Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.12785/jsap/040104

An improved Exponential-Type Estimator for Finite Population Variance Using Transformation of Variables

Mursala Khan∗

Department of Mathematics, COMSATS Institute of Information Technology, Abbottabad 22060, Pakistan

Received: 29 Oct. 2014, Revised: 30 Dec. 2014, Accepted: 2 Jan. 2015
Published online: 1 Mar. 2015

Abstract: In this article we have proposed an improved exponential-type for estimating the unknown population variance of the study variable, using transformation on both the study as well as on the auxiliary variable. The properties of the suggested estimator have been discussed up to first order of approximation. We also have derived some efficiency comparison conditions under which the proposed variance estimator performed better than the usual unbiased estimator, traditional ratio estimator and [1] estimators. Theoretical efficiency conditions are verified numerically by taking some real data sets taken from the literature, with a result in increasing the efficiency.

Keywords: Auxiliary information, exponential-type estimator, transformation, bias, mean squared error and efficiency.

1 Introduction

It is known fact, that in many real situations auxiliary information is available or may be made available in cheap cost in surveys. These auxiliary information is frequently used both at the design as well as at the estimation stage, if this information is carefully used it may give better results in terms of efficiency. Auxiliary information such as population mean, population variance, population coefficient of variation, population quartiles, population deciles, population percentiles, population inter quartile range, population coefficient of kurtosis and population Coefficient of skewness, etc. In the present study, we will use these auxiliary informations to estimate the population variance of the study variable efficiently.
In order to estimate the population mean we use survey data but in many cases the mean is not a suitable average because it fluctuate from small or large observations or outliers in a set of data, to overcome this difficulty we come to variance. Many statisticians worked on the estimation of population variance using auxiliary information, [2] suggested the usual ratio for estimating population variance of the study variable y, using auxiliary information of the auxiliary variable x. Das and [3], [4], [5,6], [7,8], [9], [10], [11], [12], [13], [14], [15], [1] and [16] proposed different estimators for population variance using auxiliary information.Let us consider a finite population of size N of different units . Let Y and X be the study and the auxiliary variable with corresponding values x_i and y_i respectively for i_{th} unit is defined on a finite population U. Let $\overline{Y}$ and $\overline{X}$ be the corresponding population means of the study as well as auxiliary variable respectively. Also let S_y^2 and S_x^2 be the corresponding population variances of the study as well as auxiliary variable respectively and let C_y and C_x be the coefficient of variation of the study as well as auxiliary variable respectively, and ρ_{yx} be the correlation coefficient between x and y.
In order to estimate the unknown population variance we take a sample of size n units from the population U by using simple random sample without replacement. Let y and x be the study and the auxiliary variable with corresponding values y_i and x_i respectively for $i-th$ unit in the sample. Let $\overline{y}$ and $\overline{x}$ be the corresponding sample means of the study as well as auxiliary variable respectively.
Also let $\hat{S}_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\overline{y})$ and $\hat{S}_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\overline{x})$ be the corresponding sample variances of the study as well as auxiliary variable respectively.
Some notations that we use are, given as below.

∗ Corresponding author e-mail: mursala.khan@yahoo.com

$\lambda_{40}=\frac{\mu_{40}}{\mu_{20}^2}$ and $\lambda_{04}=\frac{\mu_{04}}{\mu_{02}^2}$ be the coefficient of kurtosis of the study variable y and the auxiliary variable x respectively. Also let $\lambda_{rs}=\frac{\mu_{rs}}{\mu_{20}^{\frac{r}{2}}\mu_{02}^{\frac{s}{2}}}$, $\mu_{rs}=\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{Y})^r(x_i-\overline{X})^s$, $\lambda_{22}=\frac{\mu_{22}}{\mu_{20}\mu_{02}}$, $\lambda_{21}=\frac{\mu_{21}}{\mu_{20}\mu_{02}^{\frac{1}{2}}}$, $\lambda_{40}^*=\lambda_{04}-1$, $\lambda_{22}^*=\lambda_{22}-1$, $\theta=\frac{1}{n}$

2 Existing Estimators

In this section we will discuss some of the existing estimators that are available in the literature. When there is no auxiliary information the usual unbiased estimator to estimate the population variance of the study variable is

$$t_0=s_y^2$$

The bias and variance, of the estimator t_0 up to first order of approximation are given by

$$MSE(t_0)=\theta S_y^4\lambda_{40}^* \tag{1}$$

When the population variance of the auxiliary variable is known the usual ratio estimator to estimate the population variance of the study variable suggested by Isaki [2] and is, given by

$$t_r=S_x^2\frac{s_y^2}{s_x^2}$$

The bias and mean squared error, up to first order of approximation are given by

$$Bias(t_r)=\lambda S_y^2(\lambda_{04}^*-\lambda_{22}^*) \tag{2}$$

$$MSE(t_r)=\lambda S_y^4(\lambda_{04}^*+\lambda_{40}^*-2\lambda_{22}^*) \tag{3}$$

