Generalized Estimator of Finite population Variance

A.K.P.C. Swain

Former Professor of Statistics, Utkal University
Bhubaneswar, India
e-mail:akpcs@rediffmail.com

Abstract

In this paper a generalized class of finite population variance is suggested and its large sample bias and mean square error are derived. Further some special cases of this class are considered along with a numerical example to compute and compare their biases and mean square errors.

Key words: Auxiliary variable; Simple random sampling; Ratio estimator; Finite population variance; Generalized class of estimators; Bias ; Mean square error.

AMS Classification:62 D 05

1. Introduction

In finite population sampling it is an usual practice for the researchers to look for some auxiliary variables correlated with the main character under study to improve upon the efficiency of estimators formulated without the use of auxiliary variables. Ratio and regression methods of estimation are two classical methods which make use of auxiliary information to derive improved estimators of the population mean or total of the main variable under study (Cochran,1953)[1]. Although the estimation of finite population variance arises while estimating the variance of the estimates of finite population mean or total , it was only through the work of Evans(1951)[2] and Liu(1974)[3] separate attention was given to the estimation of finite population variance. Liu (1974) considered the problem of estimating variance in a general set up and presented a general class of quadratic functions and obtained a class of unbiased estimators under some conditions. Isaki (1983)[4] proposed the ratio method of estimation to estimate the finite population variance. Some early work on the estimation of finite population variance are due to Das and Tripathy (1978)[5] and Srivastava and Jhajj (1980)[6] among others. Swain and Mishra (1992)[7] suggested some unbiased estimators of finite population variance using auxiliary information. Further, Swain and Mishra (1994a)[8] studied the limiting distribution of ratio estimator of finite population variance under some general conditions. Also, Swain and Mishra (1994b)[9] suggested an alternative method of estimation of finite population variance under unequal probability sampling.

Let there be a finite population $U = (U_1, U_2, ..., U_N)$ consisting of N identifiable units. To each unit U_i in the population a paired value $(y_i, x_i), i = 1, 2, ..., N$,corresponding to the study variable y and the auxiliary variable x is attached. Denote $\bar{Y}$ and $\bar{X}$ as the population means of

y and x respectively; S_y^2 and S_x^2 as the finite population variances of y and x respectively;, where

$$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2 \quad \text{and} \quad S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{X})^2$$

ρ as the correlation coefficient between y and x, which is assumed to be positive. Further , $\bar{X}$ is assumed to be known in advance. For a simple random sample without replacement of size n, define the sample means $\bar{y}$ and $\bar{x}$ of y and x respectively as

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n}y_i, \bar{x} = \frac{1}{n}\sum_{i=1}^{n}x_i$$

The sample variances of y and x are defined by $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ and

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ respectively. To first order of approximation,

$$V(s_y^2) \cong \theta S_y^4(\beta_{2y} - 1) \quad , \quad V(s_x^2) \cong \theta S_x^4(\beta_{2x} - 1) \text{ and } Cov(s_y^2, s_x^2) \cong \theta S_y^2 S_x^2(h-1)$$

where $\beta_{2y} = \frac{\mu_{40}}{\mu_{20}^2}$, $\beta_{2x} = \frac{\mu_{04}}{\mu_{02}^2}$, $h = \frac{\mu_{22}}{\mu_{20}\mu_{02}}$ $\mu_{rs} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{Y})^r(x_i - \bar{X})^s$,

$f = \frac{n}{N}, \theta = \frac{1-f}{n}$

A ratio estimator of the population variance S_y^2 due to Isaki (1983)is given by

$$\hat{S}_{yr}^2 = \frac{s_y^2}{s_x^2}S_x^2$$

To first order of approximations,

$$MSE(\hat{S}_{yr}^2) = \theta S_y^4(\beta_{2y} + \beta_{2x} - 2h)$$

$$Bias(\hat{S}_{yr}^2) = \theta S_y^2(\beta_{2x} - h)$$

Under Bivariate Normality of (y, x) with means $(\bar{Y}, \bar{X})$, variances (S_y^2, S_x^2) and correlation coefficient ρ,

$$Var(\hat{S}_{yr}^2) = \frac{4}{n}S_y^4(1-\rho^2)$$

$$Bias(\hat{S}_{yr}^2) = \frac{2}{n}S_y^2(1-\rho^2)$$

In the following we propose a generalized class of estimator of finite population variance S_y^2 and derive its bias and mean square error and discuss its some special cases.

