

A Family of Difference-cum- Exponential Type Estimators for Estimating the Population Variance Using Auxiliary Information in Sample Surveys

H.S.Jhaji¹ and Kusam Lata²

¹ Professor, Dept. of Statistics, Punjabi University Patiala, 147002

²Asst. Professor, Swami Devi Dyal College, Panchkula, 134115.

Keywords: Regression-type estimator, Mean squared error, Double sampling, Auxiliary variable, Efficiency

Abstract. Using auxiliary information, a family of difference-cum-exponential type estimators for estimating the population variance of variable under study have been proposed under double sampling design. Expressions for bias, mean squared error and its minimum values have been obtained. The comparisons have been made with the regression-type estimator by using simple random sampling at both occasions in double sampling design. It has also been shown that better estimators can be obtained from the proposed family of estimators which are more efficient than the linear regression type estimator. Results have also been illustrated numerically as well as graphically

1. Introduction

In survey sampling, taking the advantage of the correlation between the auxiliary variable x and study variable y , the construction of efficient ratio and product type estimators for the population variance has widely been made by many statisticians to assess the variations in population. Because it is well known that the use of auxiliary information in the construction of estimators can increase the efficiency of the estimators of parameter of interest. When the population variance S_x^2 of auxiliary variable x , which is highly correlated with the study variable y , is known in advance, several estimators have been defined by different authors such as Das and Tripathi (1978), Isaki (1983), Ahmed et al.(2003), Jhaji et al.(2005,2011), Kadilar and Cingi (2006), Pradhan et al.(2010) in the literature for estimating the population variance of study variable y . Sometimes information on population variance S_x^2 of auxiliary variable x is not known in advance then we generally use two phase (double) sampling design. In the two- phase sampling design, a large preliminary random sample (called first phase sample) is drawn and auxiliary information are taken on sample units which are used to estimate the value of unknown population variance S_x^2 of auxiliary variable x . Then second phase sample is drawn either from the first phase sample or independently from the population and observations are taken on study as well as auxiliary variables.

In present paper we have proposed a family of Difference-Cum-Exponential type estimators for estimating the population variance of study variable using auxiliary information under double sampling design. The efforts have been made to compare the proposed family of estimators under simple random sampling design as a special case.

2. Notations and Results

A preliminary large random sample (first phase sample) of size n' is drawn from a finite population of size N and both auxiliary variable x and study variable y are measured on it. Then second phase random sample of size $n (< n')$ is drawn from the first phase sample.

Let Y_i and X_i denote the respective values of variables y and x on the i^{th} ($i=1,2,\dots,N$) unit of the population and the corresponding small letters denote the values in the samples.

Denoting

$$\begin{aligned} s_y^2 &= (n-1)^{-1} \sum_{i=1}^n (y_i - \bar{y})^2 & s_y'^2 &= (n'-1)^{-1} \sum_{i=1}^{n'} (y_i - \bar{y})^2 \\ s_x^2 &= (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 & s_x'^2 &= (n'-1)^{-1} \sum_{i=1}^{n'} (x_i - \bar{x})^2 \\ S_Y^2 &= (N-1)^{-1} \sum_{i=1}^N (Y_i - \bar{Y})^2 & S_X^2 &= (N-1)^{-1} \sum_{i=1}^N (X_i - \bar{X})^2 \\ S_{XY} &= (N-1)^{-1} \sum_{i=1}^N (Y_i - \bar{Y})(X_i - \bar{X}) \\ \mu_{ab} &= \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^a (Y_i - \bar{Y})^b & \lambda_{ab} &= \frac{\mu_{ab}}{\mu_{20}^{a/2} \mu_{02}^{b/2}} \end{aligned}$$

where $s_x'^2, s_x^2$ and $s_y'^2, s_y^2$ are the sample variances of variable x and y based on the sampling units of first and second phase samples of size n' and n respectively.

