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.
Rajaguru 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 \( P = (P_1, P_2, \ldots, P_N) \) 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 \( i \)th unit in the population \( (i = 1, 2, \ldots, N) \). We define:

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

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

\[
S^2_y = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \text{ as the population mean square for the study variable}
\]

\[
S^2_x = \frac{1}{N-1} \sum_{i=1}^{N} (X_i - \bar{X})^2 \text{ as the population mean square for the auxiliary variable}
\]

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

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:

\[
\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \text{ as the sample mean for the study variable } Y
\]
\( \bar{x} = \frac{1}{n} \sum_{i=1}^{N} x_i \) are the sample mean for the auxiliary variable \( X \)

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

\( s^2_x = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \) 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

\[ \varepsilon_0 = \frac{\bar{y}}{\bar{Y}} - 1, \varepsilon_1 = \frac{\bar{x}}{\bar{X}} - 1, \varepsilon_2 = \frac{s^2_y}{S^2_y} - 1 \text{ and } \varepsilon_3 = \frac{s^2_x}{S^2_x} - 1 \]

such that

\[
E(\varepsilon_0) = E(\varepsilon_1) = E(\varepsilon_2) = E(\varepsilon_3) = 0
\]

\[
E(\varepsilon_0^2) = \left( \frac{1-f}{n} \right) C_y^2, \quad E(\varepsilon_1^2) = \left( \frac{1-f}{n} \right) C_x^2, \quad E(\varepsilon_2^2) = \left( \frac{1-f}{n} \right) (\lambda_{40} - 1),
\]

\[
E(\varepsilon_3^2) = \left( \frac{1-f}{n} \right) (\lambda_{04} - 1)
\]

\[
E(\varepsilon_0 \varepsilon_1) = \left( \frac{1-f}{n} \right) \rho C_y C_x, \quad E(\varepsilon_0 \varepsilon_2) = \left( \frac{1-f}{n} \right) C_y \lambda_{30}, \quad E(\varepsilon_0 \varepsilon_3) = \left( \frac{1-f}{n} \right) C_y \lambda_{12},
\]

\[
E(\varepsilon_1 \varepsilon_2) = \left( \frac{1-f}{n} \right) C_x \lambda_{21}, \quad E(\varepsilon_1 \varepsilon_3) = \left( \frac{1-f}{n} \right) C_x \lambda_{03},
\]

\[
E(\varepsilon_2 \varepsilon_3) = \left( \frac{1-f}{n} \right) (\lambda_{22} - 1)
\]

Here, \( f = \frac{n}{N} \): Sampling fraction, \( C_y = \frac{S_y}{\bar{Y}} \) and \( C_x = \frac{S_x}{\bar{X}} \) are the population coefficient of variation for the study variable \( Y \) and auxiliary variable \( X \), respectively. Also \( \rho_{xy} \) denotes the correlation coefficient between \( X \) and \( Y \).
In general,
\[ \mu_{rs} = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y}) (x_i - \bar{X})^{s} \quad \text{and} \quad \lambda_{rs} = \frac{\mu_{rs}}{\mu_{20}^{\frac{s}{2}} \mu_{02}^{\frac{s}{2}}} \]
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:

\[ t_0 = \hat{C}_y = \frac{s_y}{\bar{y}} = \frac{S_y (1 + \varepsilon_2)^{\frac{1}{2}}}{\bar{Y}(1 + \varepsilon_0)} \]

\[ \approx \left( 1 - \varepsilon_0 + \varepsilon_0^2 + \frac{\varepsilon_2}{2} - \frac{\varepsilon_0 \varepsilon_2}{2} - \frac{\varepsilon_2^2}{8} \right) C_y \] (2.1)

Its mean squared error (MSE) is given by:

