Chilean Journal of Statistics
Vol. 5, No. 2, September 2014, 45–54

Sampling Theory
Research Paper

Improved Dual to Variance Ratio Type Estimators for Population Variance

Prayas Sharma and Rajesh Singh*

Department of Statistics, Banaras Hindu University, Varanasi, India

(Received: 12 July 2013 · Accepted in final form: 15 September 2014)

Abstract

In this paper we have suggested three improved dual to variance ratio type estimators for estimating the unknown population variance S_y^2 using auxiliary information. The optimum estimator in the suggested method has been identified along with its mean square error formula and it has been seen that the suggested estimator performs better than other existing estimators. Theoretical comparisons and empirical comparisons based on four real populations are carried out to judge the merits of proposed estimators over other traditional estimators and estimators considered here.

Keywords: Finite population variance · auxiliary variables · mean squared error · dual to variance ratio type estimator.

Mathematics Subject Classification: Primary 62D05.

1. Introduction

Estimation of the finite population variance has great significance in various fields such as agriculture, industry, medical and biological sciences where we come across population which are likely to be skewed. Variations are present in our daily life. It is a law of nature that no two things or individuals are exactly alike. For instance, a physician needs a full understanding of variation in the degree of human blood pleasure, body temperature and pulse rate for adequate prescription. A manufacturer needs constant knowledge of the level of in people's reaction about a particular product to know whether to increase or decrees his price of level of quality (see Singh and Solanki (2012)), many more situations can be seen in practice where the estimation of population variance assumes importance. Many authors have used the information on auxiliary information for increasing the precision of an estimator in estimating the population mean $(\bar{Y})$ and population variances (S_y^2) of study variable y. Das and Tripathi (1978), Isaki (1983) and Singh et al. (1988) studied the estimation of variance by using information on variance (S_x^2) for an auxiliary variable x. The other important contributions in this area are by Singh and Singh (2001, 2003), Kadilar and Cingi (2006), Grover (2010), Singh et al. (2011), Singh and Solanki (2012), Sharma and Singh (2013), and Singh and Malik (2014). Srivenkataramana (1980) first

*Corresponding author. Email: rsinghstat@gmail.com

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

time proposed the dual to ratio estimator for estimating the population mean. Singh and Tailor (2005), Tailor and Sharma (2009) and Sharma and Tailor (2010) proposed some ratio cum-product estimators. In addition, Shabbir (2006) suggested a dual to variance ratio type estimator.

Let $U = (U_1, U_2, \ldots, U_N)$ be the finite population of size N out of which a sample of size n is drawn according to simple random sampling without replacement technique. Let y and x be the study and the auxiliary variables respectively and y is positively correlated with x. Let

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i \quad \text{and} \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$$

be the population means of study and auxiliary variables and

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

be the respective sample means. Let

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

denote the population variances of y and x respectively. Similarly, one can obtain the sample variances

$$s_y^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(y_i - \bar{y})^2 \quad \text{and} \quad s_x^2 = \frac{1}{(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

of y and x respectively. Let

$$C_y = \frac{S_y}{\bar{Y}} \quad \text{and} \quad C_x = \frac{S_x}{\bar{X}}$$

denote the coefficient of variations of y and x, respectively. Here to estimate S_y^2 , it is assumed that S_x^2 is known because information on auxiliary variable x is known. Let us define

$$e_0 = \frac{(s_y^2 - S_y^2)}{S_y^2} \quad \text{and} \quad e_1 = \frac{(s_x^2 - S_x^2)}{S_x^2}$$

, therefore

$$E(e_0) = E(e_1) = 0, \qquad E(e_0^2) = f(\lambda_{40} - 1),$$
$$E(e_1^2) = f(\lambda_{04} - 1) \quad \text{and} \quad E(e_0 e_1) = f(\lambda_{22} - 1)$$

where

$$f = \left(\frac{1}{n} - \frac{1}{N}\right), \qquad \lambda_{pq} = \frac{\mu_{pq}}{\mu_{20}^{\frac{p}{2}}\mu_{02}^{\frac{q}{2}}} \quad \text{and} \quad \mu_{pq} = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^p (x_i - \bar{X})^q.$$