Defining

$$\begin{aligned} \epsilon_0 &= \frac{s_y^2}{S_y^2} - 1 & \epsilon_1 &= \frac{s_y'^2}{S_y^2} - 1 \\ \epsilon_2 &= \frac{s_x^2}{S_x^2} - 1 & \epsilon_3 &= \frac{s_x'^2}{S_x^2} - 1 \end{aligned}$$

Assuming that

$$\begin{aligned} E(\epsilon_0) &= E(\epsilon_1) = E(\epsilon_2) = E(\epsilon_3) = 0 \\ \left. \begin{aligned} E(\epsilon_0^2) &= \frac{\text{var}(s_y^2)}{S_y^4} & E(\epsilon_1^2) &= \frac{\text{var}(s_y'^2)}{S_y^4} \\ E(\epsilon_2^2) &= \frac{\text{var}(s_x^2)}{S_x^4} & E(\epsilon_3^2) &= \frac{\text{var}(s_x'^2)}{S_x^4} \\ E(\epsilon_0 \epsilon_1) &= \frac{\text{cov}(s_y^2, s_y'^2)}{S_y^4} & E(\epsilon_2 \epsilon_3) &= \frac{\text{cov}(s_x^2, s_x'^2)}{S_x^4} \\ E(\epsilon_0 \epsilon_2) &= \frac{\text{cov}(s_x^2, s_y^2)}{S_x^2 S_y^2} & E(\epsilon_1 \epsilon_2) &= \frac{\text{cov}(s_x^2, s_y'^2)}{S_x^2 S_y^2} \\ E(\epsilon_0 \epsilon_3) &= \frac{\text{cov}(s_x'^2, s_y^2)}{S_x^2 S_y^2} & E(\epsilon_1 \epsilon_3) &= \frac{\text{cov}(s_x'^2, s_y'^2)}{S_x^2 S_y^2} \end{aligned} \right\} \end{aligned} \tag{2.1}$$

3. The Proposed Family of Estimators and Its Results

When we don't have an information about any parameter of the auxiliary variable, then we propose a family of estimators of the population variance S_x^2 of the study variable y under the sampling design defined in section 2 as

$$\hat{S}_{y_{\text{est}}}^2 = \left[s_y^2 + \theta (s_y'^2 - s_y^2) \right] \left[\left\{ 1 - \frac{\theta (s_x'^2 - s_x^2)}{s_x'^2} \right\} \left\{ \exp 2 \left(\frac{s_x'^2 - s_x^2}{s_x'^2 + s_x^2} \right) \right\} \right]^\alpha \tag{3.1}$$

Where $\theta > 0$ and α are any constants.

To obtain the bias and mean square error of estimator $\hat{S}_{y_{\text{est}}}^2$, upto first order of approximation, we expand $\hat{S}_{y_{\text{est}}}^2$ in terms of ε 's and retaining terms upto second degree of approximation.

$$\begin{aligned} \hat{S}_{y_{\text{est}}}^2 = & S_y^2 \left[1 + \varepsilon_0 + \theta (\varepsilon_1 - \varepsilon_0) + \alpha (1 - \theta) (\varepsilon_3 - \varepsilon_2) + \alpha \theta (1 - \theta) (\varepsilon_1 \varepsilon_3 - \varepsilon_1 \varepsilon_2 - \varepsilon_0 \varepsilon_3 + \varepsilon_0 \varepsilon_2) \right. \\ & + \alpha (1 - \theta) (\varepsilon_0 \varepsilon_3 - \varepsilon_0 \varepsilon_2) + \alpha \left(\left(\frac{1}{2} - \theta \right) (\varepsilon_3 - \varepsilon_2)^2 - \frac{1}{2} (\varepsilon_2^2 - \varepsilon_3^2) \right) \\ & \left. - \alpha \theta (\varepsilon_2 \varepsilon_3 - \varepsilon_3^2) + \frac{\alpha (\alpha - 1)}{2} (\theta - 1)^2 (\varepsilon_3 - \varepsilon_2)^2 \right] \end{aligned} \tag{3.2}$$

