A Generalized Multivariate Ratio and Regression Type Estimator for Population Mean Using A Linear Combination of Two Auxiliary Variables

Sanjay Kumar* and Priyanka Chhaparwal
Department of Statistics, Central University of Rajasthan, Bandarsindri, Kishangarh-305817, Ajmer, Rajasthan
*Corresponding Author: rahibhu@gmail.com

Received: 15th August 2015 / Revised: 01st December 2015 / Accepted: 09th February 2016

ABSTRACT

In this paper, we propose generalized multivariate ratio and regression type estimators of a finite population mean using a linear combination of two auxiliary variables and obtain the expressions for biases and mean square errors for the proposed estimators. The conditions under which the proposed estimators are more efficient then the relevant estimators have been obtained. An empirical study has been done for the support of the problem.

Keywords: mean square error, generalized ratio and regression type estimator, coefficient of variation, auxiliary variable.

1. Introduction

An important objective in any statistical estimation procedure is to obtain the estimators of parameters of interest with more precision. It is also well known that incorporation of more information in the estimation procedure yields better estimators, provided the information is valid and proper. Information on variables correlated with the study variable is known as auxiliary information that may be utilized either at preparation stage or at design stage or at the estimation stage to arrive at improved estimator compared to those, not utilizing such auxiliary information. For example, if the study variable is the quantity of fruits produced in each plot, then auxiliary variable can be the area of the each plot or the production of fruit in the same plot in previous year, another auxiliary variable can be the number of workers in each plot.

Use of such auxiliary information is made through the ratio and regression method of estimation to obtain improved estimators of population mean. The ratio method of estimation uses the auxiliary information to improve the precision which results in improved estimators when the regression of the study variable (y) on the auxiliary variable (x) is linear and passes through origin. When the regression of y on x is linear, it is not necessary that the line should always pass through origin. Under such condition, it is more appropriate to use the regression type estimators.

The use of auxiliary information in sample surveys is widely studied in the books written by Cochran (1977) and Sukhatme, Sukhatme and Asok (1984). Further, the use of supplementary information provided by auxiliary variables in survey sampling was discussed by several authors (Upadhyaya et. al, 1999; Kadilar et. al, 2005, 2006, 2007, 2009; Bacanli et. al, 2008; Gupta et. al, 2007; Al-Omari et. al, 2009; Tailor et. al, 2011; Khare et. al, 2011, 2015)

In this paper, we propose generalized multivariate ratio and regression type estimators using a linear combination of two auxiliary variables. The expressions for biases and mean square errors of the proposed estimators are obtained and a comparison of the proposed estimators has been made with the relevant estimators.

2. Materials and Methods

The classical ratio and regression estimators for the population mean $\bar{Y}$ of the study variables (y) using an auxiliary variable (x_1) whose population mean $\bar{X}_1$ is known are given by

$$T_r = \frac{\bar{y}}{\bar{x}_1}\bar{X}_1 \tag{1}$$

and

$$T_{lr} = \bar{y} + b_{yx_1}(\bar{X}_1 - \bar{x}_1), \tag{2}$$

where, $\bar{x}_1 = \sum_{i=1}^{n} x_{1i}$, $\bar{y} = \sum_{i=1}^{n} y_i$, $b_{yx_1} = \frac{S_{yx_1}}{S_{x_1}^2}$ is the regression coefficient of y on x_1, n is the number of the units in the sample (Cochran 1977), S_y^2 and $S_{x_1}^2$ are the population mean squares of y and x_1 respectively. S_{yx_1} is the population covariance between y and x_1 (Cochran, 1977).

The bias and mean square errors of the T_r are given by

$$MSE(T_r) = \left(\frac{1}{n} - \frac{1}{N}\right)\left\{S_y^2 + R^2 S_{x_1}^2 - 2RS_{yx_1}\right\}, \tag{3}$$

$$B(T_r) = \frac{1}{\bar{X}_1}\left(\frac{1}{n} - \frac{1}{N}\right)\left\{RS_{x_1}^2 - S_{yx_1}\right\}, \tag{4}$$

where, N is the number of units in the population, $R_1 = \frac{\bar{Y}}{\bar{X}_1}$ is the population ratio and $\bar{Y}$ and $\bar{X}_1$ are the population means of the study variable y and the auxiliary variable x_1 respectively.

The mean square error of the regression estimator T_{lr} is

$$MSE(T_{lr}) = \left(\frac{1}{n} - \frac{1}{N}\right)S_y^2\left(1 - \rho_{yx_1}^2\right), \tag{5}$$

where, $\rho_{yx_1} = \frac{S_{yx_1}}{S_y S_{x_1}}$ is the population correlation coefficient between y and x_1.

Kadilar and Cingi (2003) proposed the chain ratio estimator using one auxiliary variable for $\bar{Y}$ which is given by

$$T_{KC} = \bar{y}\left(\frac{\bar{X}_1}{\bar{x}_1}\right)^{\alpha}, \tag{6}$$

where, α is an arbitrary constant.

Lu J (2013) proposed the multivariate chain ratio type estimator and regression type estimator using a linear combination of two auxiliary variables which is given as follows:

$$t_1 = \bar{y}\left[\frac{w_1\bar{X}_1 + w_2\bar{X}_2}{w_1\bar{x}_1 + w_2\bar{x}_2}\right]^{\alpha_1} \tag{7}$$

and

$$t_{lr} = \bar{y} + b_1\left[\begin{pmatrix}K_1\\K_2\end{pmatrix}'\begin{pmatrix}\bar{X}_1\\\bar{X}_2\end{pmatrix} - \begin{pmatrix}K_1\\K_2\end{pmatrix}'\begin{pmatrix}\bar{x}_1\\\bar{x}_2\end{pmatrix}\right], \tag{8}$$

where, α_1 is an arbitrary constant, $\bar{X}_{lc} = K_1\bar{X}_1 + K_2\bar{X}_2$and $\bar{x}_{lc} = K_1\bar{x}_1 + K_2\bar{x}_2$, $b_1 = \frac{S_{yx_{lc}}}{S_{x_{lc}}^2}$,

$$S_{yx_{lc}} = \frac{1}{N-1}\sum_{i=1}^{N}[K_1(Y_i - \bar{Y})(X_{1i} - \bar{X}_1) + K_2(Y_i - \bar{Y})(X_{2i} - \bar{X}_1)]$$
$$= K_1S_{yx_1} + K_2S_{yx_2},$$

