STATISTICS IN TRANSITION new series, December 2019
Vol. 20, No. 4, pp. 89–111, DOI 10.21307/stattrans-2019-036
Submitted – 28.08.2018; Paper ready for publication – 01.10.2019

ESTIMATING POPULATION COEFFICIENT OF VARIATION USING A SINGLE AUXILIARY VARIABLE IN SIMPLE RANDOM SAMPLING

Rajesh Singh[1], **Madhulika Mishra**[2]

ABSTRACT

This paper proposes an improved estimation method for the population coefficient of variation, which uses information on a single auxiliary variable. The authors derived the expressions for the mean squared error of the proposed estimators up to the first order of approximation. It was demonstrated that the estimators proposed by the authors are more efficient than the existing ones. The results of the study were validated by both empirical and simulation studies.

Key words: coefficient of variation, simple random sampling, auxiliary variable, mean square error.

1. Introduction

It is a prominent fact in the theory of sample surveys that suitable use of auxiliary information increases the efficiency of the estimators used for estimating the unknown population parameters. Some important works illustrating use of auxiliary information at estimation stage are Singh et al. (2005), Singh et al. (2007), Khoshnevisan et al. (2007), Singh et al. (2009), Singh and Kumar (2011), Malik and Singh (2013) and Singh et al. (2018). Over a vast period of time a substantial amount of work has been done by several authors for the estimation of population mean, population variance but little attention has been given to the estimation of the population coefficient of variation. Das and Tripathi (1992–93) first proposed the estimator for the coefficient of variation when samples were selected using simple random sampling without replacement (SRSWOR) scheme. Other works include Patel and Shah (2009) and Ahmed, S.E. (2002). Breunig (2001) suggested an almost unbiased estimator of the coefficient of variation. Sisodia and Dwivedi (1981) suggested a modified ratio estimator using the coefficient of variation of auxiliary variable.

[1] Department of Statistics, Banaras Hindu University, Varanasi-221005, India. ORCID ID: http://orcid.org/0000-0002-9274-8141.
[2] Corresponding Author: Department of Statistics, Banaras Hindu University, Varanasi-221005, India. E-mail: madhulika1707@gmail.com. ORCID ID: http://orcid.org/0000-0002-6408-0746.

Rajyaguru and Gupta (2005) also worked on the problem of estimation of the coefficient of variation under simple random sampling and stratified random sampling.

The coefficient of variation is extensively used in biology, agriculture and environmental sciences.

A brief summary of the paper is as follows.

Section 1 is introductory in nature, comprises the works that have been already done in the sampling literature. In Section 2 we considered five estimators for comparison purposes and their properties. In Section 3, we proposed two log type estimators for the coefficient of variation, one general type estimator and one wider type. In Section 4, an empirical study was carried out in support of our results. In Section 5, we carried out a simulation study to validate our theoretical results and have presented them with the help of bar graphs. In Section 6 we finally concluded our results.

Let us consider a finite population P = (P_1, P_2……… P_N) of size 'N' consisting of distinct and identifiable units. Let the study and auxiliary variables be denoted by Y and X, and let Y_i and X_i be their values corresponding to ith unit in the population (i = 1, 2……… N). We define:

$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ as the population mean for the study variable

$\bar{X} = \frac{1}{N}\sum_{i=1}^{N} X_i$ as the population mean for the auxiliary variable

$S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2$ as the population mean square for the study variable

$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2$ as the population mean square for the auxiliary variable

$S_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)\left(X_i - \bar{X}\right)$ as the population covariance between the study and auxiliary variable, X and Y.

Let us suppose that a sample of size 'n' has been drawn from this population of size 'N' units using SRSWOR technique. For this sample let y_i and x_i denote values of the i^{th} sample unit corresponding to study variable Y and auxiliary variable X respectively.

For the sample observations, we define:

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ as the sample mean for the study variable Y

$\bar{x} = \frac{1}{n}\sum_{i=1}^{N} x_i$ are the sample mean for the auxiliary variable X

$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$ as the sample mean square for the study variable

$s_x^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ as the sample mean square for the auxiliary variable

$s_{xy} = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})$ as the sample covariance term.

Now, let us define

$\in_0 = \frac{\bar{y}}{\bar{\bar{Y}}} - 1, \in_1 = \frac{\bar{x}}{\bar{\bar{X}}} - 1, \in_2 = \frac{s_y^2}{S_y^2} - 1$ and $\in_3 = \frac{s_x^2}{S_x^2} - 1$

such that

$$E(\in_0) = E(\in_1) = E(\in_2) = E(\in_3) = 0$$

$$E(\in_0^2) = \left(\frac{1-f}{n}\right)C_y^2, \quad E(\in_1^2) = \left(\frac{1-f}{n}\right)C_x^2, \quad E(\in_2^2) = \left(\frac{1-f}{n}\right)(\lambda_{40} - 1),$$

$$E(\in_3^2) = \left(\frac{1-f}{n}\right)(\lambda_{04} - 1)$$

$$E(\in_0 \in_1) = \left(\frac{1-f}{n}\right)\rho C_y C_x, \quad E(\in_0 \in_2) = \left(\frac{1-f}{n}\right)C_y \lambda_{30} \quad E(\in_0 \in_3) = \left(\frac{1-f}{n}\right)C_y \lambda_{12},$$

$$E(\in_1 \in_2) = \left(\frac{1-f}{n}\right)C_x \lambda_{21}, \quad E(\in_1 \in_3) = \left(\frac{1-f}{n}\right)C_x \lambda_{03},$$

$$E(\in_2 \in_3) = \left(\frac{1-f}{n}\right)(\lambda_{22} - 1)$$

Here, $f = \frac{n}{N}$: Sampling fraction, $C_y = \frac{S_y}{\bar{\bar{Y}}}$ and $C_x = \frac{S_x}{\bar{\bar{X}}}$ are the population coefficient of variation for the study variable Y and auxiliary variable X, respectively. Also ρ_{xy} denotes the correlation coefficient between X and Y.

