Proc. Pakistan Acad. Sci. 43(4):279-283.2006

A DUAL TO VARIANCE RATIO-TYPE ESTIMATOR IN SIMPLE RANDOM SAMPLING

Javid Shabbir

Department of Statistics, Quaid-i-Azam University, Islamabad 45320, Pakistan

Received February 2006, accepted July 2006

Communicated by Prof. Dr. Q. K. Ghori

Abstract: We propose a dual to variance ratio-type estimator to estimate the finite population variance. The optimum mean square error of this estimator is equal to variance of a linear regression estimator and is better than the usual unbiased variance estimator and Isaki (Jour Amer. Statis Assoc. 78:117-123, 1983) estimator. We use the jackknife technique to make the proposed estimator unbiased. The validity of the proposed estimator is examined by using the various data sets.

Keywords: Ratio estimator, auxiliary variable, bias, mean square error (*MSE*), efficiency, Jacknife technique

Introduction

Let U be a finite population consisting of N units $U_1, U_2, ..., U_N$ from which a sample of size n is to be drawn by simple random sampling without replacement (SRSWOR). Let y and x denote the study and auxiliary variables respectively and y is positively correlated with x. Let $S_y^2 = \sum_{i=1}^{N}(y_i - \bar{Y})^2/(N-1)$ and $S_x^2 = \sum_{i=1}^{N}(x_i - \bar{X})^2/(N-1)$ denote the population variances of y and x respectively. Similarly, one can obtain the sample variances $s_y^2 = \sum_{i=1}^{n}(y_i - \bar{y})^2/(n-1)$ and $s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$ of y and x respectively. Let $C_y = S_y/\bar{Y}$ and $C_x = S_x/\bar{X}$ denote the coefficient of variations of y and x, respectively. To estimate S_y^2, it is assumed that S_x^2 is known.

Definition:

Let $\Delta_0 = (s_y^2 - S_y^2)/S_y^2$ and $\Delta_1 = (s_x^2 - S_x^2)/S_x^2$, therefore $E(\Delta_0) = E(\Delta_1) = 0$, $E(\Delta_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)(\lambda_{04} - 1)$, $E(\Delta_0\Delta_1) = \left(\frac{1}{n} - \frac{1}{N}\right)(\lambda_{22} - 1)$, where $\lambda_{pq} = \mu_{pq}/(\mu_{20}^{p/2}\mu_{02}^{q/2})$,

and $\mu_{pq} = \sum_{i=1}^{N}(y_i - \bar{Y})^p(x_i - \bar{X})^q/(N-1)$.

The conventional unbiased variance estimator is defined as

$$\hat{S}_0^2 = s_y^2 \qquad (1)$$

The variance of $\hat{S}_0^2$ is given by

$$Var(\hat{S}_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^4(\lambda_{40} - 1). \qquad (2)$$

Isaki [2] suggested the following variance ratio estimator

$$\hat{S}_I^2 = s_y^2\left(\frac{S_x^2}{s_x^2}\right). \qquad (3)$$

The bias and *MSE* to first order of approximation are given by

$$Bias(\hat{S}_I^2) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^2[(\lambda_{04} - 1) - (\lambda_{22} - 1)] \quad (4)$$

and

$$MSE(\hat{S}_I^2) = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^4[(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)]. \quad (5)$$

Now we propose the following dual to variance ratio-type estimator.

Proposed Estimator

The following proposed estimator is the combination of usual unbiased variance estimator given in (1) and dual to variance ratio estimator with weights w_1 and w_2 such that $w_1+w_2=1$ as

$$\hat{S}_P^2 = w_1 s_y^2 + w_2 s_y^2 \left(\frac{s_x^{*2}}{S_x^2}\right), \quad (6)$$

where $s_x^{*2} = \frac{NS_x^2 - ns_x^2}{N-n}$ is due to Srivenkataramana [6].

From (6), we have

$$\hat{S}_P^2 = w_1 S_y^2(1+\Delta_0) + w_2 S_y^2(1+\Delta_0)(1+\eta\Delta_1), \quad (7)$$

where $\eta = n/(N-n)$.