$$S_{x_{lc}}^2 = \frac{1}{N-1}\sum_{i=1}^{n}[K_1(X_{1i} - \bar{X}_1) + K_2(X_{2i} - \bar{X}_1)]^2$$
$$= K_1^2S_{x_1}^2 + K_2^2S_{x_2}^2 + 2K_1K_2S_{x_1x_2},$$

w_1, w_2 and K_1, K_2 are weights that satisfy the conditions $w_1 + w_2 = 1$ and $K_1 + K_2 = 1$, $\left(\bar{Y}, \underline{\bar{X}} = \begin{bmatrix}\bar{X}_1\\\bar{X}_2\end{bmatrix}\right)$ are the population means of the non negative values $\left(Y_i, \underline{X_i} = \begin{bmatrix}X_{1i}\\X_{2i}\end{bmatrix}\right), i = 1, 2, 3, \dots, N$ for the i^{th} unit of the population $U = (U_1, U_2, \dots, U_N)$ on the study variable (y) and auxiliary variables $\left(\underline{x} = \begin{bmatrix}x_1\\x_2\end{bmatrix}\right)$.

The mean square errors and the bias of the estimators t_1 and t_{lr} are given by

$$MSE(t_1) = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + \alpha_1^2\left(A_1^2R_1^2S_{x_1}^2 + A_2^2R_2^2S_{x_2}^2 + 2A_1A_2R_1R_2S_{x_1x_2}\right) - 2\alpha_1\left(A_1R_1S_{yx_1} + A_2R_2S_{yx_2}\right)\right], \quad (9)$$

$$B(t_1) = \left(\frac{1}{n}-\frac{1}{N}\right)\left[\frac{\alpha_1(\alpha_1+1)}{2}\left(A_1^2\frac{R_1}{\bar{X}_1}S_{x_1}^2 + A_2^2\frac{R_2}{\bar{X}_1}S_{x_2}^2 + 2A_1A_2\frac{R_2}{\bar{X}_1}S_{x_1x_2}\right) - \alpha_1\left(\frac{A_1}{\bar{X}_1}S_{yx_1} + \frac{A_2}{\bar{X}_2}S_{yx_2}\right)\right], \quad (10)$$

$$MSE(t_{lr}) = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2(1-\rho_1^2)\right], \quad (11)$$

$$B(t_{lr}) = -\beta_1\left[\left(\frac{N-n}{N-2}\right)\frac{1}{n}\left(\frac{\mu_{21}}{\mu_{11}} - \frac{\mu_{30}}{\mu_{20}}\right)\right], \quad (12)$$

where, $\mu_{rs} = \sum_{i=1}^{N}(X_i-\bar{X})^r\,(Y_i-\bar{Y})^s, A_1 = \frac{w_1\bar{X}_1}{w_1\bar{x}_1+w_2\bar{x}_2}$, $A_2 = \frac{(1-w_1)\bar{X}_1}{w_1\bar{x}_1+w_2\bar{x}_2}$,

$$R_1 = \frac{\bar{Y}}{\bar{X}_1}, R_2 = \frac{\bar{Y}}{\bar{X}_2}, \beta_1 = \frac{K_1S_{yx_1}+K_2S_{yx_2}}{K_1^2S_{x_1}^2+K_2^2S_{x_2}^2+2K_1K_2S_{x_1x_2}},$$

$$\rho_1^2 = \frac{\left(K_1S_{yx_1}+K_2S_{yx_2}\right)^2}{S_y^2\left(K_1^2S_{x_1}^2+K_2^2S_{x_2}^2+2K_1K_2S_{x_1x_2}\right)}$$

Differentiating (9) with respect to α_1 and w_1 separately, equating them to zero and solving the equations, we get

$$\alpha_{1(opt.)} = \frac{S_{yx_1}\left(R_2S_{x_2}^2 - R_1S_{x_1x_2}\right)+S_{yx_2}\left(R_1S_{x_1}^2 - R_2S_{x_1x_2}\right)}{R_1R_2\left(S_{x_1}^2S_{x_2}^2 - S_{x_1x_2}^2\right)} = \alpha_1', \quad (13)$$

$$w_{1(opt.)} = \frac{S_{yx_1}S_{x_2}^2 - S_{yx_2}S_{x_1x_2}}{S_{yx_1}S_{x_2}^2 + S_{yx_2}S_{x_1}^2 - S_{x_1x_2}\left(S_{yx_1}+S_{yx_2}\right)} = w_1', \quad (14)$$

Similarly, differentiating (11) with respect to K_1, equating it to zero and solving the equation, we get

$$K_{1(opt)} = \frac{S_{yx_1}S_{x_2}^2 - S_{yx_2}S_{x_1x_2}}{S_{yx_2}S_{x_1}^2 + S_{yx_1}S_{x_2}^2 - S_{yx_1}S_{x_1x_2} - S_{yx_2}S_{x_1x_2}} = K_1' \quad (15)$$

Putting the optimum value $w_{1(opt.)} = w_1'$ in the A_1 and A_2, we have

$$A_{1(opt)} = \frac{w_1'\bar{X}_1}{w_1'\bar{x}_1+(1-w_1')\bar{x}_2} = A_1', \quad (16)$$

$$A_{2(opt)} = \frac{(1-w_1')\bar{X}_2}{w_1'\bar{x}_1+(1-w_1')\bar{x}_2} = A_2', \quad (17)$$

Now putting the optimum values $\alpha_{1(opt.)} = \alpha_1'$, $A_{1(opt)} = A_1', A_{2(opt)} = A_2'$ and $K_{1(opt)} = K_1'$ in equation (9) and (11) respectively, we have

$$MSE\left(t_{1.}\right)_{min} = \left(\frac{1}{n}-\frac{1}{N}\right)\left[S_y^2 + \alpha_1'^2\left\{\left(A_1'^2R_1^2S_{x_1}^2 + A_2'^2R_2^2S_{x_2}^2 + 2A_1'A_2'R_1R_2S_{x_1x_2}\right)\right\} - 2\alpha_1'\left(A_1'R_1S_{yx_1} + A_2'R_2S_{yx_2}\right)\right], \quad (18)$$

$$MSE(t_{lr})_{min} = \left(\frac{1}{n} - \frac{1}{N}\right)\left[S_y^2(1-\rho_{11}^2)\right], \quad (19)$$

where, $\rho_{11}^2 = \frac{(K_1' S_{yx_1} + K_2' S_{yx_2})^2}{S_y^2(K_2'^2 S_{x_1}^2 + K_2'^2 S_{x_2}^2 + 2K_1'K_2' S_{x_1x_2})}$, $K_2' = 1 - K_1'$.