2. Existing estimators

The usual unbiased variance estimator is defined as

$$t_0 = s_y^2 \tag{1}$$

The variance of the estimator t_0 is given by

$$Var(t_0) = S_y^4\, f(\lambda_{40} - 1) \tag{2}$$

When the population variance of auxiliary variable S_x^2 is known, Isaki (1983) proposed a ratio type estimator for estimating S_y^2 given by

$$t_{Ik} = s_y^2 \left(\frac{S_x^2}{s_x^2}\right) \tag{3}$$

The bias and MSE of the estimator t_{Ik} up to first order of approximation are, respectively given by

$$Bias(t_{Ik}) = S_y^2\, f[(\lambda_{40} - 1) - (\lambda_{22} - 1)] \tag{4}$$

$$MSE(t_{Ik}) = S_y^4\, f[(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)] \tag{5}$$

Isaki (1983) suggested a regression estimator t_{reg}, given by

$$t_{reg} = s_y^2 + b(S_x^2 - s_x^2) \tag{6}$$

where b is the sample regression coefficient between s_y^2 and s_x^2. To the first order of approximation, the estimator t_{reg} is unbiased and its variance is given by

$$Var(t_{reg}) = S_y^4 f[(\lambda_{40} - 1) + b^2(\lambda_{04} - 1)S_x^4 - 2b(\lambda_{22} - 1)S_y^2\, S_x^2] \tag{7}$$

Substituting the optimum value of b, say b^* in equation (7), we get the minimum variance of t_{reg} as

$$Var(t_{reg})_{min} = S_y^4\, f(\lambda_{40} - 1)[1 - \rho^2(s_y^2, s_x^2)] \tag{8}$$

where

$$b^* = \frac{S_y^2(\lambda_{22} - 1)}{S_x^2(\lambda_{04} - 1)} \quad \text{and} \quad \rho^2(s_y^2, s_x^2) = \frac{(\lambda_{22} - 1)}{\sqrt{(\lambda_{40} - 1)(\lambda_{04} - 1)}}.$$

Shabbir (2006) suggested a dual to variance ratio estimator given by,

$$t_J = n_1 s_y^2 + n_2 s_y^2 \left(\frac{s_x^{*2}}{S_x^2}\right) \tag{9}$$

where $s_x^{*2} = \dfrac{NS_x^2 - ns_x^2}{N - n}$ due to Srivenkataramana (1980) and $s_x^{*2} = (1 + g)S_x^2 - gs_x^2$, with $g = \dfrac{n}{N - n}$

Bias and MSE of estimator t_J up to the first order of approximation are respectively given by

$$B(t_J) = -fn_2 S_y^2 \eta(\lambda_{22} - 1) \tag{10}$$

$$MSE(t_J) = fS_y^4 \left[(\lambda_{40} - 1) - \frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{11}$$

The equation (11) can be expressed as

$$MSE(t_J) = S_y^4 \left[1 + f(\lambda_{40} - 1) + f\frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)} - 1 - 2f\frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{12}$$

Or

$$MSE(t_J) = S_y^4 \left[(D - 1) - 2f\frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \tag{13}$$

Or

$$MSE(t_J) = S_y^4 Z \tag{14}$$

where

$$D = \left[1 + f(\lambda_{40} - 1) + f\frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right] \quad \text{and} \quad Z = \left[(D - 1) - 2f\frac{(\lambda_{22} - 1)^2}{(\lambda_{04} - 1)}\right].$$

The MSE of estimator t_J and t_{reg} are same which is less than the usual ratio estimator t_0 and Isaki (1983) estimator t_{Ik}.

3. Proposed estimators

A dual to variance regression type estimator t_1 is defined as

$$t_1 = w_1 s_y^2 + w_2(S_x^2 - s_x^{*2}) \tag{15}$$

where w_1 and w_2 are suitably chosen constants.
Expressing (15) in terms of e's,we get

$$t_1 - S_y^2 = (w_1 - 1)S_y^2 + w_1 S_y^2 e_0 + w_2 g S_x^2 e_1 \tag{16}$$

Taking expectations both sides of (16) we get the bias of estimator t_1 as

$$Bias(t_1) = S_y^2(w_1 - 1) \tag{17}$$

Squaring both sides of (16), we have

$$(t_1 - S_y^2)^2 = \Big[S_y^4(w_1 - 1)^2 + w_1^2 S_y^4 e_0^2 + w_2^2 g^2 S_x^4 e_1^2 + 2(w_1 - 1)w_1 S_y^4 e_0 + 2(w_1 - 1)w_2 g \sigma_y^2 \sigma_x^2 e_1 + 2w_1 w_2 g S_y^2 S_x^2 e_0 e_1\Big] \tag{18}$$

Taking expectations both sides, we get the mean square error of the estimator t_1 to first order of approximation, as

$$MSE(t_1) = \left[(1-2w_1)S_y^4 + w_1^2 A + w_2^2 B + w_1 w_2 C\right] \tag{19}$$

where $A = S_y^4\{1 + f(\lambda_{40-1})\}$, $B = g^2 f(\lambda_{04-1})S_x^4$, $C = 2gf(\lambda_{22}-1)S_y^2 S_x^2$.

