

RESEARCH ARTICLE

Enhancing ratio estimators for estimating population mean using maximum value of auxiliary variable

Nasir Abbas¹, Muhammad Abid^{1,2*}, Muhammad Tahir¹, Nasir Abbas³ and Zawar Hussain⁴

¹ Department of Statistics, Government College University, Faisalabad 38000, Pakistan.

² Department of Mathematics, Institute of Statistics, Zhejiang University, Hangzhou 310027, China.

³ Department of Statistics, Government Postgraduate College, Jhang, Pakistan.

⁴ Department of Statistics, Quaid-i-Azam University, Islamabad 44000, Pakistan.

Revised: 19 May 2018; Accepted: 25 May 2018

Abstract: This article proposes some newly modified ratio estimators for the estimation of population mean of the study variable by incorporating the maximum value of the auxiliary variable. The expressions of the bias and mean squared error were derived. The conditions in which the suggested estimators have minimum values of mean squared error in comparison to the existing estimators were also derived. An empirical and simulation study was also conducted. From the empirical and simulation study, it was found that the suggested estimators perform more efficiently as compared to the existing estimators considered in this study.

Keywords: Bias, coefficient of variation, maximum value, ratio estimator, simulation study.

INTRODUCTION

In survey research, if information is available in every unit of population and this information is correlated with the study variable, then such information is called auxiliary information. By using this auxiliary information one can propose numerous types of estimators for estimating the population mean by incorporating product, ratio and regression methods of estimation. The ratio and product methods of estimation are helpful for estimating the population mean when the correlation between the study variable and the auxiliary variable is positive and negative, respectively (Cochran, 1940; Murthy, 1967).

Suppose a finite population $V = \{V_1, V_2, \dots, V_N\}$ consists of N different and identifiable units. Let Y be a measurable variable of interest with values Y_i being ascertained on V_i ; $i = 1, 2, \dots, N$ resulting in a set of observations $Y = \{Y_1, Y_2, \dots, Y_N\}$. The main purpose of this study is to estimate the population mean $\mu_y = \frac{1}{N} \sum_{i=1}^N Y_i$ by drawing a simple random sample (SRS) from the population. Suppose that the information on auxiliary variable X , which is correlated with the study variable Y , is also available against every unit of the population. With this additional information one can develop various types of estimators for population variance by using the ratio, regression, product and ratio-cum product methods of estimation.

Therefore, the ratio estimator for estimating the population mean \bar{Y} of the study variable Y is defined as:

$$\hat{Y}_R = \frac{\bar{y}}{\bar{x}} \mu_x = \hat{R} \mu_x, \quad \dots(1)$$

where $\hat{R} = \frac{\bar{y}}{\bar{x}}$ is the estimate of $R = \frac{\mu_y}{\mu_x}$.

The ratio estimator represented in equation (1) produced efficient results for estimating μ_y in comparison to the usual sample mean for a high positive correlation.

* Corresponding author (mabid@gcuf.edu.pk;  <https://orcid.org/0000-0001-8996-1120>)



Generally, estimation of μ_y by using auxiliary information is dealt in the context of augmenting the conventional parameters of an auxiliary variable through ratio or regression methods of estimation to achieve greater efficiency. Mostly, coefficient of skewness, coefficient of kurtosis, coefficient of variation of the auxiliary variable and coefficient of correlation between the study and the auxiliary variable are used in linear combination with some other conventional parameters of the auxiliary variable to estimate the population mean. For instance, refer Sisodia and Dwivedi (1981), Rao (1991), Upadhyaya and Singh (1999), Singh *et al.* (2004), Kadilar and Cingi (2004; 2006), Yan and Tian (2010), Jeelani *et al.* (2013) and the references therein. For the ratio estimator for estimating \bar{Y} based on the non-conventional measures of location and dispersion, refer the work of Abid *et al.* (2016a; 2016b; 2016c; 2016d). The present study was focused on estimation of the population mean by incorporating information on a maximum value of the auxiliary variable with a linear combination of the coefficient of variation and correlation coefficient under ratio type structure of estimation. The main motivation behind suggesting the proposed estimators was to use additional information contained in the maximum value of the auxiliary to get relatively more precise estimators.

