

# Almost unbiased estimators for population coefficient of variation using auxiliary information

---

## ABSTRACT

The objective of the paper is to propose an almost unbiased ratio estimator for the finite coefficient of variation(CV). In this paper, we have proposed an exponential ratio type and log ratio type estimators for estimating population coefficient of variation. Two real data sets and one simulation study is carried out in support of the theoretical results. Mean squared error and Percent relative efficiency criteria is used to assess the performance of the estimators. It has been shown that the proposed class of estimators are almost unbiased up to the first order of approximation. Also proposed estimators are better in efficiency to other estimators consider in this study.

*Keywords:* Auxiliary information; bias; mean squared error; coefficient of variation; log type estimator.

## 1. INTRODUCTION

Research of Cochran(1977)[1] is generally associated with the idea of incorporating auxiliary information to improve estimator's efficiency. Using auxiliary information, we can improve the accuracy or efficacy of estimators by incorporating more data with the sampled data. Researchers might be enabled to reduce variability in samples and provide more precise estimates of population parameters by utilizing auxiliary information. The basic principles for this methodology were established by Cochran(1977)[1] and it is now commonly employed in many different kinds of domains, such as survey sampling, econometrics, and statistics.

Ratio and product estimators are widely used in survey sampling and other fields where auxiliary information is available and can be utilized to improve the accuracy and efficiency of estimators. When there is a positive correlation between an auxiliary variable and the study variable of interest, Cochran(1940)[2] established the concept of ratio estimators as an approach to utilize auxiliary information. Ratio estimators calculate ratios among the study variable's sample means or totals and the auxiliary variables, taking into consideration any known population variables. Ratio estimators enable us to analyze an estimated value with other known information to estimate a value more accurately. To get more accurate estimates, they consider the correlation between various variables. For example, we may accurately estimate the total income of a neighborhood if we know the average income of that neighborhood and the population of the entire city. This can be done by using a ratio estimator. By utilizing more information, this approach increases the precision of our estimations.

On the other hand, Robson(1957)[3] and Murthy(1964)[4] proposed "the product estimator, which is another method for incorporating auxiliary information into estimation. The product estimator involves forming the product of the study variable and the auxiliary variable and then using this product as the basis for estimation". Similar to ratio estimators, the product estimator

seeks to capitalize on the association between the auxiliary variable and the study variable to enhance the precision of the estimates. Number of authors, including Solanki et al.(2012)[5], Ray and Sahai(1980)[6], and Srivastava and Jhaji(1981)[7], have made significant contributions to “the utilization of auxiliary information for estimating population parameters such as the population mean, variance, standard deviation, and other related statistics”. Some important works illustrating use of auxiliary information at estimation stage are Singh et al.(2018) [8], Singh and Kumar(2011) [9], Malik and Singh [10] etc.

Very less work has been done for estimating population coefficient of variation. Das and Tripathi(1992)[11] were first to suggest “an estimator for the coefficient of variation when samples were chosen using simple random sampling without replacement (SRSWOR)”. Other researchers, such as Patel and Rina(2009)[12], have also explored into this area. Breunig(2001)[13] suggested “an almost unbiased estimator of the coefficient of variation”. Additionally, Rajyaguru and Gupta(2005)[14] explored “estimating the coefficient of variation under different sampling methods like simple random sampling and stratified random sampling”. Adejumobi and Yunusa(2022)[15] proposed “ratio estimators for finite population variance with the use of known parameters”. Yunusa et al.(2021) [16] proposed “logarithmic ratio type estimator for the estimating population coefficient of variation”. Audu et al.(2021)[17] proposed “three difference-cum-ratio estimators for estimating finite population coefficient of variation”.

In this paper, using exponential and log type estimators we have proposed an almost unbiased estimator for estimation of population coefficient of variation utilizing information on a single auxiliary variable in SRSWOR.

Let's consider a finite population  $P = (P_1, P_2, P_3, \dots, P_N)$  of size `N` and each unit are uniquely defined. Let Y and X defined as study and auxiliary variable and  $Y_i$  and  $X_i$  are the values corresponding their unit  $i$  ( $i = 1, 2, 3, \dots, N$ ).

Let us consider a SRS of size  $n$  drawn from the population of `N` units and corresponding unit  $y_i$  and  $x_i$ .

$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$  and  $\bar{X} = \frac{1}{N} \sum_{i=1}^N X_i$  are the population means of the study and auxiliary variables Y and X,

$S_y^2 = \frac{1}{(N-1)} \sum_{i=1}^N (Y_i - \bar{Y})^2$  is the population variance of the study variable Y,

$S_x^2 = \frac{1}{(N-1)} \sum_{i=1}^N (X_i - \bar{X})^2$  is the population variance of the auxiliary variable X,

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

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$  and  $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$  - are respectively the sample means of the study and auxiliary variables Y and X.

$s_y^2 = \frac{1}{(n-1)} \sum_{i=1}^n (y_i - \bar{y})^2$  - is the sample variance of the study variable y,

$s_x^2 = \frac{1}{(n-1)} \sum_{i=1}^n (x_i - \bar{x})^2$  - is the sample variance of the auxiliary variable x.