Recently [1] suggested a family of estimators to estimate the population variance of the study variable using transformations on both the study as well as the auxiliary variable when coefficient of variation of an auxiliary variable is known.

$$t_{rp}=s_u^2\left[\alpha\left(\frac{C_v^2}{\hat{C}_v^2}\right)^{\gamma}+(1-\alpha)\left(\frac{\hat{C}_v^2}{C_v^2}\right)^{\eta}\right]-a \tag{4}$$

where α,γ,a,b,d and η are constants. Let also $s_u^2=s_y^2+a$, $\hat{C}_v^2=\left(b\hat{C}_v^2+dC_v^2\right)=\left(b\frac{s_x^2}{x^2}+d\frac{S_x^2}{X^2}\right)$ and $C_v^2=(b+d)C_v^2=(b+d)\frac{S_x^2}{X^2}$ The bias and mean squared error, up to first order of approximation are given by

$Bias(t_{rp}) = \left(\frac{S_y^2+a}{2n}\right)\theta_1\left[\theta_1\{\eta(\eta-1)+\alpha(\gamma+\eta)(\gamma-\eta+1)\}\{\lambda_{04}^*+4C_x^2-4\lambda_{03}C_x\}+2\{\eta-\alpha(\gamma+\eta)\}\{(3C_x^2-2\lambda_{03}C_x)+\theta_0(\lambda_{22}^*-2\lambda_{21}C_x)\}\right]$

Where $\theta_0=\frac{S_y^2}{S_y^2+a}$ and $\theta_1=\frac{b}{b+a}$

$$MSE(t_{rp})=\theta S_y^4\left[\lambda_{40}^*-\frac{(\lambda_{22}^*-2\lambda_{21}C_x)^2}{\lambda_{04}^*+4C_x^2-4\lambda_{03}C_x}\right] \tag{5}$$

Some members of the suggested family of [1] estimators $t_{rp}$are

$$MSE(t_1)=\theta S_y^4\left[\lambda_{40}^*+(\lambda_{04}^*+4C_x^2-4\lambda_{03}C_x)-(\lambda_{22}^*-2\lambda_{21}C_x)\right] \tag{6}$$

$$MSE(t_i)=\theta S_y^4\left[\lambda_{40}^*+\lambda_j^2(\lambda_{04}^*+4C_x^2-4\lambda_{03}C_x)+2\lambda_j(\lambda_{22}^*-2\lambda_{21}C_x)\right] \tag{7}$$

where $i=2,3,4$ and $j=1,2,3$. $\lambda_1=-\left(\frac{\bar{X}C_x^2}{XC_x^2-\lambda_{04}}\right)$, $\lambda_2=-\left(\frac{C_x^2}{C_x^2-\rho}\right)$, $\lambda_3=-\left(\frac{\lambda_{04}C_x^2}{\lambda_{04}C_x^2-\rho}\right)\left(\frac{S_y^2+\lambda_{04}}{S_y^2}\right)$

3 The proposed Estimator

On the lines of [1], we proposed the following exponential-type estimator using transformations on both the study as well as the auxiliary variable, when variance of the auxiliary variable is known.

$$t_m = s_u^2 \exp\left[\frac{S_v^2}{\alpha S_v^2 + (1-\alpha) S_v^2} - \frac{\alpha S_v^2 + (1-\alpha) S_v^2}{S_v^2}\right] - a \tag{8}$$

Here ss_u^2 and s_v^2 denote the population variances of the transformed variables u and v respectively. where $s_u^2 = s_y^2 + a$, $\hat{S}_v^2 = b\hat{S}_x^2 + dS_x^2$ and $S_v^2 = (b+d) S_x^2$ Also a, b, d and α are constants and the value of α is to be determined for the purpose of minimum mean squared error.
Rewriting equation 8, we have

$$t_m = \left(s_y^2 + a\right) \exp\left[\frac{(b+d) S_x^2}{\alpha\left(bs_x^2 + dS_x^2\right) + (1-\alpha)(b+d) S_x^2} - \frac{\alpha\left(bs_x^2 + dS_x^2\right) + (1-\alpha)(b+d) S_x^2}{(b+d) S_x^2}\right] - a$$

