Scientia Iranica E (2019) 26(4), 2484–2494

SCIENTIA IRANICA

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

Improved ratio estimators of variance based on robust measures

M. Abid[a,*], S. Ahmad[b], M. Tahir[a], H. Zafar Nazir[c], and M. Riaz[d]

a. *Department of Statistics, Government College University, Faisalabad, 38000, Pakistan.*
b. *Department of Mathematics, COMSATS Institute of Information Technology, Wah Cantt, 47040, Pakistan.*
c. *Department of Statistics, University of Sargodha, Sargodha, 40100, Pakistan.*
d. *Department of Mathematics and Statistics, King Fahad University of Petroleum and Minerals, Dhahran, 31261, Saudi, Arabia.*

Received 31 July 2017; received in revised form 27 February 2018; accepted 23 July 2018

KEYWORDS
Auxiliary variable;
Bias;
Correlation coefficient;
Inter decile range;
Mean squared error;
Midrange.

Abstract. This study developed some new estimators for estimating population variance by utilizing available information on midrange and an interdecile range of auxiliary variables. A general class of estimators was also suggested. The derivations of the bias and the mean squared error were presented. Conditions were determined to verify the efficiency of the proposed estimators over existing estimators considered in this study. An empirical study was also provided for illustration and verification. Moreover, a robust study was carried out to evaluate the performance of the proposed estimators in comparison to existing estimators in case of extreme values. According to the theoretical and empirical research, it was found that the suggested estimators performed more efficiently than the existing estimators.

1. Introduction

In survey research, if information is available on every unit of population and, also, correlates with the study variable, then such information is called auxiliary information. Utilizing this auxiliary information, one can propose numerous types of estimators for estimating the population variance by incorporating the product, the ratio, and the regression methods of estimation. To the best of our knowledge, Neyman [1] initially considered applying auxiliary information in his study. Then, Cochran [2] initiated the application of auxiliary information in the estimation stage and proposed a ratio estimator for estimating the population mean. The ratio estimator is the most effective one in estimating the population mean or variance when there exists a high positive correlation between the variables of interest and an auxiliary variable.

Suppose that a finite population $U = \{U_1, U_2, U_3, ..., U_N\}$ consists of N different and identifiable units. Let Y be a measurable variable of interest with values Y_i ascertained in U_i, $i = 1, 2, ..., N$, resulting in a set of observations $Y = \{Y_1, Y_2, ..., Y_N\}$. The purpose of the measurement process is to estimate the population variance $S_Y^2 = N^{-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ by drawing a random sample from the population.

The traditional ratio estimator for evaluating the population variance, S_y^2, of the variable of interest Y is

*. *Corresponding author. Tel.: 92 333 6289936*
E-mail addresses: mabid@zju.edu.cn (M. Abid); shabbiarahmad786@yahoo.com (S. Ahmad); tahirqaustat@yahoo.com (M. Tahir); hafizzafarnazir@yahoo.com (H. Zafar Nazir); riaz76qau@yahoo.com (M. Riaz).

doi: 10.24200/sci.2018.20604

described as follows:

$$\hat{S}_r^2 = \frac{s_y^2}{s_x^2} S_x^2. \tag{1}$$

The bias and Mean Squared Error (MSE) of the estimator given in Eq. (1) are respectively as follows:

$$B(\hat{S}_r^2) = \gamma S_y^2[(\beta_{2(x)}-1)-(\lambda_{22}-1)],$$

$$MSE(\hat{S}_r^2) = \gamma S_y^4[(\beta_{2(y)}-1)+(\beta_{2(x)}-1)-2(\lambda_{22}-1)],$$

where:

$$\beta_{2(x)}=\mu_{04}/\mu_{02}^2, \quad \beta_{2(y)}=\mu_{40}/\mu_{20}^2, \quad \lambda_{22}=\mu_{22}/(\mu_{20}\mu_{02}),$$

$$\mu_{rs} = N^{-1}\sum_{i=1}^{N}(Y_i-\bar{Y})^r(X_i-\bar{X})^s.$$

Isaki [3] was the first to propose variance-based ratio and regression estimators by using one auxiliary variable for estimating the population variance. Then, Prasad and Singh [4] suggested new estimators, and showed that their proposed estimators were more capable than those suggested by Isaki [3]. Arcos et al. [5] also proposed some new estimators that performed better than the usual estimators and the other variance ratio estimators considered in this study. Comprehensive details regarding the problem of creating competent estimators for estimating population variance can be found in [6-18].