Let us define sampling errors for both mean and variance of study and auxiliary variables as-

$$e_0 = \frac{\bar{y} - \bar{Y}}{\bar{Y}}, e_1 = \frac{\bar{x} - \bar{X}}{\bar{X}}, e_2 = \frac{(s_y^2 - S_y^2)}{S_y^2}, e_3 = \frac{(s_x^2 - S_x^2)}{S_x^2} \text{ such that}$$

$$\bar{y} = \bar{Y}(1 + e_0), \bar{x} = \bar{X}(1 + e_1), s_y^2 = S_y^2(1 + e_2), s_x^2 = S_x^2(1 + e_3)$$

$$E(e_0) = E(e_1) = E(e_2) = E(e_3) = 0,$$

$$E(e_0^2) = \gamma C_y^2, E(e_1^2) = \gamma C_x^2, E(e_2^2) = \gamma(\lambda_{40} - 1), E(e_3^2) = \gamma(\lambda_{04} - 1),$$

$$E(e_0 e_1) = \gamma \rho C_y C_x, E(e_0 e_2) = \gamma C_y \lambda_{30}, E(e_0 e_3) = \gamma C_y \lambda_{12},$$

$$E(e_1 e_2) = \gamma C_x \lambda_{21}, E(e_1 e_3) = \gamma C_x \lambda_{03}, E(e_2 e_3) = \gamma(\lambda_{22} - 1).$$

Here,  $\gamma = \frac{1}{n}(1-f)$ ,  $f = \frac{n}{N}$ ,  $f$  is known as sampling fraction,  $C_y$  and  $C_x$  are the population coefficient of variations of study variable Y and auxiliary variable X, respectively,

defined as,  $C_y = \frac{S_y}{\bar{Y}}$  and  $C_x = \frac{S_x}{\bar{X}}$ .  $\rho$  is the correlation coefficient between X and Y.

In general form,

$$\mu_{rs} = \frac{\sum_{i=1}^N (y_i - \bar{y})^r (x_i - \bar{x})^s}{(N-1)} \text{ and } \lambda_{rs} = \frac{\mu_{rs}}{(\mu_{20}^{r/2} \mu_{02}^{s/2})}.$$

## 2. EXISTING ESTIMATORS

Usual estimator  $t_0$  for estimating  $C_y$  is given by

$$t_0 = \hat{C}_y = \frac{s_y}{y} \quad (1)$$

The bias of the estimator  $t_0$  is given by:

$$\text{Bias}(t_0) = C_y \gamma \left( C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2} C_y \lambda_{30} \right) \quad (2)$$

The Mean square error (MSE) expression of the estimator  $t_0$  is given by:

$$\text{MSE}(t_0) = C_y^2 \gamma (C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30}) \quad (3)$$

Archana and Rao (2014)[18] introduced estimators  $t_1$  and  $t_2$  for calculating the finite population coefficient of variation as follows:

$$t_1 = C_y \left( \frac{S_x^2}{s_x^2} \right) \quad (4)$$

$$t_2 = C_y \left( \frac{S_x^2}{S_x^2} \right) \quad (5)$$

The bias of the estimators  $t_1$  and  $t_2$  are, respectively, given as

$$Bias(t_1) = C_y \gamma \left( C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2}C_y \lambda_{30} + C_y \lambda_{12} + (\lambda_{04} - 1) - \frac{1}{2}(\lambda_{22} - 1) \right) \quad (6)$$

$$Bias(t_2) = C_y \gamma \left( C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2}C_y \lambda_{30} + \frac{1}{2}(\lambda_{22} - 1) - C_y \lambda_{12} \right) \quad (7)$$

MSE of the estimators  $t_1$  and  $t_2$  are, respectively, given as

$$MSE(t_1) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + 2C_y \lambda_{12} + (\lambda_{04} - 1) - (\lambda_{22} - 1) \right) \quad (8)$$

$$MSE(t_2) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} - 2C_y \lambda_{12} + (\lambda_{04} - 1) + (\lambda_{22} - 1) \right) \quad (9)$$

### 3. PROPOSED ALMOST UNBIASED ESTIMATOR

Let,

$$t_0 = C_y, t_1 = C_y \left( \frac{S_x^2}{S_x^2} \right), t_2 = C_y \left( \frac{S_x^2}{S_x^2} \right) \quad (10)$$

such that  $t_0, t_1, t_2 \in L$ , where  $L$  denotes the set of all possible estimators for estimating the population coefficient of variation  $C_y$ .

By definition, the set  $L$  is a linear variety if

$$t_g = \sum_{i=0}^2 g_i t_i \in L \quad (11)$$

$$t_g = g_0 C_y + g_1 C_y \left( \frac{S_x^2}{S_x^2} \right) + g_2 C_y \left( \frac{S_x^2}{S_x^2} \right)$$

$$\text{For } \sum_{i=0}^2 g_i = 1, g_i \in R \quad (12)$$

where  $g_i$  ( $i = 0, 1, 2$ ) denotes the statistical constants and  $R$  denotes the set of real numbers.