Serals (1967) defined an estimator $\bar{y}_s = k_1\bar{y}$, which has $MSE(\bar{y}_s) \leq MSE(\bar{y})$, for the value of $k_1 = \left(1 + \frac{f}{n}C_y^2\right)^{-1}$; where $f = \frac{N-n}{N}$

The $V(\bar{y}_s)$ is given by

$$V(\bar{y}_s) = k_1^2 \frac{f}{n} S_y^2 = (1-2B)\frac{f}{n}S_y^2, \quad (20)$$

where, $B = \frac{f}{n}C_y^2$ and C_y is the coefficient of variation of y.

3. The Proposed Estimators and Their Mean Square Error (MSE)

We propose a generalized multivariate ratio type estimator (T_1) and regression type estimator (T_2) for the population mean $\bar{Y}$ in simple random sampling using a linear combination of two auxiliary variables, which are given as follows

$$T_1 = \bar{y}_s\left(\frac{\underline{\bar{X}}_{11}}{\underline{\bar{x}}_{11}}\right)^{\alpha_2} = \bar{y}_s\left[\frac{\binom{w_3}{w_4}'\binom{\bar{X}_1}{\bar{X}_2}}{\binom{w_3}{w_4}'\binom{\bar{x}_1}{\bar{x}_2}}\right]^{\alpha_2} \quad (21)$$

and

$$T_2 = \bar{y}_s + b_2\left[\binom{K_3}{K_4}'\binom{\bar{X}_1}{\bar{X}_2} - \binom{K_3}{K_4}'\binom{\bar{x}_1}{\bar{x}_2}\right], \quad (22)$$

where, α_2 is an arbitrary constant, $\bar{X}_{lc} = K_3\bar{X}_1 + K_4\bar{X}_2$ and $\bar{x}_{lc} = K_3\bar{x}_1 + K_4\bar{x}_2$, $b_2 = \frac{S_{yx_{lc}}}{S_{x_{lc}}^2}$,

$S_{yx_{lc}} = \frac{1}{N-1}\sum_{i=1}^{N}[K_3(Y_i - \bar{Y})(X_{1i} - \bar{X}_1) + K_4(Y_i - \bar{Y})(X_{2i} - \bar{X}_1)] = K_3S_{yx_1} + K_4S_{yx_2}$,

$S_{x_{lc}}^2 = \frac{1}{N-1}\sum_{i=1}^{N}[K_3(X_{1i} - \bar{X}_1) + K_4(X_{2i} - \bar{X}_1)]^2 = K_3^2S_{x_1}^2 + K_4^2S_{x_2}^2 + 2K_3K_4S_{x_1x_2}$, (w_3, w_4) and $(K_3,\ K_4)$ are weights that satisfy the condition $w_3 + w_4 = 1$ and $K_3 + K_4 = 1$

The expressions for the MSE of the proposed estimators $T_i, i = 1, 2$ up to the term of order $n^{-1}$are given as follows:

$$MSE(T_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[(1-2B)S_y^2 + \alpha_2^2\left(A_3^2R_1^2S_{x_1}^2 + A_4^2R_2^2S_{x_2}^2 + 2A_3A_4R_1R_2S_{x_1x_2}\right) - 2\alpha_2(1-B)\left(A_3R_1S_{yx_1} + A_4R_2S_{yx_2}\right)\right], \quad (23)$$

$$B(T_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[\frac{\alpha_1(\alpha_1+1)}{2}\left(A_1^2 \frac{R_1}{\bar{X}_1} S_{x_1}^2 + A_2^2 \frac{R_2}{\bar{X}_1} S_{x_2}^2 + 2A_1 A_2 \frac{R_2}{\bar{X}_1} S_{x_1 x_2}\right) - \alpha_1(1-B)\left(\frac{A_1}{\bar{X}_1} S_{yx_1} + \frac{A_2}{\bar{X}_2} S_{yx_2}\right)\right] \quad (24)$$

$$MSE(T_2) = \left(\frac{1}{n} - \frac{1}{N}\right)(1-2B)S_y^2(1-\rho_2^2), \quad (25)$$

where,

$$A_3 = \frac{w_3\bar{X}_1}{w_3\bar{X}_1 + (1-w_4)\bar{X}_2}, A_4 = \frac{(1-w_3)\bar{X}_2}{w_3\bar{X}_1 + (1-w_4)\bar{X}_2}, \rho_2^2 = \frac{\left(K_3 S_{yx_1} + K_4 S_{yx_2}\right)^2}{S_y^2\left(K_3^2 S_{x_1}^2 + K_4^2 S_{x_2}^2 + 2K_3K_4 S_{x_1x_2}\right)} = \rho_1^2,$$

Differentiating (23) with respect to α_2 and w_3 separately, equating them to zero and solving the equations, we get

$$\alpha_{2(opt.)} = \frac{(1-B)\{S_{yx_1}(R_2 S_{x_2}^2 - R_1 S_{x_1x_2}) + S_{yx_2}(R_1 S_{x_1}^2 - R_2 S_{x_1x_2})\}}{R_1R_2(S_{x_1}^2 S_{x_2}^2 - S_{x_1x_2}^2)} = \alpha_2', \quad (26)$$

$$w_{3(opt.)} = \frac{S_{yx_1}S_{x_2}^2 - S_{yx_2}S_{x_1x_2}}{S_{yx_1}S_{x_2}^2 + S_{yx_2}S_{x_1}^2 - S_{x_1x_2}(S_{yx_1} + S_{yx_2})} = w_3', \quad (27)$$

Similarly, Differentiating (24) with respect to K_3, and equating it to zero and solving the equation, we get

$$K_{3(opt)} = \frac{S_{yx_1}S_{x_2}^2 - S_{yx_2}S_{x_1x_2}}{S_{yx_2}S_{x_1}^2 + S_{yx_1}S_{x_2}^2 - S_{yx_1}S_{x_1x_2} - S_{yx_2}S_{x_1x_2}} = K_3' = K_1', \quad (28)$$

