"Science Stays True Here"
Journal of Mathematics and Statistical Science, Volume 2016, 178-188 | Science Signpost Publishing

An Efficient Dual to Ratio and Product Estimator of Population Variance in Sample Surveys

Subhash Kumar Yadav

Department of Mathematics and Statistics (A Centre of Excellence), Dr. R M L Avadh University, Faizabad-224001, India.

Sheela Misra

Department of and Statistics, University of Lucknow, Lucknow-226007, India.

Ravendra Kumar*

Department of Mathematics, Dr Ram Manohar Lohiya Govt. Degree College, Aonla, Bareilly, U.P., India.

Shailendra Verma

Department of Business Management and Entrepreneurship, Dr. R M L Avadh University, Faizabad-224001, U.P., India.

Shailendra Kumar

Department of Microbiology, Dr. R M L Avadh University, Faizabad-224001, U.P., India.

Abstract

The present manuscript deals with the estimation of population variance using auxiliary variable. Here we proposed an efficient dual to ratio and product type estimator of population variance of study variable utilizing auxiliary information in the form of coefficient of kurtosis and the population mean of the auxiliary variable. To the first order of approximations, the bias and the mean squared error of the proposed variance have been obtained. The optimum value of the characterizing scalar has been obtained. This optimum value of characterizing scalar minimizes the mean squared error of the proposed estimator. The minimum value of the mean squared error has been obtained for this optimum value of the characterizing scalar. A comparison of the proposed estimator has been made with the mentioned existing estimators of population variance. Through an empirical study, the performances of different estimators of population variance are judged by calculating mean squared errors of different estimators. It has been seen that the proposed estimators has minimum mean squared error among all mentioned estimators.

Keywords: Main variable, auxiliary variable, bias, mean squared error, efficiency.

Corresponding author: Ravendra Kumar, Department of Mathematics, Dr Ram Manohar Lohiya Govt. Degree College, Aonla, Bareilly, U.P., India. E-mail: rkpmbly@gmail.com.

1. Introduction

It is well known that for estimating any parameter, the most suitable estimator is the corresponding statistic. So for estimation the population variance the most suitable estimator is the sample variance. Population variance is one of the important measures of dispersion. Its role in our day today life is very important such as we are interested in knowing the estimate of the variation in the production of a particular crop, blood pressure, temperature etc.

The role of auxiliary information is of very much importance as it enhances the efficiency of the estimator for estimating the population parameter of the main variable under study. It is used at both the stages of designing and estimation of survey sampling. In the present manuscript, we have used it at the estimation stage only. In the present study, the study variable is *Y* and the auxiliary variable is *X* and both the variables are closely related with each other. When the study variable and the auxiliary variables have positively association with each other and the line of regression of the study variable *Y* on the auxiliary variable *X* passes through origin, then the ratio type estimators are used for improved estimation of the parameters of the population under study. Whereas the product type estimators are used for improved estimation of parameters under consideration when the auxiliary variable *X* and the study variable *Y* are negatively correlation with each other. The regression type estimators are used for the improvement in estimation, when the line of regression does not pass through the origin.

Let the population to be investigated is finite and consists of *N* distinct and identifiable units. Let $(x_i, y_i),\ i = 1, 2, ..., n$ be a random sample of size n from this bivariate population (*X,* Y) using a Simple Random Sampling without Replacement (*SRSWOR)* scheme. Let $\bar{X}$ and $\bar{Y}$ be the population means of the auxiliary and the study variables respectively, and let $\bar{x}$ and $\bar{y}$ are the corresponding sample means which are unbiased estimators of population means $\bar{X}$ and $\bar{Y}$ respectively.

2. Review of Literature of Variance Estimators

As mentioned earlier, the most appropriate estimator of population variance is the corresponding sample variance defined by,

$$t_0 = s_y^2 \tag{2.1}$$

The above estimator is unbiased and up to the first order of approximation, its variance is:

$$V(t_0) = \gamma S_y^4 (\lambda_{40} - 1) \tag{2.2}$$

where,

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2, \quad \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i, \quad \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\mu_{02}^{s/2}},$$

$$\mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^r (X_i - \bar{X})^s, \quad \gamma = \frac{1-f}{n} \text{ and } f = \frac{n}{N}$$