**Table 1. Members of the proposed family  $t_g$  of estimators**

$g_0$	$g_1$	$g_2$	Estimators
1	0	0	$\hat{C}_y$
0	1	0	$C_y \left( \frac{S_x^2}{S_x^2} \right)$

To obtain the bias and MSE of the estimator  $t_g$ , we write  $t_g$  in the form of error terms as

$$t_g = C_y (1+e_2)^{1/2} (1+e_0)^{-1} \left[ g_0 + g_1 \left( \frac{1}{(1+e_3)} \right) + g_2 (1+e_3) \right] \quad (13)$$

Expanding the right hand side of equation (13) and retaining terms up to second powers of e's, we have

$$t_g = C_y \left[ \begin{array}{l} 1 - e_0 + e_0^2 + \frac{1}{2}e_2 - \frac{1}{2}e_0e_2 - \frac{1}{8}e_2^2 - (g_1 - g_2)e_3 \\ + (g_1 - g_2)e_0e_3 - (g_1 - g_2)\frac{1}{2}e_2e_3 + g_1e_3^2 \end{array} \right] \quad (14)$$

Subtracting  $C_y$  and then taking expectation both sides, we get the bias of the estimator  $t_g$ , up to the first order of approximation as

$$Bias(t_g) = C_y \gamma \left( \begin{array}{l} C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2}C_y\lambda_{30} + (g_1 - g_2)C_y\lambda_{12} \\ + g_1(\lambda_{04} - 1) - (g_1 - g_2)\frac{1}{2}(\lambda_{22} - 1) \end{array} \right) \quad (15)$$

From equation (15),

We have

$$(t_g - C_y) \cong C_y \left[ \frac{1}{2}e_2 - e_0 - (g_1 - g_2)e_3 \right] \quad (16)$$

where,

$$(g_1 - g_2) = H. \quad (17)$$

Squaring both sides of equation (16) and then taking expectations, we get MSE of the estimator  $t_g$ , up to the first order of approximation, as

$$MSE(t_g) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y\lambda_{30} + 2HC_y\lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) \quad (18)$$

Which is minimum when

$$H = \frac{1(\lambda_{22} - 1)}{2(\lambda_{04} - 1)} - \frac{C_y\lambda_{12}}{(\lambda_{04} - 1)}. \quad (19)$$

Putting this value of  $H = \frac{1(\lambda_{22} - 1)}{2(\lambda_{04} - 1)} - \frac{C_y\lambda_{12}}{(\lambda_{04} - 1)}$  in equation (2.18) we get the Min. MSE of

the estimator  $t_g$  as

$$Min.MSE(t_g) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y\lambda_{30} + 2HC_y\lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) \quad (20)$$

From equation (17) and (19) we have,

$$(g_1 - g_2) = H = \frac{1(\lambda_{22} - 1)}{2(\lambda_{04} - 1)} - \frac{C_y \lambda_{12}}{(\lambda_{04} - 1)} \quad (21)$$

From equation (12) and (17), we have only two equations in three unknowns. It is not possible to find the unique values for  $g_i$ 's, ( $i = 0, 1, 2$ ). In order to get unique values of  $g_i$ 's, we shall impose the linear restriction.

$$\sum_{i=0}^2 g_i B(t_i) = 0 \quad (22)$$

Such that

$$g_0 B(t_0) + g_1 B(t_1) + g_2 B(t_2) = 0 \quad (23)$$

where  $B(t_i)$  denotes the bias in the  $i^{\text{th}}$  estimator.

Equations (2.12), (2.17) and (2.23) can be written in the matrix form as

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & -1 \\ B(t_0) & B(t_1) & B(t_2) \end{bmatrix} \begin{bmatrix} g_0 \\ g_1 \\ g_2 \end{bmatrix} = \begin{bmatrix} 1 \\ H \\ 0 \end{bmatrix} \quad (24)$$

From the system of equation (2.24), we get the unique values of  $g_i$ 's ( $i=0, 1, 2$ ) as

$$g_0 = \frac{B(t_2) + B(t_1) - HB(t_2) - HB(t_1)}{B(t_2) + B(t_1) - 2B(t_0)} \quad (25)$$

$$g_1 = \frac{HB(t_2) - B(t_0) - HB(t_0)}{B(t_2) + B(t_1) - 2B(t_0)} \quad (26)$$

$$g_2 = \frac{HB(t_0) - B(t_0) - HB(t_1)}{B(t_2) + B(t_1) - 2B(t_0)} \quad (27)$$

such that

$$g_0 + g_1 + g_2 = 1 \quad (28)$$

Use of these  $g_i$ 's ( $i=0, 1, 2$ ) remove the bias up to terms of order  $o(n^{-1})$

### 3.1 Another Almost Unbiased Estimator

In this section we propose another almost unbiased estimator  $t_{g_1}$  for coefficient of variation.

For this we have taken three estimators  $m_0$ ,  $m_1$  and  $m_2$  which are defined as

$$m_0 = C_y = (t_0) \quad (29)$$

The bias of the estimator  $m_0$  is given by:

$$\text{Bias}(m_0) = C_y \gamma \left( C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2} C_y \lambda_{30} \right) \quad (30)$$

The Mean square error (MSE) expression of the estimator  $m_0$  is given by:

$$\text{MSE}(m_0) = C_y^2 \gamma (C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30}) \quad (31)$$