Putting the optimum value $w_{3(opt.)} = w_3'$ in the A_3 and A_4, we have,

$$A_{3(opt)} = \frac{w_3'\bar{X}_1}{w_3'\bar{X}_1 + (1-w_3')\bar{X}_2} = A_3', \quad (29)$$

$$A_{4(opt)} = \frac{(1-w_3')\bar{X}_1}{w_3'\bar{X}_1 + (1-w_3')\bar{X}_2} = A_4', \quad (30)$$

$$\rho_{22}^2 = \frac{\left(K_3' S_{yx_1} + K_4' S_{yx_2}\right)^2}{S_y^2\left(K_3'^2 S_{x_1}^2 + K_4'^2 S_{x_2}^2 + 2K_3'K_4' S_{x_1x_2}\right)} = \rho_{11}^2 \quad (31)$$

Now putting the optimum values $\alpha_{2(opt.)} = \alpha_2'$, $A_{3(opt)} = A_3'$, $A_{4(opt)} = A_4'$, and $K_{3(opt)} = K_3'$ in equation (23) and (24) respectively.

$$MSE(T_{1.})_{min} = \left(\frac{1}{n} - \frac{1}{N}\right)\left[(1-2B)S_y^2 + \alpha_2'^2\left(A_3'^2 R_1^2 S_{x_1}^2 + A_4'^2 R_2^2 S_{x_2}^2 + 2A_3'A_4' R_1R_2 S_{x_1x_2}\right) - 2\alpha_2'(1-B)\left(A_3' R_1 S_{yx_1} + A_4' R_2 S_{yx_2}\right)\right], \quad (32)$$

$$MSE(T_{2.})_{min} = \left(\frac{1}{n} - \frac{1}{N}\right)(1-2B)S_y^2(1-\rho_{11}^2), \quad (33)$$

Many estimators turn out as special cases of $T_i, i = 1, 2$ which are given as follows:

(i) Putting $\alpha_2 = 0$ in (20) then the proposed estimator T_1 reduces to Searls' estimator $\overline{y}_s$.

(ii) If α_2=1 and $w_4 = 0$ then the proposed estimator T_1 reduces to $t_2 = \bar{y}_s \frac{\bar{X}_1}{\bar{x}_1}$,

$MSE(t_2) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[(1-2B)S_y^2 + R_1^2 S_{x_1}^2 - 2(1-B)R_1 S_{yx_1}\right]$,

$B(t_2) = \frac{1}{\bar{X}_1}\left(\frac{1}{n} - \frac{1}{N}\right)\left\{R_1^2 S_{x_1}^2 - (1-B)S_{yx_1}\right\}$,

(iii) putting α_2=1 and $w_3 = 0$ then the proposed estimator T_1 reduces to $t_3 = \bar{y}_s \frac{\bar{X}_2}{\bar{x}_2}$,

$MSE(t_3) = \left(\frac{1}{n} - \frac{1}{N}\right)\left[(1-2B)S_y^2 + R_2^2 S_{x_2}^2 - 2(1-B)R_2 S_{yx_2}\right]$,

$B(t_3) = \frac{1}{\bar{X}_2}\left(\frac{1}{n} - \frac{1}{N}\right)\left\{R_2^2 S_{x_2}^2 - (1-B)S_{yx_2}\right\}$,

(iv) putting $k_1 = 1$, in equation (20), the proposed estimator T_1 reduces to Li J estimator

$$t_1 = \bar{y}\left[\frac{w_1\bar{X}_1 + w_2\bar{X}_2}{w_1\bar{x}_1 + w_2\bar{x}_2}\right]^{\alpha_1}$$

(v) putting α_2=1, $k_1 = 1$ and $w_4 = 0$, then the proposed estimator T_1 reduces to traditional ratio estimator $T_r = \frac{\bar{y}}{\bar{x}_1}\bar{X}_1$,

(vi) if $k_1 = 1$, $B = 0$ and $w_4 = 0$, the proposed estimator T_1reduces to the estimator $T_{KC} = \bar{y}\left(\frac{\bar{X}_1}{\bar{x}_1}\right)^{\alpha}$,

(vii) if $k_1 = 1$, $B = 0$ and $w_3 = 0$, the proposed estimator T_1 reduces to the estimator $T_{KC_1} = \bar{y}\left(\frac{\bar{X}_2}{\bar{x}_2}\right)^{\alpha}$,

(viii) if $w_4 = 0$, the proposed estimator T_1 reduces to the estimator $T_{KC_2} = \bar{y}_s\left(\frac{\bar{X}_1}{\bar{x}_1}\right)^{\alpha_2}$,

(ix) Putting $b_2 = 0$ in (21), the proposed estimator T_2 reduces to Searls estimator $\bar{y}_s$,

(x) putting $K_4 = 0$ in (21), the proposed estimator T_2 reduces to $T_{lr_1} = \bar{y}_s + b_{yx_1}(\bar{X}_1 - \bar{x}_1)$,

(xi) putting $K_3 = 0$ in (21), the proposed estimator T_2 reduces to $T_{lr_2} = \bar{y}_s + b_{yx_2}(\bar{X}_2 - \bar{x}_2)$,

(xii) putting $k_1 = 0$, $K_4 = 0$ in (21), the proposed estimator T_2 reduces to traditional regression estimator $T_{lr_3} = \bar{y} + b_1(\bar{X}_{lc} - \bar{x}_{lc})$.

4. Comparison of the Proposed Estimators with the Relevant Estimators

On comparing the $MSE(T_i); i = 1,2,3$ with $V(\bar{y}_s)$ and other relevant estimators, we get certain conditions under which the proposed estimators have less mean square errors than the relevant estimators and the conditions are as given below:

$$MSE(T_1) \leq V(\bar{y}_s) \ \ if \quad 0 < \alpha_2 < \frac{b}{a} \ , \tag{34}$$

$$MSE(T_1) \leq MSE(t_1) \ \ if \frac{b_1 - \sqrt{b_1^2 - 4a_1c_1}}{2a_1} < \alpha_2 < \frac{b_1 + \sqrt{b_1^2 - 4a_1c_1}}{2a_1} \ , \tag{35}$$

$$MSE(T_1) \leq MSE(T_2) \ \ if \frac{b_2 - \sqrt{b_2^2 - 4a_2c_2}}{2a_2} < \alpha_2 < \frac{b_2 + \sqrt{b_2^2 - 4a_2c_2}}{2a_2} \ , \tag{36}$$

$$MSE(T_1) \leq MSE(T_r) \ \ if \frac{b_3 - \sqrt{b_3^2 - 4a_3c_3}}{2a_3} < \alpha_2 < \frac{b_3 + \sqrt{b_3^2 - 4a_3c_3}}{2a_3} \ , \tag{37}$$