Noting that expectation exists and finite, we take expectation in (3.2) and using the results of section 2, we have

$$E(\hat{S}_{y_{\text{est}}}^2) = S_y^2 + S_y^2 \left[\left\{ \alpha (1 - \theta)^2 \right\} \left\{ \frac{\text{cov}(s_x'^2, s_y^2) - \text{cov}(s_x^2, s_y^2)}{S_x^2 S_y^2} \right\} + \left\{ \frac{\alpha (\alpha - 1)}{2} (1 - \theta)^2 + \alpha (1 - \theta) \right\} \left\{ \frac{V(s_x^2) - V(s_x'^2)}{S_x^4} \right\} \right]$$

which implies that

$$\begin{aligned} \text{Bias}(\hat{S}_{y_{\text{est}}}^2) = & S_y^2 \left[\left\{ \alpha (1 - \theta)^2 \right\} \left\{ \frac{\text{cov}(s_x'^2, s_y^2) - \text{cov}(s_x^2, s_y^2)}{S_x^2 S_y^2} \right\} \right. \\ & \left. + \left\{ \frac{\alpha (\alpha - 1)}{2} (1 - \theta)^2 + \alpha (1 - \theta) \right\} \left\{ \frac{V(s_x^2) - V(s_x'^2)}{S_x^4} \right\} \right] \end{aligned} \tag{3.3}$$

Upto first order of approximation the mean square error (MSE) of estimator $\hat{S}_{y_{\text{est}}}^2$ is obtained from (3.2) and using the results of section 2, as

$$\begin{aligned} \text{MSE}(\hat{S}_{y_{\text{est}}}^2) = & E \left[\hat{S}_{y_{\text{est}}}^2 - S_y^2 \right]^2 \\ = & V(s_y^2) + (\theta^2 - 2\theta) \{ V(s_y^2) - V(s_y'^2) \} \\ & + \alpha^2 (\theta - 1)^2 \frac{S_y^4}{S_x^4} \{ V(s_x^2) - V(s_x'^2) \} \\ & - 2\alpha (\theta - 1)^2 \frac{S_y^2}{S_x^2} \{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \} \end{aligned} \tag{3.4}$$

The expression (3.4) depends upon two unknown constants α and θ , so keeping the value of θ fixed, we differentiate (3.4) w.r.t α and equating to zero, we get

$$2\alpha (\theta - 1)^2 \frac{S_y^4}{S_x^4} \{ V(s_x^2) - V(s_x'^2) \} - 2(\theta - 1)^2 \frac{S_y^2}{S_x^2} \{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \} = 0$$

After solving the above equation, we get the optimum value of α

$$\alpha)_{\text{opt}} = \frac{S_x^2 \{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \}}{S_y^2 \{ V(s_x^2) - V(s_x'^2) \}} \quad (3.5)$$

Substituting the optimum value of α from (3.5) in (3.4), we get minimum mean square error,

Min. MSE ($\hat{S}_{y_{\theta\alpha}}^2$) as

$$\begin{aligned} \text{Min. MSE} (\hat{S}_{y_{\theta\alpha}}^2) &= V(s_y^2) + (\theta^2 - 2\theta) \{ V(s_y^2) - V(s_y'^2) \} \\ &\quad - (\theta - 1)^2 \frac{\{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \}}{\{ V(s_x^2) - V(s_x'^2) \}} \end{aligned} \quad (3.6)$$