Differentiating (19) with respect to w_1 and w_2 partially, equating them to zero, we get optimum values of w_1 and w_2 respectively, as

$$w_1^* = \frac{4S_y^4 B}{4AB - C^2}, \qquad w_2^* = \frac{-2S_y^4 C}{4AB - C^2}$$

Using the values of A, B and C we get the optimum values of w_1 and w_2 in terms of D, as

$$w_1^* = \frac{1}{D} \qquad \text{and} \qquad w_2^* = \frac{-S_y^2(\lambda_{22}-1)}{gS_x^2(\lambda_{04-1})D}. \tag{20}$$

Substituting these values of w_1^* and w_2^* from (20) in (19), we get the minimum MSE of the estimator t_1 as

$$MSE(t_1)_{min} = \frac{S_y^4}{D}\left[(D-1) - \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)D}\right] \tag{21}$$

Equation (21) can be rewrite in terms of Z, as

$$MSE(t_1)_{min} = \frac{S_y^4}{D}\left[Z + \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)D}\left(1 - \frac{1}{D}\right)\right] \tag{22}$$

Motivated by Gupta and Shabbir (2008), we suggest a dual to variance estimator as

$$t_2 = k_1 s_y^2 + k_2(S_x^2 - s_x^{*2})\left[2 - \left(\frac{s_x^{*2}}{S_x^2}\right)\right] \tag{23}$$

where k_1 and k_2 are suitably chosen constants.

Expressing the estimator t_2, in terms of e's we have

$$(t_2 - S_y^2) = (k_1-1)S_y^2 + k_1 S_y^2 e_0 + k_2 g S_x^2 e_1 + k_2 \alpha g^2 S_x^2 e_1^2 \tag{24}$$

Taking expectations both sides of (24), we get the bias of the estimator t_2 up to the first order of approximation, as

$$Bias(t_2) = (k_1-1)S_y^2 + k_2\alpha g^2 f(\lambda_{04}-1) \tag{25}$$

Squaring both sides of (24) and taking expectations both sides, we get the MSE of the estimator t_2 up to the first order of approximation, as

$$MSE(t_2) = (1-2k_1)S_y^4 + k_1^2 M + k_2^2 N + k_1 k_2 O \tag{26}$$

where $M = (1 + f(\lambda_{40}-1))S_y^2$, $N = g^2 f(\lambda_{04}-1)S_x^2$ and $O = g(\lambda_{22}-1)S_y^2 S_x^2$

Differentiating (26) with respect to k_1 and k_2 partially, equating them to zero, we get optimum values of k_1 and k_2 respectively, as

$$\left.\begin{aligned} k_1^* &= \frac{4S_y^4 N}{4MN - O^2} \\ k_2^* &= \frac{-2S_y^4 O}{4MN - O^2} \end{aligned}\right\} \tag{27}$$

Substituting these optimum values of k_1 and k_2 in equation (26), we get the minimum MSE of the estimator t_2 given as

$$MSE(t_2)_{min} = \frac{S_y^4}{D}\left[(D-1) - \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)D}\right] \tag{28}$$

We suggest another improved estimator t_3 as

$$t_3 = m_1 s_y^2 \left(\frac{s_x^{*2}}{S_x^2}\right) + m_2(S_x^2 - s_x^{*2}) \tag{29}$$

where m_1 and m_2 are constants chosen so as to minimise the mean squared error of the estimator t_3.

Equation (29) can be expressed in terms of e's as

$$(t_3 - S_y^2) = \left[(m_1-1)S_y^2 + m_1\, S_y^2\, e_0 - m_1 g S_y^2 e_1 + m_2 g S_x^2 e_1 - m_1 g S_y^2 e_0 e_1\right] \tag{30}$$

Taking expectations both sides of (30), we get the bias of the estimator t_3 up to the first order of approximation, as

$$Bias(t_3) = (m_1-1)S_y^2 - m_1 g S_y^2 f(\lambda_{22}-1) \tag{31}$$

Squaring both sides of (30) and taking expectations both sides we get the MSE of the estimator t_3 up to the first order of approximation, as

$$MSE(t_3) = \left[(1-2m_1)\sigma_y^4 + m_1^2 P + m_2^2 Q + m_1 m_2 R\right] \tag{32}$$

where $P = (1 + f(\lambda_{40}-1) + g^2 f(\lambda_{04}-1) - gf(\lambda_{22}-1))S_y^4$, $Q = g^2 S_x^2 f(\lambda_{04}-1)$ and $R = 2gf\{(\lambda_{22}-1) - g(\lambda_{04}-1)\}$