In general,

$$\mu_{rs} = \frac{1}{N-1}\sum_{i=1}^{N}\left(y_i - \bar{Y}\right)^r\left(x_i - \bar{X}\right)^s \text{ and } \lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{r/2}\,\mu_{02}^{s/2}} \text{ respectively.}$$

2. Existing estimators

- The usual unbiased estimator to estimate the population coefficient of variation using information on a single auxiliary variable is defined below:

$$t_0 = \hat{C}_y = \frac{s_y}{\bar{y}} = \frac{S_y\left(1+\in_2\right)^{1/2}}{\bar{Y}\left(1+\in_0\right)}$$

$$\approx\left(1-\in_0 + \in_0^2 + \frac{\in_2}{2} - \frac{\in_0\in_2}{2} - \frac{\in_2^2}{8}\right)C_y \tag{2.1}$$

Its mean squared error (MSE) is given by:

$$MSE(t_0) \approx C_y^2\left(\frac{1-f}{n}\right)\left[C_y^2 + \frac{\lambda_{40}-1}{4} - C_y\lambda_{30}\right] \tag{2.2}$$

- Solanki et al. (2015) introduced a difference type estimator for the population coefficient of variation C_y as:

$$C_d = \hat{C}_y + \alpha_2\left(C_x - \hat{C}_x\right) \tag{2.3}$$

MSE of C_d is given by:

$$MSE(C_d) \approx C_y^2\left(\frac{1-f}{n}\right)\left[\left\{C_y^2 + \frac{\lambda_{40}-1}{4} - C_y\lambda_{30}\right\} - C_y^2\,\frac{\left\{\rho C_y C_x - \frac{C_y\lambda_{12}}{2} - \frac{C_x\lambda_{21}}{2} + \frac{\lambda_{22}-1}{4}\right\}}{\left\{C_x^2 + \frac{\lambda_{04}-1}{4} - C_x\lambda_{03}\right\}}\right]$$

(2.4)

- Solanki et al. (2015) defined another class of estimator for the population coefficient of variation C_y as:

$$C_d^* = \alpha_1 \hat{C}_y + \alpha_2 \left(\hat{C}_x - C_x\right) \tag{2.5}$$

MSE of C_d^* is given by:

$$MSE\left(C_d^*\right) \approx \alpha_1^2 C_y^2 A + \alpha_2^2 C_x^2 B + C_y^2 + 2\alpha_1\alpha_2 C_y C_x C - 2\alpha_1 C_y^2 D - 2\alpha_2 C_y C_x E \tag{2.6}$$

Here,

$$\left.\begin{aligned}
A &= 1 + \left(\frac{1-f}{n}\right)\left\{3C_y^2 - 2C_y\lambda_{30}\right\} \\
B &= \left(\frac{1-f}{n}\right)\left\{C_x^2 + \frac{\lambda_{04}-1}{4} - C_x\lambda_{03}\right\} \\
C &= \left(\frac{1-f}{n}\right)\left\{C_x^2 - \frac{\lambda_{04}-1}{8} - \frac{C_x\lambda_{03}}{2} + \rho C_y C_x - \frac{C_y\lambda_{12}}{2} - \frac{C_x\lambda_{21}}{2} + \frac{\lambda_{22}-1}{4}\right\} \\
D &= 1 + \left(\frac{1-f}{n}\right)\left\{C_y^2 - \frac{\lambda_{40}-1}{8} - \frac{C_y\lambda_{30}}{2}\right\} \\
E &= \left(\frac{1-f}{n}\right)\left\{C_x^2 - \frac{\lambda_{04}-1}{8} - \frac{C_x\lambda_{03}}{2}\right\}.
\end{aligned}\right\} \tag{2.7}$$

On differentiating equation (2.6) with respect to α_1 and α_2, we obtain their optimum values as:

$$\alpha_{1opt} = \frac{BD - CE}{AB - C^2} \tag{2.8}$$

$$\alpha_{2opt} = \frac{C_y}{C_x}\left(\frac{AE - CD}{AB - C^2}\right) \tag{2.9}$$

On substituting these optimum values of α_1 and α_2, in equation (2.6), we obtain the Minimum MSE for the estimator C_d^* as:

$$\min MSE\left(C_d^*\right) \approx \alpha_{1opt}^2 C_y^2 A + \alpha_{2opt}^2 C_x^2 B + C_y^2 + 2\alpha_{1opt}\alpha_{2opt} C_y C_x C - 2\alpha_{1opt} C_y^2 D - 2\alpha_{2opt} C_y C_x E \tag{2.10}$$

- Adichwal et al. (2016) proposed a two-parameter ratio-product-ratio estimator for the population coefficient of variation as:

$$t_{r1}=\alpha\left[\frac{(1-\beta)\bar{x}+\beta\bar{X}}{\beta\bar{x}+(1-\beta)\bar{X}}\right]\hat{C}_y+(1-\alpha)\left[\frac{\beta\bar{x}+(1-\beta)\bar{X}}{(1-\beta)\bar{x}+\beta\bar{X}}\right]\hat{C}_y \tag{2.11}$$

$$t_{r2}=\gamma\left[\frac{(1-\delta)s_x^2+\delta S_x^2}{\delta s_x^2+(1-\delta)S_x^2}\right]\hat{C}_y+(1-\gamma)\left[\frac{\delta s_x^2+(1-\delta)S_x^2}{(1-\delta)s_x^2+\delta S_x^2}\right]\hat{C}_y \tag{2.12}$$

MSE of the estimators t_{r1} and t_{r2} are respectively given by:

$$MSE(t_{r1})\approx \mathrm{MSE}(\mathrm{C_y})-\frac{1}{4}\left(\frac{1-f}{n}\right)\{2\rho_{xy}C_y-\lambda_{21}\}^2C_y^2 \tag{2.13}$$

$$MSE(t_{r2})\approx \mathrm{MSE}(\mathrm{C_y})-\frac{1}{4}\left(\frac{1-f}{n}\right)\frac{\{(\lambda_{22}-1)-2\lambda_{12}C_y\}^2}{(\lambda_{04}-1)}C_y^2 \tag{2.14}$$

3. Proposed estimators

We have proposed some estimators for the coefficient of variation based on information on a single auxiliary variable.

Motivated by Mishra and Singh (2017), we propose improved log type estimators for estimating the population coefficient of variation given by:

estimators t_1 and t_2 as:

a) $$t_1=\hat{\mathrm{C}}_\mathrm{y}+\alpha\log\left(\frac{\hat{C}_x}{C_x}\right) \tag{3.1}$$

b) $$t_2=\hat{\mathrm{C}}_\mathrm{y}(1+w_1)+w_2\log\left(\frac{\hat{C}_x}{C_x}\right) \tag{3.2}$$