Theorem1: Upto first order of approximation, the bias of estimator $\hat{S}_{y_{\theta\alpha}}^2$ is

$$\begin{aligned} \text{Bias} (\hat{S}_{y_{\theta\alpha}}^2) &= S_y^2 \left[\{ \alpha(1-\theta)^2 \} \left\{ \frac{\text{cov}(s_x'^2, s_y^2) - \text{cov}(s_x^2, s_y^2)}{S_x^2 S_y^2} \right\} \right. \\ &\quad \left. + \left\{ \frac{\alpha(\alpha-1)}{2} (1-\theta)^2 + \alpha(1-\theta) \right\} \left\{ \frac{V(s_x^2) - V(s_x'^2)}{S_x^4} \right\} \right] \end{aligned}$$

And its mean square error is

$$\begin{aligned} \text{MSE} (\hat{S}_{y_{\theta\alpha}}^2) &= V(s_y^2) + (\theta^2 - 2\theta) \{ V(s_y^2) - V(s_y'^2) \} \\ &\quad + \alpha^2 (\theta - 1)^2 \frac{S_y^4}{S_x^4} \{ V(s_x^2) - V(s_x'^2) \} \\ &\quad - 2\alpha (\theta - 1)^2 \frac{S_y^2}{S_x^2} \{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \} \end{aligned}$$

Theorem2: Upto first order of approximation, the MSE of estimator $\hat{S}_{y_{\theta\alpha}}^2$ is minimized for

$$(\alpha)_{\text{opt}} = \frac{S_x^2 \{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \}}{S_y^2 \{ V(s_x^2) - V(s_x'^2) \}}$$

And its minimum value is given by

$$\begin{aligned} \text{Min. MSE} (\hat{S}_{y_{\theta\alpha}}^2) &= V(s_y^2) + (\theta^2 - 2\theta) \{ V(s_y^2) - V(s_y'^2) \} \\ &\quad - (\theta - 1)^2 \frac{\{ \text{cov}(s_x^2, s_y^2) - \text{cov}(s_x'^2, s_y^2) \}}{\{ V(s_x^2) - V(s_x'^2) \}} \end{aligned}$$

Special case: when simple random sampling is applied for selection of samples in both phases of double sampling design we have

$$\left. \begin{aligned} \text{var}(s_x^2) &= \left(\frac{1}{n} - \frac{1}{N}\right) S_x^4 (\lambda_{40} - 1) & \text{var}(s_x'^2) &= \left(\frac{1}{n'} - \frac{1}{N}\right) S_x^4 (\lambda_{40} - 1) \\ \text{var}(s_y^2) &= \left(\frac{1}{n} - \frac{1}{N}\right) S_y^4 (\lambda_{04} - 1) & \text{var}(s_y'^2) &= \left(\frac{1}{n'} - \frac{1}{N}\right) S_y^4 (\lambda_{04} - 1) \\ \text{cov}(s_x^2, s_y^2) &= \left(\frac{1}{n} - \frac{1}{N}\right) S_x^2 S_y^2 (\lambda_{22} - 1) & \text{cov}(s_x'^2, s_y'^2) &= \left(\frac{1}{n'} - \frac{1}{N}\right) S_x^2 S_y^2 (\lambda_{22} - 1) \end{aligned} \right\} \quad (3.7)$$

Substituting results of (3.7) in (3.3) to (3.6) respectively, we have *Bias*

$$\left(\hat{S}_{y_{\theta\alpha}}^2 \right) = \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^2 \left\{ \left[\frac{\alpha(\alpha-1)}{2} (\theta-1)^2 + \alpha(1-\theta) \right] (\lambda_{40} - 1) - \alpha(1-\theta)^2 (\lambda_{22} - 1) \right\} \quad (3.8)$$

$$MSE(\hat{S}_{y_{\theta\alpha}}^2) = \frac{S_y^4}{n} (\lambda_{04} - 1) + \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 \left\{ (\theta^2 - 2\theta)(\lambda_{04} - 1) + \alpha^2 (\theta-1)^2 (\lambda_{40} - 1) - 2\alpha(\theta-1)^2 (\lambda_{22} - 1) \right\} \quad (3.9)$$