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

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

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

MSE of \( C_d \) is given by:

\[ \text{MSE}(C_d) \approx C_y^2 \left( \frac{1 - f}{n} \right) \left\{ C_y^2 + \frac{\lambda_{40} - 1}{4} - C_y \hat{\lambda}_{30} \right\} - C_y^2 \left\{ \rho C_y C_x - \frac{C_y \lambda_{12}}{2} - \frac{C_x \lambda_{21}}{2} + \frac{\lambda_{22} - 1}{4} \right\} \]

\[ \left\{ C_x^2 + \frac{\lambda_{04} - 1}{4} - C_x \hat{\lambda}_{03} \right\} \] (2.4)
Solanki et al. (2015) defined another class of estimator for the population coefficient of variation $C_y$ as:

$$C_d^* = \alpha_1 \hat{C}_y + \alpha_2 \left( \hat{C}_x - C_x \right)$$  \hspace{1cm} (2.5)

MSE of $C_d^*$ is given by:

$$MSE\left(C_d^*\right) \approx \alpha_1^2 C_y^2 A + \alpha_2^2 C_x^2 B + C_y^2 + 2\alpha_1\alpha_2 C_y C_x C - 2\alpha_1 C_y^2 D - 2\alpha_2 C_y C_x E$$  \hspace{1cm} (2.6)

Here,

$$A = 1 + \left( \frac{1-f}{n} \right) \left\{ 3C_y^2 - 2C_y \lambda_{30} \right\}$$

$$B = \left( \frac{1-f}{n} \right) \left\{ C_x^2 + \frac{\lambda_{04} - 1}{4} - C_x \lambda_{03} \right\}$$

$$C = \left( \frac{1-f}{n} \right) \left\{ C_x^2 - \frac{\lambda_{04} - 1}{8} - C_x \frac{\lambda_{03}}{2} + C_y C_x - \frac{C_y \lambda_{12}}{2} - \frac{C_x \lambda_{21}}{2} + \frac{\lambda_{22} - 1}{4} \right\}$$

$$D = 1 + \left( \frac{1-f}{n} \right) \left\{ C_y^2 - \frac{\lambda_{04}}{8} - C_y \frac{\lambda_{30}}{2} \right\}$$

$$E = \left( \frac{1-f}{n} \right) \left\{ C_x^2 - \frac{\lambda_{04}}{8} - C_x \frac{\lambda_{03}}{2} \right\}$$  \hspace{1cm} (2.7)

On differentiating equation (2.6) with respect to $\alpha_1$ and $\alpha_2$, we obtain their optimum values as:

$$\alpha_{1\text{opt}} = \frac{BD - CE}{AB - C^2}$$  \hspace{1cm} (2.8)

$$\alpha_{2\text{opt}} = \frac{C_y \left( AE - CD \right)}{C_x \left( AB - C^2 \right)}$$  \hspace{1cm} (2.9)

On substituting these optimum values of $\alpha_1$ and $\alpha_2$, in equation (2.6), we obtain the Minimum MSE for the estimator $C_d^*$ as:

$$\min MSE\left(C_d^*\right) \approx \alpha_{1\text{opt}}^2 C_y^2 A + \alpha_{2\text{opt}}^2 C_x^2 B + C_y^2 + 2\alpha_{1\text{opt}}\alpha_{2\text{opt}} C_y C_x C - 2\alpha_{1\text{opt}} C_y^2 D - 2\alpha_{2\text{opt}} C_y C_x E$$  \hspace{1cm} (2.10)
Adichwal et al. (2016) proposed a two-parameter ratio-product-ratio estimator for the population coefficient of variation as:

\[
t_{r_1} = \alpha \left( \frac{\beta \bar{X}}{\bar{Y}} \right) \hat{C}_y + (1 - \alpha) \left( \frac{\beta \bar{X}}{\bar{Y}} \right) \hat{C}_y
\]

\[
t_{r_2} = \gamma \left( \frac{(1 - \delta)\bar{s}^2 + \delta \bar{S}^2}{\bar{s}^2 + (1 - \delta)\bar{S}^2} \right) \hat{C}_y + (1 - \gamma) \left( \frac{(1 - \delta)\bar{s}^2 + \delta \bar{S}^2}{(1 - \delta)\bar{s}^2 + \delta \bar{S}^2} \right) \hat{C}_y
\]

