IMPROVED RATIO AND PRODUCT TYPE ESTIMATORS OF FINITE POPULATION MEAN IN SIMPLE RANDOM SAMPLING

Gajendra K. Vishwakarma*, Ravendra Singh**, P.C. Gupta**, Sarla Pareek**
*Department of Applied Mathematics, Indian School of Mines,
Dhanbad-826004, Jharkhand, India
vishwagk@rediffmail.com
**Department of Mathematics and Statistics, Banasthali University,
Jaipur-304022, Rajasthan, India
ravendrasingh84@gmail.com

ABSTRACT
In this paper, improved ratio and product type estimators have been developed for estimating the finite population mean of the study variable using auxiliary information in *simple random sampling (SRS)*. The expressions for the *bias* and *mean square error (MSE)* of the proposed estimators are obtained under first order of approximation. Theoretical and empirical studies have been done to demonstrate the efficiencies of the proposed estimators over other well known estimators.

KEYWORDS: Study variable, Auxiliary variable, bias, MSE, Efficiency.

MSC: 62D05

RESUMEN
En este trabajo estimadores del tipo razón y producto mejorados han sido desarrrollados para estimar la media de una población finita de la variable bajo estudio usando información auxiliar en el muestreo simple aleatorio (msa). Las expresiones del sesgo y del error cuadrático medio (ECM) de los estimadores propuestos son obtenidos bajo aproximacion de primer orden. Estudios teóricos y empíricos han sido desarrollados para demostar las eficiencias de los estimadores propuestos respecto a otros estimadores bien conocidos.

1. INTRODUCTION

The use of auxiliary information has become indispensable for improving the precision of the estimators of population parameters such as the mean and variance of a variable under study. A great variety of techniques such as the ratio, product and regression methods of estimation are commonly known in this regard. Auxiliary information can be used either at the design stage or at the estimation stage or at both the stages. Keeping this fact in view, large number of estimators have been suggested in sampling literature. Some noteworthy contributions in this direction have been made by Cochran(1940), Robson(1957), Murthy(1964), Singh (1967), Sahai (1979), Bahl and Tuteja (1991), Singh and Espejo (2003), Singh and Tailor (2005), Kadilar and Cingi (2005), Singh and Vishwakarma (2007, 2008), Shabbir *et al.* (2014), and many others.
Let U denote a finite population consisting of N units $\{U_1, U_2, ..., U_N\}$. Also, let (Y, X) denote the study variable and the auxiliary variable taking values (y_i, x_i), $(i = 1,2,..., N)$, respectively, on the i^{th} unit

U_i of the population U. On the assumption that the population mean $(\overline{X})$ of X is known, the estimate of population mean $(\overline{Y})$ of Y is obtained by selecting a sample of size n $(n < N)$ from the population U using simple random sampling without replacement (SRSWOR) scheme.

The conventional ratio and product estimators of $\overline{Y}$ are given by

$$\overline{y}_R = \overline{y}\left(\frac{\overline{X}}{\overline{x}}\right) \tag{1}$$

$$\overline{y}_P = \overline{y}\left(\frac{\overline{x}}{\overline{X}}\right) \tag{2}$$

where $\overline{y}$ and $\overline{x}$ are the sample means of Y and X, respectively.

Bahl and Tuteja (1991) suggested the following exponential type ratio and product estimators for the population mean $\overline{Y}$:

$$t_1 = \overline{y}\exp\left(\frac{\overline{X}-\overline{x}}{\overline{X}+\overline{x}}\right) \tag{3}$$

$$t_2 = \overline{y}\exp\left(\frac{\overline{x}-\overline{X}}{\overline{x}+\overline{X}}\right) \tag{4}$$

2. PROPOSED ESTIMATORS

We define the following improved ratio and product type estimators for the population mean $\overline{Y}$ in *SRSWOR*:

$$t_3 = \alpha\,\overline{y} + (1-\alpha)t_1 \tag{5}$$

$$t_4 = \beta\,\overline{y} + (1-\beta)t_2 \tag{6}$$

where α and β are real constants to be determined such that the *MSEs* of t_3 and t_4 are minimized.