$$(\alpha)_{opt} = \frac{\lambda_{22} - 1}{\lambda_{40} - 1} \quad (3.10)$$

$$Bias(\hat{S}_{y_{\theta\alpha}}^2)_{opt} = \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^2 \left[\frac{(\lambda_{22} - 1)(\lambda_{22} - \lambda_{40})(\theta-1)^2}{2(\lambda_{40} - 1)} - \frac{(\lambda_{22} - 1)^2 (\theta-1)^2}{(\lambda_{40} - 1)} + (\lambda_{22} - 1)(1-\theta) \right]$$

$$MSE_{min}(\hat{S}_{y_{\theta\alpha}}^2) = \frac{S_y^4}{n} (\lambda_{04} - 1) + \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 \left[(\theta^2 - 2\theta)(\lambda_{04} - 1) - \frac{(\lambda_{22} - 1)^2 (\theta-1)^2}{(\lambda_{40} - 1)} \right] \quad (3.11)$$

From (3.11), we see that *Min.MSE*($\hat{S}_{y_{\theta\alpha}}^2$) will vary with the change in variation of θ so range of variation of θ can be obtained at which proposed family of estimator is better than the existing ones.

4. Comparison

For comparing the proposed family of estimators with the linear regression type estimator s_{lrd}^2 considered by

$$s_{lrd}^2 = s_y^2 + \frac{\text{cov}(s_y^2, s_x^2)}{\text{var}(s_x^2)} (s_x'^2 - s_x^2)$$

we first obtain expressions of its bias and mean square error, under double sampling design using simple random sampling at both phases, upto first order of approximation,

$$MSE(s_{lrd}^2) = \frac{1}{n} S_y^4 (\lambda_{04} - 1) - \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 \frac{(\lambda_{22} - 1)^2}{(\lambda_{40} - 1)} \quad (4.1)$$

Using (3.11) and (4.1) and after some algebra, we obtain

$$MSE(s_{lrd}^2) - MSE_{min}(\hat{S}_{y_{\theta\alpha}}^2) = \left(\frac{1}{n} - \frac{1}{n'}\right) S_y^4 (\lambda_{04} - 1) (\theta^2 - 2\theta) (\rho_v^2 - 1) \geq 0 \quad (4.2)$$

$$\text{Where } \rho_v = \frac{(\lambda_{22} - 1)}{\sqrt{(\lambda_{04} - 1)(\lambda_{40} - 1)}}$$

$$\Rightarrow MSE_{min}(\hat{S}_{y_{\theta\alpha}}^2) \leq MSE(s_{lrd}^2) \quad \text{if } 0 < \theta < 2 \quad (4.3)$$

5. Numerical Illustration

For comparing the efficiency of the proposed estimator ($\hat{S}_{y\theta\alpha}^2$) with Regression-type estimate (S_{lrd}^2) under double sampling design, we take the empirical population considered in literature (source: Institute of Statistics, Republic of Turkey). This empirical data concerns the level of apple production (1 unit=100 tonnes) as the variate of interest and number of apple trees (1 unit=100 trees) as the auxiliary variate in 106 villages in the Marmarian Region respectively in 1999. The values of population parameters obtained are given in Table 1. The biases, mean squared error and relative efficiency of proposed estimator ($\hat{S}_{y\theta\alpha}^2$) w.r.t regression-type estimator (S_{lrd}^2) are given for some different values of θ in table 2.

Table 5.1: Value of population parameters

N	n'	n	S_y	λ_{04}	λ_{40}	λ_{22}	λ_{31}
106	48	20	64.25	80.13	25.71	33.30	29.67

Table 5.2: Mean Squared Error and Relative Efficiency

Values of θ	Mean Squared Error		Efficiency	
	S_{lrd}^2	$\hat{S}_{y\theta\alpha}^2$	S_{lrd}^2	$\hat{S}_{y\theta\alpha}^2$
0	46483968.52	46483968.5	100	100
0.5	46483968.52	32757889.7	100	142
1.0	46483968.52	28182530.1	100	165
1.5	46483968.52	32757889.7	100	142
2.0	46483968.52	46483968.5	100	100