$$MSE(T_2) \leq MSE(t_{lr}) \ \ if -1 \leq \rho_1 \leq 1, \tag{38}$$

$$MSE(T_2) \leq MSE(T_{lr}) \ \ if \ \rho_1^2 \geq \left(\frac{\rho^2 - 2B}{1 - 2B}\right), \tag{39}$$

$$MSE(T_2) \leq V(\bar{y}_s) \ \ if \quad \rho_1^2 \leq 1, \tag{40}$$

$$MSE(T_{KC2}) \leq MSE(T_{KC}) \ if \ \frac{-b_4 \pm \sqrt{b_4^2 - 4a_4b_4}}{a_4} < \alpha_2 < \frac{b_4 \pm \sqrt{b_4^2 - 4a_4b_4}}{a_4}, \tag{41}$$

$$MSE(t_2) \leq MSE(T_r) \ if \ R_1 \leq \frac{S_y^2}{S_{yx_1}}, \tag{42}$$

$$MSE(T_{lr1}) \leq MSE(T_{lr3}) \ if \ \rho_{33}^2 \leq 1, \tag{43}$$

where,

$a = A_1^2R_1^2S_{x_1}^2 + A_2^2R_2^2S_{x_2}^2 + 2A_1A_2R_1R_2S_{x_1x_2}$,

$b = 2(1 - B)\left(A_1R_1S_{yx_1} + A_2R_2S_{yx_2}\right)$

$a_1 = A_1^2R_1^2S_{x_1}^2 + A_2^2R_2^2S_{x_2}^2 + 2A_1A_2R_1R_2S_{x_1x_2}$,

$b_1 = 2(1 - B)\left(A_1R_1S_{yx_1} + A_2R_2S_{yx_2}\right)$,

$c_1 = (-2B)S_y^2 - \left[\alpha_1'^2\left(A_1'^2R_1^2S_{x_1}^2 + A_2'^2R_2^2S_{x_2}^2 + 2A_1'A_2'R_1R_2S_{x_1x_2}\right) - 2\alpha_1'\left(A_1'R_1S_{yx_1} + A_2'R_2S_{yx_2}\right)\right]$,

$a_2 = A_3^2R_1^2S_{x_1}^2 + A_4^2R_2^2S_{x_2}^2 + 2A_3A_4R_1R_2S_{x_1x_2}$,

$b_2 = 2(1 - B)\left(A_3R_1S_{yx_1} + A_4R_2S_{yx_2}\right)$,

$c_2 = (1 - 2B)S_y^2\rho_1^2$,

$a_3 = A_3^2 R_1^2 S_{x_1}^2 + A_4^2 R_2^2 S_{x_2}^2 + 2A_3 A_4 R_1 R_2 S_{x_1 x_2}$,
$b_3 = 2(1 - B)\left(A_3 R_1 S_{yx_1} + A_4 R_2 S_{yx_2}\right)$,
$c_3 = 2R_1 S_{yx_1} - R_1^2 S_{x_1}^2 - 2BS_y^2$, $a_4 = R_1^2 S_{x_1}^2$, $b_4 = 2(1 - B)R_1 S_{yx_1}$,
$c_4 = 2\alpha_1 R_1 S_{yx_1} - \alpha_1^2 R_1^2 S_{x_1}^2 - 2BS_y^2$, $\rho_{33}^2 = \frac{S_{yx_1}^2}{S_y^2 S_{x_1}^2}$

5. Determination of the Sample Size for a Fixed Cost

One aspect for choosing the sample size so that the available resources are used in an effective way is to maximize the precision of the estimator for a fixed cost.

Let C_0 be the total coat (fixed) of the survey apart from the overhead cost. The cost function C_0 can be written as

$$C_0 = nC_1, \tag{44}$$

where, C_1 is the cost per unit in the sample.

Since C_0 will vary from sample to sample, so the expected cost can be written as

$$C = E(C_0) = nC_1 \tag{45}$$

The expression for $MSE\{T(i)\}, i = 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11$ can be written as follows:

$$MSE\{T(i)\} = \frac{M_{0i}}{n} - \frac{M_{1i}}{N}, \tag{46}$$

where, $T(1) = \bar{y}_s, T(2) = T_1, T(3) = t_1, T(4) = T_2, T(5) = t_{lr}, T(6) = T_{KC_2}$,

$T(7) = T_{KC}$, $T(8) = t_2, T(9) = T_r, T(10) = T_{lr_1}$ and $T(11) = T_{lr_3}$.

M_{0i} and M_{1i} are the coefficients of the terms of $\frac{1}{n}$ and $\frac{1}{N}$ respectively in the expression of $MSE\{T(i)\}$.

To find the optimum values of n and minimum values of $MSE\{T(i)\}$ in case of the fixed cost $\leq C_0$, let us define a function φ which is given by

$$\varphi = MSE\{T(i)\} + \delta_i\{n(C_1) - C_0\}, i = 1, \ldots, 11, \tag{47}$$

where, δ_i is the Lagrange's multiplier.

Now, differentiating φ with respect to n and equating it to zero, we get,

$$n = \sqrt{\frac{M_{0i}}{\delta_i(C_1)}} \tag{48}$$

Putting the value of n in (44), we get

$$\sqrt{\delta_i} = \frac{1}{C_0}\sqrt{M_{0i}C_1} \tag{49}$$

Putting the value of δ_i in (48), we get

$$n = \frac{C_0}{C_1} \tag{50}$$

It has also been seen that the determinant of the matrix of second order derivative of φ with respect to n is negative for the optimum values of n. The minimum value of $MSE\{T(i)\}$ for the optimum value of n are given as follows:

$$MSE\{T(i)\}_{min} = \frac{M_{0i}C_1}{C_0} - \frac{M_{1i}}{N} \tag{51}$$

Neglecting the term of order N^{-1} , we have

$$MSE\{T(i)\}_{min} = \frac{M_{0i}C_1}{C_0}, i = 1, 2, 3, 4, 5 \tag{52}$$

6. Determination of the Sample Size for a Specified Variance

Another aspect for choosing the sample size so that the available resources are used in an effective way is to minimize the cost of the survey for a specified variance.

Let V_0 be the variance of the estimator fixed in advance, then we have,

$$V_0 = \frac{M_{0i}}{n} - \frac{M_{1i}}{N} \tag{53}$$

The total cost apart from overhead cost is minimized by obtaining the optimum value of n and for specified precision $V = V_0$. For this purpose, we defined a function φ_1 which is given as follows:

$$\varphi_1 = nC_1 + \mu_i[MSE\{T(i)\} - \mathrm{V}_0], i = 1, \ldots, 11, \tag{54}$$

where, μ_i is the Lagrange's multiplier.