2 . Generalized Class of Estimators of Finite Population Variance

Consider a generalized class of estimators of finite population variance

$$\hat{S}_{yrg}^2 = s_y^2 \left[\alpha \left(\frac{S_x^2}{s_x^2} \right)^g + (1-\alpha) \left(\frac{s_x^2}{S_x^2} \right)^h \right]^\delta ,$$
(2.1)

where α, g, h and δ are real and free parameters to be chosen suitably.

Let $e_1 = \dfrac{s_y^2 - S_y^2}{S_y^2}$ and $e_2 = \dfrac{s_x^2 - S_x^2}{S_x^2}$

Thus we write

$s_y^2 = S_y^2(1+e_1)$, $s_x^2 = S_x^2(1+e_2)$ with $E(e_1) = E(e_2) = 0$, $V(e_1) \cong \theta(\beta_{2y} -1)$, $V(e_2) \cong \theta(\beta_{2x} -1)$, $Cov(e_1, e_2) \cong \theta(h-1)$

$\hat{S}_{yr}^2$ belongs to the class specified by $\hat{S}_{yrg}^2$. An optimum ratio type estimator in this class is obtained by fixing any three of the parameters and minimizing the mean square of $\hat{S}_{yrg}^2$ with respect to unspecified one.

Assuming $|e_2| < 1$ for all samples ,we may expand $\hat{S}_{yrg}^2$ in Binomial series and keeping second degree terms in the expansion we have

$$\hat{S}_{yrg}^2 = S_y^2(1+e_1)\left\{ 1 + \delta(h - \alpha(g+h))e_2 + \delta\left(\frac{h(h-1)}{2} + \alpha(\frac{g(g+1)}{2} - \frac{h(h-1)}{2}) \right)e_2^2 + \frac{\delta(\delta-1)}{2}(h - \alpha(g+h))^2 e_2^2 \right\}$$

Thus to first order of approximations

$$\hat{S}_{yrg}^2 \cong S_y^2 \left[1 + e_1 + \delta(h - \alpha(g+h))e_2 \right]$$
(2.2)

The mean square error of $\hat{S}_{yrg}^2$ to $O(1/n)$ is given by

$$MSE(\hat{S}_{yrg}^2) = \theta S_y^4 \left[(\beta_{2y} -1) + \delta^2(h - \alpha(g+h))^2(\beta_{2x} -1) - 2\delta(h - \alpha(g+h))(h-1) \right]$$
(2.3)

Minimizing $MSE(\hat{S}_{yrg}^2)$ with respect to α gives

$$\alpha_{opt} = \frac{\delta h + k}{\delta(g+h)},$$
(2.4)

$$k = \frac{Cov(e_1, e_2)}{V(e_2)} = \frac{h-1}{\beta_{2x} - 1}$$

Substituting the optimum value of α in expression for $MSE(\hat{S}^2_{yrg})$ we have

$$MSE(\hat{S}^2_{yrg})_{opt} = \theta S_y^4 \left[(\beta_{2y} - 1) - \frac{(h-1)^2}{\beta_{2x} - 1} \right] , \tag{2.5}$$

Substituting the optimum value of α in the approximate expression of bias to $O(1/n)$, we have

$$Bias(\hat{S}^2_{yrg})_{opt} = \theta S_y^2 \left[\delta \{ h - \alpha_{opt}(g+h) \}(h-1) + \delta \left\{ \frac{h(h-1)}{2} + \alpha_{opt}(\frac{g(g+1)}{2} - \frac{h(h-1)}{2}) \right\}(\beta_{2x} - 1) + \right.$$

$$\left. \frac{\delta(\delta-1)}{2} \{ h - \alpha_{opt}(g+h) \}^2 (\beta_{2x} - 1) \right] \tag{2.6}$$

The following Table 1 gives the values of the optimum α along with the optimum approximate biases of the generalized ratio type estimators of finite population variance for different choices of g, h and δ.