MSE of the estimators \( t_{r_1} \) and \( t_{r_2} \) are respectively given by:

\[
MSE(t_{r_1}) \approx MSE(C_y) - \frac{1}{4} \left( \frac{1 - f}{n} \right) \left[ \lambda_2 \right]^2 C_y^2
\]

\[
MSE(t_{r_2}) \approx MSE(C_y) - \frac{1}{4} \left( \frac{1 - f}{n} \right) \left[ \frac{(\lambda_2 - 1) - 2\lambda_{12} C_y}{\lambda_{04} - 1} \right]^2 C_y^2
\]

### 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) \( t_1 = \hat{C}_y + \alpha \log \left( \frac{\hat{C}_s}{C_s} \right) \)

b) \( t_2 = \hat{C}_y (1 + w_1) + w_2 \log \left( \frac{\hat{C}_s}{C_s} \right) \)

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

\[
MSE(t_1) \approx C_y^2 \left[ \frac{1 - f}{n} \right] \left[ C_y^2 + \frac{(\lambda_{40} - 1)}{4} - C_y \lambda_{30} \right] + \alpha^2 \left[ \frac{1 - f}{n} \right] \left[ C_s^2 + \frac{(\lambda_{04} - 1)}{4} - C_s \lambda_{03} \right] + 2C_y \alpha \left[ \frac{1 - f}{n} \right] \left[ \rho C_y C_s + \frac{(\lambda_{22} - 1)}{4} - \left( \frac{C_y \lambda_{12}}{2} \right) - \left( \frac{C_s \lambda_{21}}{2} \right) \right]
\]
\(MSE(t_1) = C_y^2 A_1 + \alpha^2 A_2 + 2C_y \alpha A_3\) \hspace{1cm} (3.4)

Here,
\[
A_1 = \left(\frac{1-f}{n}\right) \left\{ C_y^2 + \frac{(\lambda_{40} - 1)}{4} - C_y \lambda_{30} \right\},
\]
\[
A_2 = \left(\frac{1-f}{n}\right) \left\{ C_x^2 + \frac{(\lambda_{04} - 1)}{4} - C_x \lambda_{03} \right\},
\]
\[
A_3 = \left(\frac{1-f}{n}\right) \left\{ \rho C_y C_x + \frac{(\lambda_{22} - 1)}{4} - \frac{C_x \lambda_{21}}{2} - \frac{C_y \lambda_{12}}{2} \right\}. \hspace{1cm} (3.5)
\]

To obtain the optimum value of \(\alpha\), we partially differentiate the expression (3.4) with respect to \(\alpha\) and we obtain the optimum value as:
\[
\alpha_{opt} = -\frac{C_y A_3}{A_2} \hspace{1cm} (3.6)
\]

Putting this optimum value of \(\alpha\) in equation (3.4), we get the minimum value for \(MSE(t_1)\) as:
\[
\min MSE(t_1) \approx C_y^2 \left( A_1 - \frac{A_3^2}{A_2} \right) \hspace{1cm} (3.7)
\]