The remainder of the article is structured as follows. Section 2 provides a detailed explanation of the existing variance estimators. The formation of the recommended estimators and the efficiency assessment of the proposed estimators with the usual estimator and the existing estimators are given in Section 3. Section 4 includes a practical study of the suggested estimators. Robustness study of the proposed estimators is investigated in Section 5. The final comments and conclusion are presented in Section 6.

2. The existing estimators

Upadhyaya and Singh [19] suggested the following estimator for S_y^2 by employing the information on kurtosis of an auxiliary variable. Upadhyaya and Singh [19] proposed the following estimator:

$$\hat{S}_1^2 = s_y^2\left(\frac{S_x^2+\beta_{2(x)}}{s_x^2+\beta_{2(x)}}\right).$$

Kadilar and Cingi [20] proposed some new ratio estimators of variance, and proved that these estimators performed more efficiently than the usual estimator of variance. The estimators proposed by Kadilar and Cingi [20] are specified as in the following:

$$\hat{S}_2^2 = s_y^2\left(\frac{S_x^2+C_x}{s_x^2+C_x}\right), \quad \hat{S}_3^2 = s_y^2\left(\frac{S_x^2+\beta_{2(x)}}{s_x^2+\beta_{2(x)}}\right)$$

$$\hat{S}_4^2 = s_y^2\left(\frac{S_x^2\beta_{2(x)}+C_x}{s_x^2\beta_{2(x)}+C_x}\right),$$

$$\hat{S}_5^2 = s_y^2\left(\frac{S_x^2C_x+\beta_{2(x)}}{s_x^2C_x+\beta_{2(x)}}\right).$$

Subramani and Kumarapandiyan [21] suggested an estimator of variance by adopting the information on the median of an auxiliary variable. The estimator proposed by Subramani and Kumarapandiyan [21] is written as follows:

$$\hat{S}_6^2 = s_y^2\left(\frac{S_x^2+M_d}{s_x^2+M_d}\right).$$

Motivated by Upadhyaya and Singh [19] and Kadilar and Cingi [20], Subramani and Kumarapandiyan [22] developed the following estimators:

$$\hat{S}_7^2 = s_y^2\left(\frac{S_x^2+Q_1}{s_x^2+Q_1}\right), \quad \hat{S}_8^2 = s_y^2\left(\frac{S_x^2+Q_3}{s_x^2+Q_3}\right),$$

$$\hat{S}_9^2 = s_y^2\left(\frac{S_x^2+Q_r}{s_x^2+Q_r}\right), \quad \hat{S}_{10}^2 = s_y^2\left(\frac{S_x^2+Q_d}{s_x^2+Q_d}\right), \text{and}$$

$$\hat{S}_{11}^2 = s_y^2\left(\frac{S_x^2+Q_a}{s_x^2+Q_a}\right).$$

Making use of the information related to the deciles of an auxiliary variable, Subramani and Kumarapandiyan [23] introduced the following estimators for estimating the population variance:

$$\hat{S}_{12}^2 = s_y^2\left(\frac{S_x^2+D_1}{s_x^2+D_1}\right), \quad \hat{S}_{13}^2 = s_y^2\left(\frac{S_x^2+D_2}{s_x^2+D_2}\right)$$

$$\hat{S}_{14}^2 = s_y^2\left(\frac{S_x^2+D_3}{s_x^2+D_3}\right), \quad \hat{S}_{15}^2 = s_y^2\left(\frac{S_x^2+D_4}{s_x^2+D_4}\right),$$

$$\hat{S}_{16}^2 = s_y^2\left(\frac{S_x^2+D_5}{s_x^2+D_5}\right), \quad \hat{S}_{17}^2 = s_y^2\left(\frac{S_x^2+D_6}{s_x^2+D_6}\right),$$

$$\hat{S}_{18}^2 = s_y^2\left(\frac{S_x^2+D_7}{s_x^2+D_7}\right), \quad \hat{S}_{19}^2 = s_y^2\left(\frac{S_x^2+D_8}{s_x^2+D_8}\right),$$

$$\hat{S}_{20}^2 = s_y^2\left(\frac{S_x^2+D_9}{s_x^2+D_9}\right), \quad \hat{S}_{21}^2 = s_y^2\left(\frac{S_x^2+D_{10}}{s_x^2+D_{10}}\right).$$