Expressing the estimator t_1 and in terms of $\in's$ and then taking expectations up to the first order of approximation, we get MSE of the estimator t_1 as:

$$MSE(t_1)\approx \mathrm{C_y^2}\left(\frac{1-f}{n}\right)\left[\mathrm{C_y^2}+\frac{(\lambda_{40}-1)}{4}-C_y\lambda_{30}\right]+\alpha^2\left(\frac{1-f}{n}\right)\left[\mathrm{C_x^2}+\frac{(\lambda_{04}-1)}{4}-C_x\lambda_{03}\right]+$$
$$2C_y\alpha\left(\frac{1-f}{n}\right)\left[\rho C_yC_x+\frac{(\lambda_{22}-1)}{4}-\frac{(C_y\lambda_{12})}{2}-\frac{(C_x\lambda_{21})}{2}\right] \tag{3.3}$$

$$MSE(t_1) = C_y^2 A_1 + \alpha^2 A_2 + 2C_y \alpha A_3 \tag{3.4}$$

Here,

$$\left.\begin{aligned} A_1 &= \left(\frac{1-f}{n}\right)\left\{C_y^2 + \frac{(\lambda_{40}-1)}{4} - C_y\lambda_{30}\right\}, \\ A_2 &= \left(\frac{1-f}{n}\right)\left\{C_x^2 + \frac{(\lambda_{04}-1)}{4} - C_x\lambda_{03}\right\}, \\ A_3 &= \left(\frac{1-f}{n}\right)\left\{\rho C_y C_x + \frac{(\lambda_{22}-1)}{4} - \frac{C_x\lambda_{21}}{2} - \frac{C_y\lambda_{12}}{2}\right\}. \end{aligned}\right\} \tag{3.5}$$

To obtain the optimum value of α, we partially differentiate the expression (3.4) with respect to α and we obtain the optimum value as:

$$\alpha_{opt} = -C_y \frac{A_3}{A_2} \tag{3.6}$$

Putting this optimum value of α in equation (3.4), we get the minimum value for $MSE(t_1)$ as:

$$\min MSE(t_1) \approx C_y^2\left(A_1 - \frac{A_3^2}{A_2}\right) \tag{3.7}$$

Expressing the estimators t_2 in terms of $\in$'s and then taking expectations up to the first order of approximation we get MSE of the estimator t_2 as:

$$MSE(t_2) \approx C_y^2\left(\frac{1-f}{n}\right)\left[C_y^2 + \frac{\lambda_{40}-1}{4} - C_y\lambda_{30}\right] + C_y^2 w_1^2\left[1 + 3\left(\frac{1-f}{n}\right)C_y^2 - 2\left(\frac{1-f}{n}\right)C_y\lambda_{30}\right] +$$

$$w_2^2\left(\frac{1-f}{n}\right)\left[C_x^2 + \frac{\lambda_{04}-1}{4} - C_x\lambda_{03}\right] + 2C_y^2 w_1\left(\frac{1-f}{n}\right)\left[2C_y^2 + \frac{\lambda_{40}-1}{8} - \frac{3}{2}C_y\lambda_{30}\right] +$$

$$2C_y w_1 w_2\left(\frac{1-f}{n}\right)\left[\frac{C_x^2}{2} + \rho C_y C_x + \frac{\lambda_{22}-1}{4} - \frac{\lambda_{04}-1}{4} - \frac{C_y\lambda_{12}}{2} - \frac{C_x\lambda_{21}}{2}\right] +$$

$$2C_y w_2\left(\frac{1-f}{n}\right)\left[\rho C_y C_x + \frac{\lambda_{22}-1}{4} - \frac{C_y\lambda_{12}}{2} - \frac{C_x\lambda_{21}}{2}\right] \tag{3.8}$$

$$MSE(t_2) = C_y^2 B_1 + C_y^2 w_1^2 B_2 + w_2^2 B_3 + 2C_y^2 w_1 B_4 + 2C_y w_1 w_2 B_5 + 2C_y w_2 B_6 \tag{3.9}$$

Here,

$$\left.\begin{aligned}
B_1 &= \left(\frac{1-f}{n}\right)\left[C_y^2 + \frac{\lambda_{40}-1}{4} - C_y\lambda_{30}\right]\\
B_2 &= 1 + 3\left(\frac{1-f}{n}\right)C_y^2 - 2\left(\frac{1-f}{n}\right)C_y\lambda_{30}\\
B_3 &= \left(\frac{1-f}{n}\right)\left[C_x^2 + \frac{\lambda_{04}-1}{4} - C_x\lambda_{03}\right]\\
B_4 &= \left(\frac{1-f}{n}\right)\left[2C_y^2 + \frac{\lambda_{40}-1}{8} - \frac{3}{2}C_y\lambda_{30}\right]\\
B_5 &= \left(\frac{1-f}{n}\right)\left[\frac{C_x^2}{2} + \rho C_y C_x + \frac{\lambda_{22}-1}{4} - \frac{\lambda_{04}-1}{4} - \frac{C_y\lambda_{12}}{2} - \frac{C_x\lambda_{21}}{2}\right]\\
B_6 &= \left(\frac{1-f}{n}\right)\left[\rho C_y C_x + \frac{\lambda_{22}-1}{4} - \frac{C_y\lambda_{12}}{2} - \frac{C_x\lambda_{21}}{2}\right]
\end{aligned}\right\} \tag{3.10}$$

To obtain the optimum value of w_1 and w_2, we differentiate the expression (2.21) with respect to w_1 and w_2 and obtain the optimum values as:

$$w_{1opt} = \left(\frac{B_5B_6 - B_3B_4}{B_2B_3 - B_5^2}\right) \tag{3.11}$$

$$w_{2opt} = C_y\left(\frac{B_6B_2 - B_4B_5}{B_5^2 - B_2B_3}\right) \tag{3.12}$$

Putting these optimum values of w_1 and w_2 in equation (2.21), we get the minimum value for $MSE(t_2)$ as:

$$MSE(t_2) = C_y^2 B_1 + C_y^2 w_{1opt}^2 B_2 + w_{2opt}^2 B_3 + 2C_y^2 w_{1opt} B_4 + 2C_y w_{1opt} w_{2opt} B_5 + 2C_y w_{2opt} B_6 \quad (3.13)$$

c) Following Srivastava and Jhajj (1981), we propose a general class of estimators to estimate the population coefficient of variation C_y of the study variable Y using known mean and known variance of auxiliary variable X as:

$$t_3 = \hat{C}_y H(u,v) \quad (3.14)$$

where $u = \frac{\bar{x}}{\bar{X}}$, $v = \frac{s_x^2}{S_x^2}$ and $H(u,v)$ is a function of u and v such that the point (u,v) assumes the value in a closed convex subset R_2 of two-dimensional real space containing the point $(1,1)$;

The function $H(u,v)$ is continuous and bounded in R_2;
$H(1,1) = 1$;

The first and the second order partial derivatives of $H(u,v)$ exist and are continuous and bounded in R_2.