Table 1   Optimum biases of some selected ratio type estimators

Sl.No.	g	h	δ	α	Ratio type estimators	Optimum biases excepting the Constant multiplier θS_y^2
1	1	1	1	$(\frac{k+1}{2})$	$\hat{S}^2_{yrg1} = s_y^2 \left[\alpha \frac{S_x^2}{s_x^2} + (1-\alpha) \frac{s_x^2}{S_x^2} \right]$	$B_1 = -k(h-1) + \frac{k+1}{2}(\beta_{2x} - 1)$
2	1	0	1	k	$\hat{S}^2_{yrg2} = s_y^2 \left[\alpha \frac{S_x^2}{s_x^2} + (1-\alpha) \right]$	$B_2 = -k(h-1) + k(\beta_{2x} - 1)$
3	0	1	1	$k+1$	$\hat{S}^2_{yrg3} = s_y^2 \left[\alpha + (1-\alpha) \frac{s_x^2}{S_x^2} \right]$	$B_3 = -k(h-1)$
4	1	1	-1	$(k-1)/2$	$\hat{S}^2_{yrg4} = s_y^2 / \left[\alpha \frac{S_x^2}{s_x^2} + (1-\alpha) \frac{s_x^2}{S_x^2} \right]$	$B_4 = -k(h-1)$ $+(\frac{k-1}{2} + k^2)(\beta_{2x} - 1)$
5	1	0	-1	$-k$	$\hat{S}^2_{yrg5} = s_y^2 / \left[\alpha \frac{S_x^2}{s_x^2} + (1-\alpha) \right]$	$B_5 = -k(h-1) +$ $k(k+1)(\beta_{2x} - 1)$
6	0	1	-1	$-(k-1)$	$\hat{S}^2_{yrg6} = s_y^2 / \left[\alpha + (1-\alpha) \frac{s_x^2}{S_x^2} \right]$	$B_6 = -k(h-1) + k^2(\beta_{2x} - 1)$

Under Bivariate Normality ,

$$MSE(\hat{S}^2_{yri}) = \frac{2}{n} S_y^4 (1 - \rho^4) , i=1,2,\ldots 6$$

$$B_1 = Bias(\hat{S}^2_{yr1}) = \frac{1}{n} S^2_y (1 + \rho^2 - 2\rho^4)$$

$$B_2 = Bias(\hat{S}^2_{yr2}) = \frac{2}{n} S^2_y \rho^2 (1 - \rho^2)$$

$$B_3 = Bias(\hat{S}^2_{yr3}) = -\frac{2}{n} S^2_y \rho^4$$

$$B_4 = Bias(\hat{S}^2_{yr4}) = -\frac{1}{n} S^2_y (1 - \rho^2)$$

$$B_5 = Bias(\hat{S}^2_{yr5}) = \frac{2}{n} S^2_y \rho^2$$

$$B_6 = Bias(\hat{S}^2_{yr6}) = 0$$

3. Some Alternative Modified Ratio type Estimators of Finite population Variance

Tailor and Sharma (2012) considered some alternative modified ratio type estimators of finite population variance based on weighted combination of s^2_y and $s^2_y \frac{S^2_x}{s^2_x}$ where weights are some arbitrary functions of sampling fraction. We consider only the simplest one given by

$$\hat{S}^2_{rts} = f s^2_y + (1 - f) s^2_y \frac{S^2_x}{s^2_x} \tag{3.1}$$

To first order approximation ,

$$Bias(\hat{S}^2_{rts}) = \theta S^2_y (1 - f)(\beta_{2x} - h) \tag{3.2}$$

$$MSE(\hat{S}^2_{rts}) = \theta S^4_y \left[(\beta_{2y} - 1) + (1 - f)^2 (\beta_{2x} - 1) - 2(1 - f)(h - 1) \right] \tag{3.3}$$