METHODOLOGY

Existing ratio estimators

Kadilar and Cingi (2004) proposed the ratio estimators based on the C_x and β_2 of an auxiliary variable and showed that their suggested estimators are more efficient in comparison to the usual ratio estimator of population mean. Kadilar and Cingi (2004) estimators are mentioned by:

$$\hat{Y}_1 = \frac{\bar{y} + b(\mu_x - \bar{x})}{\bar{x}} \mu_x$$

$$\hat{Y}_2 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + C_x)} (\mu_x + C_x)$$

$$\hat{Y}_3 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + \beta_2)} (\mu_x + \beta_2)$$

$$\hat{Y}_4 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\beta_2 + C_x)} (\mu_x\beta_2 + C_x)$$

$$\hat{Y}_5 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}C_x + \beta_2)} (\mu_x C_x + \beta_2)$$

The values of biases, constants and mean squared errors (MSEs) of the above-given estimators are:

$$B(\bar{Y}_1) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_1^2 \quad R_1 = \frac{\mu_y}{\mu_x}$$

$$MSE(\bar{Y}_1) = \frac{(1-f)}{n} \left(R_1^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_2) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_2^2 \quad R_2 = \frac{\mu_y}{(\mu_x + C_x)}$$

$$MSE(\bar{Y}_2) = \frac{(1-f)}{n} \left(R_2^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_3) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_3^2 \quad R_3 = \frac{\mu_y}{(\mu_x + \beta_2)}$$

$$MSE(\bar{Y}_3) = \frac{(1-f)}{n} \left(R_3^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_4) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_4^2 \quad R_4 = \frac{\mu_y \beta_2}{(\mu_x \beta_2 + C_x)}$$

$$MSE(\bar{Y}_4) = \frac{(1-f)}{n} \left(R_4^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_5) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_5^2 \quad R_5 = \frac{\mu_y C_x}{\mu_x C_x + \beta_2}$$

$$MSE(\bar{Y}_5) = \frac{(1-f)}{n} \left(R_5^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

where N is the population size, n is the sample size, $f = n/N$ represent the sampling fraction, μ_x is the population mean of an auxiliary variable X , μ_y is the population mean of study variable Y , S_x is the population standard deviation of X , S_y is the population standard deviation of Y , C_x is the coefficients of variation of X , C_y is the coefficients of variation of Y and ρ is the coefficient of correlation between X and Y .

Kadilar and Cingi (2006) suggested some ratio estimators by using the correlation coefficient of the auxiliary variable. It was proved that the suggested estimators are smaller values of MSEs compared to the other existing estimators (Kadilar & Cingi, 2006). The estimators are represented below:

$$\hat{Y}_6 = \frac{\bar{y} + b(\mu_x - \bar{x})}{\bar{x}} (\mu_x + \rho)$$

$$\hat{Y}_7 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}C_x + \rho)} (\mu_x C_x + \rho)$$

$$\hat{Y}_8 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\rho + C_x)} (\mu_x \rho + C_x)$$

$$\hat{Y}_9 = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\beta_2 + \rho)} (\mu_x \beta_2 + \rho)$$

$$\hat{Y}_{10} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\rho + \beta_2)} (\mu_x \rho + \beta_2)$$

The values of biases, constants and MSEs for Kadilar and Cingi (2006) are represented as:

$$B(\bar{Y}_6) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_6^2 \quad R_6 = \frac{\mu_y}{\mu_x + \rho}$$

$$MSE(\bar{Y}_6) = \frac{(1-f)}{n} \left(R_6^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_7) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_7^2 \quad R_7 = \frac{\mu_y C_x}{\mu_x C_x + \rho}$$

$$MSE(\bar{Y}_7) = \frac{(1-f)}{n} \left(R_7^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_8) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_8^2 \quad R_8 = \frac{\mu_y \rho}{\mu_x \rho + C_x}$$

$$MSE(\bar{Y}_8) = \frac{(1-f)}{n} \left(R_8^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_9) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_9^2 \quad R_9 = \frac{\mu_y \beta_2}{\mu_x \beta_2 + \rho}$$

$$MSE(\bar{Y}_9) = \frac{(1-f)}{n} \left(R_9^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_{10}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{10}^2 \quad R_{10} = \frac{\mu_y \rho}{\mu_x \rho + \beta_2}$$

$$MSE(\bar{Y}_{10}) = \frac{(1-f)}{n} \left(R_{10}^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

Yan and Tian (2010) proposed the estimators based on the values of β_1 and β_2 of an auxiliary variable and these estimators are given as:

$$\hat{Y}_{11} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + \beta_1)} (\mu_x + \beta_1)$$

$$\hat{Y}_{12} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\beta_1 + \beta_2)} (\mu_x \beta_1 + \beta_2)$$

where β_1 and β_2 are the coefficient of skewness and coefficient of kurtosis of X , respectively.