The exponential and logarithmic estimator for estimating population coefficient of variation is given as follows-

$$m_1 = C_y \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right) \quad (32)$$

$$m_2 = C_y \left\{1 + \log\left(\frac{s_x^2}{S_x^2}\right)\right\} \quad (33)$$

The bias of the estimators  $m_1$  and  $m_2$  are respectively given as

$$Bias(m_1) = C_y \gamma \left( C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2}C_y \lambda_{30} + \frac{1}{2}C_y \lambda_{12} + \frac{3}{8}(\lambda_{04} - 1) - \frac{1}{4}(\lambda_{22} - 1) \right) \quad (34)$$

$$Bias(m_2) = C_y \gamma \left( C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2}C_y \lambda_{30} + \frac{1}{2}(\lambda_{22} - 1) - C_y \lambda_{12} - \frac{1}{2}(\lambda_{04} - 1) \right) \quad (35)$$

MSE of the estimators  $m_1$  and  $m_2$  are respectively given as

$$MSE(m_1) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + C_y \lambda_{12} + \frac{1}{4}(\lambda_{04} - 1) - \frac{1}{2}(\lambda_{22} - 1) \right) \quad (36)$$

$$MSE(m_2) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} - 2C_y \lambda_{12} + (\lambda_{04} - 1) + (\lambda_{22} - 1) \right) \quad (37)$$

$m_0, m_1$  and  $m_2 \in L$ , where  $L$  denotes the set of all possible estimators for estimating the population coefficient of variation  $C_y$ .

By definition, the set  $L$  is a linear variety if

$$t_{g1} = \sum_{i=0}^2 l_i m_i \in L \quad (38)$$

$$t_{g1} = l_0 m_0 + l_1 m_1 + l_2 m_2 \quad (39)$$

$$t_{g1} = l_0 C_y + l_1 C_y \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right) + l_2 C_y \left\{1 + \log\left(\frac{s_x^2}{S_x^2}\right)\right\}$$

$$\text{For } \sum_{i=0}^2 l_i = 1, l_i \in R \quad (40)$$

where  $l_i$  ( $i = 0, 1, 2$ ) denotes the statistical constants and  $R$  denotes the set of real numbers.

**Table 2. Members of the proposed family  $t_{g1}$  of estimators**

$l_0$	$l_1$	$l_2$	Estimators
1	0	0	$C_y$
0	1	0	$C_y \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right)$
0	0	1	$C_y \left\{1 + \log\left(\frac{s_x^2}{S_x^2}\right)\right\}$

To obtain the bias and MSE of the  $t_{g1}$ , we write  $t_{g1}$  in the form of error terms as

$$t_{g1} = C_y (1+e_2)^{1/2} (1+e_0)^{-1} \left[ l_0 + l_1 \exp\left(\frac{-e_3}{2+e_3}\right) + l_2 \{1 + \log(1+e_3)\} \right] \quad (41)$$

Expanding the right hand side of (41) and retaining terms up to second powers of e's we have

$$t_{g1} = C_y \left[ \begin{aligned} &1 - e_0 + e_0^2 + \frac{1}{2}e_2 - \frac{1}{2}e_0e_2 - \frac{1}{8}e_2^2 - \left(\frac{1}{2}l_1 - l_2\right)e_3 \\ &+ \left(\frac{3}{8}l_1 - \frac{1}{2}l_2\right)e_3^2 + \left(\frac{1}{2}l_1 - l_2\right)e_0e_3 - \left(\frac{1}{4}l_1 - \frac{1}{2}l_2\right)e_2e_3 \end{aligned} \right] \quad (42)$$

Subtracting  $C_y$  from both sides of equation (42) and then taking expectations, we get the bias of the estimator  $t_{g1}$ , up to the first order of approximation as

$$Bias(t_{g1}) = C_y \gamma \left( \begin{aligned} &C_y^2 - \frac{1}{8}(\lambda_{40} - 1) - \frac{1}{2}C_y\lambda_{30} + \left(\frac{1}{2}l_1 - l_2\right)C_y\lambda_{12} \\ &+ \left(\frac{3}{8}l_1 - \frac{1}{2}l_2\right)(\lambda_{04} - 1) - \left(\frac{1}{4}l_1 - \frac{1}{2}l_2\right)\frac{1}{2}(\lambda_{22} - 1) \end{aligned} \right) \quad (43)$$

From (42), we have

$$(t_{g1} - C_y) \cong C_y \left[ \frac{1}{2}e_2 - e_0 - \left(\frac{1}{2}l_1 - l_2\right)e_3 \right] \quad (44)$$

where,

$$\frac{1}{2}l_1 - l_2 = H_1 \quad (45)$$

Squaring both sides of equation (43) and then taking expectations, we get MSE of the estimator  $t_{g1}$ , up to the first order of approximation, as

$$MSE(t_{g1}) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y\lambda_{30} + 2H_1C_y\lambda_{12} + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \right) \quad (46)$$

which is minimum when

$$H_1 = \frac{1}{2} \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)} - \frac{C_y\lambda_{12}}{(\lambda_{04} - 1)}. \quad (47)$$

Putting the value of  $H_1 = \frac{1}{2} \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)} - \frac{C_y\lambda_{12}}{(\lambda_{04} - 1)}$  in equation (46) the minimum MSE value of

the estimator  $t_{g1}$  is given by

$$Min.MSE(t_{g1}) = C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y\lambda_{30} + 2H_1C_y\lambda_{12} + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \right) \quad (48)$$

From equations (45) and (47), we have

$$\left(\frac{1}{2}l_1 - l_2\right) = H_1 = \frac{1}{2} \frac{(\lambda_{22} - 1)}{(\lambda_{04} - 1)} - \frac{C_y\lambda_{12}}{(\lambda_{04} - 1)} \quad (49)$$

From equation (38) and (45), we have only two equations in three unknowns. It is not possible to find the unique values for  $l_i$ 's,  $i = 0, 1, 2$ . In order to get unique values of  $l_i$ 's, we shall impose the linear restriction.

$$\sum_{i=0}^2 l_i B(m_i) = 0 \quad (50)$$

such that

$$l_0 B(m_0) + l_1 B(m_1) + l_2 B(m_2) = 0 \quad (51)$$

here  $B(m_i)$  denotes the bias in the  $i^{th}$  ( $i=0,1,2$ ) estimator.