Expressing the estimators \(t_2\) in terms of \(\varepsilon_i\)'s and then taking expectations up to the first order of approximation we get MSE of the estimator \(t_2\) as:
\[
MSE(t_2) \approx C_y^2 \left( \frac{1-f}{n} \right) \left[ C_y^2 + \frac{\lambda_{40} - 1}{4} - C_y \lambda_{30} \right] + C_y^2 w_1^2 \left[ 1 + 3 \left( \frac{1-f}{n} \right) C_y^2 - 2 \left( \frac{1-f}{n} \right) C_y \lambda_{30} \right] + \]
\[
w_2^2 \left( \frac{1-f}{n} \right) \left[ C_x^2 + \frac{\lambda_{04} - 1}{4} - C_x \lambda_{03} \right] + 2C_y^2 w_1 \left( \frac{1-f}{n} \right) \left[ 2C_y^2 + \frac{\lambda_{40} - 1}{8} - \frac{3}{2} C_y \lambda_{30} \right] + \]
\[
2C_y w_1 w_2 \left( \frac{1-f}{n} \right) \left[ \frac{C_x^2}{2} + \rho C_y C_x + \frac{\lambda_{22} - 1}{4} - \frac{\lambda_{04} - 1}{4} - \frac{C_y \lambda_{12}}{2} - \frac{C_x \lambda_{21}}{2} \right] + \]
\[ 2C_y w_2 \left( \frac{1-f}{n} \right) \left[ \rho C_y C_x + \frac{\lambda_{22} - 1}{4} - \frac{C_y \lambda_{12}}{2} - \frac{C_x \lambda_{21}}{2} \right] \]  

\[ \text{MSE}(t_2) = C_y^2 B_1 + C_y w_1^2 B_2 + w_2^2 B_3 + 2C_y^2 w_1 B_4 + 2C_y w_1 w_2 B_5 + 2C_y w_2 B_6 \]  

Here,

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

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:

\[
 w_{1\text{opt}} = \left( \frac{B_5 B_6 - B_3 B_4}{B_2 B_3 - B_5^2} \right) \]  

\[
 w_{2\text{opt}} = C_y \left( \frac{B_6 B_2 - B_4 B_5}{B_5^2 - B_2 B_3} \right) \]
Putting these optimum values of \( w_1 \) and \( w_2 \) in equation (2.21), we get the minimum value for \( MSE(t_2) \) as:

\[
MSE(t_2) = C_y^2 B_1 + C_y^2 w_{1opt}^2 B_2 + w_{2opt}^2 B_3 + 2C_y w_{1opt} w_{2opt} B_4 + 2C_y w_{1opt} w_{2opt} B_5 + 2C_y w_{2opt} B_6
\]  
(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:

\[
t_3 = \hat{C}_y H(u, v)
\]  
(3.14)

where \( u = \frac{\bar{X}}{X} \), \( v = \frac{s_x^2}{S_x^2} \) 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

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

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

\[+ (u - 1)(v - 1) \frac{1}{2} \frac{\partial^2 H}{\partial u \partial v} \bigg|_{(1,1)} \right]
\]  
(3.16)

\[
t_3 = \hat{C}_y \left[ 1 + \varepsilon_1 H_1 + \varepsilon_3 H_2 + \varepsilon_4^2 H_3 + \varepsilon_4 H_4 + \varepsilon_3 H_5 \right]
\]  
(3.17)
Here,
\[ H_1 = \left. \frac{\partial H}{\partial u} \right|_{(1,1)}, \quad H_2 = \left. \frac{\partial H}{\partial v} \right|_{(1,1)}, \quad H_3 = \left. \frac{\partial^2 H}{\partial u^2} \right|_{(1,1)}, \quad H_4 = \left. \frac{\partial^2 H}{\partial v^2} \right|_{(1,1)}, \quad H_5 = \left. \frac{\partial^2 H}{2 \partial u \partial v} \right|_{(1,1)} \]

Substituting the value of \( \hat{C}_y \) in the above expression (2.28), we get
\[
t_3 = C_y \left( 1 - \varepsilon_0 + \varepsilon_2^2 - \frac{\varepsilon_0 \varepsilon_2}{2} - \frac{\varepsilon_2^2}{8} \right) \left( 1 + \varepsilon_1 H_1 + \varepsilon_3 H_2 + \varepsilon_1^2 H_3 + \varepsilon_3^2 H_4 + \varepsilon_1 \varepsilon_3 H_5 \right)
\]

(3.18)

