STATISTICS IN TRANSITION new series, December 2016
Vol. 17, No. 4, pp. 591–604

A NEW MEDIAN BASED RATIO ESTIMATOR FOR ESTIMATION OF THE FINITE POPULATION MEAN

J. Subramani [1]

ABSTRACT

The present paper deals with a new median based ratio estimator for the estimation of finite population means in the absence of an auxiliary variable. The bias and mean squared error of the proposed median based ratio estimator are obtained. The performance of the median based ratio estimator is compared with that of the SRSWOR sample mean, ratio estimator and linear regression estimator for certain natural population. It is shown from the numerical comparisons that the proposed median based ratio estimator outperforms the SRSWOR sample mean, ratio estimator and also the linear regression estimator.

Key words: bias, linear regression estimator, mean squared error, natural population, simple random sampling.

1. Introduction

Let $U = \{U_1, U_2, \ldots, U_N\}$ be a finite population with N distinct and identifiable units. Let Y be the study variable with value Y_i measured on $U_i, i = 1, 2, 3, \ldots, N$ giving a vector $Y = \{Y_1, Y_2, \ldots, Y_N\}$. The problem is to estimate the population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ with some desirable properties like unbiased, minimum variance of the estimators on the basis of a random sample of size n selected from the population U. Suppose that there exists an auxiliary variable X and it is positively correlated with the study variable Y, then one can propose estimators like ratio estimator, linear regression estimator and their modifications, which will perform better than the SRSWOR sample mean for estimating the population mean of the study variable as stated in Cochran (1977) and Murthy (1967). In the absence of the auxiliary variable, the above estimators are not possible. However one may think of getting additional information on the study variable and one can propose ratio and linear regression type estimators to improve the performance of the estimator. The idea of this paper is to use such information, namely the

[1] Department of Statistics, Pondicherry University, R. V. Nagar, Kalapet, Puducherry – 605 014, India. E-mail: drjsubramani@yahoo.co.in.

median of the study variable, in the proposed ratio estimator. It is reasonable to assume that the median of the study variable is known since this parameter does not require the complete information on the population units of the study variable unlike the other parameters like population mean, population variance, etc. In particular, in the surveys involving the estimation of average income, average marks, etc. it is very reasonable to assume that the population mean is unknown whereas the population median is known. For more details, consider the examples given below, in which the interests are to estimate the population mean of the respective populations.

Example 1. In an Indian university 5000 students entered for the university examination. The results are given below. The problem is to estimate the average marks scored by the students (population mean). Here, it is reasonable to assume that the median of the marks is known since we have the following information.

Table 1. Results of University Examination

Passed with	Percentage of marks	Number of students	Cumulative total
Distinction	[75-100]	850	850
First Class	[60-75)	3100	3950
Second Class	[50-60)	600	4550
Failed	[0-50)	450	5000
Total		5000	5000

The median value will be between 60 and 75. Approximately one can assume the population median value as 67.5.

Example 2. In an Indian university 800 faculty members are working in different categories and the basic salary drawn by different categories of the faculty members are given in Table 2. The problem is to estimate the average salary drawn by the faculty members (population mean) per month. Here, it is reasonable to assume that the median of the salary is known based on the information given in Table 2.

Table 2. Salary of University faculty members

Category	Basic Salary in Indian Rupees (IRs) Per month*	Number of faculty members	Cumulative total
Senior Professor	56000+10000**	20	20
Professor - Grade I	43000+10000	40	60
Professor - Grade II	37400+10000	60	120

Table 2. Salary of University faculty members (cont.)

Category	Basic Salary in Indian Rupees (IRs) Per month*	Number of faculty members	Cumulative total
Associate Professor - Grade I	37400+10000	80	200
Associate Professor - Grade II	37400+9000	100	300
Assistant Professor - Grade I	15100+8000	110	410
Assistant Professor - Grade II	15100+7000	140	550
Assistant Professor - Grade III	15100+6000	250	800
Total		800	800

*Actual salary depends on their experience in their designation and other allowances.

**The Basic salary is the sum of the basic (the first value) and the academic grade pay (the second value), which will differentiate people with same designation but different grades.

The population median value will be assumed as IRs. 15100+8000 = IRs. 23100.

Example 3. In the estimation of body mass index (BMI) of the 350 patients of a Hospital, it is reasonable to assume that the population median of the BMI is known based on the information given in Table 3.

Table 3. Body mass index of 350 patients of a hospital

Category	BMI range – kg/m^2	Number of patients	Cumulative total
Very severely underweight	less than 15	15	15
Severely underweight	from 15.0 to 16.0	35	50
Underweight	from 16.0 to 18.5	67	117
Normal (healthy weight)	from 18.5 to 25	92	209
Overweight	from 25 to 30	47	256
Obese Class I (Moderately obese)	from 30 to 35	52	308
Obese Class II (Severely obese)	from 35 to 40	27	335
Obese Class III (Very severely obese)	over 40	15	350
Total		350	350

The median value will be between 18.5 and 25. Approximately one can assume the population median of the BMI value as 21.75.