Alternatively , we may form general weighted combination of s^2_y and $s^2_y \frac{S^2_x}{s^2_x}$, given by

$$\hat{S}^2_{rw} = w s^2_y + (1 - w) s^2_y \frac{S^2_x}{s^2_x} \tag{3.4}$$

The optimum weight is obtained by minimizing the approximate mean square error of $\hat{S}^2_{rw}$ with respect to w. Thus, the optimum weight is given by

$$w_{opt} = 1 - k = 1 - \frac{h - 1}{\beta_{2x} - 1} = \beta_{2x} - h \tag{3.5}$$

Hence, the optimum mean square error of $\hat{S}^2_{rw}$ is given by

$$MSE(\hat{S}^2_{rw})_{opt} = \theta S_y^4 \left[(\beta_{2y} - 1) - \frac{(h-1)^2}{\beta_{2x} - 1} \right] \qquad (3.6)$$

and the optimum bias is given by

$$Bias(\hat{S}^2_{rw})_{opt} = \theta S_y^2 (\frac{h-1}{\beta_{2x} - 1})(\beta_{2x} - h) \qquad (3.7)$$

Another alternative adjusted ratio type variance estimator is proposed as

$$\hat{S}^2_{r(adj)} = s_y^2 \frac{S_x^2}{s_x^2} + \mu(1 - \frac{S_x^2}{s_x^2}) \ , \qquad (3.8)$$

where μ is a real constant to be suitably chosen.

Expanding $\hat{S}^2_{r(adj)}$ in binomial series , and keeping terms up to second degree terms, we have

$$\hat{S}^2_{r(adj)} = S_y^2 (1 + e_1 - e_2 + e_2^2 - e_1 e_2) + \mu(e_2 - e_2^2)$$

Thus to $O(1/n)$, $E(\hat{S}^2_{r(adj)}) = S_y^2 + \theta S_y^2 \left\{ (\beta_{2x} - 1) - (h-1) - \frac{\mu}{S_y^2}(\beta_{2x} - 1) \right\} \qquad (3.9)$

$$MSE(\hat{S}^2_{r(adj)}) = \theta S_y^4 \left\{ (\beta_{2y} - 1) + (1 - \frac{\mu}{S_y^2})^2 (\beta_{2x} - 1) - 2(1 - \frac{\mu}{S_y^2})(h-1) \right\} \quad (3.10)$$

Minimizing $MSE(\hat{S}^2_{r(adj)})$ with respect to μ we have

$$\mu_{opt} = (1 - \frac{h-1}{\beta_{2x} - 1}) \qquad (3.11)$$

Hence, $\qquad MSE(\hat{S}^2_{r(adj)})_{opt} = \theta S_y^4 \left[(\beta_{2y} - 1) - \frac{(h-1)^2}{\beta_{2x} - 1} \right] \qquad (3.12)$

and $\qquad Bias(\hat{S}^2_{r(adj)})_{opt} = 0 \qquad (3.13)$

Note: $\hat{S}^2_{rw}, \hat{S}^2_{r(adj)}$ and $\hat{S}^2_{yrgi}(i = 1, 2, ..., 6)$ are more efficient than $\hat{S}^2_{yr}$ and $\hat{S}^2_{rts}$. However, $\hat{S}^2_{r(adj)}$ and $\hat{S}^2_{yrg6}$ are to be preferred over others because their first order biases are zero. Further, $\hat{S}^2_{rts}$ is less biased than $\hat{S}^2_{yr}$; $\hat{S}^2_{rw}$ and $\hat{S}^2_{yrg2}$ have the same first order bias.