Further, it is observed that the estimators t_3 and t_4 reduces to a set of estimators $\{\overline{y}, t_1, t_2\}$ by assigning suitable values to the constants α and β as follows:

(i) Usual unbiased estimator: $\overline{y}$ for $\alpha = \beta = 1$

(ii) Bahl and Tuteja (1991) ratio estimator: t_1 for $\alpha = 0$

(iii) Bahl and Tuteja (1991) product estimator: t_2 for $\beta = 0$

To obtain the *bias* and *MSE* of the estimators t_3 and t_4, we consider

$$\overline{y} = \overline{Y}(1+e_0), \overline{x} = \overline{X}(1+e_1)$$

Then, we have

$$\left.\begin{aligned} &E(e_0) = E(e_1) = 0 \\ &E(e_0^2) = \lambda C_y^2, E(e_1^2) = \lambda C_x^2 \\ &E(e_0 e_1) = \lambda \rho_{yx} C_y C_x \end{aligned}\right\} \tag{7}$$

where, $\lambda = \dfrac{1-f}{n}, f = \dfrac{n}{N},\ C_y^2 = \dfrac{S_y^2}{\overline{Y}^2}, C_x^2 = \dfrac{S_x^2}{\overline{X}^2}, \rho_{yx} = \dfrac{S_{yx}}{S_y S_x},$

$$S_y^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(y_i - \overline{Y})^2, S_x^2 = \frac{1}{(N-1)}\sum_{i=1}^{N}(x_i - \overline{X})^2,$$

$$S_{yx} = \frac{1}{(N-1)}\sum_{i=1}^{N}(y_i - \overline{Y})(x_i - \overline{X}).$$

Now, expressing (5) and (6) in terms of e_0, e_1, and retaining the terms of *e's* upto the second degree, we obtain

$$t_3 - \overline{Y} = \overline{Y}\left[e_0 - \left(\frac{e_1}{2} - \frac{3}{8}e_1^2 + \frac{e_0 e_1}{2}\right) + \alpha\left(\frac{e_1}{2} - \frac{3}{8}e_1^2 + \frac{e_0 e_1}{2}\right)\right] \tag{8}$$

$$t_4 - \overline{Y} = \overline{Y}\left[e_0 + \left(\frac{e_1}{2} - \frac{e_1^2}{8} + \frac{e_0 e_1}{2}\right) - \beta\left(\frac{e_1}{2} - \frac{e_1^2}{8} + \frac{e_0 e_1}{2}\right)\right] \tag{9}$$

Taking the expectations in (8), (9), and using results in (7), we obtain the *bias* of the estimators t_3 and t_4 to the terms of order $O(n^{-1})$ as

$$B(t_3) = \lambda\overline{Y}C_x^2\left[\left(\frac{3}{8} - \frac{k_{yx}}{2}\right) - \alpha\left(\frac{3}{8} - \frac{k_{yx}}{2}\right)\right] \tag{10}$$

$$B(t_4) = \lambda\overline{Y}C_x^2\left[\left(\frac{k_{yx}}{2} - \frac{1}{8}\right) - \beta\left(\frac{k_{yx}}{2} - \frac{1}{8}\right)\right] \tag{11}$$

where, $k_{yx} = \dfrac{\rho_{yx}C_y}{C_x}$.

Again, from (8) and (9), by neglecting the terms of *e's* having degree greater than one, we have

$$t_3 - \overline{Y} = \overline{Y}\left(e_0 - \frac{e_1}{2} + \frac{\alpha e_1}{2}\right) \tag{12}$$

$$t_4 - \overline{Y} = \overline{Y}\left(e_0 + \frac{e_1}{2} - \frac{\beta e_1}{2}\right) \tag{13}$$

Squaring both sides of (12) and (13), taking the expectation, and using results in (7), we obtain the *MSE* of the estimators t_3 and t_4 to the terms of order $O(n^{-1})$ as

$$MSE(t_3) = \lambda\overline{Y}^2\left[C_y^2 + \frac{C_x^2(1-\alpha)^2}{4} - \rho_{yx}C_yC_x(1-\alpha)\right] \tag{14}$$

$$MSE(t_4) = \lambda\overline{Y}^2\left[C_y^2 + \frac{C_x^2(1-\beta)^2}{4} + \rho_{yx}C_yC_x(1-\beta)\right] \tag{15}$$