Example 4. In the problem of estimating the blood pressure of the 202 patients of a hospital, it is reasonable to assume that the median of the blood pressure is known based on the information available in Table 4.

Table 4. Blood pressure of 202 patients of a hospital

Category	Systolic, mmHg	Number of patients	Cumulative No. of patients
Hypotension	< 90	10	10
Desired	90–119	112	122
Pre-hypertension	120–139	40	162
Stage 1 Hypertension	140–159	20	182
Stage 2 Hypertension	160–179	13	195
Hypertensive Emergency	≥ 180	7	202
Total		202	202

The median value will be between 90 and 119. Approximately one can assume the population median value as 104.5.

Before discussing further about the existing estimators and the proposed median based ratio estimator the notations and the formulae to be used in this paper are described below:

- N - Population size
- n - Sample size
- $N_{C_n} = \binom{N}{n}$ - Number of possible samples of size n from the population of size N
- Y - Study variable
- M - Median of the Study variable
- X - Auxiliary variable
- $\overline{\mathrm{X}}, \overline{\mathrm{Y}}$ - Population means
- $\overline{\mathrm{x}}, \overline{\mathrm{y}}$ - Sample means
- ρ - Correlation coefficient between X and Y
- β - Regression coefficient of Y on X
- $\overline{\mathrm{M}}$ - Average of sample medians of Y
- m - Sample median of Y
- $\mathrm{B}(.)$ - Bias of the estimator
- $\mathrm{V}(.)$ - Variance of the estimator
- $\mathrm{MSE}(.)$ - Mean squared error of the estimator
- $\mathrm{PRE}(\mathrm{e},\mathrm{p}) = \frac{\mathrm{MSE}(\mathrm{e})}{\mathrm{MSE}(\mathrm{p})} * 100$ – Percentage relative efficiency of the proposed estimator(p) with respect to the existing estimator (e)

The formulae for computing various measures including the variance and the covariance of the SRSWOR sample mean and sample median are as follows:

$$V(\bar{y}) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(\bar{y}_i - \bar{Y})^2 = \frac{1-f}{n}S_y^2\ , V(\bar{x}) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(\bar{x}_i - \bar{X})^2 = \frac{1-f}{n}S_x^2,$$

$$MSE(m) = V(m) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(m_i - M)^2,$$

$$Cov(\bar{y},\bar{x}) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(\bar{x}_i - \bar{X})\,(\bar{y}_i - \bar{Y}) = \frac{1-f}{n}\frac{1}{N-1}\sum_{i=1}^{N}(\bar{x}_i - \bar{X})\,(\bar{y}_i - \bar{Y})\ ,$$

$$Cov(\bar{y},m) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(m_i - M)\,(\bar{y}_i - \bar{Y}) \text{ where } f = \frac{n}{N};\ S_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y})^2,$$

$$S_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(X_i - \bar{X})^2,\ C'_{xx} = \frac{V(\bar{x})}{\bar{X}^2}, C'_{mm} = \frac{V(m)}{M^2}, C'_{ym} = \frac{Cov(\bar{y},m)}{M\bar{Y}}, C'_{yx} = \frac{Cov(\bar{y},\bar{x})}{\bar{X}\bar{Y}}$$

If there is no auxiliary variable available, the simplest estimator of population mean $\bar{Y}$ is the sample mean $\bar{y}$ of size n drawn by using simple random sampling without replacement. The variance of the SRSWOR sample mean $\hat{\bar{Y}}$=$\hat{\bar{Y}}_r$ is given by

$$V(\hat{\bar{Y}}_r) = \frac{1-f}{n}S_y^2 \tag{1}$$

The ratio estimator for estimating the population mean $\bar{Y}$ of the study variable Y is defined as $\hat{\bar{Y}}_R = \frac{\bar{y}}{\bar{x}}\bar{X} = \hat{R}\bar{X}$. The bias and the mean squared error are as given below:

$$B\left(\hat{\bar{Y}}_R\right) = \bar{Y}\{C'_{xx} - C'_{yx}\} \tag{2}$$

$$MSE\left(\hat{\bar{Y}}_R\right) = V(\bar{y}) + R^2V(\bar{x}) - 2RCov(\bar{y},\bar{x}), \text{where } R = \frac{\bar{Y}}{\bar{X}} \tag{3}$$

The other important and optimum estimator for estimating the population mean $\bar{Y}$ of the study variable Y using the auxiliary information is the linear regression estimator. The linear regression estimator and its variance are given below:

$$\hat{\bar{Y}}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x}) \tag{4}$$

$$V\left(\hat{\bar{Y}}_{lr}\right) = V(\bar{y})(1 - \rho^2) \text{ where } \rho = \frac{Cov(\bar{y},\bar{x})}{\sqrt{V(\bar{x})*V(\bar{y})}} \tag{5}$$