Mean square error of the estimator \( t_3 \) is given by
\[
MSE(t_3) = \mathbb{E} \left[ (t_3 - C_y)^2 \right] = C_y^2 \mathbb{E} \left[ -\varepsilon_0 + \varepsilon_1 H_1 + \frac{\varepsilon_2^2}{2} + \varepsilon_3 H_2 + O(\varepsilon) \right]^2
\]

(3.19)

Simplifying the expression (2.30), we get
\[
MSE(t_3) = C_y^2 \left[ 1 - \frac{f}{n} \left( C_y^2 + C_x^2 H_1^2 + 2H_1 \rho C_x, C_x + \frac{(\lambda_{40} - 1)}{4} + (\lambda_{04} - 1)H_2^2 + (\lambda_{22} - 1)H_2 + \frac{C_y \lambda_{30}}{2} - C_y \lambda_{12} H_2 + \frac{C_x \lambda_{21}}{2} H_1 + C_x \lambda_{03} H_1 H_2 \right) \right]
\]

(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

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

(3.21)

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

(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 \( t_3 \).
\[ MSE(t_3) = C_y^2 \left( 1 - \frac{f}{n} \right) \left[ C_y^2 + C_x^2 \hat{y}_{opt}^2 - 2H_{opt} \rho C_y C_x + \frac{\lambda_{40} - 1}{4} + (\lambda_{44} - 1)H_{2opt}^2 + (\lambda_{22} - 1)H_{2opt}^2 \right] + 2 \left[ -\frac{C_y \lambda_{30}}{2} - C_y \lambda_{12}H_{2opt} + \frac{C_x \lambda_{21}}{2} H_{1opt} + C_x \lambda_{03}H_{1opt}H_{2opt} \right] \]

(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:

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

(3.24)

where \( u = \frac{\bar{x}}{\bar{X}} \), \( v = \frac{S_x^2}{S_y^2} \) 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 \( (C_y, 1, 1) \);

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 \( (C_y, 1, 1) \) in a second order Taylor’s series, we have

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

\[ \frac{(\hat{C}_y - C_y)^2}{2} \frac{\partial^2 H^*}{\partial \hat{C}_y^2} \bigg|_{(C_y, 1, 1)} + (u - 1)^2 \frac{\partial^2 H^*}{\partial u^2} \bigg|_{(C_y, 1, 1)} + (v - 1)^2 \frac{\partial^2 H^*}{\partial v^2} \bigg|_{(C_y, 1, 1)} + \]
\[
(\hat{C}_y - C_y)(u-1)\left.\frac{\partial^2 H^*}{\partial \hat{C}_y \partial u}\right|_{(c_y,1,1)} + (u-1)(v-1)\left.\frac{\partial^2 H^*}{\partial u \partial v}\right|_{(c_y,1,1)} + (v-1)(\hat{C}_y - C_y)\left.\frac{\partial^2 H^*}{\partial \hat{C}_y \partial v}\right|_{(c_y,1,1)}
\]

\[
t_4 = C_y + (\hat{C}_y - C_y) + e_1 H^*_1 + e_3 H^*_2 + (\hat{C}_y - C_y)H^*_3 + e_1^2 H^*_4 + e_3^2 H^*_5 + (\hat{C}_y - C_y)e_1 H^*_6 +
\]

\[
e_1 e_3 H^*_7 + (\hat{C}_y - C_y)e_3 H^*_8
\]

(3.26)

Here,

\[
H^*(C_y,1,1) = C_y, \quad \left.\frac{\partial H^*}{\partial \hat{C}_y}\right|_{(c_y,1,1)} = 1, \quad \left.\frac{\partial H^*}{\partial u}\right|_{(c_y,1,1)} = H^*_1, \quad \left.\frac{\partial H^*}{\partial v}\right|_{(c_y,1,1)} = H^*_2
\]

\[
H^*_3 = \left.\frac{1}{2}\frac{\partial^2 H^*}{\partial \hat{C}_y^2}\right|_{(c_y,1,1)}
\]