Equations (40), (45) and (51) can be written in the matrix form

$$\begin{bmatrix} 1 & 1 & 1 \\ 0 & \frac{1}{2} & -1 \\ B(m_0) & B(m_1) & B(m_2) \end{bmatrix} \begin{bmatrix} l_0 \\ l_1 \\ l_2 \end{bmatrix} = \begin{bmatrix} 1 \\ H_1 \\ 0 \end{bmatrix} \quad (52)$$

Where,

$B(m_0)$ ,  $B(m_1)$  and  $B(m_2)$  are defined in equation (31), (34) and (35).

From the system of equation (52), we get the unique values of  $l_i$ 's ( $i=0, 1, 2$ ) respectively as

$$l_0 = \frac{\frac{1}{2} B(m_2) + B(m_1) - H_1 B(m_2) + H_1 B(m_1)}{\frac{1}{2} B(m_2) + B(m_1) - \frac{3}{2} B(m_0)} \quad (53)$$

$$l_1 = \frac{H_1 B(m_2) - B(m_0) - H_1 B(m_0)}{\frac{1}{2} B(m_2) + B(m_1) - \frac{3}{2} B(m_0)} \quad (54)$$

$$l_2 = \frac{H_1 B(m_0) - H_1 B(m_1) - \frac{1}{2} B(m_0)}{\frac{1}{2} B(m_2) + B(m_1) - \frac{3}{2} B(m_0)} \quad (55)$$

Where,

$$l_0 + l_1 + l_2 = 1 \quad (56)$$

Use of these  $l_i$ 's ( $i=0, 1, 2$ ) remove the bias up to terms of order  $o(n^{-1})$ .

#### 4. THEORETICAL EFFICIENCY COMPARISON

In this section, efficiency conditions of  $t_g$  and  $t_{g_1}$  over sample coefficient of variation  $t_0$ ,  $t_1$ ,  $t_2$ ,  $m_1$  and  $m_2$  are established.

##### 4.1 Efficiency comparison for the estimators $t_g$

(i).  $t_g$  is more efficient than  $t_0$

$$\text{Min.MSE}(t_g) < \text{MSE}(t_0)$$

$$C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \right. \\ \left. + 2HC_y \lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) < C_y^2 \gamma (C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30}) \quad (57)$$

(ii).  $t_g$  is more efficient than  $t_1$

$$\text{Min.MSE}(t_g) < \text{MSE}(t_1)$$

$$C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \right. \\ \left. + 2HC_y \lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) < C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + \right. \\ \left. 2C_y \lambda_{12} + (\lambda_{04} - 1) - (\lambda_{22} - 1) \right) \quad (58)$$

(iii).  $t_g$  is more efficient than  $t_2$

$$\text{Min.MSE}(t_g) < \text{MSE}(t_2)$$

$$C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \right. \\ \left. + 2HC_y \lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) < C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} - 2C_y \lambda_{12} \right. \\ \left. + (\lambda_{04} - 1) + (\lambda_{22} - 1) \right) \quad (59)$$

(iv).  $t_g$  is more efficient than  $m_1$

$$\text{Min.MSE}(t_g) < \text{MSE}(m_1)$$

$$C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \right. \\ \left. + 2HC_y \lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) < C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \right. \\ \left. + C_y \lambda_{12} + \frac{1}{4}(\lambda_{04} - 1) - \frac{1}{2}(\lambda_{22} - 1) \right) \quad (60)$$

(v).  $t_g$  is more efficient than  $m_2$

$$\text{Min.MSE}(t_g) < \text{MSE}(m_2)$$

$$C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \right. \\ \left. + 2HC_y \lambda_{12} + H^2(\lambda_{04} - 1) - H(\lambda_{22} - 1) \right) < C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} - 2C_y \lambda_{12} \right. \\ \left. + (\lambda_{04} - 1) + (\lambda_{22} - 1) \right) \quad (61)$$

## 4.2 Efficiency comparison for the estimators $t_{g_1}$

(i).  $t_{g_1}$  is more efficient than  $t_0$

$$\text{Min.MSE}(t_{g_1}) < \text{MSE}(t_0)$$

$$C_y^2 \gamma \left( C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + 2H_1 C_y \lambda_{12} \right. \\ \left. + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \right) < C_y^2 \gamma (C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30}) \quad (62)$$

(ii).  $t_{g_1}$  is more efficient than  $t_1$

$$\text{Min.MSE}(t_{g_1}) < \text{MSE}(t_1)$$

$$C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + 2H_1 C_y \lambda_{12} \\ + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \end{array} \right) < C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + \\ 2C_y \lambda_{12} + (\lambda_{04} - 1) - (\lambda_{22} - 1) \end{array} \right) \quad (63)$$