The ratio estimator has been extended and introduced several modified ratio estimators in the literature if the population parameters like coefficient of variation, skewness, kurtosis, correlation coefficient, quartiles, etc. of the auxiliary variable are known. For further details on the modified ratio estimators with the known population parameters of the auxiliary variable such as coefficient of variation, skewness, kurtosis, correlation coefficient, quartiles and their linear combinations, the readers are referred to see the following papers: Kadilar and Cingi (2004, 2006a,b, 2009), Koyuncu and Kadilar (2009), Singh and Kakran (1993), Singh and Tailor (2003, 2005), Singh (2003), Sisodia and Dwivedi (1981), Subramani (2013), Subramani and Kumarapandiyan (2012a,b,c, 2013a,b), Tailor and Sharma (2009), Tin (1965), Yan and Tian (2010) and the references cited therein.

In general it has been established that the ratio estimator and the linear regression estimator performs better than the SRSWOR sample mean under certain conditions, of which that provided the auxiliary information $\overline{X}$ is known. It is to be noted that both the ratio and regression estimators use the population mean of the auxiliary variable as auxiliary information. In the absence of the auxiliary variable X, these estimators are not possible. This point is motivated to look for an alternative method for this problem. Further, it is observed from various discussions and studies that the median of the study variable may be known well in advance in several situations. Hence, an attempt is made in this paper to propose a ratio estimator making use of the population median of the study variable as auxiliary information and as a result a new ratio estimator, namely median based ratio estimator has been proposed. The performance of the median based ratio estimator has been compared with that of the SRSWOR sample mean, ratio estimator and linear regression estimator for certain natural populations. It is shown that the proposed median based ratio estimator outperformed not only the SRSWOR sample mean, ratio estimator but also the linear regression estimator. The proposed median based ratio estimator together with the bias and mean squared error are given in section 2.

2. Proposed median based ratio estimator

In this section a new median based ratio estimator for estimating population mean $\overline{Y}$ has been proposed if the median M of the study variable *Y* is known. The proposed median based ratio estimator together with the bias and mean squared error are given below:

$$\widehat{\overline{Y}}_M = \overline{y}\left[\frac{M}{m}\right] \tag{6}$$

$$B\left(\widehat{\overline{Y}}_M\right) = \overline{Y}\left\{C'_{mm} - C'_{ym} - \frac{Bias(m)}{M}\right\} \tag{7}$$

$$MSE\left(\widehat{\overline{Y}}_M\right) = V(\overline{y}) + R'^2 V(m) - 2R'Cov(\overline{y}, m) \text{ where } R' = \frac{\overline{Y}}{M} \tag{8}$$

The detailed derivations of the bias and the mean squared error are given below:

Consider $\widehat{\overline{Y}}_M = \overline{y}\left[\frac{M}{m}\right]$

Let $e_0 = \frac{\overline{y}-\overline{Y}}{\overline{Y}}$ and $e_1 = \frac{m-M}{M}$

$$\Longrightarrow \mathrm{E}(\mathrm{e}_0) = 0$$
$$\Longrightarrow \mathrm{E}(\mathrm{e}_1) = \frac{\bar{\mathrm{M}}-\mathrm{M}}{\mathrm{M}} = \frac{\mathrm{Bias(m)}}{\mathrm{M}}$$
$$\Longrightarrow \mathrm{E}(\mathrm{e}_0^2) = \frac{\mathrm{V}(\bar{\mathrm{y}})}{\bar{\mathrm{Y}}^2}; \mathrm{E}(\mathrm{e}_1^2) = \frac{\mathrm{V(m)}}{\mathrm{M}^2}; \mathrm{E}(\mathrm{e}_0\mathrm{e}_1) = \frac{\mathrm{Cov}(\bar{\mathrm{y}},\mathrm{m})}{\bar{\mathrm{Y}}\mathrm{M}}$$

The estimator $\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}$ can be written in terms of e_0 and e_1 as

$$\hat{\bar{\mathrm{Y}}}_{\mathrm{M}} = \bar{\mathrm{Y}}(1+\mathrm{e}_0)\left(\frac{\mathrm{M}}{\mathrm{M}(1+\mathrm{e}_1)}\right)$$
$$\Longrightarrow \hat{\bar{\mathrm{Y}}}_{\mathrm{M}} = \bar{\mathrm{Y}}(1+\mathrm{e}_0)\left(\frac{1}{1+\mathrm{e}_1}\right)$$
$$\Longrightarrow \hat{\bar{\mathrm{Y}}}_{\mathrm{M}} = \bar{\mathrm{Y}}(1+\mathrm{e}_0)(1+\mathrm{e}_1)^{-1}$$

It is worth to mention that the computation of bias and mean squared error is based on the Taylor series of the function $1/(1+x)$, rounded up to second order term and neglecting the terms $-2(\bar{Y}e_0 - \bar{Y}e_1)(\bar{Y}e_0e_1 - \bar{Y}e_1^2)$ and $(\bar{Y}e_0e_1 - \bar{Y}e_1^2)^2$.