Subramani and Kumarapandiyan [24] also suggested an estimator based on the median and coefficient of variation of an auxiliary variable. The estimator of Subramani and Kumarapandiyan [24] is given as follows:

$$\hat{S}^2_{22} = s^2_y \left(\frac{S^2_x C_x + M_d}{s^2_x C_x + M_d} \right).$$

Khan and Shabbir [25] suggested an estimator as follows:

$$\hat{S}^2_{23} = s^2_y \left(\frac{S^2_x \rho + Q_3}{s^2_x \rho + Q_3} \right).$$

3. The suggested estimator

Encouraged by the work stated in Section 2, some new ratio estimators of variance through the information on the midrange and an interdecile range of auxiliary variables are recommended. The midrange (MR) is found as $MR = \frac{X_{(1)}+X_{(N)}}{2}$, where $X_{(1)}$ and $X_{(N)}$ are the smallest and largest order statistics in a population of size N. Ferrell [26] verified that since the midrange is based on only extreme values of data, this measure is sensitive to outliers. The next measure included in this study is the interdecile range (IDR), which is the difference between the 9th and 1st deciles of an auxiliary variable and is defined as:

$$\mathrm{IDR} = \mathrm{D}_9 - \mathrm{D}_1.$$

The major edge of IDR is its robustness against outliers (cf., [27-31]). The proposed estimators can be specified as follows:

$$\hat{S}^2_{p1} = s^2_y \left(\frac{S^2_x + MR}{s^2_x + MR} \right), \quad \hat{S}^2_{p2} = s^2_y \left(\frac{S^2_x + IDR}{s^2_x + IDR} \right),$$

$$\hat{S}^2_{p3} = s^2_y \left(\frac{S^2_x \rho + MR}{s^2_x \rho + MR} \right), \quad \hat{S}^2_{p4} = s^2_y \left(\frac{S^2_x \rho + IDR}{s^2_x \rho + IDR} \right),$$

$$\hat{S}^2_{p5} = s^2_y \left(\frac{S^2_x C_x + MR}{s^2_x C_x + MR} \right), \quad \hat{S}^2_{p6} = s^2_y \left(\frac{S^2_x C_x + \mathrm{IDR}}{s^2_x C_x + IDR} \right).$$

The proposed and the existing estimators also belong to the following general class of estimators for S^2_y defined as follows:

$$\hat{S}^2_i = s^2_y \left(\frac{k_1 S^2_x + k_2}{k_1 s^2_x + k_2} \right),$$

where k_1, and k_2 are either constant or functions of known parameters of the population. The bias and the Mean Squared Error (MSE) of $\hat{S}^2_i$ can be obtained as follows:

$$e_0 = \frac{(s^2_y - S^2_y)}{(S^2_y)}, \quad e_1 = \frac{(s^2_x - S^2_x)}{(S^2_x)}.$$

Further:

$$s^2_y = S^2_y (1 + e_0), \qquad s^2_x = S^2_x (1 + e_1).$$

Based on the definition of e_0 and e_1, we obtain:

$$E(e_0) = E(e_1) = 0, \qquad E(e^2_0) = \gamma(\beta_{2(y)} - 1),$$

$$E(e^2_1) = \gamma(\beta_{2(x)} - 1), \qquad E(e_0 e_1) = \gamma(\lambda_{22} - 1).$$

The bias of the proposed estimators, $\hat{S}^2_i$, $i = p1, p2, ..., p6$, is derived as follows:

$$\hat{S}^2_i = s^2_y \left(\frac{k_1 S^2_x + k_2}{k_1 s^2_x + k_2} \right), \quad i = p1, p2, \ldots, p6.$$

$$\hat{S}^2_i = S^2_y (1 + e_0) \left(\frac{k_1 S^2_x + k_2}{k_1 S^2_x (1 + e_1) + k_2} \right),$$

$$\hat{S}^2_i = S^2_y (1 + e_0) \left(\frac{1}{(1 + R_i e_1)} \right),$$

where:

$$R_i = \left(\frac{k_1 S^2_x}{k_1 S^2_x + k_2} \right),$$

$$\hat{S}^2_i = S^2_y (1 + e_0) (1 + R_i e_1)^{-1}.$$

Assuming $|R_i e_1| < 1$ so that $(1+R_i e_1)^{-1}$ is expandable, expanding the right-hand side of the above equation, and neglecting the terms of e's having power greater than two, we get:

$$\hat{S}^2_i - S^2_y = S^2_y e_0 - S^2_y R_i e_1 + S^2_y R^2_i e^2_1 - S^2_y R_i e_0 e_1. \quad (2)$$

