Cingi (Appl. Math. & Comput., 173, 1047-1059), Upadhyaya and Singh (Vikram Math. J. 19, 14-17, 1999a) and Lone and Tailor (Pak. J. Stat. Oper.res. Vol.XI, No.2, pp 213-220, 2015).

REFEREENCES

1. Isaki, C. T. (1983). Variance estimation using auxiliary information. J. Amer. Statist. Assoc. 78, 117-123.

2. Kadilar, C. and Cingi, H. (2006). Ratio estimators for the population in simple and stratified random sampling. Applied Maths. & Comp., 173, 1047-1059.

3. S. Singh (2003). Advanced sampling theory with applications, Kluwer Academic Press.

4.Tailor, R. and Lone, H. A. (2015). A family of estimators for estimating population variance using auxiliary information in sample survey. Pak. J. Stat. Oper. Res. Vol.XI, No.2, pp 213-220.

5. Upadhyaya, L. N. and Singh, H. P. (1999a). An estimator for finite population variance that utilizes the kurtosis of an auxiliary variable in sample survey. Vikram Math. J. 19, 14-17.

6. Singh, R., Chouhan, P., Sawan, N., Smarandache, F., (2009a). Improvement in estimating the population mean using exponential estimator in simple random sampling. Bulletin of Statistics & Economics, Vol.3, A09, 13-18.