We have

$$\hat{\bar{\mathrm{Y}}}_{\mathrm{M}} = \bar{\mathrm{Y}}(1+\mathrm{e}_0)(1-\mathrm{e}_1+\mathrm{e}_1^2)$$
$$\Longrightarrow \hat{\bar{\mathrm{Y}}}_{\mathrm{M}} = \bar{\mathrm{Y}} + \bar{\mathrm{Y}}\mathrm{e}_0 - \bar{\mathrm{Y}}\mathrm{e}_1 - \bar{\mathrm{Y}}\mathrm{e}_0\mathrm{e}_1 + \bar{\mathrm{Y}}\mathrm{e}_1^2$$
$$\Longrightarrow \hat{\bar{\mathrm{Y}}}_{\mathrm{M}} - \bar{\mathrm{Y}} = \bar{\mathrm{Y}}\mathrm{e}_0 - \bar{\mathrm{Y}}\mathrm{e}_1 - \bar{\mathrm{Y}}\mathrm{e}_0\mathrm{e}_1 + \bar{\mathrm{Y}}\mathrm{e}_1^2$$

Taking expectations on both sides we have

$$\mathrm{E}(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}} - \bar{\mathrm{Y}}) = \bar{\mathrm{Y}}\mathrm{E}(\mathrm{e}_0) - \bar{\mathrm{Y}}\mathrm{E}(\mathrm{e}_1) - \bar{\mathrm{Y}}\mathrm{E}(\mathrm{e}_0\mathrm{e}_1) + \bar{\mathrm{Y}}\mathrm{E}(\mathrm{e}_1^2)$$
$$\Longrightarrow \mathrm{E}\left(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}} - \bar{\mathrm{Y}}\right) = \bar{\mathrm{Y}}\{\mathrm{E}(\mathrm{e}_1^2) - \mathrm{E}(\mathrm{e}_0\mathrm{e}_1) - \mathrm{E}(\mathrm{e}_1)\}$$
$$\Longrightarrow \mathrm{Bias}(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}) = \bar{\mathrm{Y}}\left\{\frac{\mathrm{V(m)}}{\mathrm{M}^2} - \frac{\mathrm{Cov}(\bar{\mathrm{y}},\mathrm{m})}{\bar{\mathrm{Y}}\mathrm{M}} - \frac{\bar{\mathrm{M}}-\mathrm{M}}{\mathrm{M}}\right\}$$
$$\Longrightarrow \mathrm{Bias}(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}) = \bar{\mathrm{Y}}\left\{\mathrm{C}'_{\mathrm{mm}} - \mathrm{C}'_{\mathrm{ym}} - \frac{\mathrm{Bias(m)}}{\mathrm{M}}\right\}$$

The mean squared error of $\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}$ is obtained as given below:

$$\mathrm{MSE}\left(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}\right) = \mathrm{E}(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}} - \bar{\mathrm{Y}})^2 = \mathrm{E}(\bar{\mathrm{Y}}\mathrm{e}_0 - \bar{\mathrm{Y}}\mathrm{e}_1)^2$$
$$\Longrightarrow \mathrm{MSE}\left(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}\right) = \bar{\mathrm{Y}}^2\{\mathrm{E}(\mathrm{e}_0^2) + \mathrm{E}(\mathrm{e}_1^2) - 2\mathrm{E}(\mathrm{e}_0\mathrm{e}_1)\}$$
$$\Longrightarrow \mathrm{MSE}\left(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}\right) = \bar{\mathrm{Y}}^2\left\{\frac{\mathrm{V}(\bar{\mathrm{y}})}{\bar{\mathrm{Y}}^2} + \frac{\mathrm{V(m)}}{\mathrm{M}^2} - 2\frac{\mathrm{Cov}(\bar{\mathrm{y}},\mathrm{m})}{\bar{\mathrm{Y}}\mathrm{M}}\right\}$$
$$\Longrightarrow \mathrm{MSE}\left(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}\right) = \mathrm{V}(\bar{\mathrm{y}}) + \frac{\bar{\mathrm{Y}}^2}{\mathrm{M}^2}\mathrm{V(m)} - 2\frac{\bar{\mathrm{Y}}}{\mathrm{M}}\mathrm{Cov}(\bar{\mathrm{y}},\mathrm{m})$$
$$\Longrightarrow \mathrm{MSE}\left(\hat{\bar{\mathrm{Y}}}_{\mathrm{M}}\right) = \mathrm{V}(\bar{\mathrm{y}}) + \mathrm{R}'^2\mathrm{V(m)} - 2\mathrm{R}'\mathrm{Cov}(\bar{\mathrm{y}},\mathrm{m})$$