Considering both sides of the above equation, we will obtain:

$$E(\hat{S}^2_i - S^2_y) = S^2_y E(e_0) - S^2_y R_i E(e_1) + S^2_y R^2_i E(e^2_1) - S^2_y R_i E(e_0 e_1).$$

So,

$$B\left(\hat{S}^2_i\right) = \gamma S^2_y R_i \left[R_i \left(\beta_{2(x)} - 1 \right) - (\lambda_{22} - 1) \right].$$

The MSE of the proposed set of estimators, $\hat{S}^2_i$, $i = p1, p2, ..., p6$, is derived as follows:

Squaring both sides of Eq. (2):

$$(\hat{S}^2_i - S^2_y)^2 = (S^2_y e_0 - S^2_y R_i e_1 + S^2_y R^2_i e^2_1 - S^2_y R_i e_0 e_1)^2.$$

Expanding the right-hand side of the above equation, and neglecting the terms of e's having power greater than two, we get:

$$(\hat{S}^2_i - S^2_y)^2 = S^4_y e^2_0 + S^4_y R^2_i e^2_1 - 2 S^4_y R_i e_0 e_1.$$

Considering both sides, we will obtain:

$$E(\hat{S}_i^2 - S_y^2)^2 = \left(S_y^4 E(e_0^2) + S_y^4 R_i^2 E\left(e_1^2\right) - 2S_y^4 R_i E\left(e_0 e_1\right)\right)^2.$$

As is clear, the definition of MSE is:

$$MSE(\hat{S}_i^2) = E(\hat{S}_i^2 - S_y^2)^2.$$

Hence,

$$MSE(\hat{S}_i^2) = \gamma S_y^4[(\beta_{2(y)} - 1) + R_i^2(\beta_{2(x)} - 1) - 2R_i(\lambda_{22} - 1)],$$

where $i = 1, 2, ..., 23$ for existing estimators and $i = p1, p2, ..., p6$ for the proposed estimators and where $R_i = \left(\frac{k_1 S_x^2}{k_1 S_x^2 + k_2}\right)$, $\beta_{2(x)} = \mu_{04}/\mu_{02}^2$, $\beta_{2(y)} = \mu_{40}/\mu_{20}^2$, and $\lambda_{22} = \mu_{22}/(\mu_{20}\mu_{02})$.

The proper selections of k_1 are k_2 are given in Table 1.

3.1. Efficiency comparison

This section presents the conditions in which the proposed estimators perform more capably than the usual and the existing ratio estimators of variance. This study used the notations of $MSE(\hat{S}_{pj}^2)$ and $MSE(\hat{S}_i^2)$ (Table 1) for the proposed and the existing estimators, respectively, for comparison purposes.

3.1.1. Comparison with usual ratio estimator of variance

If the suggested estimators have lower MSE values than those of the usual estimator, then they are more efficient. Mathematically, it is expressed as follows:

$$MSE\left(\hat{S}_{pj}^2\right) < MSE\left(\hat{S}_r^2\right),$$

$$R_{pj}^2\left(\beta_{2(x)} - 1\right) - 2R_{pj}\left(\lambda_{22} - 1\right) \le \left(\beta_{2(x)} - 1\right) - 2\left(\lambda_{22} - 1\right). \quad (3)$$

After solving Eq. (3), we get $R_{pj} < \frac{2(\lambda_{22}-1)}{(\beta_{2(x)}-1)} - 1$. Hence:

$$MSE\left(\hat{S}_{pj}^2\right) < MSE\left(\hat{S}_r^2\right),$$

if:

$$R_{pj} < \frac{2\left(\lambda_{22} - 1\right)}{\left(\beta_{2(x)} - 1\right)} - 1, \quad (4)$$

where $j = 1, 2, ..., 6$.

Table 1. The suitable choices of k_1 and k_2 for the existing and proposed estimators.