The values of biases, constants and MSEs for Yan and Tian (2010) are as follows:

$$B(\bar{Y}_{11}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{11}^2 \quad R_{11} = \frac{\mu_y}{(\mu_x + \beta_1)}$$

$$MSE(\bar{Y}_{11}) = \frac{(1-f)}{n} \left(R_{11}^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

$$B(\bar{Y}_{12}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{12}^2 \quad R_{12} = \frac{\mu_y \beta_1}{(\mu_x \beta_1 + \beta_2)}$$

$$MSE(\bar{Y}_{12}) = \frac{(1-f)}{n} \left(R_{12}^2 S_x^2 + S_y^2 (1-\rho^2) \right)$$

Subramani and Kumarapandiyam (2012a; 2012b; 2012c; 2012d) introduced some new estimators using the linear combination of M_d , C_x , β_1 , β_2 and deciles of an auxiliary variable for estimating the population mean. Subramani and Kumarapandiyam (2012a; 2012b; 2012c; 2012d) estimators are given as:

$$\hat{Y}_{13} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + M_d)} (\mu_x + M_d)$$

$$\hat{Y}_{14} = \frac{\bar{y} + b(\bar{X} - \bar{x})}{(C_x \bar{x} + M_d)} (C_x \mu_x + M_d)$$

$$\hat{Y}_{15} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(B_1 \bar{x} + M_d)} (B_1 \mu_x + M_d)$$

$$\hat{Y}_{16} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(B_2 \bar{x} + M_d)} (B_2 \mu_x + M_d)$$

$$\hat{Y}_{17} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_1)} (\mu_x + D_1)$$

$$\hat{Y}_{18} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_2)}(\mu_x + D_2)$$

$$\hat{Y}_{19} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_3)}(\mu_x + D_3)$$

$$\hat{Y}_{20} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_4)}(\mu_x + D_4)$$

$$\hat{Y}_{21} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_5)}(\mu_x + D_5)$$

$$\hat{Y}_{22} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_6)}(\mu_x + D_6)$$

$$\hat{Y}_{23} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_7)}(\mu_x + D_7)$$

$$\hat{Y}_{24} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_8)}(\mu_x + D_8)$$

$$\hat{Y}_{25} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_9)}(\mu_x + D_9)$$

$$\hat{Y}_{26} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + D_{10})}(\mu_x + D_{10})$$

where M_d is median of X and $D_i, i=1,2,\dots,10$ are the deciles of X .

The values of biases, constants and MSEs of Subramani and Kumarapandiyan (2012a; 2012b; 2012c; 2012d) estimators are shown below:

$$B(\hat{Y}_{13}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{13}^2, \quad R_{13} = \frac{\mu_y}{(\mu_x + M_d)}$$

$$MSE(\hat{Y}_{13}) = \frac{(1-f)}{n} (R_{13}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{14}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{14}^2, \quad R_{14} = \frac{C_x \mu_y}{(C_x \mu_x + M_d)}$$

$$MSE(\hat{Y}_{14}) = \frac{(1-f)}{n} (R_{14}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{15}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{15}^2, \quad R_{15} = \frac{B_1 \mu_y}{(B_1 \mu_x + M_d)}$$

$$MSE(\hat{Y}_{15}) = \frac{(1-f)}{n} (R_{15}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{16}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{16}^2, \quad R_{16} = \frac{B_2 \mu_y}{(B_2 \mu_x + M_d)}$$

$$MSE(\hat{Y}_{16}) = \frac{(1-f)}{n} (R_{16}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{17}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{17}^2, \quad R_{17} = \frac{\mu_y}{(\mu_x + D_1)}$$

$$MSE(\hat{Y}_{17}) = \frac{(1-f)}{n} (R_{17}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{18}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{18}^2, \quad R_{18} = \frac{\mu_y}{(\mu_x + D_2)}$$

$$MSE(\hat{Y}_{18}) = \frac{(1-f)}{n} (R_{18}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{19}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{19}^2, \quad R_{19} = \frac{\mu_y}{(\mu_x + D_3)}$$

$$MSE(\hat{Y}_{19}) = \frac{(1-f)}{n} (R_{19}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{20}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{20}^2, \quad R_{20} = \frac{\mu_y}{(\mu_x + D_4)}$$

$$MSE(\hat{Y}_{20}) = \frac{(1-f)}{n} (R_{20}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{21}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{21}^2, \quad R_{21} = \frac{\mu_y}{(\mu_x + D_5)}$$

$$MSE(\hat{Y}_{21}) = \frac{(1-f)}{n} (R_{21}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{22}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{22}^2, \quad R_{22} = \frac{\mu_y}{(\mu_x + D_6)}$$

$$MSE(\hat{Y}_{22}) = \frac{(1-f)}{n} (R_{22}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{23}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{23}^2, \quad R_{23} = \frac{\mu_y}{(\mu_x + D_7)}$$

$$MSE(\hat{Y}_{23}) = \frac{(1-f)}{n} (R_{23}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{24}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{24}^2, \quad R_{24} = \frac{\mu_y}{(\mu_x + D_8)}$$

$$MSE(\hat{Y}_{24}) = \frac{(1-f)}{n} (R_{24}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{25}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{25}^2, \quad R_{25} = \frac{\mu_y}{(\mu_x + D_9)}$$

$$MSE(\hat{Y}_{25}) = \frac{(1-f)}{n} \left(R_{25}^2 S_x^2 + S_y^2 (1 - \rho^2) \right)$$

$$B(\hat{Y}_{26}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{26}^2, \quad R_{26} = \frac{\mu_y}{(\mu_x + D_{10})}$$

$$MSE(\hat{Y}_{26}) = \frac{(1-f)}{n} \left(R_{26}^2 S_x^2 + S_y^2 (1 - \rho^2) \right)$$

Jeelani *et al.* (2013) developed the ratio estimator based on the value of quartile deviation and β_1 of the auxiliary variable for estimating the population mean and it is given as:

$$\hat{Y}_{27} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\beta_1 + QD)} (\mu_x\beta_1 + QD)$$

where QD is the quartile deviation of X .