Remark 2.1. When $\overline{M} > M$, the use of the average of the sample medians $\overline{M}$ in the median based ratio estimator in the place of population median M reduces the variance of the proposed estimator due to the following reasons:

1. In general, the sample median m is not an unbiased estimator for the population median M, which leads the mean squared error of m and the value is larger than the variance of m.
2. The covariance $Cov(\bar{y}, m) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(m_i - \overline{M})(\bar{y}_i - \overline{Y})$ based on $\overline{M}$ and the covariance $Cov(\bar{y}, m) = \frac{1}{N_{C_n}}\sum_{i=1}^{N_{C_n}}(m_i - M)(\bar{y}_i - \overline{Y})$ based on the population median M are the same.

Remark 2.2. When $n = 2$ the sample mean becomes the sample median, i.e. $\bar{y} = m$, the mean squared error of $\hat{\overline{Y}}_M = MSE\left(\hat{\overline{Y}}_M\right) = 0$ and hence the SRSWOR becomes the trend free sampling (see Mukerjee and Sengupta (1990)).

3. Efficiency comparison

In this section we have derived the algebraic conditions for which the proposed median based ratio estimator will have minimum mean squared error compared to the SRSWOR sample mean, ratio estimator and linear regression estimator for estimating the finite population mean.

3.1. Comparison with that of SRSWOR sample mean

From the expressions given in (8) and (1), the conditions for which the proposed estimator $\hat{\overline{Y}}_M$ is more efficient than the existing estimator $\hat{\overline{Y}}_r$ are derived and given below:

$$MSE\left(\hat{\overline{Y}}_M\right) \leq V\left(\hat{\overline{Y}}_r\right), \text{if } 2C'_{ym} \geq C'_{mm} \tag{9}$$

The detailed derivation is given below:

Consider $MSE\left(\hat{\overline{Y}}_M\right) \leq V\left(\hat{\overline{Y}}_r\right)$

$$\Rightarrow V(\bar{y}) + R'^2\, V(m) - 2R'\, Cov(\bar{y}, m) \leq V(\bar{y})$$

$$\Rightarrow R'^2\, V(m) - 2R'\, Cov(\bar{y}, m) \leq 0$$

$$\Rightarrow R'^2 V(m) \leq 2R'\, Cov(\bar{y}, m)$$

$$\Rightarrow Cov(\bar{y}, m) \geq \frac{R'\, V(m)}{2}$$

$$\Rightarrow Cov(\bar{y}, m) \geq \frac{\overline{Y} M C'_{mm}}{2}$$

$$\Rightarrow 2C'_{ym} - C'_{mm} \geq 0$$

3.2. Comparison with that of Ratio Estimator

From the expressions given in (8) and (3) the conditions for which the proposed estimator $\hat{\hat{\bar{Y}}}_M$ is more efficient than the usual ratio estimator $\hat{\hat{\bar{Y}}}_R$ are derived and given below:

$$MSE\left(\hat{\hat{\bar{Y}}}_M\right) \leq MSE\left(\hat{\hat{\bar{Y}}}_R\right), \text{if } C'_{mm} - C'_{xx} \leq 2\left\{C'_{ym} - C'_{xy}\right\} \quad (10)$$

The detailed derivation is given below:

Consider $MSE(\hat{\hat{\bar{Y}}}_M) \leq MSE(\hat{\hat{\bar{Y}}}_R)$

$$\Rightarrow V(\bar{y}) + R'^2 V(m) - 2R' Cov(\bar{y}, m) \leq V(\bar{y}) + R^2 V(\bar{x}) - 2R Cov(\bar{y}, \bar{x})$$

$$\Rightarrow R'^2 V(m) - 2R' Cov(\bar{y}, m) \leq R^2 V(\bar{x}) - 2R Cov(\bar{y}, \bar{x})$$

$$\Rightarrow R'^2 V(m) - R^2 V(\bar{x}) \leq 2R' Cov(\bar{y}, m) - 2R Cov(\bar{y}, \bar{x})$$

$$\Rightarrow \frac{\bar{Y}^2}{M^2} V(m) - \frac{\bar{Y}^2}{\bar{X}^2} V(\bar{x}) \leq 2\frac{\bar{Y}}{M} Cov(\bar{y}, m) - 2\frac{\bar{Y}}{\bar{X}} Cov(\bar{y}, \bar{x})$$