(iii).  $t_{g_1}$  is more efficient than  $t_2$

$$\text{Min.MSE}(t_{g_1}) < \text{MSE}(t_2)$$

$$C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + 2H_1 C_y \lambda_{12} \\ + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \end{array} \right) < C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} - 2C_y \lambda_{12} \\ + (\lambda_{04} - 1) + (\lambda_{22} - 1) \end{array} \right) \quad (64)$$

(iv).  $t_{g_1}$  is more efficient than  $m_1$

$$\text{Min.MSE}(t_{g_1}) < \text{MSE}(m_1)$$

$$C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + 2H_1 C_y \lambda_{12} \\ + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \end{array} \right) < C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} \\ + C_y \lambda_{12} + \frac{1}{4}(\lambda_{04} - 1) - \frac{1}{2}(\lambda_{22} - 1) \end{array} \right) \quad (65)$$

(v).  $t_{g_1}$  is more efficient than  $m_2$

$$\text{Min.MSE}(t_{g_1}) < \text{MSE}(m_2)$$

$$C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} + 2H_1 C_y \lambda_{12} \\ + H_1^2(\lambda_{04} - 1) - H_1(\lambda_{22} - 1) \end{array} \right) < C_y^2 \gamma \left( \begin{array}{l} C_y^2 + \frac{1}{4}(\lambda_{40} - 1) - C_y \lambda_{30} - 2C_y \lambda_{12} \\ + (\lambda_{04} - 1) + (\lambda_{22} - 1) \end{array} \right) \quad (66)$$

## 5. EMPIRICAL STUDY

In this section, we will carry out empirical study to demonstrate the performance of the proposed estimator over existing ones using two real data sets.

**Population 1:** [Source: Murthy (1967)[19], p.399]

X: Area under wheat in 1963, Y: Area under wheat in 1964 N=34, n=15,  $C_x = 0.72$ ,

$C_y = 0.75$ ,  $\rho = 0.98$ ,  $\lambda_{21} = 1.0045$ ,  $\lambda_{12} = 0.9406$ ,  $\lambda_{40} = 3.6161$ ,  $\lambda_{04} = 2.8266$ ,  $\lambda_{30} = 1.1128$ ,

$\lambda_{03} = 0.9206$ ,  $\lambda_{22} = 3.01133$ ,  $\bar{Y} = 199.44$ ,  $\bar{X} = 208.88$

**Population 2:** [Source: S. Singh(2003) [20], p.1116]

X: Number of fish caught in year 1993, Y: Number of fish caught in year 1995

N = 69, n=40,  $C_x = 1.38$ ,  $C_y = 1.35$ ,  $\rho = 0.96$ ,  $\lambda_{21} = 2.19$ ,  $\lambda_{12} = 2.3$ ,  $\lambda_{40} = 7.66$ ,  $\lambda_{04} = 9.84$ ,

$\lambda_{30} = 1.11$ ,  $\lambda_{03} = 2.52$ ,  $\lambda_{22} = 8.19$ ,  $\bar{Y} = 4514.89$ ,  $\bar{X} = 4591.07$ .

**Table 3 Values of  $g_i$ 's (i=0,1,2)**

S.N.	Scalars	Population I	Population II
1	$g_0$	<b>3.23672</b>	-0.01106
2	$g_1$	<b>-1.03618</b>	0.533243

3

 $g_2$ **-1.20054**

0.477814

Using these values of  $g_i$ 's ( $i=0,1,2$ ) given in the Table, one can reduce the bias to the order  $o(n^{-1})$  in the estimator  $t_g$ .

**Table 4 Values of  $l_i$ 's ( $i=0,1,2$ )**

S.N.	Scalars	Population I	Population II
1.	$l_0$	<b>0.631909</b>	1.677949
2.	$l_1$	<b>0.354966</b>	-0.41501
3.	$l_2$	<b>0.013126</b>	-0.26294

Using these values of  $l_i$ 's ( $i=0,1,2$ ) given in the Table, one can reduce the bias to the order  $o(n^{-1})$  in the estimator  $t_{g_1}$ .

**Table 5: Biases of the existing estimators and proposed estimators**

Estimators	Population 1 Bias	Population 2 Bias
$t_0 (= m_0)$	-0.00508	0.003414986
$t_1$	0.037569	0.121857921
$t_2$	0.003308	0.01036553
$m_1$	0.009865	0.046962269
$m_2$	-0.02221	-0.05233121
$t_g$	0	0
$t_{g_1}$	0	0

**Table 6 The MSE and PRE of the existing and the proposed estimators**

Estimators	MSE	PRE
$t_0 (= m_0)$	0.03808827	100.00
$t_1$	0.188603	20.1948
$t_2$	0.2261359	16.84297
$m_1$	0.071025	53.62631

$m_2$	0.226136	16.84309
$t_g$	0.037568	101.3844
$t_{g1}$	0.037568	101.3844

**Table 7 The MSE and PRE of the existing and the proposed estimators**

Estimators	MSE	PRE
$t_0 (= m_0)$	0.008003	100.00
$t_1$	0.03365	23.78054
$t_2$	0.05890	13.58789
$m_1$	0.01128	70.94231
$m_2$	0.05886	13.59669
$t_g$	0.00697	114.8289
$t_{g1}$	0.00697	114.8289

## 6. SIMULATION STUDY

In this section we have done simulation analysis.