The value of bias, constant and the MSE are specified as:

$$B(\bar{Y}_{27}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{27}^2$$

$$R_{27} = \frac{\mu_y\beta_1}{(\mu_x\beta_1 + QD)}$$

$$MSE(\bar{Y}_{27}) = \frac{(1-f)}{n} \left(R_{27}^2 S_x^2 + S_y^2 (1 - \rho^2) \right)$$

All the above estimators utilise auxiliary information in one way or the other. Good auxiliary variables are those which are similar to the study variable and can be taken as good proxies of the study variable. In addition, if correlation between the study and auxiliary variable is high (positive or negative), then based on the auxiliary information the characteristics of the study variable (such as mean, variance, coefficient of variation, maximum, minimum) can be precisely predicted. If the auxiliary variable is available in advance, in the past experience or in the pilot study, its extreme values (minimum and maximum) are easily available. When correlation between the study variable and the auxiliary variable is positive, then selecting the larger (smaller) value of the auxiliary variable will more likely result in the larger (smaller) value of study variable in the sample.

Therefore, in many populations there exist some large or small values, and to estimate the population parameters without considering this information is sensitive. As is evident, in case of presence of an extremely high (low) outlier in the distribution of auxiliary variable, its mean is larger (smaller). In either case the result will be overestimation or underestimation of the mean. In the next section, we intend to propose estimators of mean by giving the same increment (the maximum of auxiliary variable) to population and sample means of auxiliary variable.

Proposed ratio estimators

Generally when the correlation between the study variable and the auxiliary variable is positive, the selection of the maximum value of the auxiliary variable is highly associated with the selection of the maximum value of the study variable. In this section, ratio estimators are proposed for the estimation of finite population mean using information of the maximum value of the auxiliary variable with the linear combination of the coefficient of variation and coefficient of correlation of an auxiliary variable. The proposed estimators are given by:

$$\hat{Y}_{p1} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x} + M_{(x)})} (\mu_x + M_{(x)})$$

$$\hat{Y}_{p2} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}C_x + M_{(x)})} (\mu_x C_x + M_{(x)})$$

$$\hat{Y}_{p3} = \frac{\bar{y} + b(\mu_x - \bar{x})}{(\bar{x}\rho + M_{(x)})} (\mu_x\rho + M_{(x)})$$

where $M_{(x)}$ is the maximum value of an auxiliary variable.

It is to be noted that in all the proposed estimators the ration part is augmented by using the population maximum of the auxiliary variable. It is anticipated that using the maximum of auxiliary variable will result in improved estimation of mean.

The values of biases, constants and MSEs of the proposed estimators are specified below:

$$B(\hat{Y}_{p1}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{p1}^2, \quad R_{p1} = \frac{\mu_y}{(\mu_x + M_{(x)})}$$

$$MSE(\hat{Y}_{p1}) = \frac{(1-f)}{n} \left(R_{p1}^2 S_x^2 + S_y^2 (1 - \rho^2) \right)$$

$$B(\hat{Y}_{p2}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{p2}^2 \quad R_{p2} = \frac{\mu_y}{(\mu_x C_x + M_{(x)})} \quad R_{pk} < R_l \quad \dots(2)$$

$$MSE(\hat{Y}_{p2}) = \frac{(1-f)}{n} (R_{p2}^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$B(\hat{Y}_{p3}) = \frac{(1-f)}{n} \frac{S_x^2}{\mu_y} R_{p3}^2 \quad R_{p3} = \frac{\mu_y}{(\mu_x \rho + M_{(x)})}$$

$$MSE(\hat{Y}_{p3}) = \frac{(1-f)}{n} (R_{p3}^2 S_x^2 + S_y^2 (1-\rho^2))$$

Efficiency comparisons

In this section, we derive the conditions in which the suggested estimators perform more efficiently in comparison to the existing estimators. The proposed estimator performs more efficiently if and only if;

$$MSE(\hat{Y}_{pk}) < MSE(\hat{Y}_l)$$

$$\frac{1-f}{n} (R_{pk}^2 S_x^2 + S_y^2 (1-\rho^2)) < \frac{1-f}{n} (R_l^2 S_x^2 + S_y^2 (1-\rho^2))$$

$$R_{pk}^2 S_x^2 < R_l^2 S_x^2$$

where $k = 1,2,3$ for the proposed estimators and $l = 1,2,\dots,27$ for the existing estimators.