$$\Rightarrow \frac{V(m)}{M^2} - \frac{V(\bar{x})}{\bar{X}^2} \leq 2\left\{\frac{Cov(\bar{y}, m)}{\bar{Y}M} - \frac{Cov(\bar{y}, \bar{x})}{\bar{Y}\bar{X}}\right\}$$

$$\Rightarrow 2C'_{ym} - C'_{mm} \geq 2C'_{xy} - C'_{xx}$$

3.3. Comparison with that of Linear Regression Estimator

From the expressions given in (8) and (5) the conditions for which the proposed estimator $\hat{\hat{\bar{Y}}}_M$ is more efficient than the usual linear regression estimator $\hat{\hat{\bar{Y}}}_{lr}$ are derived and given below:

$$MSE\left(\hat{\hat{\bar{Y}}}_M\right) \leq V\left(\hat{\hat{\bar{Y}}}_{lr}\right), \text{if } 2\,C'_{ym} - C'_{mm} \geq \frac{[C'_{yx}]^2}{C'_{xx}} \quad (11)$$

The detailed derivation is given below:

Consider $MSE(\hat{\hat{\bar{Y}}}_M) \leq V(\hat{\hat{\bar{Y}}}_{lr})$

$$\Rightarrow V(\bar{y}) + R'^2 V(m) - 2R' Cov(\bar{y}, m) \leq V(\bar{y})(1 - \rho^2)$$

$$\Rightarrow R'^2 V(m) - 2R' Cov(\bar{y}, m) \leq -V(\bar{y})\left(\frac{[Cov(\bar{y}, \bar{x})]^2}{V(\bar{x}) * V(\bar{y})}\right)$$

$$\Rightarrow 2R' Cov(\bar{y}, m) - R'^2 V(m) \geq \frac{[Cov(\bar{y}, \bar{x})]^2}{V(\bar{x})}$$

$$\Rightarrow 2\frac{\bar{Y}}{M} Cov(\bar{y}, m) - \frac{\bar{Y}^2}{M^2} V(m) \geq \frac{[Cov(\bar{y}, \bar{x})]^2}{V(\bar{x})}$$

$$\Rightarrow 2\bar{Y}^2 C'_{ym} - \bar{Y}^2 C'_{mm} \geq \frac{[Cov(\bar{y}, \bar{x})]^2}{V(\bar{x})}$$

$$\Rightarrow 2\,C'_{ym} - C'_{mm} \geq \frac{[C'_{yx}]^2}{C'_{xx}}$$

Remark 3.1. It is well known that the ratio estimator is more efficient than the SRSWOR sample mean if $2C'_{xy} - C'_{xx} \geq 0$. Similar results hold good here. That is, the median based ratio estimator is more efficient than

(i) SRSWOR sample mean if $2C'_{ym} - C'_{mm} \geq 0$

(ii) Ratio estimator if $2C'_{ym} - C'_{mm} \geq 2C'_{xy} - C'_{xx}$

(iii) Linear regression estimator if $2C'_{ym} - C'_{mm} \geq \frac{[C'_{yx}]^2}{C'_{xx}}$

4. Numerical study

In section 3, the conditions are derived for which the proposed median based ratio estimator performed better than the SRSWOR sample mean, ratio estimator and linear regression estimator. However, it has not been proved explicitly by algebraic expressions that the proposed estimators are better than the estimators mentioned above. Alternatively one has to resort for numerical comparisons to determine the efficiencies of the proposed estimators. In this view, three natural populations available in the literature are used for comparing the efficiencies of the proposed median based ratio estimators with that of the existing estimators. The population 1 and 2 are taken from Daroga Singh and Chaudhary (1986, page no. 177) and population 3 is taken from Mukhopadhyay (2005, page no. 96). The populations 1 and 2 pertain to estimate the area of cultivation under wheat in the year 1974, whereas the auxiliary information is the cultivated areas under wheat in 1971 and 1973 respectively. The population 3 pertains to the number of labourers (the auxiliary variable, in thousand) and the quantity of raw materials (study variable, in lakhs of bales) for 20 jute mills. The parameter values and constants computed for the above populations are presented in Table 5; the bias for the proposed modified ratio estimators and the existing estimators computed for the three populations discussed above are presented in Table 6, whereas the mean squared errors are presented in Table 7. It is worth noting that since the computation of mean squared error is involved in computing all possible N_{C_n} samples of size n, the computation requires strong capacity, which, for very large samples, exceeds the possibilities of the computation facilities. Hence only small sample sizes are considered. However, the same results will hold good for large sample sizes too.

Table 5. Parameter values and constants computed from the 3 populations

Para-meters	For sample size $n = 3$			For sample size $n = 5$		
	Popln-1	**Popln-2**	**Popln-3**	**Popln-1**	**Popln-2**	**Popln-3**
N	34	34	20	34	34	20
n	3	3	3	5	5	5
N_{C_n}	5984	5984	1140	278256	278256	15504
$\bar{Y}$	856.4118	856.4118	41.5	856.4118	856.4118	41.5

Table 5. Parameter values and constants computed from the 3 populations (cont.)