The following steps have been used for the simulation:

1. We have generated bivariate random observations of size  $N=1000$  units from a bivariate normal distribution with parameters  $\mu_x = 3$ ,  $\sigma_x = 4$ , and  $\mu_y = 5$ ,  $\sigma_y = 9$  and  $\rho = 0.95$ .
2. Sample of sizes  $n = 150, 200$  and  $250$  have been selected from this simulated population.
3. Sample statistics that is the sample mean, sample variance, and the values of the suggested and existing estimators of population CV are calculated for these samples.
4. Steps (3) and (4) are repeated  $m=10,000$  times.

$$PRE(estimator) = \frac{MSE(t_0)}{MSE(estimator)} * 100$$

**Table 8: Values of Scalars by which one can reduce the bias of the estimator:**

Using the values of  $g_i$ 's ( $i=0,1,2$ ) given in the table, one can reduce the bias to the order

Scalars	N=1000, n=150	N=1000, n=200	N=1000, n=250
$g_0$	5.882077	6.881528	6.355966
$g_1$	-2.23049	-2.69344	-2.45441
$g_2$	-2.65159	-3.18809	-2.90156
$(g_1 - g_2)$	0.421105	0.494646	0.447154
H = H1	0.4211047758	0.4946458684	0.4471535012

$o(n^{-1})$  in the estimator  $t_g$ .

**Table 9: Values of Scalars by which one can reduce the bias of the estimator:**

Scalars	N=1000, n=150	N=1000, n=200	N=1000, n=250
$l_0$	0.622759	0.450548	0.534127
$l_1$	0.53223	0.696065	0.608684
$l_2$	-0.15499	-0.14661	-0.14281
$\left(\frac{1}{2}l_1 - l_2\right)$	0.421105	0.494646	0.447154
<b>H = H1</b>	0.4211047758	0.4946458684	0.4471535012

Similarly using the values of  $l_i$ 's ( $i=0,1,2$ ) given in the table, one can reduce the bias to the order  $o(n^{-1})$  in the estimator  $t_{g_1}$ .

**Table 10: Biases of the existing estimators and proposed estimators**

Estimators	N=1000, n=150	N=1000, n=200	N=1000, n=250
	Bias	Bias	Bias
$t_0 (= m_0)$	0.000372935	0.000286	0.000142
$t_1$	0.00428133	0.002819	0.001959
$t_2$	-0.00277411	-0.00177	-0.00135
$m_1$	0.001483199	0.000926	0.000639
$m_2$	0.006591748	0.005272	0.003254
$t_g$	0	0	0
$t_{g1}$	0	0	0

**Table 11 MSE and PRE values of existing and proposed estimators.**

Estimators	N=1000, n=150		N=1000, n=200		N=1000, n=250	
	MSE	PRE	MSE	PRE	MSE	PRE
$t_0 (= m_0)$	0.001688	100.00	0.001355	100.00	0.000793	100.00
$t_1$	0.002342	72.08397	0.001386	97.73829	0.000992	79.97066
$t_2$	0.00932	18.11233	0.007178	18.87754	0.004357	18.20195
$m_1$	0.000979	172.381	0.000639	212.0795	0.000422	187.9839
$m_2$	0.00932	18.11233	0.007178	18.87754	0.004357	18.20195
$t_g$	0.000953	177.043	0.000639	212.1073	0.000417	190.3548
$t_{g1}$	0.000953	177.043	0.000639	212.1073	0.000417	190.3548

## 7. DISCUSSION

We have proposed two almost unbiased estimators  $t_g$  and  $t_{g_1}$  for the estimation of population coefficient of variation utilizing information on a single auxiliary variable in srswor and compared them with some existing estimators. From Table 5 we observe that in the class of estimators  $t_g$ , the estimator  $t_2$  is least biased followed by the estimator  $t_0$  and the estimator  $t_1$  has the highest bias. Our proposed class of estimators  $t_g$  is almost unbiased for the proper choices of the constants  $g_i (i = 0, 1, 2)$ . Similarly, from the Table 5 we observe that for the class of the estimators  $t_{g_1}$ , the estimator  $m_0$  is least biased, followed by the estimator  $m_1$  and  $m_2$  has the highest bias. Our second proposed class of estimators  $t_{g_1}$  is almost unbiased for the proper choice of constants  $l_i (i = 0, 1, 2)$ . From Table 6 and Table 7, we observe that for population.1 and population.2, the proposed class of estimators  $t_g$  and  $t_{g_1}$  are having highest PRE. In the simulation study from Table 11 we observe that the proposed class of estimators  $t_g$  and  $t_{g_1}$  are having highest PRE.

## 8. CONCLUSION

In this paper we have proposed almost unbiased estimators for population coefficient of variation  $C_y$ . We have derived the bias and MSE expressions up to the first order of approximations for the estimators considered in this paper. In efficiency comparison section, we have derived the conditions under which our proposed class of estimators  $t_g$  and  $t_{g_1}$  will be better than the estimators considered in this paper.