2.1. Optimal Values of α and β

The optimal values of α and β, for which the *MSE* of the estimators t_3 and t_4 are minimized, are obtained by using the following conditions:

$$\frac{\partial}{\partial\alpha}MSE(t_3) = 0 \tag{16}$$

$$\frac{\partial}{\partial\beta}MSE(t_4) = 0 \tag{17}$$

On solving (16) and (17), we have

$$\alpha^* = 1 - 2k_{yx} \tag{18}$$

$$\beta^* = 1 + 2k_{yx} \tag{19}$$

where α^* and β^* denote the respective optimal values of α and β. Also, using these optimal values of α and β in (14) and (15), respectively, we obtain the minimum attainable *MSE* of the estimators t_3 and t_4 as

$$MSE(t_3)_{min} = MSE(t_4)_{min} = \lambda \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2) \tag{20}$$

Remark. The minimum attainable *MSEs* in (20) corresponds to the *MSEs* of *asymptotic optimum estimators (AOEs)* t_3^* and t_4^*, which are obtained on replacing α and β, in (5) and (6) by their respective optimal values, i.e., α^* and β^*. So, we have

$$t_3^* = (1 - 2k_{yx})\bar{y} + 2k_{yx} t_1$$

$$t_4^* = (1 + 2k_{yx})\bar{y} - 2k_{yx} t_2$$

and $MSE(t_3^*) = MSE(t_4^*) = \lambda \bar{Y}^2 C_y^2 (1 - \rho_{yx}^2)$

To the first degree of approximation, the *MSE* of the various estimators listed above are:

$$V(\bar{y}) = \lambda \bar{Y}^2 C_y^2 \tag{21}$$

$$MSE(\bar{y}_R) = \lambda \bar{Y}^2 \left[C_y^2 + C_x^2 (1 - 2k_{yx}) \right] \tag{22}$$

$$MSE(\bar{y}_P) = \lambda \bar{Y}^2 \left[C_y^2 + C_x^2 (1 + 2k_{yx}) \right] \tag{23}$$

$$MSE(t_1) = \lambda \bar{Y}^2 \left[C_y^2 + \frac{C_x^2}{4} - \rho_{yx} C_y C_x \right] \tag{24}$$

$$MSE(t_2) = \lambda \bar{Y}^2 \left[C_y^2 + \frac{C_x^2}{4} + \rho_{yx} C_y C_x \right] \tag{25}$$

3. EFFICIENCY COMPARISONS

For making efficiency comparisons of the estimators t_3 and t_4 with the existing estimators, we have from (14), (15), and (21) to (25),

(i) $MSE(t_3) < V(\bar{y})$ if

$$\alpha > 1 - 4k_{yx} \tag{26}$$

(ii) $MSE(t_4) < V(\bar{y})$ if

$$\beta > 1 + 4k_{yx} \tag{27}$$

(iii) $MSE(t_3) < MSE(\bar{y}_R)$ if

$$\alpha < 3 - 4k_{yx} \tag{28}$$

(iv) $MSE(t_4) < MSE(\bar{y}_P)$ if

$$\beta < 3 + 4k_{yx} \tag{29}$$

(v) $MSE(t_3) < MSE(t_1)$ if

$$\alpha < 2 - 4k_{yx} \tag{30}$$

(vi) $MSE(t_4) < MSE(t_2)$ if

$$\beta < 2 + 4k_{yx} \tag{31}$$

4. EMPIRICAL STUDY

To examine the merits of the proposed estimators t_3 and t_4 over other existing estimators, we have considered three natural population data sets as follows:

Population I - [Source: Johnston (1972)]

Y : Percentage of hives affected by disease

X : Mean January temperature

Z : Date of flowering of a particular summer species (number of days from January 1)

$N = 10, n = 4, \bar{Y} = 52, \bar{X} = 42, \bar{Z} = 200, \rho_{YX} = 0.80, \rho_{YZ} = -0.94, \rho_{XZ} = -0.73,$

$C_Y^2 = 0.0244, C_X^2 = 0.0170, C_Z^2 = 0.0021$

Population II - [Source: Singh (1969)]

Y : Number of females employed

X : Number of females in service

Z : Number of educated females

$N = 61, n = 20, \bar{Y} = 7.46, \bar{X} = 5.31, \bar{Z} = 179.00, \rho_{YX} = 0.7737, \rho_{YZ} = -0.2070,$

$\rho_{XZ} = -0.0033, C_Y^2 = 0.5046, C_X^2 = 0.5737, C_Z^2 = 0.0633$

Population III - [Source: Steel and Torrie (1960)]

Y : Log of leaf burn in sec

X : Potassium percentage

Z : Chlorine percentage

$N = 30, n = 6, \bar{Y} = 0.6860, \bar{X} = 4.6537, \bar{Z} = 0.8077, \rho_{YX} = 0.1794, \rho_{YZ} = -0.4996,$
$\rho_{XZ} = 0.4074, C_Y = 0.4803, C_X = 0.2295, C_Z = 0.7493$

The *percentage relative efficiencies (PREs)* are obtained for various suggested estimators of $\bar{Y}$ with respect to the usual unbiased estimator $\bar{y}$ and the findings are presented in Table 1.