Expanding $H(u,v)$ about the point $(1,1)$ in a second order Taylor's series we obtain

$$t_3 = \hat{C}_y H(u,v) = \hat{C}_y H[1 + (u-1), 1 + (v-1)] \quad (3.15)$$

$$t_3 = \hat{C}_y \left[H(1,1) + (u-1)\left.\frac{\partial H}{\partial u}\right|_{(1,1)} + (v-1)\left.\frac{\partial H}{\partial v}\right|_{(1,1)} + (u-1)^2 \frac{1}{2}\left.\frac{\partial^2 H}{\partial u^2}\right|_{(1,1)} + (v-1)^2 \left.\frac{\partial^2 H}{\partial v^2}\right|_{(1,1)} + (u-1)(v-1)\frac{1}{2}\left.\frac{\partial^2 H}{\partial u \partial v}\right|_{(1,1)} \right] \quad (3.16)$$

$$t_3 = \hat{C}_y \left[1 + \in_1 H_1 + \in_3 H_2 + \in_1^2 H_3 + \in_3^2 H_4 + \in_1 \in_3 H_5\right] \quad (3.17)$$

Here,

$$H_1 = \left.\frac{\partial H}{\partial u}\right|_{(1,1)},\quad H_2 = \left.\frac{\partial H}{\partial v}\right|_{(1,1)},\quad H_3 = \frac{1}{2}\left.\frac{\partial^2 H}{\partial u^2}\right|_{(1,1)},\quad H_4 = \left.\frac{\partial^2 H}{\partial v^2}\right|_{(1,1)},\quad H_5 = \frac{1}{2}\left.\frac{\partial^2 H}{\partial u \partial v}\right|_{(1,1)}$$

Substituting the value of $\hat{C}_y$ in the above expression (2.28), we get

$$t_3 = C_y\left(1 - \in_0 + \in_0^2 + \frac{\in_2}{2} - \frac{\in_0 \in_2}{2} - \frac{\in_2^2}{8}\right)\left(1 + \in_1 H_1 + \in_3 H_2 + \in_1^2 H_3 + \in_3^2 H_4 + \in_1 \in_3 H_5\right) \tag{3.18}$$

Mean square error of the estimator t_3 is given by

$$MSE(t_3) = \mathrm{E}\left[t_3 - C_y\right]^2 = C_y^2 \mathrm{E}\left[- \in_0 + \in_1 H_1 + \frac{\in_2}{2} + \in_3 H_2 + \mathrm{O}(\in)\right]^2 \tag{3.19}$$

Simplifying the expression (2.30), we get

$$MSE(t_3) = C_y^2\left(1 - \frac{f}{n}\right)\left[C_y^2 + C_x^2 H_1^2 - 2H_1 \rho C_y C_x + \frac{(\lambda_{40} - 1)}{4} + (\lambda_{04} - 1)H_2^2 + (\lambda_{22} - 1)H_2 + 2\left\{-\frac{C_y \lambda_{30}}{2} - C_y \lambda_{12} H_2 + \frac{C_x \lambda_{21}}{2} H_1 + C_x \lambda_{03} H_1 H_2\right\}\right] \tag{3.20}$$

In order to obtain the minimum MSE for the estimator t_3, we partially differentiate the expression (2.31) with respect to H_1 and H_2 to get the optimum values as

$$H_{1opt} = \frac{1}{2}\left[\frac{(\lambda_{04} - 1)\{2\rho C_y - \lambda_{21}\} - \lambda_{03}\{2C_y \lambda_{12} - (\lambda_{22} - 1)\}}{C_x\{(\lambda_{04} - 1) - \lambda_{03}^2\}}\right] \tag{3.21}$$

$$H_{2opt} = \frac{1}{2}\left[\frac{\lambda_{03}\{2\rho C_y - \lambda_{21}\} - \{2C_y \lambda_{12} - (\lambda_{22} - 1)\}}{\lambda_{03}^2 - (\lambda_{04} - 1)}\right] \tag{3.22}$$

Substituting these optimum values of H_1 and H_2 in equation (2.31), we obtain the expression for the minimum MSE of t_3

$$MSE(t_3) = C_y^2\left(1-\frac{f}{n}\right)\left[C_y^2 + C_x^2 H_{1opt}^2 - 2H_{1opt}\rho C_y C_x + \frac{(\lambda_{40}-1)}{4} + (\lambda_{04}-1)H_{2opt}^2 + (\lambda_{22}-1)H_{2opt} + 2\left\{-\frac{C_y\lambda_{30}}{2} - C_y\lambda_{12}H_{2opt} + \frac{C_x\lambda_{21}}{2}H_{1opt} + C_x\lambda_{03}H_{1opt}H_{2opt}\right\}\right] \quad (3.23)$$

d) Again, following Srivastava and Jhajj (1981), we propose a wider class of estimators to estimate the population coefficient of variation C_y as:

$$t_4 = H^*\left(\hat{C}_y, u, v\right) \quad (3.24)$$

where $u = \frac{\bar{x}}{\bar{X}}$, $v = \frac{s_x^2}{S_x^2}$ and $H^*\left(\hat{C}_y, u, v\right)$ is a function of $\hat{C}_y$, u and v such that the point $\left(\hat{C}_y, u, v\right)$ assumes the value in a closed convex subset R_3 of three-dimensional real space containing the point $\left(C_y, 1, 1\right)$;

The function $H^*\left(\hat{C}_y, u, v\right)$ is continuous and bounded in R_3;

$H^*\left(C_y, 1, 1\right) = C_y$;

The first and the second order partial derivatives of $H^*\left(\hat{C}_y, u, v\right)$ exist and are continuous and bounded in R_3.

Expanding $H^*\left(\hat{C}_y, u, v\right)$ about the point $\left(C_y, 1, 1\right)$ in a second order Taylor's series, we have

$$t_4 = H^*\left(\hat{C}_y, u, v\right) = H^*\left[C_y + \left(\hat{C}_y - C_y\right), 1 + (u-1), 1 + (v-1)\right]$$

$$= H^*\left(C_y, 1, 1\right) + \left(\hat{C}_y - C_y\right)\left.\frac{\partial H^*}{\partial \hat{C}_y}\right|_{\left(C_{y,1,1}\right)} + (u-1)\left.\frac{\partial H^*}{\partial u}\right|_{\left(C_{y,1,1}\right)} + (v-1)\left.\frac{\partial H^*}{\partial v}\right|_{\left(C_{y,1,1}\right)} +$$