## REFERENCES

- [1] W. G. Cochran, *Sampling techniques*. John Wiley & Sons, 1977. Accessed: Apr. 23, 2024. [Online]. Available: <https://books.google.com/books?hl=en&lr=&id=xbNn41DUrNwC&oi=fnd&pg=PA1&dq=Sampling+Techniques+Book+by+William+Cochran&ots=TYrUiAwmkZ&sig=V1VYvy7LsjGVVBuUtFfbRsczbOo>
- [2] W. G. Cochran, "The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce," *J. Agric. Sci.*, vol. 30, no. 2, pp. 262–275, 1940.
- [3] D. S. Robson, "Applications of Multivariate Polykeys to the Theory of Unbiased Ratio-Type Estimation," *J. Am. Stat. Assoc.*, vol. 52, no. 280, pp. 511–522, Dec. 1957, doi: 10.1080/01621459.1957.10501407.
- [4] M. N. Murthy, "Product method of estimation," *Sankhyā Indian J. Stat. Ser. A*, pp. 69–74, 1964.
- [5] R. S. Solanki, H. P. Singh, and A. Rathour, "An Alternative Estimator for Estimating the Finite Population Mean Using Auxiliary Information in Sample Surveys," *ISRN Probab. Stat.*, vol. 2012, pp. 1–14, May 2012, doi: 10.5402/2012/657682.
- [6] S. K. Ray and A. Sahai, "Efficient Families of Ratio and Product-Type Estimators".
- [7] S. K. Srivastava and H. S. Jhaji, "A class of estimators of the population mean in survey sampling using auxiliary information".
- [8] R. Singh, M. Mishra, B. P. Singh, P. Singh, and N. K. Adichwal, "Improved estimators for population coefficient of variation using auxiliary variable," *J. Stat. Manag. Syst.*, vol. 21, no. 7, pp. 1335–1355, Oct. 2018, doi: 10.1080/09720510.2018.1503405.

- [9] R. Sing and M. Kumar, "A note on transformations on auxiliary variable in survey sampling," *Model Assist. Stat. Appl.*, vol. 6, no. 1, pp. 17–19, Mar. 2011, doi: 10.3233/MAS-2011-0154.
- [10] S. Malik and R. Singh, "An improved estimator using two auxiliary attributes," *Appl. Math. Comput.*, vol. 219, no. 23, pp. 10983–10986, Aug. 2013, doi: 10.1016/j.amc.2013.05.014.
- [11] A. K. Das and T. P. Tripathi, "Use of auxiliary information in estimating the coefficient of variation," *Alig J Stat.*, vol. 12, pp. 51–58, 1992.
- [12] P. A. Patel and S. Rina, "A Monte Carlo comparison of some suggested estimators of Co-efficient of variation in finite population," *J. Stat. Sci.*, vol. 1, no. 2, pp. 137–147, 2009.
- [13] R. Breunig, "An almost unbiased estimator of the coefficient of variation," *Econ. Lett.*, vol. 70, no. 1, pp. 15–19, Jan. 2001, doi: 10.1016/S0165-1765(00)00351-7.
- [14] A. Rajyaguru and P. C. Gupta, "On the estimation of the coefficient of variation from finite population-II," *Model Assist. Stat. Appl.*, vol. 1, no. 1, pp. 57–66, Dec. 2005, doi: 10.3233/MAS-2006-1110.
- [15] A. Adejumobi and M. A. Yunusa, "Some Improved Class of Ratio Estimators for Finite Population Variance with the Use of Known Parameters," *LC Int. J. STEM ISSN 2708-7123*, vol. 3, no. 3, pp. 35–45, 2022.
- [16] M. A. Yunusa *et al.*, "Logarithmic ratio-type estimator of population coefficient of variation," *Asian J. Probab. Stat.*, vol. 14, no. 2, pp. 13–22, 2021.
- [17] A. Audu *et al.*, "Difference-cum-ratio estimators for estimating finite population coefficient of variation in simple random sampling," *Asian J. Probab. Stat.*, vol. 13, no. 3, pp. 13–29, 2021.
- [18] V. Archana and A. Rao, "Some improved estimators of co-efficient of variation from bi-variate normal distribution: A Monte Carlo comparison," *Pak. J. Stat. Oper. Res.*, pp. 87–105, 2014.

- [19] M. N. Murthy, "Sampling theory and methods.," *Sampl. Theory Methods*, 1967,  
Accessed: Apr. 19, 2024. [Online]. Available:  
<https://www.cabdirect.org/cabdirect/abstract/19702700466>
- [20] S. Singh, *Advanced Sampling Theory With Applications: How Michael"" Selected"" Amy*, vol. 2. Springer Science & Business Media, 2003. Accessed: Apr. 19, 2024.  
[Online]. Available:  
[https://books.google.com/books?hl=en&lr=&id=8XGDJFotX\\_QC&oi=fnd&pg=PA615&dq=SINGH,+S.,+\(2003\).+Advanced+Sampling+Theory+With+Applications:+How+Michael+Selected+Amy+\(Vol.+2\),+Springer+Science+and+Business+Media.&ots=2t drEnZykt&sig=Fvsjf34tjMfw\\_pD-v4mVFDyJuE0](https://books.google.com/books?hl=en&lr=&id=8XGDJFotX_QC&oi=fnd&pg=PA615&dq=SINGH,+S.,+(2003).+Advanced+Sampling+Theory+With+Applications:+How+Michael+Selected+Amy+(Vol.+2),+Springer+Science+and+Business+Media.&ots=2t drEnZykt&sig=Fvsjf34tjMfw_pD-v4mVFDyJuE0)

UNDER PEER REVIEW