Isaki (1983) used the auxiliary information in the form of population variance and sample mean of auxiliary variable and proposed the following traditional ratio estimator of population variance as:

$$t_R = s_y^2 \left(\frac{S_x^2}{s_x^2} \right) \tag{2.3}$$

where

$$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2, \quad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2, \quad \bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i .$$

The *Bias* and Mean Square Error (*MSE*) of the estimator t_R, up to the first order of approximations respectively are,

$$B(t_R) = \gamma S_y^2 [(\lambda_{04} - 1) - (\lambda_{22} - 1)] \tag{2.4}$$

$$MSE(t_R) = \gamma S_y^4 [(\lambda_{40} - 1) + (\lambda_{04} - 1) - 2(\lambda_{22} - 1)] \tag{2.5}$$

Singh *et al*. (2011), based on Bahl and Tuteja (1991) exponential ratio type estimator for the population mean, proposed the following exponential ratio type estimator for the population variance as,

$$t_1 = s_y^2 \exp\left[\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right] \tag{2.6}$$

The MSE of the estimator t_1, up to the first order of approximation is given by,

$$MSE(t_1) = \gamma S_y^4 \left[(\lambda_{40} - 1) + \frac{(\lambda_{04} - 1)}{4} - (\lambda_{22} - 1) \right] \tag{2.7}$$

The product type estimator for the population variance, based on Bahl and Tuteja (1991) product type estimator can be defined as,

$$t_2 = s_y^2 \exp\left[\frac{s_x^2 - S_x^2}{s_x^2 + S_x^2} \right] \tag{2.8}$$

The MSE of the estimator t_2, up to the first order of approximation is given by,

$$MSE(t_2) = \gamma S_y^4 \left[(\lambda_{40} - 1) + \frac{(\lambda_{04} - 1)}{4} + (\lambda_{22} - 1) \right] \tag{2.9}$$

Upadhyaya and Singh (2001) utilized the mean of the auxiliary variable and proposed the following modified ratio estimator of population variance as,

$$t_3 = s_y^2 \left[\frac{\bar{X}}{\bar{x}} \right] \tag{2.10}$$

The bias and the mean squared error of the estimator t_3 up to the first order of approximation respectively are,

$$B(t_3) = \gamma S_y^2 [C_x^2 - \lambda_{21} C_x] \tag{2.11}$$

$$MSE(t_3) = \gamma S_y^4 [(\lambda_{40} - 1) + C_x^2 - 2\lambda_{21} C_x] \tag{2.12}$$

where $\lambda_{21} = \dfrac{\mu_{21}}{\mu_{20}\sqrt{\mu_{02}}}$

The product type estimator based on the Upadhyaya and Singh (2001) estimator in (2.10) may be proposed as,

$$t_4 = s_y^2 \left[\frac{\bar{x}}{\bar{X}} \right] \tag{2.13}$$

The bias and the mean squared error of the estimator t_4 up to the first order of approximation respectively are,

$$B(t_4) = \gamma S_y^2[C_x^2 + \lambda_{21} C_x] \tag{2.14}$$

$$MSE(t_4) = \gamma S_y^4[(\lambda_{40} - 1) + C_x^2 + 2\lambda_{21} C_x] \tag{2.15}$$

Asgar et al. (2014) by using the population mean of the auxiliary variable, proposed the following improved ratio and product type estimators of population variance respectively as,

$$t_5 = s_y^2 \exp\left[\frac{\bar{X} - \bar{x}}{\bar{X} + \bar{x}}\right] \tag{2.16}$$

$$t_6 = s_y^2 \exp\left[\frac{\bar{x} - \bar{X}}{\bar{x} + \bar{X}}\right] \tag{2.17}$$

The bias and mean squared error of above estimators up to the first order of approximation respectively are,

$$B(t_5) = \gamma S_y^2[\frac{C_x^2}{8} - \frac{\lambda_{21}}{2} C_x] \tag{2.18}$$

$$MSE(t_5) = \gamma S_y^4[(\lambda_{40} - 1) + \frac{C_x^2}{4} - \lambda_{21} C_x] \tag{2.19}$$

$$B(t_6) = \gamma S_y^2[\frac{C_x^2}{8} + \frac{\lambda_{21}}{2} C_x] \tag{2.20}$$

$$MSE(t_6) = \gamma S_y^4[(\lambda_{40} - 1) + \frac{C_x^2}{4} + \lambda_{21} C_x] \tag{2.21}$$

Many more estimators of population variance in the literature have been proposed by different authors utilizing various parameters of the auxiliary variable. In the present study we have also proposed a dual to ratio and product type estimator of population variance using coefficient of kurtosis and the population mean of the auxiliary variable.