$$\left(\hat{C}_y - C_y\right)^2 \frac{1}{2}\left.\frac{\partial^2 H^*}{\partial \hat{C}_y^2}\right|_{\left(C_{y,1,1}\right)} + (u-1)^2\frac{1}{2}\left.\frac{\partial^2 H^*}{\partial u^2}\right|_{\left(C_{y,1,1}\right)} + (v-1)^2\frac{1}{2}\left.\frac{\partial^2 H^*}{\partial v^2}\right|_{\left(C_{y,1,1}\right)} +$$

$$\left(\hat{C}_y - C_y\right)(u-1)\frac{1}{2}\frac{\partial^2 H^*}{\partial \hat{C}_y \partial u}\bigg|_{(C_y,1,1)} + (u-1)(v-1)\frac{1}{2}\frac{\partial^2 H^*}{\partial u \partial v}\bigg|_{(C_y,1,1)} + (v-1)\left(\hat{C}_y - C_y\right)\frac{1}{2}\frac{\partial^2 H^*}{\partial v \partial \hat{C}_y}\bigg|_{(C_y,1,1)} \tag{3.25}$$

$$t_4 = \mathrm{C_y} + \left(\hat{C}_y - C_y\right) + \in_1 H_1^* + \in_3 H_2^* + \left(\hat{C}_y - C_y\right)^2 H_3^* + \in_1^2 H_4^* + \in_3^2 H_5^* + \left(\hat{C}_y - C_y\right)\in_1 H_6^* + \in_1\in_3 H_7^* + \left(\hat{C}_y - C_y\right)\in_3 H_8^* \tag{3.26}$$

Here,

$$H^*(C_y,1,1) = \mathrm{C_y}, \quad \frac{\partial H^*}{\partial \hat{C}_y}\bigg|_{(C_y,1,1)} = 1, \quad H_1^* = \frac{\partial H^*}{\partial u}\bigg|_{(C_y,1,1)}, \quad H_2^* = \frac{\partial H^*}{\partial v}\bigg|_{(C_y,1,1)},$$

$$H_3^* = \frac{1}{2}\frac{\partial^2 H^*}{\partial \hat{C}_y^2}\bigg|_{(C_y,1,1)}$$

$$H_4^* = \frac{1}{2}\frac{\partial^2 H^*}{\partial u^2}\bigg|_{(C_y,1,1)}, \quad H_5^* = \frac{1}{2}\frac{\partial^2 H^*}{\partial v^2}\bigg|_{(C_y,1,1)}, \quad H_6^* = \frac{1}{2}\frac{\partial^2 H^*}{\partial \hat{C}_y \partial u}\bigg|_{(C_y,1,1)},$$

$$H_7^* = \frac{1}{2}\frac{\partial^2 H^*}{\partial u \partial v}\bigg|_{(C_y,1,1)}, \quad H_8^* = \frac{1}{2}\frac{\partial^2 H^*}{\partial v \partial \hat{C}_y}\bigg|_{(C_y,1,1)}$$

Now, substituting the value of $\hat{C}_y$ in equation (2.37), we have

$$t_4 = \mathrm{C_y}\left(1 - \in_0 + \in_0^2 + \frac{\in_2}{2} - \frac{\in_0\in_2}{2} - \frac{\in_2^2}{8}\right) + \in_1 H_1^* + \in_3 H_2^* + \left\{C_y\left(1 - \in_0 + \in_0^2 + \frac{\in_2}{2} - \frac{\in_0\in_2}{2} - \frac{\in_2^2}{8}\right) - C_y\right\}^2 H_3^*$$

$$\in_1^2 H_4^* + \in_3^2 H_5^* + \left\{C_y\left(1 - \in_0 + \in_0^2 + \frac{\in_2}{2} - \frac{\in_0\in_2}{2} - \frac{\in_2^2}{8}\right) - C_y\right\}\in_1 H_6^* + \in_1\in_3 H_7^* +$$

$$\left\{C_y\left(1 - \in_0 + \in_0^2 + \frac{\in_2}{2} - \frac{\in_0\in_2}{2} - \frac{\in_2^2}{8}\right) - C_y\right\}\in_3 H_8^* \tag{3.27}$$

$$MSE(t_4) = \mathrm{E}\left[\mathrm{t}_4 - C_y\right]^2 = \mathrm{E}\left[C_y\left(-\in_0 + \frac{\in_2}{2}\right) + \in_1 H_1^* + \in_3 H_2^* + \mathrm{O}(\in)\right]^2 \tag{3.28}$$

After simplifying the expression (2.39), we get:

$$MSE(t_4)=\left(\frac{1-f}{n}\right)\left[C_y^2\left\{C_y^2+\frac{(\lambda_{40}-1)}{4}-C_y\lambda_{30}\right\}+C_x^2H_1^{*2}+(\lambda_{04}-1)H_2^{*2}+2C_y\left(\frac{C_x\lambda_{21}}{2}-\rho C_yC_x\right)H_1^*+\right.$$
$$\left.2C_x\lambda_{03}H_1^*H_2^*+2C_y\left(\frac{(\lambda_{22}-1)}{2}-C_y\lambda_{12}\right)H_2^*\right] \quad (3.29)$$

In order to obtain the minimum MSE for the estimator t_4 we partially differentiate the expression (2.40) with respect to H_1^* and H_2^* and obtain optimum values as:

$$\left.\begin{aligned} H_{1opt}^* &= \frac{C_y}{2C_x}\left[\frac{(\lambda_{04}-1)\{2\rho C_y-\lambda_{21}\}-\lambda_{03}\{2C_y\lambda_{12}-(\lambda_{22}-1)\}}{\lambda_{04}-\lambda_{03}^2-1}\right] \\ H_{2opt}^* &= \frac{C_y}{2}\left[\frac{\lambda_{03}(2\rho C_y-\lambda_{21})-\{2C_y\lambda_{12}-(\lambda_{22}-1)\}}{\lambda_{03}^2-(\lambda_{04}-1)}\right] \end{aligned}\right\} \quad (3.30)$$

Substituting these optimum values of H_1^* and H_2^* in equation (2.40), we obtain the expression for the minimum MSE of t_4

$$MSE(t_4)=\left(\frac{1-f}{n}\right)\left[C_y^2\left\{C_y^2+\frac{(\lambda_{40}-1)}{4}-C_y\lambda_{30}\right\}+C_x^2H_{1opt}^{*\,2}+(\lambda_{04}-1)H_{2opt}^{*\,2}+2C_y\left(\frac{C_x\lambda_{21}}{2}-\rho C_yC_x\right)H_{1opt}^*+\right.$$
$$\left.2C_x\lambda_{03}H_{1opt}^*H_{2opt}^*+2C_y\left(\frac{(\lambda_{22}-1)}{2}-C_y\lambda_{12}\right)H_{2opt}^*\right] \quad (3.31)$$

4. Empirical study

In this section, we have carried out an empirical study to explicate the performance of our proposed estimator. We used the following data sets:

Population I: [Source : Murthy (1967), p. 399].