\[
H^*_4 = \left.\frac{1}{2}\frac{\partial^2 H^*}{\partial u^2}\right|_{(c_y,1,1)}, \quad H^*_5 = \left.\frac{1}{2}\frac{\partial^2 H^*}{\partial v^2}\right|_{(c_y,1,1)}, \quad H^*_6 = \left.\frac{1}{2}\frac{\partial^2 H^*}{\partial \hat{C}_y \partial u}\right|_{(c_y,1,1)}
\]

\[
H^*_7 = \left.\frac{1}{2}\frac{\partial^2 H^*}{\partial \hat{C}_y \partial v}\right|_{(c_y,1,1)}, \quad H^*_8 = \left.\frac{1}{2}\frac{\partial^2 H^*}{\partial v \partial \hat{C}_y}\right|_{(c_y,1,1)}
\]

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

\[
t_4 = C_y \left(1-e_0 + e_0^2 + e_1^2 + e_3 + \frac{e_1}{2} - \frac{e_0 e_2}{2} - \frac{e_1^2}{8}\right) + e_1 H^*_1 + e_3 H^*_2 + \left[C_y \left(1-e_0 + e_0^2 + e_1 + e_3 - \frac{e_0 e_2}{2} - \frac{e_1^2}{8}\right) - C_y\right] H^*_3
\]

\[
e_1^2 H^*_4 + e_3^2 H^*_5 + \left[C_y \left(1-e_0 + e_0^2 + \frac{e_2}{2} - \frac{e_0 e_2}{2} - \frac{e_3^2}{8}\right) - C_y\right] e_1 H^*_6 + e_1 e_3 H^*_7 + \left[C_y \left(1-e_0 + e_0^2 + \frac{e_2}{2} - \frac{e_0 e_2}{2} - \frac{e_3^2}{8}\right) - C_y\right] e_3 H^*_8
\]

(3.27)

\[
MSE(t_4) = E[t_4 - C_y]^2 = E \left[C_y \left(- e_0 + \frac{e_2}{2}\right) + e_1 H^*_1 + e_3 H^*_2 + O(\varepsilon)\right]^2
\]

(3.28)
After simplifying the expression (2.39), we get:

\[
MSE(t_4) = \left(1 - \frac{f}{n}\right) \left[ C_y^2 \left( \frac{C_y}{4} - C_y A_{30} \right) + C_x^2 H_1^{*2} + \left( \lambda_{32} - 1 \right) H_2^{*2} + 2C_y \left( \frac{C_x A_{21}}{2} - \rho C_y C_x \right) H_1^{*} + 2C_x \lambda_{03} H_1^{*} H_2^{*} \right] \]

(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:

\[
H_{1opt} = \frac{C_y}{2C_x} \left[ \frac{\lambda_{04} - 1}{\lambda_{04} - \lambda_{02}^2 - 1} \left( \lambda_{04} - 1 \right) \right]
\]

\[
H_{2opt} = \frac{C_y}{2} \left[ \frac{\lambda_{04} - 1}{\lambda_{04} - \lambda_{02}^2 - 1} \left( \lambda_{04} - 1 \right) \right]
\]

(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 \( t_4 \):

\[
MSE(t_4) = \left(1 - \frac{f}{n}\right) \left[ C_y^2 \left( \frac{C_y}{4} - C_y A_{30} \right) + C_x^2 H_{1opt}^{*2} + \left( \lambda_{32} - 1 \right) H_{2opt}^{*2} + 2C_y \left( \frac{C_x A_{21}}{2} - \rho C_y C_x \right) H_{1opt}^{*} + 2C_x \lambda_{03} H_{1opt}^{*} H_{2opt}^{*} \right] \]

(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,
- \( N=34 \), \( n=15 \),
\(\bar{X} = 208.88, \quad \bar{Y} = 199.44,\)

\[C_X = 0.72, \quad C_Y = 0.75, \quad \rho_{xy} = 0.98,\]