Para-meters	For sample size $n = 3$			For sample size $n = 5$		
	Popln-1	**Popln-2**	**Popln-3**	**Popln-1**	**Popln-2**	**Popln-3**
$\bar{M}$	747.7223	747.7223	40.2351	736.9811	736.9811	40.0552
M	767.5	767.5	40.5	767.5	767.5	40.5
$\bar{X}$	208.8824	199.4412	441.95	208.8824	199.4412	441.95
R	4.0999	4.2941	0.0939	4.0999	4.2941	0.0939
R'	1.1158	1.1158	1.0247	1.1158	1.1158	1.0247
$V(\bar{y})$	163356.4086	163356.4086	27.1254	91690.3713	91690.3713	14.3605
$V(\bar{x})$	6884.4455	6857.8555	2894.3089	3864.1726	3849.248	1532.2812
$V(m)$	101518.7738	101518.7738	26.1307	59396.2836	59396.2836	10.8348
$Cov(\bar{y}, m)$	90236.2939	90236.2939	21.0918	48074.9542	48074.9542	9.0665
$Cov(\bar{y}, \bar{x})$	15061.4011	14905.0488	182.7425	8453.8187	8366.0597	96.7461
ρ	0.4491	0.4453	0.6522	0.4491	0.4453	0.6522

Table 6. Bias of the existing and proposed estimators

Estimators	For sample size $n = 3$			For sample size $n = 5$		
	Popln-1	**Popln-2**	**Popln-3**	**Popln-1**	**Popln-2**	**Popln-3**
$\hat{\bar{Y}}_R$	63.0241	72.9186	0.2015	35.3748	40.9285	0.1067
$\hat{\bar{Y}}_M$	52.0924	52.0924	0.4118	57.7705	57.7705	0.5061

Table 7. Variance / Mean squared error of the existing and proposed estimators

Estimators		For sample size $n = 3$			For sample size $n = 5$		
		Popln-1	**Popln-2**	**Popln-3**	**Popln-1**	**Popln-2**	**Popln-3**
Existing	$\hat{\bar{Y}}_r$	163356.4086	163356.4086	27.1254	91690.3713	91690.3713	14.3605
	$\hat{\bar{Y}}_R$	155579.7064	161801.6355	18.3265	87325.3836	90817.6922	9.7023
	$\hat{\bar{Y}}_{lr}$	130405.9256	130961.3720	15.5873	73195.5841	73508.8959	8.2521
Proposed	$\hat{\bar{Y}}_M$	88379.0666	88379.0666	11.3372	58356.9234	58356.9234	7.1563

The percentage relative efficiencies of the proposed estimator with respect to the existing estimators are also obtained and given in the following table:

Table 8. PRE of the proposed estimator $\hat{\bar{Y}}_M$ with respect to existing estimators

Existing Estimators	For sample size $n = 3$			For sample size $n = 5$		
	Popln-1	Popln-2	Popln-3	Popln-1	Popln-2	Popln-3
$\hat{\bar{Y}}_r$	184.84	184.84	239.26	157.12	157.12	200.67
$\hat{\bar{Y}}_R$	176.04	183.08	161.65	149.64	155.62	135.58
$\hat{\bar{Y}}_{lr}$	147.55	148.18	137.49	125.43	125.96	115.31

From the Table 8 it is observed that the percentage relative efficiencies of the proposed estimator with respect to existing estimators are in general ranging from 115.31 to 239.26. Particularly, the PRE is ranging from 157.12 to 239.26 for comparison with the SRSWOR sample mean; ranging from 135.58 to 183.08 for comparison with ratio estimator; ranging from 115.31 to 148.18 for comparison with linear regression estimator. This shows that the proposed estimator performs better than the existing SRSWOR sample mean, ratio and linear regression estimator for all the three populations considered here. Further, it is observed from the numerical comparisons that the following inequalities hold good.

$$\mathrm{MSE}(\hat{\bar{Y}}_M) \leq \mathrm{V}(\hat{\bar{Y}}_{lr}) \leq \mathrm{MSE}\left(\hat{\bar{Y}}_R\right) \leq \mathrm{V}\left(\hat{\bar{Y}}_r\right)$$

5. Conclusion

This paper deals with a new median based ratio estimator for estimation of the finite population mean. The conditions are derived for which the proposed estimator is more efficient than the existing estimators. Further, it is shown that the percentage relative efficiencies of the proposed estimators with respect to existing estimators are in general ranging from 115.31 to 239.26 for certain natural populations available in the literature. It is usually believed that the linear regression estimator is the best linear unbiased estimator or the optimum estimator for estimating the population mean whenever there exists an auxiliary variable, which is positively correlated with that of the study variable. However, it is shown that the proposed median based ratio estimator outperformed not only the SRSWOR sample mean, ratio estimator but also the linear regression estimator. Hence, the proposed modified ratio estimators are recommended for the practical applications. Further, it is to be noted that the median based ratio estimator can be easily extended to median based modified ratio estimators in line with the modified ratio estimators available in the literature and one of the author's research students Prabavathy, G is working at present in this direction.