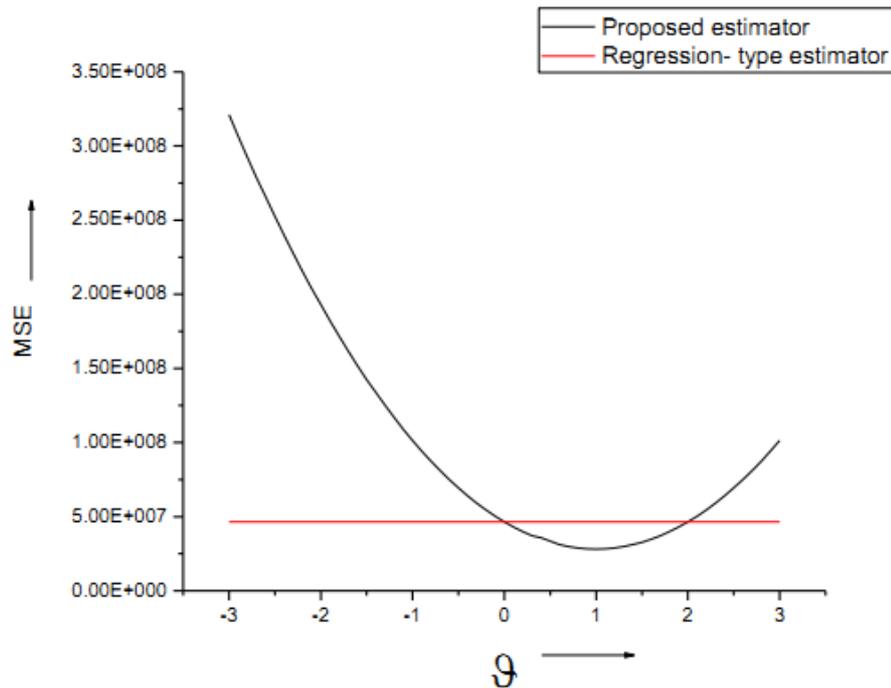


Figure 5.1: Graphical Representation of MSE of Estimators

6. Conclusion

From table 5.2, we observe that in double sampling for $0 < \theta < 2$ the efficiency of proposed estimator $(\hat{S}_{y_{ba}}^2)$ got increased as compared to the Regression type estimator (s_{lrd}^2) . The graphical representation also predicts that there is gain in efficiency of proposed estimator $(\hat{S}_{y_{ba}}^2)$ over the Regression-type estimator (s_{lrd}^2) for $0 < \theta < 2$. Hence we conclude that the proposed estimator will always be better than existing Regression-type estimator under double sampling design for $0 < \theta < 2$

REFERENCES:

- [1] A.K. Das and B. Tripathi, "Use of auxiliary information in estimating the finite population variance", Sankhya, C40,139-148,1978.
- [2] M.S. Ahmed, A.D. Walid and A.O.H. Ahmed," Some estimators for finite population variance under two-phase sampling", Statistics in Transtition, 6(1),143-150,2003.
- [3] H.S.Jhajj, M.K.Sharma and L.K.Grover,"An efficient class of chain estimators of population variance under sub-sampling scheme", Jour Japan Statistical Society, 35(2),273-286,2005.
- [4] C.Kadilar and H.Cingi,"Ratio estimator for the population variance in simple and stratified random sampling", Applied Mathematics and Computation, 173, 1047-1059, 2006.
- [5] B.K.Pradhan, "Regression type estimators of finite population variance under multiphase sampling", Bull.Malays.Math. Sci. Soc.,(2) 33(2), 345-353,2010.
- [6] H.S.Jhajj and G.S.Walia," A generalized difference-cum-ratio type estimator for the population variance in double sampling", International journal of Applied Mathematics, 41:4,2011.