After differentiating φ_1 with respect to n and equating it to zero, we get,

$$n = \sqrt{\frac{\mu_i M_{0i}}{C_1}} \tag{55}$$

Putting the value of n in (53), we get

$$\mu_i = \frac{M_{0i}C_1}{V_0^2} \tag{56}$$

Putting the value of μ_i in (55), we get

$$n = \frac{M_{0i}}{V_0} \tag{57}$$

It has also been seen that the determinant of the matrix of second order derivative of φ_1 with respect to n is negative for the optimum values of n, which shows the solution for n given by (55).

We can obtain the value of n for which the estimator $MSE\{T(i)\}$ attains the variance V_0 with expected cost given by

$$C[MSE\{T(i)\}_{min}] = \frac{M_{0i}C_1}{V_0},\ i = 1, \ldots, 11. \tag{58}$$

7. Empirical Study

The data on physical growth of 100 fish (Laengelmavesi on Lake, near Tampere, Finland) has been taken under study. Our goal is to estimate the average weight of fish and it is also well understood that incorporation of more information in the estimation procedure yields better estimators. So for the purpose, the study variable (y), auxiliary variable (x_1) and the additional auxiliary variable (x_2) are taken as follows:

y- weight (in g) of the fish,

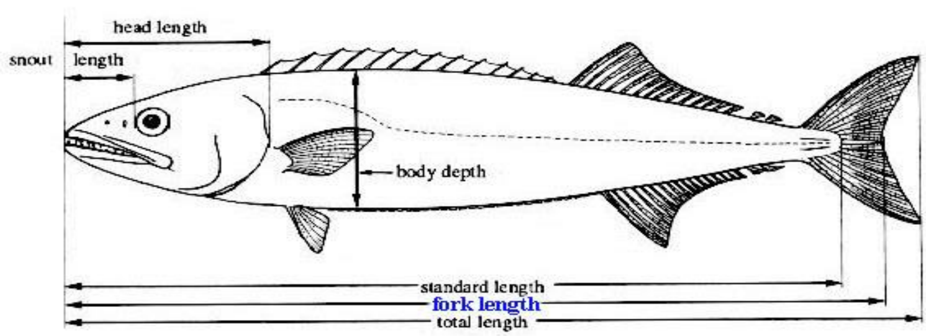
x_1- Length from the nose to the beginning of the tail (in cm),

x_2- Length from the nose to the notch of the tail (in cm).

Since the auxiliary variables are positively correlated with the weight of the fish, the information on these two auxiliary variables will yield better estimators.

The values of parameters of the population are given as follows:

$\bar{Y} = 399.39, \bar{X}_1 = 26.25, \bar{X}_2 = 28.40, S_y^2 = 119777.85, S_{x_1}^2 = 107.14, S_{x_2}^2 = 124.78, S_{yx_1} = 3199.20, S_{yx_2} = 3477.15, S_{x_1x_2} = 115.56.$



From Table 1, we see that the proposed estimator T_1 has minimum mean square error than $\bar{y}_s, t_1$ and other special cases of the estimators. The proposed estimator T_2 has minimum mean square error than $\bar{y}_s$, t_{lr} and other special cases of the estimators. Also the proposed estimator T_1 has minimum mean square error than t_{lr}. Further, we observe that the special estimator T_{KC_2} has minimum mean square error than the estimators T_{KC}, t_2 and T_r. Also the estimator T_{lr_1} has minimum mean square error than T_{lr_3}.

Table 1: Relative efficiencies of different estimators $\bar{y}_s, T_1, t_1, T_2, t_{lr}, T_{KC_2}, T_{KC}, t_2,$ T_r, T_{lr_1} and T_{lr_3} (for the fixed values of $N = 100, n = 60$).

Estimators	Efficiency (MSE)
$\bar{y}_s$ (eq. no. 20)	100.00 (790.52)
T_1 (eq. no. 21)	625.212 (126.44, $\alpha_{2(opt.)} = 1.99$)
t_1 (eq. no. 7)	618.873 (127.74)
T_2 (eq. no. 22)	625.137 (126.46, $K_{3(opt.)} = -11.97$)
t_{lr} (eq. no. 8)	618.878 (127.74)
T_{KC_2}(special case no. viii)	492.98 (160.38)
T_{KC}(special case no. vi)	487.99 (162.00)
t_2(special case no. ii)	254.95 (310.07)
T_r(special case no. v)	251.11 (314.82)
T_{lr_1}(special case no. x)	493.93 (160.05)
T_{lr_3}(special case no. xii)	488.99 (161.67)

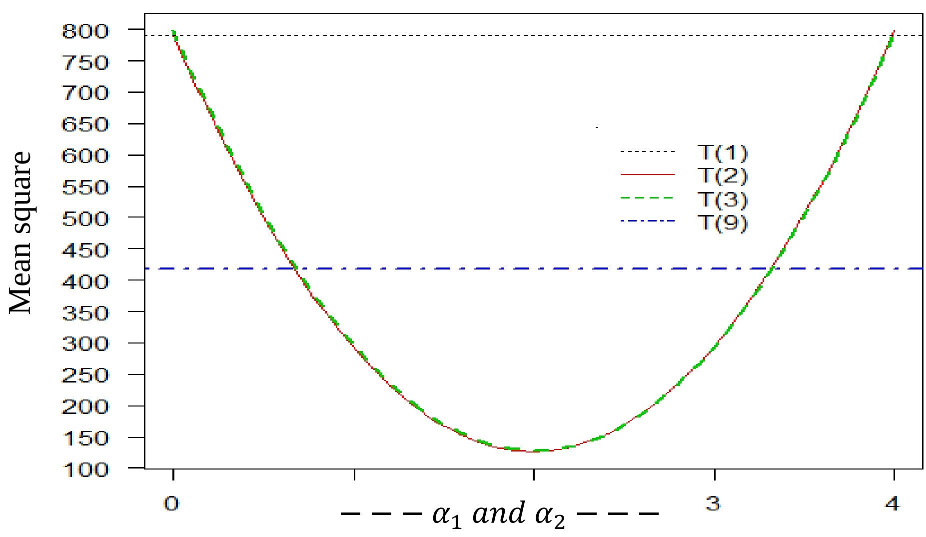


Figure1: MSE of different estimators for different value of α_1 and α_2.