RESULTS AND DISCUSSION

Practical study

To evaluate the performance of the proposed estimators against their competing estimators which are given in the Methodology, five natural populations have been selected. Populations I and II were obtained from Singh and Chaudhary (1986), population III was taken from Cochran (1940) and populations IV and V were obtained from Murthy (1967). The characteristics of the populations considered in this study are presented in Table 1. The values of biases and constants are given in Tables 2 and 3 for existing and proposed estimators, respectively. The MSE values are represented in Tables 4 and 5 for the existing and proposed estimators, respectively.

It was revealed that the values of constants of the proposed estimators, i.e., R_{pk} are smaller as compared to the values of the constants of the existing estimators i.e., R_l which satisfied equation (2) (Table 2 vs Table 3).

Table 1: Values of population characteristics

Parameter	Population I	Population II	Population III	Population IV	Population V
N	34	34	49	80	80
n	20	20	20	20	20
μ_y	856.411	856.411	127.796	5182.637	5182.637
μ_x	208.883	199.441	103.143	285.125	1126.463
ρ	0.449	0.445	0.98174	0.915	0.941
S_y	733.141	733.141	123.121	1835.659	1835.659
C_y	0.857	0.856	0.963	0.354	0.354
S_x	150.506	150.215	104.405	279.429	845.610
C_x	0.721	0.753	1.012	0.948	0.751
β_2	0.098	1.045	5.141	1.301	-0.063
β_1	0.978	1.182	2.255	0.698	1.050
M_d	150	142.500	64	148	757.500
QD	80.25	89.375	38.50	179.375	588.125
$M_{(x)}$	564	634	507	1095	3485

From Tables 2 and 3, it was noted that the suggested estimators have minimum values of biases compared to the existing estimators. As expected, MSE values of all the proposed estimators were smaller in comparison to the existing estimators for all the populations considered in this study, which revealed the supremacy of the proposed estimators (Table 4 vs Table 5). It was also observed that the suggested estimators based on the linear combination

of correlation coefficient and the maximum value of auxiliary variable, i.e., \hat{Y}_{p3} , have the minimum MSE value as compared to the other proposed estimators, i.e., \hat{Y}_{p1} and \hat{Y}_{p2} (Table 5). Hence, we can say that the suggested estimators perform more efficiently compared to the existing estimators based on the five real datasets considered in this study.

Table 2: The values of constants and biases for existing estimators

Estimator	Constant					Bias				
	Population I	Population II	Population III	Population IV	Population V	Population I	Population II	Population III	Population IV	Population V
\hat{Y}_1	4.100	4.294	1.239	18.177	4.601	9.154	10.002	3.875	174.83	109.52
\hat{Y}_2	4.086	4.278	1.227	18.116	4.598	9.091	9.927	3.800	173.67	109.37
\hat{Y}_3	4.098	4.272	1.180	18.132	4.601	9.145	9.898	3.516	173.98	109.53
\hat{Y}_4	3.960	4.279	1.237	18.090	4.650	8.539	9.930	3.860	173.18	111.86
\hat{Y}_5	4.097	4.264	1.181	18.130	4.601	9.142	9.865	3.520	173.93	109.53
\hat{Y}_6	4.091	4.285	1.227	18.119	4.597	9.115	9.957	3.802	173.71	109.34
\hat{Y}_7	4.088	4.281	1.228	18.115	4.596	9.099	9.943	3.803	173.65	109.27
\hat{Y}_8	4.069	4.258	1.227	18.111	4.598	9.015	9.834	3.799	173.57	109.36
\hat{Y}_9	4.012	4.285	1.237	18.094	4.662	8.763	9.960	3.861	173.23	112.46
\hat{Y}_{10}	4.096	4.244	1.179	18.128	4.601	9.135	9.771	3.509	173.90	109.53
\hat{Y}_{11}	4.081	4.269	1.213	18.094	4.597	9.069	9.885	3.711	173.24	109.31
\hat{Y}_{12}	4.098	4.275	1.212	18.143	4.601	9.145	9.914	3.709	174.17	109.53
\hat{Y}_{13}	2.3863	2.5046	0.765	11.966	2.751	3.101	3.403	1.476	75.76	39.15
\hat{Y}_{14}	2.053	2.204	0.768	11.747	2.427	2.296	2.634	1.489	73.02	30.47
\hat{Y}_{15}	2.364	2.676	0.971	12.991	2.804	3.043	3.886	2.383	89.31	40.69
\hat{Y}_{16}	0.489	2.550	1.105	10.422	-0.478	0.130	3.526	3.085	57.48	1.186
\hat{Y}_{17}	3.068	3.293	0.926	14.823	3.464	5.124	5.883	2.1664	116.275	62.082
\hat{Y}_{18}	2.998	3.032	0.893	14.288	3.266	4.893	4.987	2.0118	108.028	55.188
\hat{Y}_{19}	2.701	2.834	0.860	13.605	3.007	3.9725	4.358	1.8682	97.952	46.786
\hat{Y}_{20}	2.532	2.757	0.828	12.686	2.874	3.4902	4.122	1.7305	85.163	42.737
\hat{Y}_{21}	2.386	2.505	0.765	11.966	2.751	3.1010	3.402	1.4756	75.764	39.154
\hat{Y}_{22}	1.964	2.091	0.715	10.613	2.622	2.1003	2.370	1.2903	59.604	35.567
\hat{Y}_{23}	1.865	1.846	0.650	8.047	1.985	1.8934	1.8484	1.0650	34.268	20.385
\hat{Y}_{24}	1.573	1.700	0.533	6.588	1.765	1.3472	1.5673	0.7160	22.964	16.117
\hat{Y}_{25}	1.328	1.496	0.366	5.211	1.429	0.9601	1.2133	0.3389	14.367	10.567
\hat{Y}_{26}	1.108	1.028	0.209	3.755	1.124	0.6686	0.5727	0.1107	7.462	6.535
\hat{Y}_{27}	2.9438	3.1138	1.063	12.251	3.073	4.719	5.260	2.852	79.42	48.85