Table 1: *Percentage Relative Efficiencies (PREs)* of various estimators with respect to $\bar{y}$

Estimators	Auxiliary variables used	Population I	Population II	Population III
$\bar{y}$	-	100.00	100.00	100.00
$\bar{y}_R$	X	276.85	205.34	94.62
$\bar{y}_P$	Z	187.08	102.16	53.33
t_1	X	197.46	217.74	102.95
t_2	Z	134.09	104.38	120.62
t_3	X	277.78	249.14	103.33
t_4	Z	859.11	104.48	133.26

5. CONCLUSIONS

From Table 1, it is observed that:

(i) For all the population data sets, the *PRE* of the proposed ratio estimator t_3 is more than that of the usual unbiased estimator $\bar{y}$, the ratio estimator $\bar{y}_R$ and the Bahl and Tuteja (1991) ratio estimator t_1.

(ii) For all the population data sets, the *PRE* of the proposed product estimator t_4 is more than that of the usual unbiased estimator $\bar{y}$, the product estimator $\bar{y}_P$ and the Bahl and Tuteja (1991) product estimator t_2.

So, the proposed estimators t_3 and t_4 outperforms the other existing estimators of the sampling literature, and hence can be preferred for practical applications.

Acknowledgement: The authors are thankful to the editor and the learned referees for their valuable comments and suggestions towards the improvement of the paper.

RECEIVED SEPTEMBER, 2014
REVISED MAY, 2015

REFERENCES

[1] BAHL, S. and TUTEJA, R.K. (1991): Ratio and product type exponential estimator. **Information and Optimization Sciences**, 12, 159-163.
[2] COCHRAN, W.G. (1940): The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce. **The Journal of Agricultural Science**, 30, 262-275.
[3] JOHNSTON, J. (1972): **Econometric methods (2nd edn.)**. Mc Graw Hill Book Co., Tokyo.
[4] KADILAR, C. and CINGI, H. (2005): A new estimator using two auxiliary variables. **Applied Mathematics and Computation**, 162, 901-908.
[5] MURTHY, M.N. (1964): Product method of estimation. **The Indian Journal of Statistics,** Series A, 26, 69-74.
[6] ROBSON, D. S. (1957): Application of multivariate polykays to the theory of unbiased ratio-type estimation. Journal of American Statistical Association, 52, 511-522.
[7] SAHAI, A. (1979): An efficient variant of the product and ratio estimators. **Statistica Neerlandica**, 33, 1, 27-35.
[8] SHABBIR, J., HAQ, A. and GUPTA, S. (2014): A new difference-cum-exponential type estimator of finite population mean in simple random sampling. **Revista Colombiana de Estadistica**, 37, 197-209.
[9] SINGH, H.P. and ESPEJO, M.R.(2003): On linear regression and ratio-product estimation of a finite population mean. **Journal of the Royal Statistical Society**, 52(1), 59-67.

[10] SINGH, H. P. and TAILOR, R.(2005): Estimation of finite population mean using known correlation coefficient between auxiliary characters. **Statistica**, LXV, 4, 407-418.
[11] SINGH, H.P. and VISHWAKARMA, G.K. (2007): Modified exponential ratio and product estimators for finite population mean in double sampling. **Austrian Journal of Statistics**, 36, 3, 217-225,
[12] SINGH, H.P. and VISHWAKARMA, G.K. (2008): Some estimators of finite population mean using auxiliary information in sample surveys. **Journal of Applied Statistical Sciences**, 16, 4, 11-22.
[13] SINGH, M.P. (1967): Ratio-cum-product method of estimation. **Metrika**, 12, 34-72.
[14] SINGH, M.P. (1969): Comparison of some ratio-cum-product estimators. **Sankhya**, B, 31, 375-378.
[15] STEEL, R.G.D. and TORRIE, J.H. (1960): **Principles and Procedures of Statistics**. Mc Graw Hill Book Co.