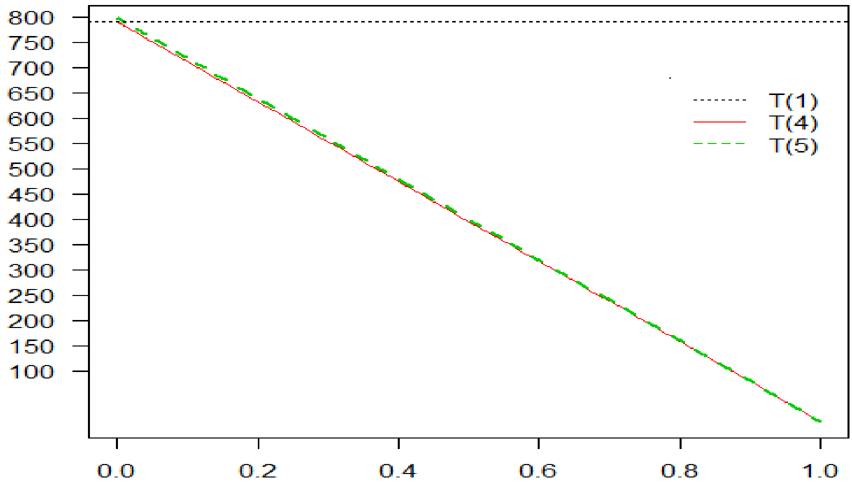


Figure2: MSE of diff $---\rho_1^2---$ ferent value of ρ_1^2.

We obtained the values of mean square error of the proposed estimator T_1 and t_1 for different values of α_2 and α_1 respectively. Similarly, the values of mean square error of the proposed estimator T_2 for different values of ρ_1^2. After plotting the mean square errors of the estimators $T(1) = \bar{y}_s, T(2) = T_1, T(3) = t_1, T(4) = T_2$ and $T(5) = t_{lr}$, we observe from figure 1 that the proposed estimator T_1 has minimum mean square error than $\bar{y}_s, T_r$ and t_1. Similarly from figure 2, we observe that the proposed estimator T_2 has minimum mean square error than $\bar{y}_s, T_{lr}$ and t_{lr}.

From table 2, we observe that the proposed estimator T_1 is more efficient than $\bar{y}_s, t_1$ and other special cases of the estimators. The proposed estimator T_2 is more efficient than $\bar{y}_s$, t_{lr} and other special cases of the estimators. Also the proposed estimator T_1 has minimum mean square error than t_{lr}. Further, we observe that the special estimator T_{KC_2} has minimum mean square error than the estimators T_{KC}, t_2 and T_r. Also the estimator T_{lr_1} has minimum mean square error than T_{lr_3}.

Table 2: Relative efficiencies of different estimators $\bar{y}_s, T_1, t_1, T_2, t_{lr}, T_{KC_2}, T_{KC}, t_2,$ $T_r,\ T_{lr_1}$ and T_{lr_3} (for the fixed cost $C_0 = Rs.100$ and $C_1 = Rs.2$)

Estimators	n_{opt}	Efficiency (MSE)
$\bar{y}_s$ (eq. no. 20)	50	100.00 (2371.573)
T_1 (eq. no. 21)	50	625.212 (379.323)
t_1 (eq. no. 7)	50	618.870 (383.210)
T_2 (eq. no. 22)	50	625.136 (379.369)
t_{lr} (eq. no. 8)	50	618.876 (383.206)
T_{KC_2}(special case no. viii)	50	492.98 (481.07)
T_{KC}(special case no. vi)	50	487.99 (485.99)
t_2(special case no. ii)	50	254.95 (930.21)
T_r(special case no. v)	50	251.11 (944.45)
T_{lr_1}(special case no. x)	50	493.93 (480.14)
T_{lr_3}(special case no. xii)	50	488.99 (484.99)

From table 3, we observe that the proposed estimator T_1 has minimum cost than $\bar{y}_s, t_1$ and other special cases of the estimators. The proposed estimator T_2 has minimum cost than $\bar{y}_s$, t_{lr} and other special cases of the estimators. Also the proposed estimator T_1 has less cost than t_{lr}. Further, we observe that the special estimator T_{KC_2} has minimum cost than the estimators T_{KC}, t_2 and T_r. Also the estimator T_{lr_1} has minimum cost than T_{lr_3}.

8. Monte Carlo Simulation Study

In the present Monte Carlo simulation study, we consider another data set on physical growth of upper socio-economic group of 100 school going children of Varanasi under an ICMR study, Department of Pediatrics, B.H.U., during 1983-84 has been taken under study. The study variable (y), auxiliary variable (x_1) and the additional auxiliary variable (x_2) are taken as follows:

y- weight (in kg) of the children,

x_1- chest circumference (in cm) of the children,

x_2- skull circumference (in cm) of the children.

The values of the parameters of the y, x_1 and x_2 variables for the given data are taken as follows:

$\bar{X}_1 = 55.76, \bar{Y} = 19.46, \bar{X}_2 = 51.09, S^2_{x_1} = 10.63, S^2_y = 8.92, S^2_{x_2} = 2.42,$

$\rho_{yx_1} = 0.85, \rho_{yx_2} = 0.30, \rho_{x_1x_2} = 0.31, S_{yx_1} = 8.31, S_{yx_2} = 1.44.$

Table 3: Expected cost of different estimators $\bar{y}_s, T_1, t_1, T_2, t_{lr}, T_{KC_2}, T_{KC}, t_2, T_r, T_{lr_1}$ and T_{lr_3} (for the fixed variance $V_0 = 1500,\ C_1 = Rs.2$)

Estimators	n_{opt}	**Expected cost (Rs.)**
$\bar{y}_s$ (eq. no. 20)	79	158.10
T_1 (eq. no. 21)	13	25.28
t_1 (eq. no. 7)	13	25.54
T_2 (eq. no. 22)	13	25.29
t_{lr} (eq. no. 8)	13	25.54
T_{KC_2}(special case no. viii)	16	32.07
T_{KC}(special case no. vi)	16	32.40
t_2(special case no. ii)	31	62.01
T_r(special case no. v)	32	62.96
T_{lr_1}(special case no. x)	16	32.00
T_{lr_3}(special case no. xii)	16	32.33

From the population of 100 schools going children of Varanasi, a sample of different sizes 25, 35, 45 and 60 are taken by simple random sampling without replacement. The above process is replicated 1000 times. Simulated mean square error of $T(i), i = 1,2,3,4,7,8,9,10,11$ are calculated as follows:

$$MSE\{T(i)\} = \frac{1}{1000}\sum_{j=1}^{1000}\{T(i)_j - \bar{Y}\}^2 ,$$

Table 4: Simulated percent relative efficiencies (with respect to $\bar{y}_s$) of different estimators $T(i); i = 1,2, \ldots, 11$ for the fixed values of n $(n = 25, 35, 45, 60)$.