3. Proposed Estimator

Motivated by Vishwakarma et al (2014) and Upadhyaya and Singh (2001), we propose an improved ratio estimator of population variance as,

$$\hat{S}_{\alpha}^{2*} = s_y^2\left[\frac{\bar{x}^* + \alpha\bar{X}}{\bar{X} + \alpha\,\bar{x}^*}\right] \tag{3.1}$$

where α is a suitably chosen characterizing constant and is obtained by minimizing the MSE of the proposed estimator $\hat{S}_{\alpha}^{2*}$ and $\bar{x}^* = (N\bar{X} - n\bar{x})/(N-n) = (1+g)\bar{X} - g\bar{x}$, $g = n/(N-n)$.

In order to study the large sample properties of the proposed estimator $\hat{S}_{\alpha}^{2*}$, we define $s_y^2 = S_y^2(1+e_0)$ and $\bar{x} = \bar{X}(1+e_1)$ such that $E(e_i) = 0$ for $(i = 0,1)$ and $E(e_0^2) = \gamma(\lambda_{40} - 1)$, $E(e_1^2) = \gamma C_x^2$, $E(e_0 e_1) = \gamma\,\lambda_{21} C_x$.

Expressing $\hat{S}_{\alpha}^{2*}$ in terms of e_i's ($i = 0,1$), we have

$$\hat{S}_{\alpha}^{2*} = S_y^2(1+e_0)\left[\left\{1 - \frac{ge_1}{(1+\alpha)}\right\}\left\{1 - \frac{\alpha g e_1}{(1+\alpha)}\right\}^{-1}\right] \tag{3.2}$$

Expanding the right hand side of (3.2), simplifying and retaining the terms up to the first order of approximation, we have

$$\hat{S}_{\alpha}^{2*} = S_y^2\left[1 + e_0 - Fge_1 - Fge_0e_1 - \frac{F\alpha}{(1+\alpha)}g^2e_1^2\right] \tag{3.3}$$

where, $F = \left(\dfrac{1-\alpha}{1+\alpha}\right)$.

Subtracting S_y^2 on both sides of (3.3), we get

$$\hat{S}_{\alpha}^{2*} - S_y^2 = S_y^2\left[e_0 - Fge_1 - Fge_0e_1 - \frac{F\alpha}{(1+\alpha)}g^2e_1^2\right] \tag{3.4}$$

Taking expectations on both sides of (3.4) and putting the values of different expectations, we get the bias of $\hat{S}_{\alpha}^{2*}$, up to the first order of approximations as,

$$B(\hat{S}_{\alpha}^{2*}) = -\gamma S_y^2 \left[Fg\lambda_{21}C_x + \frac{F\alpha}{(1+\alpha)} g^2 C_x^2 \right] \tag{3.5}$$

From (3.4), we have up to the first order of approximation as,

$$\hat{S}_{\alpha}^{2*} - S_y^2 \approx S_y^2 (e_0 - Fge_1) \tag{3.6}$$

Squaring both sides of (3.6), we have

$$(\hat{S}_{\alpha}^{2*} - S_y^2)^2 = S_y^4 (e_0 - Fge_1)^2 = S_y^4 [e_0^2 + F^2 g^2 e_1^2 - 2Fge_0 e_1] \tag{3.7}$$

Taking expectations on both sides of (3.7) and putting the values of different expectations, up to the first order of approximation, we get the mean squared error of $\hat{S}_{\alpha}^{2*}$ as,

$$MSE(\hat{S}_{\alpha}^{2*}) = \gamma S_y^4 [(\lambda_{40} - 1) + F^2 g^2 C_x^2 - 2Fg\lambda_{21}C_x] \tag{3.8}$$

Which is minimum for,

$$F = \left(\frac{\lambda_{21}}{gC_x} \right) \tag{3.9}$$

The minimum mean squared error of $\hat{S}_{\alpha}^{2*}$ for this optimum value of F is,

$$MSE_{\min}(\hat{S}_{\alpha}^{2*}) = \gamma S_y^4 [(\lambda_{40} - 1) - \lambda_{21}^2] \tag{3.10}$$

4. Efficiency Comparison

From equation (3.10) and (2.2), we have

$$V(t_0) - MSE_{\min}(\hat{S}_{\alpha}^{2*}) = \gamma S_y^4 \lambda_{21}^2 > 0 \tag{4.1}$$

From above equation, we see that the proposed estimator $\hat{S}_{\alpha}^{2*}$ is better that the estimator t_0 as it has lesser mean squared error than the estimator t_0.