In order to study large sample properties of t_m , we write
$s_y^2 = S_y^2(1+\zeta_0)$, $s_x^2 = S_x^2(1+\zeta_1)$, $\bar{x} = \bar{X}(1+\zeta_2)$, Such that $E(\zeta_0) = E(\zeta_1) = E(\zeta_2) = 0$, $E\left(\zeta_0^2\right) = \theta\lambda_{40}^*$, $E\left(\zeta_1^2\right) = \theta\lambda_{04}^*$, $E\left(\zeta_2^2\right) = \theta C_x^2$, $E(\zeta_0\zeta_1) = \theta(\lambda_{22} - 1)$, $E(\zeta_0\zeta_2) = \theta\lambda_{21}C_x$ and $E(\zeta_1\zeta_2) = \theta\lambda_{03}C_x$
In term of ζ_i^s, we have

$$t_m = S_y^2\left(1 - 2\alpha A\zeta_1 + 3\alpha^2 A^2\zeta_1^2 + \zeta_0 - 2\alpha A\zeta_0\zeta_1\right) + a - 2a\alpha A\zeta_1 + 3a\alpha^2 A^2\zeta_1^2 - a \tag{9}$$

The bias and minimum mean squared error up to first order of approximation are given by

$$Bias(t_m) = \theta\alpha A\left[3\alpha A\lambda_{04}^*\left(S_y^2 - a\right) - 2S_y^2\lambda_{22}^*\right] \tag{10}$$

$$MSE(t_m)_{min} = \theta S_y^4\lambda_{40}^*\left(1 - \frac{\lambda_{22}^{2*}}{\lambda_{40}^*\lambda_{04}^*}\right) \tag{11}$$

After finding the MSE of the proposed estimator the constants have no role but it provided to us Minimum MSE. Where the optimum value of α is given by.

$$\alpha_{opt} = \frac{S_y^2\lambda_{22}^*}{2A\left(S_y^2 + a\right)\lambda_{04}^*}$$

4 Efficiency Comparison

In this section we have compared the efficiency of the proposed estimator t_m with the usual estimator $t_0, t_r, t_{rp}, t_1, t_2, t_3$ and t_4
(1)- By 1 and 11,

$$MSE(t_0) - MSE(t_m)_{min} \geq 0 \text{ if } \frac{\lambda_{22}^{2*}}{\lambda_{04}^{2*}} \geq 0$$

(2)- By 3 and 11,

$$MSE(t_r) - MSE(t_m)_{min} \geq 0 \text{ if } \left[\lambda_{04}^* - \lambda_{22}^*\left(2\lambda_{04}^* - \lambda_{22}^*\right)\right] \geq 0$$

(3)- By 5 and 11,

$$MSE(t_r) - MSE(t_m)_{min} \geq 0 \text{ if } \left[\frac{\lambda_{22}^{2*}}{\lambda_{04}^{2*}} - \frac{\left(\lambda_{22}^* - 2\lambda_{21}C_x\right)^2}{\lambda_{04}^* + 4C_x^2 - 4\lambda_{03}C_x}\right] \geq 0$$

(4)- By 6 and 11,

$$MSE(t_r) - MSE(t_m)_{min} \geq 0 \text{ if } \left(\lambda_{04}^* + 4C_x^2 - 4\lambda_{03}C_x - 2\lambda_{22}^* + 4\lambda_{21}C_x + \frac{\lambda_{22}^{2*}}{\lambda_{04}^{2*}}\right] \geq 0$$

(4)- By 6 and 11,

$$MSE\left(t_r\right) - MSE\left(t_m\right)_{min} \geq 0 \text{ if } \left[\lambda_j^2\left(\lambda_{04}^* + 4C_x^2 - 4\lambda_{03}C_x\right) + 2\lambda_j\left(\lambda_{22}^* - 24\lambda_{21}C_x\right) + \frac{\lambda_{22}^{2*}}{\lambda_{04}^{2*}}\right] \geq 0$$

for $i = 2,3,4$ and $j = 1,2,3$. The proposed estimator is more efficient than the other existing estimator, if the above conditions are satisfied.