Est.	n			
	25	35	45	60
$\bar{y}_s$	100.00 (4185.74)	100.00 (2249.81)	100.00 (2132.10)	100.00 (1034.99)
T_1	364.36 (1148.78)	366.74(613.46)	396.32 (537.97)	433.14 (238.95)
t_1	361.43 (1158.12)	357.98(628.48)	387.64(550.02)	427.61 (242.04)
T_2	361.62 (1157.51)	372.31(604.28)	402.44(529.79)	437.47 (236.46)
t_{lr}	360.68 (1160.52)	366.27(614.28)	394.69(540.19)	432.39 (239.37)
T_{KC_2}	295.82 (1414.95)	271.59(828.37)	320.99(664.23)	334.96 (308.99)
T_{KC}	291.45 (1436.17)	265.61 (847.03)	316.87(672.87)	333.30 (310.53)
t_2	269.93 (1550.67)	255.53(880.46)	308.87(690.28)	330.05 (313.59)
T_r	268.65 (1558.07)	251.02(896.28)	306.26 (696.17)	329.05(314.54)
T_{lr_1}	314.55 (1330.71)	307.99 (730.47)	382.10 (557.99)	396.85 (260.80)
T_{lr_3}	313.56 (1334.91)	306.63(733.72)	380.75 (559.97)	395.78 (261.50)

*Figures in parenthesis give mean square errors of the estimators.

From table 4, we see that the proposed estimator T_1 has minimum mean square error than $\bar{y}_s$, t_1 and other special estimators. The proposed estimator T_2 has minimum mean square error than $\bar{y}_s$, t_{lr} and other special estimators. Also the proposed estimator T_1 has minimum mean square error than t_{lr}. The special estimator T_{KC2} has minimum mean square error than the estimators T_{KC}, t_2 and T_r. Also the estimator T_{lr1} has minimum mean square error than T_{lr3}. Further, we observe that when the sample size increases, the mean square errors of the estimators decrease.

9. Conclusions

In this work, we have proposed generalized multivariate chain ratio and regression type estimators with a linear combination of two auxiliary variables and using known coefficient of variation of study variable. The conditions under which the proposed estimators have minimum mean square errors are mentioned in the section 5. A generalization can be made with more than two variables also. Here, we conclude that the information on coefficient of variation of study variable is fruitful in increasing the precision of the estimators compared to those, not utilizing such information. However, these values depend on the population parameters where prior information can be used. In lack of prior information, the estimated values of the parameters based on sample values can be used. For the support of the problem, an empirical study as well as a Monte Carlo simulation study has been made. The results obtained from the Monte Carlo simulation study are found to be similar to the results based on the empirical study. On the basis of the empirical study, we observe that use of known coefficient of variation of the study variable in the proposed estimators for population mean is found to be more useful in increasing the precision of the proposed estimators with respect to the relevant estimators for the fixed cost $C \leq C_0$. The total cost for the proposed estimators is also less than the relevant estimators for the specified variance $V = V_0$.

Acknowledgements

Authors wish to thank to the learned referees for their critical and constructive comments regarding improvement of the paper.

References

1. Al-Omari A. I., Jemain A. A. and Ibrahim K. (2009). New ratio estimators of the mean using simple random sampling and ranked set sampling methods. *Revista Investigacion Operacional,* 30:97–108.

2. Bacanli S. and Kadilar C. (2008). Ratio estimators with unequal probability designs. *Pakistan Journal of Statistics,* 24:167–172.

3. Cochran W. G. (1977): *Sampling Techniques*. New-York: John Wiley and Sons.

4. Gupta S. and Shabbir J. (2007). On the use of transformed auxiliary variables in estimating population mean by using two auxiliary variables. *Journal Statistical Planning Inference*, 137:1606–1611. DOI: 10.1016/j.jspi.2006.09.008.

5. Kadilar C., Candan M. and Cingi H. (2007). Ratio estimators using robust regression. *Hacettepe Journal of Mathematics and Statistics*, 36:181–188.

6. Kadilar C. and Cingi H. (2003). A study on the chain ratio-type estimator. *Hacettepe Journal of Mathematics and Statistics*, 32:105–108.

7. Kadilar C. and Cingi H. (2005). A new estimator using two auxiliary variables. *Applied Mathematics and Computation*, 162:901–908. DoI:10.1016/j.amc.2003.12.130.

8. Kadilar C. and Cingi H. (2006). Improvement in estimating the population mean in simple random sampling. *Applied Mathematics Letters*, 19:75–79. DOI:10.1016/j.aml.2005.02.039.

9. Kadilar C., Unyazici Y. and Cingi H. (2009). Ratio estimator for the population mean using ranked set sampling. *Statistical Papers*, 50:301–309. DOI 10.1007/s00362-007-0079-y.

10. Khare B. B. and Kumar S (2011). A generalised chain ratio type estimator for population mean using coefficient of variation of the study variable. *National Academy Science Letters, India*, 34(9 &10):353-358.

11. B. B. and Kumar S (2015). Chain ratio-to-regression estimators in two-phase sampling in the presence of non-response. *ProbStat forum*, 8:95-102.

12. Lu J. (2013). The Chain Ratio Estimator and Regression Estimator with Linear Combination of Two Auxiliary Variables, *PLoS ONE*, 8(11):1-4. DOI:10.1371/journal.pone.0081085.

13. Searls D. T. (1967). A note on the use of an approximately known coefficient of variation. *Journal of the American Statistical Association*, 21:20-21.

14. Sukhatme B. V., Sukhatme S., and Asok C. (1984). *Sampling Theory of Surveys with Applications*. 3rd Edition, Iowa State University Press.

15. Tailor R., Parmar R., Kim J. M. and Tailor R. (2011). Ratio-cum-Product estimators of population mean using known population parameters of auxiliary variates. *Communications of the Korean Statistical Society*, 18:155–164. DOI: 10.5351/CKSS.2011.18.2.155.

16. Upadhyaya L. N. and Singh H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal*, 41:627–636. DOI: 10.1002/(SICI)1521-4036(199909)41:5<627::AID- BIMJ627>3.0.CO;2-W.

17. http://www.amstat.org/publications/jse/jse_data_archive.htm.