From (7), we have

$$\hat{S}_P^2 - S_y^2 = S_y^2[\Delta_0 - w_2\eta(\Delta_1 + \Delta_0\Delta_1)]. \quad (8)$$

Solving (8), we get the bias and *MSE of* $\hat{S}_P^2$ which are given by

$$Bias(\hat{S}_P^2) = E(\hat{S}_{P1}^2 - S_y^2) = -\left(\frac{1}{n} - \frac{1}{N}\right) w_2 S_y^2 \eta(\lambda_{22} - 1) \quad (9)$$

and

$$MSE(\hat{S}_P^2) = E(\hat{S}_{P1}^2 - S_y^2)^2 = S_y^4 E[\Delta_0 - w_2\eta\Delta_1)^2.$$

$$= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^4[(\lambda_{40} - 1) + w_2^2\eta^2(\lambda_{04} - 1) - 2w_2\eta(\lambda_{22} - 1)]. \quad (10)$$

From (2) and (10), it follows that

(i) $MSE(\hat{S}_P^2) < Var(\hat{S}_0^2)$ if either $\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} > \frac{1}{2} w_2\eta$. (11)

From (5) and (10), it follows that

(ii) $MSE(\hat{S}_P^2) < MSE(\hat{S}_I^2)$ if either $\frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} < \frac{1}{2}(w_2\eta+1)$. (12)

Thus, from (12), the estimator $\hat{S}_P^2$ is more efficient than $\hat{S}_P^2$ and $\hat{S}_I^2$ if

$$\frac{1}{2} w_2\eta < \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)} < \frac{1}{2}(w_2\eta + 1) \quad (13)$$

The above inequality is obviously true.

The proposed estimator $\hat{S}_P^2$ is efficient if the following conditions are satisfied.

Cond. (i) $w_2^2\eta^2 = \frac{(\lambda_{22}-1)^2}{(\lambda_{04}-1)^2} \le 1$ if $(\lambda_{22}-1) \le (\lambda_{04}-1)$.

Cond. (ii) $w_2\eta = \frac{(\lambda_{22}-1)}{(\lambda_{04}-1)} \le 1$ if $(\lambda_{22}-1) \le (\lambda_{04}-1)$.

Cond. (iii) $\frac{MSE(\hat{S}_P^2)}{MSE(S_0^2)} \le 1$.

Cond. (iv) $\frac{MSE(\hat{S}_P^2)}{MSE(S_I^2)} \le 1$.

The optimum choice of w_2 minimizing (10) is given by

$$w_2 = \frac{\lambda_{22} - 1}{\eta(\lambda_{04} - 1)} = w_2^* \text{ (say) and } w_1^* = 1 - w_2^*.$$

Substitution of optimum value of w_2 in (10), we get the minimum MSE of $\hat{S}_P^2$ which is given by

$$MSE(\hat{S}_P^2)_{\min} = \left(\frac{1}{n} - \frac{1}{N}\right) S_y^4 [(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}]. \quad (14)$$

The *MSE* of $\hat{S}_P^2$ in (14) is equal to variance of the linear regression estimator, $\hat{S}_{lr}^2 = s_y^2 + b(S_x^2 - s_x^2)$ where $b = \frac{s_y^2(\hat{\lambda}_{22} - 1)}{s_x^2(\hat{\lambda}_{04} - 1)}$ is sample regression coefficient. It is noted that Equation (14) is independent of η.

Efficiency comparison based on optimum *MSE*

We compare the proposed estimator with the usual variance estimator and Isaki [2] estimator.

(i) By (2) and (12),

$$Var(\hat{S}_0^2) - MSE(\hat{S}_P^2)_{\min} = \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} > 0; \; (\lambda_{04} - 1) > 0.$$

(ii) By (5) and (12),

$$Var(\hat{S}_I^2)_{\min} - MSE(\hat{S}_P^2)_{\min} > 0 \text{ if } \left(\frac{1}{n} - \frac{1}{N}\right) S_y^4 \left(\sqrt{(\lambda_{04} - 1)} - \frac{(\lambda_{22} - 1)}{\sqrt{(\lambda_{04} - 1)}}\right)^2 > 0.$$

Both of the above conditions are obviously true.

Data and Results

For comparison, we consider the following seven data sets from various sources.