Acknowledgement

The author is very thankful to the Reviewers and the Editor for their constructive comments helpful in improving the presentation of this paper.

REFERENCES

COCHRAN, W. G., (1977). Sampling techniques, Third Edition, Wiley Eastern Limited, USA.

KADILAR, C., CINGI, H., (2004). Ratio estimators in simple random sampling, Applied Mathematics and Computation, 151, pp. 893–902.

KADILAR, C., CINGI, H., (2006a). An improvement in estimating the population mean by using the correlation co-efficient, Hacettepe Journal of Mathematics and Statistics, 35 (1), pp. 103–109.

KADILAR, C., CINGI, H., (2006b). Improvement in estimating the population mean in simple random sampling, Applied Mathematics Letters, 19, pp. 75–79.

KADILAR, C., CINGI, H., (2009). Advances in sampling theory - Ratio method of estimation, Bentham Science Publishers.

KOYUNCU, N., KADILAR, C., (2009). Efficient estimators for the population mean, Hacettepe Journal of Mathematics and Statistics, 38(2), pp. 217–225.

MUKERJEE, R., SENGUPTA, S., (1990). Optimal estimation of a finite population mean in the presence of linear trend, Biometrika, 77, pp. 625–630.

MUKHOPADHYAY, P., (2005). Theory and methods of survey sampling, PHI Learning, 2nd edition, New Delhi.

MURTHY, M. N., (1967). Sampling theory and methods, Statistical Publishing Society, Calcutta, India.

SINGH, G.N., (2003). On the improvement of product method of estimation in sample surveys, Journal of Indian Society of Agricultural Statistics, 56 (3), pp. 267–265.

SINGH, D., CHAUDHARY, F. S., (1986). Theory and analysis of sample survey designs, New Age International Publisher, New Delhi.

SINGH, H. P., KAKRAN, M., (1993). A modified ratio estimator using known coefficient of kurtosis of an auxiliary character, Revised version submitted to Journal of Indian Society of Agricultural Statistics.

SINGH, H. P., TAILOR, R., (2003). Use of known correlation coefficient in estimating the finite population means, Statistics in Transition, 6 (4), pp. 555–560.

SINGH, H. P., TAILOR, R., (2005). Estimation of finite population mean with known co-efficient of variation of an auxiliary variable, Statistica, anno LXV, 3, pp. 301–313.

SISODIA, B. V. S., DWIVEDI, V. K., (1981). A modified ratio estimator using co-efficient of variation of auxiliary variable, Journal of the Indian Society of Agricultural Statistics, 33 (2), pp. 13–18.

SUBRAMANI, J., (2013). Generalized modified ratio estimator of finite population mean, Journal of Modern Applied Statistical Methods, 12 (2), pp. 121–155.

SUBRAMANI, J., KUMARAPANDIYAN, G., (2012a). Estimation of population mean using coefficient of variation and median of an auxiliary variable, International Journal of Probability and Statistics, 1 (4), pp. 111–118.

SUBRAMANI, J., KUMARAPANDIYAN, G., (2012b). Modified ratio estimators using known median and coefficient of kurtosis, American Journal of Mathematics and Statistics, 2 (4), pp. 95–100.

SUBRAMANI, J., KUMARAPANDIYAN, G., (2012c). Estimation of population mean using known median and coefficient of skewness, American Journal of Mathematics and Statistics, 2 (5), pp. 101–107.

SUBRAMANI, J., KUMARAPANDIYAN, G., (2013a). Estimation of population mean using deciles of an auxiliary variable, Statistics in Transition-New Series, 14 (1), pp. 75–88.

SUBRAMANI, J., KUMARAPANDIYAN, G., (2013b). A new modified ratio estimator of population mean when median of the auxiliary variable is known, Pakistan Journal of Statistics and Operation Research, Vol. 9 (2), pp. 137–145.

TAILOR, R., SHARMA, B., (2009). A modified ratio-cum-product estimator of finite population mean using known coefficient of variation and coefficient of kurtosis, Statistics in Transition-New Series, 10 (1), pp. 15–24.

TIN, M., (1965). Comparison of some ratio estimators, Journal of the American Statistical Association, 60, pp. 294–307.

YAN, Z., TIAN, B., (2010). Ratio method to the mean estimation using coefficient of skewness of auxiliary variable, ICICA 2010, Part II, CCIS 106, pp. 103–11.