Estimators	k_1	k_2
$\hat{S}_1^2$	1	$\beta_{2(x)}$
$\hat{S}_2^2$	1	C_x
$\hat{S}_3^2$	1	$\beta_{2(x)}$
$\hat{S}_4^2$	$\beta_{2(x)}$	C_x
$\hat{S}_5^2$	C_x	$\beta_{2(x)}$
$\hat{S}_6^2$	1	M_d
$\hat{S}_7^2$	1	Q_1
$\hat{S}_8^2$	1	Q_3
$\hat{S}_9^2$	1	Q_r
$\hat{S}_{10}^2$	1	Q_d
$\hat{S}_{11}^2$	1	Q_a
$\hat{S}_{12}^2$	1	D_1
$\hat{S}_{13}^2$	1	D_2
$\hat{S}_{14}^2$	1	D_3
$\hat{S}_{15}^2$	1	D_4
$\hat{S}_{16}^2$	1	D_5
$\hat{S}_{17}^2$	1	D_6
$\hat{S}_{18}^2$	1	D_7
$\hat{S}_{19}^2$	1	D_8
$\hat{S}_{20}^2$	1	D_9
$\hat{S}_{21}^2$	1	D_{10}
$\hat{S}_{22}^2$	C_x	M_d
$\hat{S}_{23}^2$	ρ	Q_s
$\hat{S}_{p1}^2$	1	MR
$\hat{S}_{p2}^2$	1	IDR
$\hat{S}_{p3}^2$	ρ	MR
$\hat{S}_{p4}^2$	ρ	IDR
$\hat{S}_{p5}^2$	C_x	MR
$\hat{S}_{p6}^2$	C_x	IDR

3.1.2. Comparisons with existing ratio estimators of variance

The proposed estimators are considered to perform better if the MSE values of the suggested estimators are lower than those of existing estimators; in addition, they are written in the algebraic form below:

$$MSE\left(\hat{S}_{pj}^2\right) < MSE\left(\hat{S}_i^2\right),$$

$$R_{pj}^2(\beta_{2(x)} - 1) - 2R_{pj}(\lambda_{22} - 1) \le R_i^2(\beta_{2(x)} - 1) - 2R_i(\lambda_{22} - R_i). \quad (5)$$

After solving Eq. (5), we get $R_{pj} < \frac{2(\lambda_{22}-1)}{(\beta_{2(x)}-1)} - R_i$. Hence:

$$MSE\left(\hat{S}_{pj}^2\right) < MSE\left(\hat{S}_r^2\right),$$

if:

$$R_{pj} < \frac{2(\lambda_{22}-1)}{(\beta_{2(x)}-1)} - R_i, \quad (6)$$

where $i = 1, 2, ..., 23$ and $j = 1, 2, ..., 6$.

4. Practical study

The performances of the recommended estimators with respect to the usual and the existing ratio estimators of variance are judged by the use of 3 real populations. Populations 1 and 2 are obtained from Murthy study ([32], p. 228) and Population 3 is taken from Cochran study ([33], p. 152).

Table 2 presents the characteristics of 3 real populations. The values of the constants and biases of the existing and suggested estimators are reported in Tables 3 and 4, respectively, whereas the MSE values are reported in Tables 5 and 6, correspondingly.

According to Tables 3 and 4, it is found that the recommended estimators have the smallest values of constant in comparison with the usual and existing ratio estimators of variance. Therefore, they also satisfy the conditions given in Eq. (4) and (6), meaning that the proposed estimators perform better than the usual and the existing estimators.

The values of the biases of the recommended estimators are also lower than those of the usual estimator and the existing estimators (cf. Table 3 versus Table 4). Based on the comparison made between the estimators (suggested by Isaki [3], Upadhyaya and Singh [19], Kadilar and Cingi [20], Subramani and Kumarapandiyan [21-24], and Khan and Shabbir [25]) and the proposed estimators, it is revealed that the recommended estimators have the smallest values of MSE, compared to the existing estimators; thus, it is confirmed that the proposed estimators are more efficient (cf. Table 5 versus Table 6).

5. Robustness study of the proposed estimators

As in the previous sections, it was mentioned that the measures used in this study, such as midrange and interdecile range, were robust against outliers. Thus, when there exist outliers in the data, these measures perform more efficiently than other measures of locations. Therefore, in this section, the efficiency of the recommended estimators in the case of outliers is evaluated. For this purpose, two real data sets obtained from the Italian Bureau of the Environment Protection (IBEP) are considered [34]. This dataset consists of three variables: the entire quantity (tons) of reusable waste collection in Italy in 2003 (Y), the entire quantity of reusable waste collection in Italy in 2002 (X_1), and the number of residents in 2003 (X_2).