\(\lambda_{21} = 1.0045, \quad \lambda_{12} = 0.9406, \quad \lambda_{40} = 3.6161, \quad \lambda_{04} = 2.8266, \quad \lambda_{30} = 1.1128,\)

\(\lambda_{03} = 0.9206, \quad \lambda_{22} = 3.0133\)

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

X: Number of fish caught in year 1993,

Y: Number of fish caught in year 1995,

\(N=69, \quad n=40,\)

\(\bar{X} = 4591.07, \quad \bar{Y} = 4514.89,\)

\[C_X = 1.38, \quad C_Y = 1.35,\]

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

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

\[PRE(t, t_0) = \frac{Var(t_0)}{MSE(t)} \times 100\]

where \(t = C_d, C^*_d, t_r_1, t_r_2, t_1, t_2, t_3, t_4.\)

**Table 1.** MSE and PRE of the estimators

<table>
<thead>
<tr>
<th>ESTIMATOR</th>
<th>POPULATION-1</th>
<th>POPULATION-2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PRE</td>
</tr>
<tr>
<td>(t_0)</td>
<td>0.008016</td>
<td>100.00</td>
</tr>
<tr>
<td>(C_d)</td>
<td>0.00123</td>
<td>651.7051</td>
</tr>
<tr>
<td>(C^*_d)</td>
<td>0.00122</td>
<td>654.4814</td>
</tr>
<tr>
<td>(t_r_1)</td>
<td>0.006868</td>
<td>116.54</td>
</tr>
<tr>
<td>(t_r_2)</td>
<td>0.006963</td>
<td>114.95</td>
</tr>
</tbody>
</table>
Table 1. MSE and PRE of the estimators (cont.)

<table>
<thead>
<tr>
<th>ESTIMATOR</th>
<th>POPULATION-1</th>
<th></th>
<th>POPULATION-2</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MSE</td>
<td>PRE</td>
<td>MSE</td>
<td>PRE</td>
</tr>
<tr>
<td>( t_1 )</td>
<td>0.00123</td>
<td>651.7356</td>
<td>0.0299</td>
<td>127.4598</td>
</tr>
<tr>
<td>( t_2 )</td>
<td>0.001038</td>
<td>771.9898</td>
<td>0.0283</td>
<td>134.6127</td>
</tr>
<tr>
<td>( t_3 )</td>
<td>0.001203</td>
<td>666.4304</td>
<td>0.0297</td>
<td>128.345</td>
</tr>
<tr>
<td>( t_4 )</td>
<td>0.001203</td>
<td>666.4304</td>
<td>0.0297</td>
<td>128.345</td>
</tr>
</tbody>
</table>

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 \( 0 < \rho = 0.75, 0.85, 0.95 \leq 1 \).

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 \((y_i, x_i) (i = 1, 2, \ldots, n)\) 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:

$$MSE(i) = \frac{1}{1500} \sum_{j=1}^{1500} mse_j(i)$$

where $i = t_o$, $Cd$, $Cd^*$, $t_1$, $t_2$, $t_{r1}$ and $t_{r2}$.