From equation (3.10) and (2.5), we have

$$MSE(t_R)-MSE_{\min}(\hat{S}_{\alpha}^{2*})=\gamma S_y^4[(\lambda_{04}-1)-2(\lambda_{22}-1)+\lambda_{21}^2]>0 \text{, if, } [(\lambda_{04}-1)+\lambda_{21}^2]>2(\lambda_{22}-1) \quad (4.2)$$

When the above condition is met out, the proposed estimator is better than the estimator t_R.

From equation (3.10) and (2.7), we have

$$MSE(t_1)-MSE_{\min}(\hat{S}_{\alpha}^{2*})=\gamma S_y^4\left[\frac{(\lambda_{04}-1)}{4}-(\lambda_{22}-1)+\lambda_{21}^2\right]>0 \text{, if, } \frac{(\lambda_{04}-1)}{4}+\lambda_{21}^2>(\lambda_{22}-1) \quad (4.3)$$

When the above condition is fulfilled, the proposed estimator $\hat{S}_{\alpha}^{2*}$ performs better than the estimator t_1. Thus it will be preferable over t_1 under above condition.

From equation (3.10) and (2.9), we have

$$MSE(t_2)-MSE_{\min}(\hat{S}_{\alpha}^{2*})=\gamma S_y^4\left[\frac{(\lambda_{04}-1)}{4}+(\lambda_{22}-1)+\lambda_{21}^2\right]>0, \quad (4.4)$$

Under the above condition, the proposed estimator $\hat{S}_{\alpha}^{2*}$ is better than the estimator t_2 as it will have lesser mean squared error as compared to estimator t_2. Thus it will be preferred over t_2.

From equation (3.10) and (2.12), we have

$$MSE(t_3)-MSE_{\min}(\hat{S}_{\alpha}^{2*})=\gamma S_y^4[(C_x^2-2\lambda_{21}C_x+\lambda_{21}^2]=\gamma S_y^4[C_x-\lambda_{21}]^2>0 \quad (4.5)$$

Thus we see that the proposed estimator $\hat{S}_{\alpha}^{2*}$ has lesser mean squared error than the estimator t_3. Thus it will be more efficient than the estimator t_3.

From equation (3.10) and (2.15), we have

$$MSE(t_4)-MSE_{\min}(\hat{S}_{\alpha}^{2*})=\gamma S_y^4[(C_x^2+2\lambda_{21}C_x+\lambda_{21}^2]=\gamma S_y^4[C_x+\lambda_{21}]^2>0 \quad (4.6)$$

Thus we see that the proposed estimator $\hat{S}_{\alpha}^{2*}$ has lesser mean squared error than the estimator t_4. Thus it will be preferred over the estimator t_4.

From equation (3.10) and (2.19), we have

$$MSE(t_5) - MSE_{\min}(\hat{S}_{\alpha}^{2*}) = \gamma S_y^4[\frac{C_x^2}{4} - \lambda_{21}C_x + \lambda_{21}^2] = \gamma S_y^4[\frac{C_x}{2} - \lambda_{21}]^2 > 0 \quad (4.7)$$

It is clear from above equation that the proposed estimator $\hat{S}_{\alpha}^{2*}$ is having lesser mean squared error as compared to estimator t_4. Thus it will be more efficient than the estimator t_4.

From equation (3.10) and (2.21), we have

$$MSE(t_6) - MSE_{\min}(\hat{S}_{\alpha}^{2*}) = \gamma S_y^4[\frac{C_x^2}{4} + \lambda_{21}C_x + \lambda_{21}^2] = \gamma S_y^4[\frac{C_x}{2} + \lambda_{21}]^2 > 0 \quad (4.8)$$

Which shows that the proposed estimator $\hat{S}_{\alpha}^{2*}$ is better than the estimator t_6 as it has lesser mean squared error as compared to t_6.