Table 3: The values of constants and biases for proposed estimators

Estimator	Constant					Bias				
	Population I	Population II	Population III	Population IV	Population V	Population I	Population II	Population III	Population IV	Population V
\hat{Y}_{p1}	1.108	1.028	0.209	3.755	1.124	0.669	0.573	0.111	7.462	6.535
\hat{Y}_{p2}	0.864	0.822	0.212	3.600	0.898	0.406	0.367	0.113	6.858	4.176
\hat{Y}_{p3}	0.585	0.528	0.206	3.497	1.073	0.186	0.151	0.107	6.473	5.960

Table 4: MSE values of the existing estimators for all selected populations

Estimator	MSE				
	Population I	Population II	Population III	Population IV	Population V
\hat{Y}_1	16673.45	17437.65	511.42	926660.70	581994.20
\hat{Y}_2	16619.64	17373.31	501.84	920662.50	581238.50
\hat{Y}_3	16666.14	17348.62	465.51	922242.50	582058.10
\hat{Y}_4	16146.61	17376.04	509.53	918082.10	594119.80
\hat{Y}_5	16663.31	17319.75	466.02	922003.40	582079.30
\hat{Y}_6	16639.85	17399.52	502.12	920873.20	581046.80
\hat{Y}_7	16626.87	17387.08	502.23	920560.30	580732.70
\hat{Y}_8	16554.4	17294.19	501.66	920108.20	581191.40
\hat{Y}_9	16338.65	17401.14	509.59	918382.80	597260.90
\hat{Y}_{10}	16657.19	17239.66	464.72	921833.60	582062.10
\hat{Y}_{11}	16600.54	17336.98	490.45	918450.90	580937.60
\hat{Y}_{12}	16665.98	17362.26	490.23	923260.70	582055.10
\hat{Y}_{13}	11489.70	11785.70	204.80	413230.80	217319.80
\hat{Y}_{14}	10800.4	11127.47	206.55	399044.90	172323.80
\hat{Y}_{15}	11440.8	12199.76	320.78	483450.4	225319.5
\hat{Y}_{16}	8945.9	11892.07	410.50	318486.7	20545.47
\hat{Y}_{17}	13222.5	13910.4	293.1	623182.3	336145.8
\hat{Y}_{18}	13025.0	13142.8	273.3	580441.2	300416.7
\hat{Y}_{19}	12236.1	12604.0	255.0	528220.0	256874.4
\hat{Y}_{20}	11823.0	12402.5	237.4	461943.7	235888.6
\hat{Y}_{21}	11489.7	11785.7	204.8	413230.8	217319.8
\hat{Y}_{22}	10632.6	10902.1	181.1	329476.5	198733.1
\hat{Y}_{23}	10455.5	10454.6	152.3	198169.8	120048.6
\hat{Y}_{24}	9987.7	10213.8	107.7	139586.0	97924.6
\hat{Y}_{25}	9656.2	9910.7	59.5	95032.7	69163.6
\hat{Y}_{26}	9406.6	9362.1	30.4	59245.5	48266.6
\hat{Y}_{27}	12875.4	13376.0	380.8	432164.6	267595.2

Table 5: MSE values of the suggested estimators for all selected populations

Estimator	MSE				
	Population I	Population II	Population III	Population IV	Population V
\hat{Y}_{p1}	9406.6	9362.1	30.4	59245.5	48266.6
\hat{Y}_{p2}	9181.8	9185.9	30.7	56114.7	36039.3
\hat{Y}_{p3}	8993.4	9000.9	30.0	54117.3	45286.9