X: Area under wheat in 1963,

Y: Area under wheat in 1964,

N=34, n=15,

$\overline{X}$=208.88, $\overline{Y}$=199.44,

$C_X = 0.72$, C_Y=0.75, $\rho_{XY} = 0.98$,

$\lambda_{21} = 1.0045$, $\lambda_{12} = 0.9406$, $\lambda_{40} = 3.6161$, $\lambda_{04} = 2.8266$, $\lambda_{30} = 1.1128$, $\lambda_{03} = 0.9206$, $\lambda_{22} = 3.0133$

Population II: [Source : Sarjinder Singh (2003), p. 1116].

X: Number of fish caught in year 1993,

Y: Number of fish caught in year 1995,

N=69, n=40,

$\overline{X}$=4591.07, $\overline{Y}$=4514.89,

$C_X = 1.38$, C_Y=1.35,

$\lambda_{21} = 2.19$, $\lambda_{12} = 2.30$, $\lambda_{40} = 7.66$, $\lambda_{04} = 9.84$, $\lambda_{30} = 1.11$, $\lambda_{03} = 2.52$, $\lambda_{22} = 8.19$

In order to determine the Percent Relative Efficiency (PRE) of the estimators we have used the following formula

$$PRE(t, t_o) = \frac{Var(t_0)}{MSE(t)} \times 100$$

where $t = C_d, C_d^*, t_{r1}, t_{r2}, t_1, t_2, t_3.t_4$.

Table 1. MSE and PRE of the estimators

ESTIMATOR	POPULATION-1		POPULATION-2	
	MSE	PRE	MSE	PRE
t_0	0.008016	100.00	0.0380	100.00
C_d	0.00123	651.7051	0.0298	127.4607
C_d^*	0.00122	654.4814	0.0285	133.2609
t_{r1}	0.006868	116.54	0.037313	102.04
t_{r2}	0.006963	114.95	0.037563	101.36

Table 1. MSE and PRE of the estimators (cont.(

ESTIMATOR	POPULATION-1		POPULATION-2	
	MSE	PRE	MSE	PRE
t_1	0.00123	651.7356	0.0299	127.4598
t_2	0.001038	771.9898	0.0283	134.6127
t_3	0.001203	666.4304	0.0297	128.345
t_4	0.001203	666.4304	0.0297	128.345

We can summarize the results from Table 1 as:

All the proposed estimators t_1, t_2, t_3 and t_4 are more efficient than the usual unbiased estimator t_0.The estimator t_1 turns out to be nearly as efficient as the difference type estimator C_d while all the remaining estimators, t_2, t_3 and t_4 are more efficient than the estimators C_d, C_d^*, t_{r1} and t_{r2}. Among all the estimators, t_2 is the most efficient because of the smallest value of MSE and highest value of PRE.

5. Simulation studies

This section describes the procedure that we adopted for the simulation study. We have used R programming for calculating MSE of the existing and proposed estimators. We followed the procedure adopted by Reddy et al. (2010) and have generated bivariate population with a specified correlation coefficient between the study and auxiliary variable. The algorithm is as follows:

1. Generate two independent random variables X from $N(\mu, \sigma^2)$ and Z from $N(\mu_1, \sigma_1^2)$ using Box-Muller method (Jhonson, 1987).
2. Set $Y = \rho X + \sqrt{1-\rho^2} Z$ where $0 < \rho = 0.75, 0.85, 0.95 < 1$.
3. Consider the population with the parameters $\mu = 2.5$, $\sigma^2 = 2$, $\mu_1 = 5$ $\sigma_1^2 = 3$ and repeat the steps 1-2 2000 times.
4. From the population of size N=2000, draw 1500 simple random samples (y_i, x_i) $(i = 1, 2, ..., n)$ without replacement of size $n = 30, 50, 70$.

5. For each of the sample, compute MSE of the estimators t_o , Cd , Cd^*, t_1, t_2, t_{r1} and t_{r2}.

6. Compute the average MSE of the estimator by the following formula:

$$MSE(i) = \frac{1}{1500}\sum_{j=1}^{1500} mse_j(i) \quad \text{where } i = t_o, Cd, Cd^*, t_1, t_2, t_{r1} \text{ and } t_{r2}.$$

Table 2. Table showing MSE and PRE of the existing and proposed estimators for different values of ρ and n

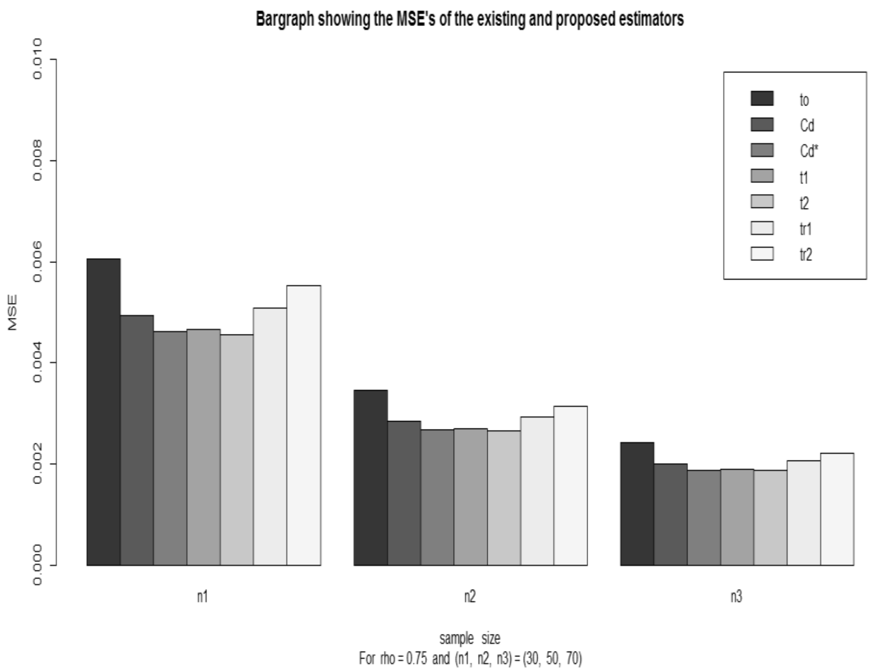
ρ	n	Estimator	MSE	PRE
		t_o	0.006053626	100.0000
		Cd	0.004924006	122.9410
		Cd^*	0.004617663	131.0970
		t_{r1}	0.005080027	119.1651
		t_{r2}	0.005532107	109.4270
		t_1	0.004668748	129.6626
		t_2	0.004552470	132.9744
	50	t_o	0.003450581	100.0000
		Cd	0.002835622	121.6869
		Cd^*	0.002671694	129.1533
		t_{r1}	0.002688403	117.5860
		t_{r2}	0.002650186	109.5933
		t_1	0.002934516	128.3506
		t_2	0.003148534	130.2015
	70	t_o	0.002412659	100.0000
		Cd	0.001990824	121.1889
		Cd^*	0.001879170	128.3896
		t_{r1}	0.002062564	116.9737
		t_{r2}	0.002200267	109.6530
		t_1	0.001887289	127.8373
		t_2	0.001868644	129.1128