Now, for minimising MSE of estimator t_3 , differentiating (32) with respect to m_1 and m_2 partially, equating them to zero, we get the optimum values of m_1 and m_2 respectively, as

$$m_1^* = \frac{4Q\sigma_y^4}{4PQ - R^2} \qquad \text{and} \qquad m_2^* = \frac{-2R\sigma_y^4}{4PQ - R^2} \tag{33}$$

Substituting these optimum values of m_1 and m_2 in equation (32), we get the minimum MSE of the estimatort_3 given as

$$MSE(t_3)_{min} = \frac{S_y^4}{D}\left[(D-1) - \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)D}\right] \tag{34}$$

Now we establish the following theorem

Theorem 3.1 To the first degree of approximation

$$MSE(t_1)_{min} = MSE(t_2)_{min} = MSE(t_3)_{min} \geq \frac{S_y^4}{D}\left[(D-1) - \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)D}\right]$$

With equality holding

$$\left.\begin{matrix} w_1 = w_1^* \\ w_2 = w_2^* \end{matrix}\right\}, \qquad \left.\begin{matrix} k_1 = k_1^* \\ k_2 = k_2^* \end{matrix}\right\} \qquad \text{and} \qquad \left.\begin{matrix} m_1 = m_1^* \\ m_2 = m_2^* \end{matrix}\right\}$$

respectively for the estimators t_1,t_2 and t_3.

4. Efficiency Comparison

We compare the MSE of proposed estimator t_1 given in (22) with the MSE of Shabbir (2006) estimator given in (13) which is similar to the usual regression estimator t_{reg} , given in (8). We will have the condition as follows:

$$MSE(t_J) - MSE(t_1)_{min} > 0 \tag{35}$$

$$S_y^4 Z - \frac{S_y^4}{D}\left[Z + \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\left(1-\frac{1}{D}\right)\right] > 0$$

$$S_y^4 Z\left(1-\frac{1}{D}\right)\left[1 - \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)}\right] > 0$$

$$D > \frac{2f(\lambda_{22}-1)^2}{(\lambda_{04}-1)}$$

Or

$$\frac{f(\lambda_{22}-1)^2}{(\lambda_{04}-1)} - f(\lambda_{04}-1) < 1 \tag{36}$$

Theorem 4.1 To the first degree of approximation

$$\{MSE(t_J) = MSE(t_{reg})\} > \{MSE(t_1)_{min} = MSE(t_2)_{min} = MSE(t_3)\}$$

With equality holding if,

$$\frac{f(\lambda_{22}-1)^2}{(\lambda_{04}-1)} - f(\lambda_{04}-1) < 1$$

5. Empirical Study

Data Statistics: For comparison, we consider the following four data sets from various sources.
Population 1: Cochran (1977, p. 325)
y: Number of persons per block,
x: Number of rooms per block.
$N = 100, n = 10, S_y^2 = 214.69, S_x^2 = 56.76, \lambda_{40} = 2.2387, \lambda_{04} = 2.2523,$
$\lambda_{22} = 1.5432, X = 58.8, \rho = 0.6515$

Population 2: Cochran (1977, p. 152)
y: Number of inhabitants in 1930,
x: Number of inhabitants in 1920.
$N = 196, n = 49, S_y^2 = 151558.83, S_x^2 = 10900.42, \lambda_{40} = 8.5362, \lambda_{04} = 7.3617,$
$\lambda_{22} = 7.8780, X = 103.18, \rho = 0.9820$

Population 3: Cochran (1977, p. 203)
y: Actual weight of peaches on each tree,
x: Eye estimate of weight of peaches on each tree.
$N = 200, n = 10, S_y^2 = 99.18, S_x^2 = 85.09, \lambda_{40} = 1.9249, \lambda_{04} = 2.5932,$
$\lambda_{22} = 2.1149, X = 56.9, \rho = 0.9937$

Population 4: Sukhatme and Sukhatme (1970, p. 185)
y: Wheat acreage in 1937,
x: Wheat acreage in 1936
$N = 170, n = 10, S_y^2 = 26456.89, S_x^2 = 22355.76, \lambda_{40} = 3.1842, \lambda_{04} = 2.2030,$
$\lambda_{22} = 2.5597, X = 265.8, \rho = 0.977$