5 Numerical

For percent relative efficiency (PRE), we use the following formula.

$$PRE\left(k, t_0\right) = \frac{MES\left(t_0\right)}{MES\left(k\right)} \times 100,$$

where $k = t_0, t_r, t_{rp}, t_1, t_2, t_3, t_4$.
To look closely the gain of the proposed estimator, we consider the following real data sets from the literature of survey sampling.
Population-1: [17]
It consists of 142 cities of India with population (number of persons) 100,000 and above. The variates considered are:
X: Census population in the year 1961 and Y: Census population in the year 1971.
$N = 142, n = 15, \bar{Y} = 4015.2183, \bar{X} = 2900.3872, S_y = 8479.3380, S_x = 3672.4407, C_y = 2.1118, C_x = 2.1971, \rho_{xy} = 0.9948, \lambda_{40} = 40.8536, \lambda_{04} = 48.1567, \lambda_{03} = 40.2179, \lambda_{22} = 43.7615, \lambda_{21} = 5.9786$
Population-2: [18]
Y: Number of rooms per block and X: Number of persons per block.
$N = 100, n = 10, \bar{Y} = 101.1, \bar{X} = 58.8, S_y = 14.6595, S_x = 7.53228, C_y = 0.1450, C_x = 0.1281, \rho_{xy} = 0.6500, \lambda_{40} = 2.3523, \lambda_{04} = 2.2387, \lambda_{03} = 0.4861, \lambda_{30} = 0.3248, \lambda_{22} = 1.5432, \lambda_{21} = 0.5714$

6 Conclusion

Table 1 clearly indicates that the mean squared error of the proposed estimator is less than the mean squared errors of the existing estimators. Table 1 also shows that the proposed estimator had the largest gain in efficiency over the usual unbiased estimator , the usual ratio estimator suggested by [2] and the estimators suggested by [1]. Having the largest gain in efficiency the proposed estimator appeared to be the best one among all the estimators and would work very well in practical surveys.

References

[1] Yadav, R., Upadhyaya, L. N., Singh, H. P. and Chatterjee, S. A general procedure of estimating the population variance when coefficient of variation of an auxiliary variable is known in sample surveys, Qual Quant, 47, 2331-2339 (2013).
[2] Isaki, C. T. Variance estimation using auxiliary information, Journal of American Statistical Association, 78, 117-123 (1983).
[3] Das, A. K. and Tripathi, T. P. Use of auxiliary information in estimating the finite population variance, Sankhya, C, 40, 139-148 (1978).
[4] Srivastava, S. K. and Jhajj, H. S. A class of estimators using auxiliary information for estimating finite population variance, Sankhya, C, 42, 87-96 (1980).
[5] Upadhyaya, L. N. and Singh, H. P. Use of auxiliary information in the estimation of population variance, Mathematical Forum, 4, 33-36 (1983).
[6] Upadhyaya, L. N. and Singh, H. P. An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample survey, Vikram Mathematical Journal, 19, 14-17 (1999).
[7] Singh, H. P. and Singh, R. Improved ratio-type estimator for variance using auxiliary information, Journal of the Indian Society of Agricultural Statistics, 54(3), 276-287 (2001).
[8] Singh, H. P. and Singh, R. Estimation of variance through regression approach in two phase sampling, Aligarh Journal of Statistics, 23, 13-30 (2003).
[9] Upadhyaya, L. N., Singh, H. P. and Singh, S.. A class of estimators for estimating the variance of the ratio estimator, Journal of Japan statistical Society, 34(1), 47-63 (2004).

[10] Kadilar, C. and Cingi, H. Ratio estimators for population variance in simple and Stratified sampling, Applied Mathematics and Computation, 173, 1047-1058 (2006).

[11] Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. Almost unbiased ratio and product type estimator of finite population variance using the knowledge of kurtosis of an auxiliary variable in sample survey, Octogon Mathematical Journal, 16(1), 123-130 (2008).

[12] Gupta, S and Shabbir, J. Variance estimation in simple random sampling using auxiliary information, Hacettepe Journal of mathematics and Statistics, 37, 57-67 (2008).

[13] Grover, L. K. A correction note on improvement in variance estimation using auxiliary information, Communication in Statistics Theory and Methods, 39, 753-764 (2010).

[14] Singh, R., Chauhan, P., Sawan, N. and Smarandache, F. Improved exponential estimator for population variance using two auxiliary variables, Italian Journal of Pure and Applied Mathematics 28, 103-110 (2011).

[15] Khan, M. and Shabbir, J. A ratio type estimator for the estimation of population variance using quartiles of an auxiliary variable, Journal of Statistics Applications and Probability, 2(3), 157-162 (2013).

[16] Khan, M and Arunachalam, A. 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), 2394-2894 (2014).

[17] Das, A. K. Contributions to the theory of sampling strategies based on auxiliary information, unpublished Ph.D. Thesis, Bidhan Chandra Krishi Vishwavidyalaya, Mohanpur, Nadia, West Bengal, India, (1980).

[18] Cochran, W. G. Sampling Techniques, JohnWiley& Sons, New York, NY, USA, (1977).