0.85	30	t_o	0.006358341	100.0000
		Cd	0.004912595	129.4294
		Cd^*	0.003809327	166.9151
		t_{r1}	0.004876890	130.3769
		t_{r2}	0.005133219	123.8665
		t_1	0.003831045	165.9688
		t_2	0.003739557	170.0293
	50	t_o	0.003621428	100.0000
		Cd	0.002828737	128.0228
		Cd^*	0.002203557	164.3447
		t_{r1}	0.002825058	128.1895
		t_{r2}	0.002910006	124.4474
		t_1	0.002210627	163.8190
		t_2	0.002180249	166.1016
	70	t_o	0.002527309	100.0000
		Cd	0.001982556	127.4561
		Cd^*	0.001547597	163.3054
		t_{r1}	0.001984483	127.3535
		t_{r2}	0.002027634	124.6433
		t_1	0.001551035	162.9434
		t_2	0.001536162	164.5210

95	30	t_o	0.008647395	100.0000
		Cd	0.005851489	147.7811
		Cd^*	0.002426053	356.4389
		t_{r1}	0.005461214	158.3420
		t_{r2}	0.005113439	169.1094
		t_1	0.002430095	355.8459
		t_2	0.002364604	365.7015
	50	t_o	0.004896276	100.0000
		Cd	0.003355658	145.9111
		Cd^*	0.001397620	350.3295
		t_{r1}	0.003172583	154.3309
		t_{r2}	0.002841595	172.3073
		t_1	0.001398209	350.1820
		t_2	0.001376286	355.7601
	70	t_o	0.0034018746	100.0000
		Cd	0.0023435450	145.1593
		Cd^*	0.009782472	347.7520
		t_{r1}	0.0022234723	152.9983
		t_{r2}	0.0019631488	173.2866
		t_1	0.0009784789	347.6697
		t_2	0.0009677328	351.5304

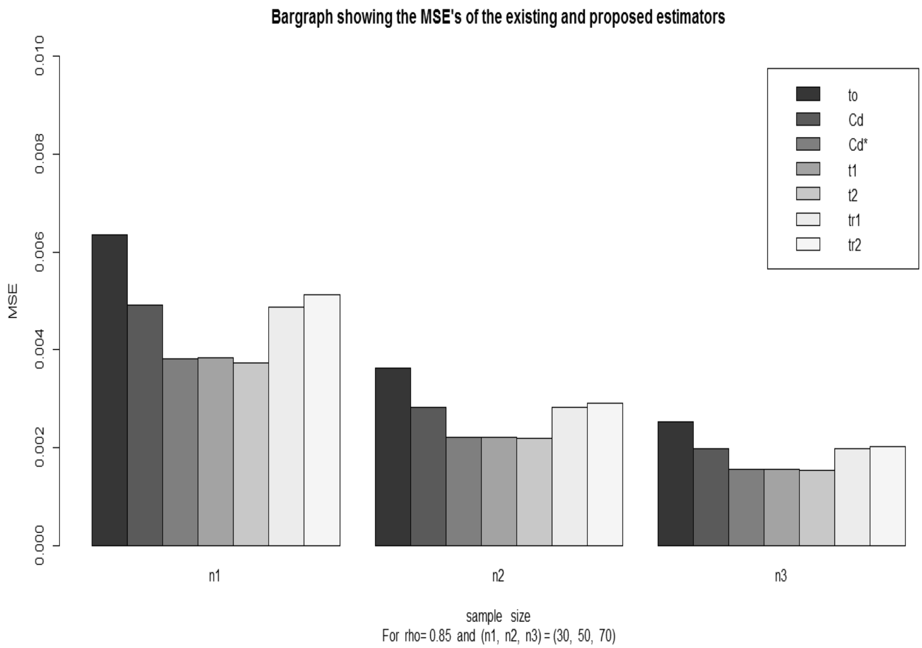
From the table, we can observe that for a particular value of ρ the value of MSE of the estimators decreases as the sample size increases. Also, we can see that in each of the cases among the proposed estimators t_1 and t_2 , t_2 is more efficient amongst all the existing estimators t_o , Cd , Cd^*, t_{r1}, t_{r2} and the proposed estimator t_1 while the estimator t_1 turns out to be more efficient than the existing estimators t_o , Cd , t_{r1}, t_{r2} and nearly as efficient as the estimator Cd^*. Hence, it turns out that the proposed estimator performs better than the existing estimators, therefore it is desirable to use the estimator in practice.

We have also shown the results through a bar diagram as below:



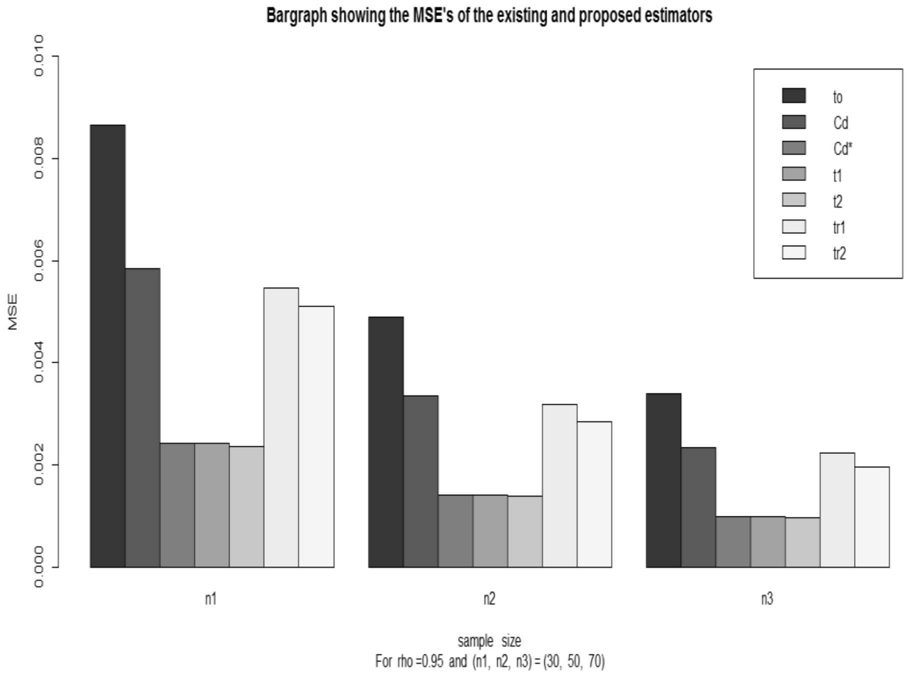
Bar graph showing MSEs of the existing and proposed estimators for $\rho = 0.75$ and (n1, n2, n3)= (30, 50, 70)