**Table 2.** Table showing MSE and PRE of the existing and proposed estimators for different values of $\rho$ and n

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>n</th>
<th>Estimator</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50</td>
<td>$t_o$</td>
<td>0.006053626</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Cd$</td>
<td>0.004924006</td>
<td>122.9410</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Cd^*$</td>
<td>0.004617663</td>
<td>131.0970</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{r1}$</td>
<td>0.005080027</td>
<td>119.1651</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>$t_o$</td>
<td>0.003450581</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Cd$</td>
<td>0.002835622</td>
<td>121.6869</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Cd^*$</td>
<td>0.002671694</td>
<td>129.1533</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{r1}$</td>
<td>0.002688403</td>
<td>117.5860</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{r2}$</td>
<td>0.002650186</td>
<td>109.5933</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_1$</td>
<td>0.002934516</td>
<td>128.3506</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_2$</td>
<td>0.003148534</td>
<td>130.2015</td>
</tr>
<tr>
<td></td>
<td>70</td>
<td>$t_o$</td>
<td>0.002412659</td>
<td>100.0000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Cd$</td>
<td>0.001990824</td>
<td>121.1889</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$Cd^*$</td>
<td>0.001879170</td>
<td>128.3896</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{r1}$</td>
<td>0.002062564</td>
<td>116.9737</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_{r2}$</td>
<td>0.002200267</td>
<td>109.6530</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_1$</td>
<td>0.001887289</td>
<td>127.8373</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$t_2$</td>
<td>0.001868644</td>
<td>129.1128</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( t_o )</td>
<td>( Cd )</td>
<td>( Cd^* )</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>0.006358341</td>
<td>0.004912595</td>
<td>0.003809327</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.0000</td>
<td>129.4294</td>
<td>166.9151</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.004876890</td>
<td>0.005133219</td>
<td>0.003831045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>130.3769</td>
<td>123.8665</td>
<td>165.9688</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.003831045</td>
<td>0.003739557</td>
<td>0.003831045</td>
</tr>
<tr>
<td></td>
<td></td>
<td>165.9688</td>
<td>170.0293</td>
<td>165.9688</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>0.003621428</td>
<td>0.002828737</td>
<td>0.002203557</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.0000</td>
<td>128.0228</td>
<td>164.3447</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.002825058</td>
<td>0.002910006</td>
<td>0.002210627</td>
</tr>
<tr>
<td></td>
<td></td>
<td>128.1895</td>
<td>124.4744</td>
<td>163.8190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.002180249</td>
<td>0.002180249</td>
<td>0.002180249</td>
</tr>
<tr>
<td></td>
<td></td>
<td>166.1016</td>
<td>166.1016</td>
<td>166.1016</td>
</tr>
<tr>
<td>70</td>
<td></td>
<td>0.002527309</td>
<td>0.001982556</td>
<td>0.001547597</td>
</tr>
<tr>
<td></td>
<td></td>
<td>100.0000</td>
<td>127.4561</td>
<td>163.3054</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001984483</td>
<td>0.002027634</td>
<td>0.001551035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>127.3535</td>
<td>124.6433</td>
<td>162.9434</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.001536162</td>
<td>0.001536162</td>
<td>0.001536162</td>
</tr>
<tr>
<td></td>
<td></td>
<td>164.5210</td>
<td>164.5210</td>
<td>164.5210</td>
</tr>
<tr>
<td></td>
<td>$t_o$</td>
<td>$C_d$</td>
<td>$C_d^*$</td>
<td>$t_{r1}$</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
<td>--------</td>
<td>---------</td>
<td>----------</td>
</tr>
<tr>
<td>30</td>
<td>0.008647395</td>
<td>0.005851489</td>
<td>0.002426053</td>
<td>0.005461214</td>
</tr>
<tr>
<td>50</td>
<td>0.004896276</td>
<td>0.003355658</td>
<td>0.001397620</td>
<td>0.003172583</td>
</tr>
<tr>
<td>70</td>
<td>0.0034018746</td>
<td>0.0023435450</td>
<td>0.009782472</td>
<td>0.0022234723</td>
</tr>
</tbody>
</table>

From the table, we can observe that for a particular value of $\rho$, 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 $t_o$, $C_d$, $C_d^*$, $t_{r1}$, $t_{r2}$ and the proposed estimator $t_1$ while the estimator $t_1$ turns out to be more efficient than the existing estimators $t_o$, $C_d$, $t_{r1}$, $t_{r2}$ and nearly as efficient as the estimator $C_d^*$. 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 \((n_1, n_2, n_3) = (30, 50, 70)\)

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 $\rho = 0.95$.

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 \( t_0, Cd, C_d^*, t_{r1}, t_{r2} \) and the proposed estimator \( t_1 \). As regards \( t_1 \), it performs better than the estimators \( t_0, Cd, t_{r1}, t_{r2} \) 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**


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