***Population 1**:* (Cochran [1], p. 325)

y: Number of persons per block,
x: Number of rooms per block.
N=100, n=10, S_y^2=214.69, S_x^2=56.76, λ_{40}=2.2387, λ_{04}=2.2523, λ_{22}=1.5432, λ_{21}=0.4536, C_x = 0.1281, C_y=0.1450, $\bar{X}$ = 58.8, ρ = 06515.

***Population 2**:* (Cochran [1], p. 152)

y: Number of inhabitants in 1930,
x: Number of inhabitants in 1920.
N=196, n=49, S_y^2=151558.83, S_x^2=10900.42, λ_{40}=8.5362, λ_{04}=7.3617, λ_{22}=7.8780, λ_{21}=0.2295, C_x = 1.0126, C_y=0.9634, $\bar{X}$ = 103.1, ρ = 0.9820.

Population 3: (Cochran [1], p. 203)

y: Actual weight of peaches on each tree,
x: Eye estimate of weight of peaches on each tree.
N=200, n=10, S_y^2=99.81, S_x^2=85.09, λ_{40}=1.9249, λ_{04}=2.5932, λ_{22}=2.1149, λ_{21}=0.1875, C_x= 0.1621, C_y=0.1840, $\bar{X}$=56.9, ρ = 0.9937.

Population 4: (Sukhatme and Sukhatme [7], p. 185)

y: Wheat acreage in 1937,
x: Wheat acreage in 1936.
N=170, n=10, S_y^2=26456.89, S_x^2=22355.76, λ_{40}=3.1842, λ_{04}=2.2030, λ_{22}=2.5597, λ_{21}=0.6665, C_x= 0.5625, C_y=0.6163, $\bar{X}$=265.8, ρ = 0.977.

Population 5: (Upadhyaya and Singh [8])

y: Census population in year 1971,
x: Census population in year 1961.
N=100.000, n=142, S_y^2=71899173.02, S_x^2=40608000.69, λ_{40}=40.8536, λ_{04}=48.1567, λ_{22}=43.7615, λ_{21}=5.9786, C_x= 2.1971, C_y=2.1118, $\bar{X}$=2900.4, ρ = 0.977.

***Population 6**:* (Singh *et al.* [4])

y: The number of agriculture labourers for 1971,
x: The number of agriculture labourers for 1961.
N=278, n=30, S_y^2=3187030, S_x^2=1654.40,

λ_{40}=24.8969, λ_{04}=37.8898, λ_{22}=25.8142, λ_{21}=3.4347, C_x= 1.61198, C_y=104451, $\bar{X}$ =250111, ρ = 0.7273.

***Population** 7*: (Singh [5])

y: Amount (in \$1000) of real estate farm loans in different states during 1997,
x: Amount (in \$1000) of non real estate farm loans in different states during 1997.
N=50, n=8, S_y^2=7342021.5, S_x^2=1176526, λ_{40}=3.5822, λ_{04}=4.5247, λ_{22}=2.8411, λ_{21}=0.9387, C_x= 1.2352, C_y=1.0529, $\bar{X}$=878.16, ρ = 0.8038.

The relative efficiency (*RB*) and relative bias (*RE*) can be obtained by using the following expressions respectively

$$RB = \frac{Bias(\hat{S}_j^2)}{\sqrt{MSE(\hat{S}_j^2)}}\ ; j = I, P.$$

and

$$RE = \frac{Var(\hat{S}_0^2)}{MSE(\hat{S}_i^2)} \times 100\ ; \text{i} = 0, I, P.$$

The results are given in Tables 1, 2 and 3.

In Table 3, conditions (i) and (ii) are true only if (λ_{22}-1)≤(λ_{04}-1). Conditions (iii), and (iv) will always remain true.

Table 1. *RB* of different estimators.

Estimator	Pop. 1	Pop. 2	Pop. 3	Pop. 4	Pop. 5	Pop.6	Pop. 7
$\hat{S}_I^2$	0.1979	−0.1696	0.2746	−0.2115	0.3022	0.6234	0.3503
$\hat{S}_P^2$	−0.0648	−0.9093	−0.6321	−1.5412	−3.1324	−1.0730	−0.2448

Table 2. *RE* of different estimators w. r. t $\hat{S}_0^2$ in percentage.