Table 6: Relative efficiency of suggested estimators with respect to existing estimators

Existing	Proposed			Existing	Proposed		
	\hat{Y}_{p1}	\hat{Y}_{p2}	\hat{Y}_{p3}		\hat{Y}_{p1}	\hat{Y}_{p2}	\hat{Y}_{p3}
\hat{Y}_1	0.9709	0.9675	0.9655	\hat{Y}_{15}	0.9881	0.9864	0.9846
\hat{Y}_2	0.9711	0.9677	0.9656	\hat{Y}_{16}	0.9865	0.9846	0.9828
\hat{Y}_3	0.9712	0.9678	0.9657	\hat{Y}_{17}	0.9869	0.9851	0.9833
\hat{Y}_4	0.9715	0.9681	0.9661	\hat{Y}_{18}	0.9801	0.9777	0.9758
\hat{Y}_5	0.9680	0.9646	0.9624	\hat{Y}_{19}	0.9824	0.9802	0.9784
\hat{Y}_6	0.9710	0.9676	0.9656	\hat{Y}_{20}	0.9842	0.9821	0.9803
\hat{Y}_7	0.9679	0.9644	0.9622	\hat{Y}_{21}	0.9849	0.9828	0.9810
\hat{Y}_8	0.9592	0.9555	0.9531	\hat{Y}_{22}	0.9871	0.9852	0.9835
\hat{Y}_9	0.9715	0.9681	0.9660	\hat{Y}_{23}	0.9907	0.9891	0.9873
\hat{Y}_{10}	0.9594	0.9557	0.9533	\hat{Y}_{24}	0.9929	0.9913	0.9896
\hat{Y}_{11}	0.9712	0.9678	0.9657	\hat{Y}_{25}	0.9941	0.9927	0.9910
\hat{Y}_{12}	0.9726	0.9693	0.9672	\hat{Y}_{26}	0.9959	0.9945	0.9928
\hat{Y}_{13}	0.9830	0.9809	0.9790	\hat{Y}_{27}	1.0000	0.9988	0.9971
\hat{Y}_{14}	0.9871	0.9852	0.9835				

Simulation study

We also conducted a simulation study between the proposed estimators and the existing estimators by using the real dataset of population IV. We used the simulation study to find the MSE of the existing estimators and the suggested estimators. The algorithm is designed in R language. The simulations were conducted according

to the following steps:

- i. Select samples of size $n = 2$ from the real dataset of population IV using simple random sampling without replacement.
- ii. Use the data from samples in (i) to find the value of \hat{Y}_l and \hat{Y}_{pk} . So, we obtain 5000 values of \hat{Y}_l and \hat{Y}_{pk} from 5000 samples for $n = 20$.

iii. For $n = 20$, the MSE of \hat{Y}_l and \hat{Y}_{pk} are computed by the formulas given below:

$$MSE(\hat{Y}_l) = E(\hat{Y}_l - \bar{Y})^2 \quad \dots(3)$$

and

$$MSE(\hat{Y}_{pk}) = E(\hat{Y}_{pk} - \bar{Y})^2 \quad \dots(4)$$

where \bar{Y} is the population mean of the study variable.

The comparison between the existing and proposed estimators were done on the basis of relative efficiency (RE). The RE of the suggested estimators with respect to existing estimators were found by using the expression given below:

$$RE(\hat{Y}_{pk}, \hat{Y}_l) = \frac{MSE(\hat{Y}_{pk})}{MSE(\hat{Y}_l)}, \text{ where } l = 1, 2, 3, \dots, 27 \text{ and}$$

$k = 1, 2, 3$.

The values of RE are reported in Table 6. From Table 6, it can be concluded that all the suggested estimators perform better based on the values of RE in comparison to the existing estimators considered in this study and also these results are in accordance with the results of the practical study.

CONCLUSION

The present study proposes some new ratio estimators by integrating the information on the maximum value of an auxiliary variable. The comparison between proposed estimators and existing estimators are done on the basis of mean square error and relative efficiency. From the empirical study, it was revealed that the suggested estimators have smaller mean square error values in comparison to the existing estimators, which indicate the supremacy of the proposed estimators compared to the existing estimators considered in this study. It is noted that the performance of the proposed estimator \hat{Y}_{p3} is relatively higher as compared to the other proposed estimators. The results of simulation study also support the theoretical findings of the study. Therefore, based on the findings obtained from this study, we strongly recommend the use of proposed estimators for the practical applications against the existing estimators.