Explanation: It can be seen from the bar graph that for $\rho = 0.75$, MSE of all the estimators decreases as the value of the sample size (n) increases. And for a particular value of n, estimator t_2 has the least MSE among all the other estimators.



Bar graph showing MSEs of the existing and proposed estimators $\rho = 0.85$ and (n1, n2, n3) = (30, 50, 70)

Explanation: It can be seen from the bar graph that for $\rho = 0.85$, MSE of all the estimators decreases as the value of the sample size (n) increases. And for a particular value of n, estimator t_2 has the least MSE among all the other estimators.



Bar graph showing MSE of the existing and proposed estimators $\rho = 0.85$ and (n1, n2, n3)= (30, 50, 70)

Explanation: It can be seen from the bar graph that for $\rho = 0.95$, MSE of all the estimators decreases as the value of the sample size (n) increases. And for a particular value of n, estimator t_2 has the least MSE among all the other estimators.

Combined Explanation: From the above three bar graphs it can be summarized that for every value of $\rho = (0.75, 0.85, 0.95)$, the increase in the sample size causes a decrease in the mean square error of all the estimators. It is also evident that for a particular value of n, t_2 has the minimum MSE as compared to the other estimators.

6. Conclusion

In this paper we have proposed estimators for the population coefficient of variation and compared them with some existing estimators and saw from the empirical and simulation studies that the proposed estimator t_2 performs better than all the existing estimators t_o , Cd , C_d^* , t_{r1}, t_{r2} and the proposed estimator t_1. As regards t_1, it performs better than the estimators t_o , Cd , t_{r1}, t_{r2} but is no more better than the estimator C_d^*. For a better understanding of our results we have also considered a graphical approach and considered bar graphs to depict our results.

Acknowledgement

The authors are grateful and obliged to the Editor-in-Chief Prof. Włodzimierz Okrasa and the anonymous referees who devoted a part of their valuable time to provide us with their fruitful recommendations, which in turn helped us to improve the manuscript.

REFERENCES

AHMED, S. E., (2002). Simultaneous estimation of Co-efficient of Variation. Journal of Statistical Planning and Inference, 104, pp. 31–51.

ARCHANA, V., RAO, A., (2014). Some Improved Estimators of Co-efficient of Variation from Bi-variate normal distribution: A Monte Carlo Comparison. Pakistan Journal of Statistics and Operation Research, 10(1).

BREUNIG, R., (2001). An almost unbiased estimator of the co-efficient of variation. Economics Letters, 70(1), pp. 15–19.

DAS, A. K., TRIPATHI, T. P., (1981). A class of estimators for co-efficient of variation using knowledge on coefficient of variation of an auxiliary character, In annual conference of Ind. Soc. Agricultural Statistics, Held at New Delhi, India.

DAS, A. K., TRIPATHI, T. P., (1992). Use of auxiliary information in estimating the coefficient of variation, Alig. J. of. Statist, 12, pp. 51–58.

FINITE POPULATION-II, Model Assisted Statistics and application, 1(1), pp. 57–66.

KHOSHNEVISAN, M., SINGH, R., CHAUHAN, P., SAWAN, N., SMARANDACHE, F. (2007). A general family of estimators for estimating population means using known value of some population parameter(s), Far East Journal of Statistics, 22(2), pp. 181–191.

MALIK, S., SINGH, R., (2013). An improved estimator using two auxiliary attributes, Appli. Math. Compt., 219, pp. 10983–10986.

MISHRA, P., SINGH, R., (2017). A new log-product type estimator using auxiliary information, Jour. Sci. Res., 61(1&2), pp. 179–183.

MURTHY, M. N., (1967). Sampling theory and methods, Sampling theory and methods.

PATEL, P. A., RINA, S., (2009). A Monte Carlo comparison of some suggested estimators of co-efficient of variation in finite population, Journal of Statistics sciences, 1(2), pp. 137–147.

RAJYAGURU, A., GUPTA, P., (2005). On the estimation of the co-efficient of variation from

SINGH, H. P., TAILOR, R., (2005). Estimation of finite population mean with known coefficient of variation of an auxiliary character, Statistica, 65(3), pp. 301–313.

SINGH, H. P., SINGH, R., (2002). A class of chain ratio- type estimators for the coefficient of variation of finite population in two phase sampling, Aligarh Journal of Statistics, Vol. 22, pp. 1–9.

SINGH, S., (2003). Advanced Sampling Theory With Applications: How Michael Selected Amy (Vol. 2), Springer Science and Business Media.

SINGH, H. P., SINGH, R., ESPEJO, M. R., PINEDA, M. D., NADRAJAH, S., (2005). On the efficiency of a dual to ratio-cum-product estimator in sample surveys. Mathematical proceedings of the Royal Irish Academy, 105A (2), pp. 51–56.

SINGH, R., CHAUHAN, P., SAWAN, N., (2007). A family of estimators for estimating population means using known correlation coefficient in two-phase sampling. Statistics in Transition, 8(1), pp. 89–96.

SINGH, R., KUMAR, M., CHAUDHARY, M. K., KADILAR, C., (2009). Improved Exponential Estimator in Stratified Random Sampling, Pak. J. Stat. Oper. Res., 5(2), pp 67–82.

SINGH, R., KUMAR, M., (2011). A note on transformations on auxiliary variable in survey sampling. Mod. Assis. Stat. Appl., 6:1, pp. 17–19.

SINGH, R., MISHRA, P., BOUZA, C. N., (2018). Estimation of population mean using information on auxiliary attribute: A review, RG, DOI: 10.13140/RG.2.2.20477.87524.

SISODIA, B. V. S., DWIVEDI, V. K., (1981). Modified ratio estimator using coefficient of variation of auxiliary variable, Journal-Indian Society of Agricultural Statistics.