Table 1. Relative efficiency of different estimators w.r.t. t_0 in percentage

Estimator	Population 1	Population 2	Population 3	Population 4
t_0	100.000	100.000	100.000	100.000
t_{Ik}	88.188	5310.923	320.810	815.608
t_{reg}	123.489	7536.324	639.149	1347.978
t_J	123.382	7536.323	639.152	1347.981
t_1	134.637	7547.859	647.931	1368.535
t_2	134.637	7547.859	647.931	1368.535
t_3	134.637	7547.859	647.931	1368.535

Table 1 shows that the comparisons of estimators on the basis of four data set. Table 1 exhibits that the all the proposed estimators t_1, t_2 and t_3 are equally efficient but more efficient than the usual unbiased estimator, ratio and regression estimators proposed by Isaki (1983) and estimator proposed by Shabbir (2006) in the sense of having least mean square error. Thus any of the proposed estimators can be picked up preferably over others estimators considered here.

6. Conclusion

The present paper considered the problem of estimating the population variance S_y^2 of study variable y using auxiliary information. We have derived the biases and mean square errors formulae to the first degree of approximation. In efficiency comparison, it has been shown that the proposed estimators are more efficient than the other usual unbiased estimator, Isaki (1983), and Shabbir (2006) estimators. These results have been satisfied empirically with the help of four population earlier considered by Shabbir (2006).

Acknowledgement

Authors are thankful to the Editor-in-Chief Manuel Galea and unknown learned referees for their helpful comments that aided in improving the article.

References

Cochran, W.G., 1977. Sampling Techniques. New-York, John Wiley and Sons.

Das, A.K., Tripathi, T. P., 1978. Use of auxiliary information in estimating the finite Population variance. Sankhya C. 4, 139-148.

Gupta, S., Shabbir, J., 2008. Variance estimation in simple random sampling using Auxiliary information. Hacettepe Journal of Mathematics and Statistics. 37(1), 57-67.

Grover, L. K ., 2010. A correction note on improvement in variance estimation using auxiliary information. Commun. Stat. Theo. Meth. 39, 753-764.

Isaki, C.T., 1983. Variance estimation using auxiliary information. Jour. Amer. Statis. Assoc. 78, 117-123.

Kadilar, C., Cingi, H., 2006. Improvement in variance estimation using auxiliary information. Hacettepe Journal of Mathematics and Statistics. 35(1), 111-115.

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

Sharma, B.,Tailor, R., 2010. A new ratio-cum-dual to ratio estimator of finite population Mean in simple random sampling. Global Journal of Science Frontier Research. 10(1), 27-31.

Sharma, P., Singh, R., 2013. A generalized class of estimators for finite population variance in presence of measurement errors. Jour. of Moder. Appl. Stat. Methd. 12(2), 231-241.

Singh, H.P., Singh, R., 2001. Improved ratio-type estimator for variance using auxiliary information. J Indian Soc Agric Stat. 54(3), 276-287.

Singh,H.P., Singh, R., 2003. Estimation of variance through regression approach in two phase sampling. Aligarh Journal of Statistics. 23, 13-30.

Singh, H. P.,Solanki, R.S., 2012. A new procedure for variance estimation in simple random sampling using auxiliary information. Stat. Paps. DOI 10.1007/s00362-012-0445-2.

Singh, H.P., Tailor, R., 2005. Estimation of finite population mean using known correlation coefficient between auxiliary characters. Statistica. 65(4), 407-418.

Singh, H.P., Upadhyaya, L.N.,Namjoshi, U.D., 1988. Estimation of the population standard deviation using auxiliary information. Am. Jour. Math. Manag. Sci. 21, 345-358.

Singh, R., Chauhan, P., Sawan, N.,Smarandache, F., 2011. Improved exponential estimator for population variance using two auxiliary variables. Italian Jour. Of Pure and Appld. Math. 28, 103-110.

Singh, R., Malik, S., 2014. Improved estimator of population variance using information on auxiliary attribute in simple random sampling. Appl. Math. Comp. 235, 43-49.

Srivenkataramana, T., 1980. A dual to ratio estimator in sample surveys. Biometrika. 67(1), 199-204.

Sukhatme, P.V.,Sukhatme, B.V., 1970. Sampling theory of surveys with applications. Asia Publishing House, New Delhi.

Tailor, R., Sharma, B., 2009. Modified ratio-cum-product estimator of finite population

Mean using known coefficient of variation and known coefficient of kurtosis. Statistics in Transition-new series. 10(1), 15-24.