Estimator	Pop. 1	Pop. 2	Pop. 3	Pop. 4	Pop. 5	Pop.6	Pop. 7
$\hat{S}_0^2$	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\hat{S}_I^2$	82.33	5310.92	320.81	815.61	2679.59	214.32	106.50
$\hat{S}_P^2$	121.38	7536.32	639.15	1347.98	3698.20	331.90	159.34

Table 3. Conditional values.

Conditions	Pop. 1	Pop. 2	Pop. 3	Pop. 4	Pop. 5	Pop.6	Pop. 7
Cond. (i)	0.1613	1.1689	0.4897	1.6809	0.8223	0.4526	0.2728
Cond. (ii)	0.4017	1.0812	0.6998	1.2965	0.9068	0.6728	0.5223
Cond. (iii)	0.8238	0.0133	0.1564	0.0742	0.0270	0.3013	0.6276
Cond. (iv)	0.6783	0.7047	0.5019	0.6051	0.7246	0.6457	0.6683

Unbiased Version of the Proposed Estimator

Since our proposed estimator $\hat{S}_P^2$ is biased, so we use the jacknifie technique to make the estimator unbiased. Following Sukhatme and Sukhatme [7], take $n = 2m$ and split the sample at random into two sub-samples of m units each. Let s_{yi}^2, s_{xi}^2, (i=1, 2) be unbiased estimator of the population variance S_y^2 and S_x^2 and s_y^2 and s_x^2 be the sample variance based on entire sample. So the unbiased version of the proposed estimator is given by

$$\hat{S}_{P(J)}^{2(U)} = \frac{(2N-n)}{N}\hat{S}_P^2 - \frac{(N-n)}{2n}\left\{\hat{S}_{P1}^2 + \hat{S}_{P2}^2\right\},$$

where $\hat{S}_{Pi}^2 = w_1 s_{yi}^2 + w_2 s_{yi}^2\left(\frac{s_{xi}^{*2}}{S_x^2}\right)$, ($i$=1, 2) and $w_1+w_2=1$.

The variance expression of the unbiased estimator $\hat{S}_{p(J)}^{2(U)}$ can be derived easily. It is to be noted that the variance expression of $\hat{S}_{p(J)}^{2(U)}$ and *MSE* expression of $\hat{S}_P^2$ are both equal. Hence we can prefer estimator $\hat{S}_{p(J)}^{2(U)}$ as compared to $\hat{S}_P^2$ because of unbiasedness.

In conclusion, the unbiased version of the proposed dual to variance ratio-type estimator is as efficient as the linear regression estimator and is more efficient than the usual variance estimator $\hat{S}_0^2$ and Isaki [2] estimator $\hat{S}_I^2$. In Table 2 (Pop. 2), the efficiency of the proposed estimator is much higher as compared to the estimators $\hat{S}_0^2$ and $\hat{S}_I^2$. Also in Table 2 (Pop.1), the Isaki [2] estimator $\hat{S}_0^2$ is inferior and even worse than the usual variance estimator $\hat{S}_0^2$. The unbiased estimator $\hat{S}_{p(J)}^{2(U)}$ is preferable as compared to the biased estimator $\hat{S}_P^2$.

References

1. **Cochran, W.G.** 1977. *Sampling Techniques*. New-York, John Wiley and Sons.
2. **Isaki, C.T.** 1983. Variance estimation using auxiliary information. *Jour. Amer. Statis. Assoc.* 78:117-123.
3. **Murthy, M.N.** 1967. *Sampling theory and methods*. Calcutta Statistical. Publising Society, Calcutta.
4. **Singh, H.P., Upadhyaya, L.N. and Namjoshi, U.D.** 1988. Estimation of finite population variance. *Curr. Sc.* 57:1331-1334.
5. **Singh, S.** 2003. *Advanced sampling theory with applications*. Kluwer Academic. Press.
6. **Srivenkataramana, T.** 1980. A dual to ratio estimator in sample surveys. *Biometrika* 67:199-204.
7. **Sukhatme, P.V., and Sukhatme, B.V.** 1970. *Sampling theory of surveys with applications*. Asia Publising House, New Delhi.
8. **Upadhyaya, L.N. and Singh, H.P.** 2001. Estimation of the population standard deviation using auxiliary information. *Am.Jour. Math. Manag. Sci.* 21:345-358.