4. Numerical illustration

The data taken from Murthy(1967)[10] used by Tailor and Sharma(2012)[11] relate to output(y) and number of workers(x) of 80 factories.

The summary computations are

$$N = 80, n = 30, f = 0.375, \beta_{2y} = 2.2667, \beta_{2x} = 3.65 , h = 2.3377$$

Table 2 : Comparison of Absolute Biases and Efficiencies of Estimators

	Estimator	$|Bias| / \theta S_y^2$	$MSE / \theta S_y^4$	Efficiency
1	s_y^2	0	1.2667	100
2	$\hat{S}_{yr}^2$	1.3123	1.2413	102.05
3	$\hat{S}_{rts}^2$	0.8202	0.6298	201.13
4	$\hat{S}_{rw}^2$	0.6624	0.5914	214.19
5	$\hat{S}_{r(adj)}^2$	0	0.5914	214.19
3	$\hat{S}_{yrg1}^2$	1.3186	0.5914	214.19
4	$\hat{S}_{yrg2}^2$	0.6624	0.5914	214.19
5	$\hat{S}_{yrg3}^2$	0.6753	0.5914	214.19
6	$\hat{S}_{yrg4}^2$	0.6561	0.5914	214.19
7	$\hat{S}_{yrg5}^2$	1.3377	0.5914	214.19
8	$\hat{S}_{yrg6}^2$	0	0.5914	214.19

Conclusions: Numerical illustration shows that $\hat{S}_{yr}^2$ and $\hat{S}_{rts}^2$ are less efficient than $\hat{S}_{rw}^2, \hat{S}_{r(adj)}^2$ and $\hat{S}_{yrgi}^2 (i = 1,2,...6)$.Further, the first order biases of $\hat{S}_{r(adj)}^2$ and $\hat{S}_{yrg6}^2$ are zero, followed by $\hat{S}_{yrg4}^2$.The biases of $\hat{S}_{rw}^2$ and $\hat{S}_{yrg2}^2$ are equal.

 References
[1]   W.G. Cochran, *Sampling Techniques* (First Edition, Wiley, New York, 1953).
[2]   D. Evans, On the variance estimators of SD and Variance, *Journ. Amer. Stat. Assoc*. 46, (1951) 220-224.
[3]   T.P. Liu, A general unbiased estimator for the variance of a finite population , *Sankhya*, C,**36** (1974) 23-32.
[4]   C.T. Isaki, Variance estimation using auxiliary information, *Journ Amer. Stat. Assoc*.**78** (1983)117-123.
[5]   A.K. Das and T.P. Tripathy, Use of auxiliary information in estimating the finite population variance, *Sankhya*,**C,40** (1978)139-148.
[6]   S.K. Srivastava and H.S. Jhajj, A class of estimators using auxiliary information for estimating finite population variance, *Sankhya*, C,**42** (1980) 87-96.
[7]   A.K.P.C. Swain and G. Mishra, Unbiased estimators of finite population variance using auxiliary information, Metron Vol. L-n, 3-4 (1992) 201-216.
[8]   A.K.P.C. Swain and G. Mishra (1994 a)Limiting distribution of the ratio estimator of finite population variance, *Sankhya*, B,56 (1994a) 11-17.
[9]   A.K.P.C. Swain and G. Mishra (1994 b).Estimation of finite population variance under unequal probability sampling, *Sankhya*, **B,56 (**1994b) 374-388.
[10] M.N. Murthy, Sampling Theory and Methods (Statistical Publishing Society, Calcutta, 1967).
 [11] R. Tailor and B. Sharma. Modified estimators of population variance in presence of auxiliary information, Statistics in Transition,Vol.13, No.1 (2012) 37-46.