5. Numerical Illustration

To judge the theoretical finding of the proposed estimator in the form of its performance of the proposed estimator regarding its mean squared error, an empirical study has been carried out using following two real populations. The Sources and the parameters for two populations are given below;

Population I- Source: Murthy [6]

X: output

Y: Number of workers

$N = 25$, $n = 25$, $\bar{X} = 283.875$, $\bar{Y} = 33.8465$, $C_y = 0.352$, $C_x = 0.746$, $\rho_{yx} = 0.9136$, $\lambda_{40} = 2.2667$, $\lambda_{04} = 3.65$, $\lambda_{21} = 1.0475$, $\lambda_{22} = 2.3377$

Population II- Source: Gujarati [3]

X: Average (miles per gallon)

Y: Top Speed (miles per hour)

$N = 81$, $n = 21$, $\bar{X} = 112.4568$, $\bar{Y} = 2137.086$, $C_y = 0.1248$, $C_x = 0.48$,

$\rho_{yx} = -0.691135$, $\lambda_{40} = 3.59$, $\lambda_{04} = 6.82$, $\lambda_{21} = 1.4137$, $\lambda_{22} = 2.110$

Table 1. Percent relative efficiency (PRE) of different estimators with respect to t_0.

Estimator	Pop-I	Pop-II
t_0	100.00	100.00
t_R	102.04	*
t_1	214.15	*
t_2	*	50.24
t_3	489.40	177.00
t_4	*	61.99
t_5	131.54	202.86
t_6	*	77.86
$\hat{S}_{\alpha}^{2*}$ (Proposed)	**747.49**	**521.28**

*Data is not applicable.

6. Result and Conclusion

The present manuscript deals with the estimation of population variance of study variable using auxiliary information. We have proposed a dual to ratio and product type estimator of population variance using coefficient of kurtosis and the mean of the auxiliary variable. Up to the first order of approximation, the bias and the mean squared error of the proposed estimator have been obtained. The optimum value of the characterizing scalar which minimizes the mean squared error of the proposed estimator is obtained up to the first order of approximation. From the table-1 of the numerical study and given in table-1 and the theoretical discussion in section-4, we see that the proposed estimator is having very less mean squared error as compared to other mentioned estimators of population variance. Thus it is the most efficient estimator of population variance among the class of all mentioned estimator of population variance. Therefore the proposed estimator should be preferred over all other estimators for the estimation of population variance.

References

[1]. Amber Asghar, Aamir Sananullah and Muhammad Hanif (2014): Generalized Exponential Type Estimators for Population variance in Survey Sampling. Revista Colombiana deEstadística, 37 (1), 213–224.

[2]. Bhal, S. and R.K. Tuteja, 1991. Ratio and product type exponential estimator. Information and Optimization Sciences. 12: 159-163.

[3]. Gujarati, D. (2004), Basic Econometrics, 4 edn, The McGraw-Hill Companies.

[4]. Isaki, C. T. (1983). Variance estimation using auxiliary information, Journal of American Statistical Association, 78, 117-123.

[5]. Khan, M. and Arunachalam, A. (2014). Estimation of Population Variance using the Knowledge of Kurtosis of an Auxiliary Variable under Simple Random Sampling, International Journal of Applied Science and Mathematics, 1, 2, 63-67.

[6]. Murthy, M. (1967), Sampling Theory and Methods, Calcutta Statistical Publishing Society,Kolkatta, India.

[7]. Singh, R. S., Chauhan, P., Sawan, N. & Smarandache, F. (2011), 'Improved exponential estimator for population variance using two auxiliary variables', Italian Journal of Pure and Applied Mathematics 28, 101–108.

[8]. Upadhyaya, L.N. and Singh, G.N. (2001) Chain-type estimators using transformed auxiliary variable in two-phase sampling. A.M.S.E., Vol. 38, N0.1, 2, 1-9.

[9]. Vishwakarma, G.K., Gangele, R.K. and Singh, R. (2014). An Efficient Variant of Dual to Ratio and Product Estimator in Sample Surveys, The Philippine Statistician, 63, 2, 21-29.