REFERENCES

- Abid M., Abbas N., Nazir Z.A. & Lin Z. (2016a). Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters. *Revista Colombiana de Estadística* **39**(1): 63 – 79.
DOI: <https://doi.org/10.15446/rce.v39n1.55139>
- Abid M., Abbas N. & Riaz M. (2016b). Improved modified ratio estimators of population mean based on deciles. *Chiang Mai Journal of Science* **43**(1): 1311 – 1323.
- Abid M., Abbas N., Sherwani R.A.K. & Nazir Z.A. (2016c). Improved ratio estimators for the population mean using non-conventional measures of dispersion. *Pakistan Journal of Statistics and Operation Research* **12**(2): 353 – 367.
DOI: <https://doi.org/10.18187/pjsor.v12i2.1182>
- Abid M., Sherwani R.A.K., Abbas N. & Nawaz T. (2016d). Some improved modified ratio estimators based on decile mean of an auxiliary variable. *Pakistan Journal of Statistics and Operation Research* **12**(4): 787 – 797.
DOI: <https://doi.org/10.18187/pjsor.v12i4.1408>
- Cochran W.G. (1940). *Sampling Techniques*, 3rd edition. Wiley Eastern Limited, India.
- Ferrell E.B. (1953). Control charts using midranges and medians. *Industrial Quality Control* **9**: 30 – 34.
- Hettmansperger T.P. & McKean J.W. (1998). *Robust Nonparametric Statistical Methods*. John Wiley and Sons, New York, USA.
- Jeelani M.I., Maqbool S. & Mir S.A. (2013). Modified ratio estimators of population mean using linear combination of coefficient of skewness and quartile deviation. *International Journal of Modern Mathematical Sciences* **6**: 174 – 183.
- Kadilar C. & Cingi H. (2004). Ratio estimators in simple random sampling. *Applied Mathematics and Computation* **151**: 893 – 902.
- Kadilar C. & Cingi H. (2006). An improvement in estimating the population mean by using the correlation coefficient. *Haceteppe Journal of Mathematics and Statistics* **35**: 103 – 109.
- Murthy M.N. (1967). *Sampling Theory and Methods*. Statistical Publishing Society, Calcutta, India.
- Rao T.J. (1991). On certain methods of improving ratio and regression estimators. *Communications in Statistics: Theory and Methods* **20**: 3325 – 3340.
DOI: <https://doi.org/10.1080/03610929108830705>
- Singh D. & Chaudhary F.S. (1986). *Theory and Analysis of Sample Survey Designs*. New Age International Publisher, India.
- Singh G.N. (2003). On the improvement of product method of estimation in sample surveys. *Journal of the Indian Society of Agricultural Statistics* **56**: 267 – 275.
- Singh H.P. & Tailor R. (2003). Use of known correlation coefficient in estimating the finite population means.

- Statistics in Transition* **6**: 555 – 560.
- Singh H.P., Tailor R., Tailor R. & Kakran M.S. (2004). An improved estimator of population mean using power transformation. *Journal of the Indian Society of Agricultural Statistics* **58**: 223 – 230.
- Singh H.P. & Tailor R. (2005). Estimation of finite population mean with known coefficient of variation of an auxiliary. *STATISTICA* **3**: 301 – 313.
- Sisodia B.V.S. & Dwivedi V.K. (1981). A modified ratio estimator using coefficient of variation of auxiliary variable. *Journal of the Indian Society of Agricultural Statistics* **33**: 13 – 18.
- Subramani J. & Kumarapandiyan G. (2012a). A class of almost unbiased modified ratio estimators for population mean with known population parameters. *Elixir Statistics* **44**: 7411 – 7415.
- Subramani J. & Kumarapandiyan G. (2012b). Estimation of population mean using co-efficient of variation and median of an auxiliary variable. *International Journal of Probability and Statistics* **1**: 111 – 118.
DOI: <https://doi.org/10.5923/j.ijps.20120103.02>
- Subramani J. & Kumarapandiyan G. (2012c). Estimation of population mean using known median and co-efficient of skewness. *American Journal of Mathematics and Statistics* **2**: 101 – 107.
DOI: <https://doi.org/10.5923/j.ajms.20120205.01>
- Subramani J. & Kumarapandiyan G. (2012d). Modified ratio estimators using known median and co-efficient of kurtosis. *American Journal of Mathematics and Statistics* **2**: 95 – 100.
DOI: <https://doi.org/10.5923/j.ajms.20120204.05>
- Upadhyaya L.N. & Singh H.P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal* **41**: 627 – 636.
- Wang T., Li Y. & Cui H. (2007). On weighted randomly trimmed means. *Journal of Systems Science and Complexity* **20**: 47 – 65.
DOI: <https://doi.org/10.1007/s11424-007-9004-7>
- Yan Z. & Tian B. (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable. In: Information computing and applications. ICICA 2010 (eds. R. Zhu, Y. Zhang, B. Liu, C. Liu). *Communications in Computer and Information Science*, volume 106. Springer, Heidelberg, Berlin, Germany.