Table 2. Characteristics of the populations.

Parameters	Population 1	Population 2	Population 3
N	80	80	49
n	20	20	20
$\bar{Y}$	51.8264	51.8264	127.7959
$\bar{X}$	2.8513	11.2646	103.1429
ρ	0.915	0.941	0.9817
S_y	18.3566	18.3566	123.1212
C_y	0.3540	0.3540	0.8508
S_x	2.7043	8.4561	104.4051
C_x	0.948	0.751	1.0435
$\beta_{2(x)}$	3.5808	2.8664	5.9878
$\beta_{2(y)}$	2.2667	2.2667	4.9245
λ_{22}	2.3234	2.2209	4.6977
M_d	1.48	7.575	64.000
Q_1	0.8650	5.1500	43.000
Q_3	4.4525	16.975	120.000
Q_r	3.5875	11.825	77.000
Q_d	1.7937	5.9125	38.500
Q_a	2.6587	11.0625	81.500
MR	5.73	17.955	254.500
IDR	6.45	21.303	210.800

Table 3. The constants and biases of the existing ratio estimators of variance.

Estimators	Constant			Bias		
	Population 1	Population 2	Population 3	Population 1	Population 2	Population 3
$\hat{S}_r^2$	–	–	–	21.1851	10.8755	977.8210
$\hat{S}_1^2$	0.6713	0.9615	0.9995	4.6273	9.2911	975.2090
$\hat{S}_2^2$	1.149	1.0106	1.0001	31.7871	11.3281	978.2760
$\hat{S}_3^2$	1.9594	1.0418	1.0006	123.2439	12.6977	980.4370
$\hat{S}_4^2$	1.0376	1.0037	1.0000	23.6767	11.0315	977.8970
$\hat{S}_5^2$	2.0671	1.0564	1.0005	139.7088	13.3630	980.3280
$\hat{S}_6^2$	0.8317	0.9042	0.9942	14.8318	8.1742	959.1820
$\hat{S}_7^2$	0.8942	0.9328	0.9961	14.8318	8.1742	959.1820
$\hat{S}_8^2$	0.6216	0.8082	0.9891	2.9403	3.9136	926.4560
$\hat{S}_9^2$	0.6709	0.8581	0.993	4.6124	5.5032	944.6300
$\hat{S}_{10}^2$	0.803	0.9236	0.9965	10.1349	7.8268	961.1200
$\hat{S}_{11}^2$	0.7334	0.866	0.9926	7.0345	5.7698	942.7160
$\hat{S}_{12}^2$	0.919	0.9508	0.9968	16.2300	8.8710	962.7160
$\hat{S}_{13}^2$	0.9041	0.9395	0.9963	15.3820	8.4310	960.4740
$\hat{S}_{14}^2$	0.8842	0.9229	0.9959	14.2780	7.8010	958.1500
$\hat{S}_{15}^2$	0.8556	0.9135	0.9953	12.7550	7.4510	955.6580
$\hat{S}_{16}^2$	0.8317	0.9042	0.9942	11.5330	7.1100	950.1750
$\hat{S}_{17}^2$	0.7826	0.8937	0.9931	9.1800	6.7330	945.2260
$\hat{S}_{18}^2$	0.6708	0.8281	0.9915	4.6090	4.5290	937.5840
$\hat{S}_{19}^2$	0.5932	0.798	0.9876	2.0730	3.6100	919.4250
$\hat{S}_{20}^2$	0.5076	0.7409	0.978	1.1150	2.0220	874.8080
$\hat{S}_{21}^2$	0.4004	0.6723	0.9556	0.9560	0.3840	773.8070
$\hat{S}_{22}^2$	0.6005	0.7986	0.9889	2.2891	3.6275	925.5200
$\hat{S}_{23}^2$	0.8242	0.8763	0.9944	11.1581	6.1228	951.3160

Table 4. The related constants and biases of the proposed ratio estimators of variance.

Estimators	Constant			Bias		
	Population 1	Population 2	Population 3	Population 1	Population 2	Population 3
$\hat{S}_{p1}^2$	0.5607	0.7993	0.9772	1.1681	3.6482	871.2280
$\hat{S}_{p2}^2$	0.5314	0.7705	0.9810	0.4293	2.8181	888.9080
$\hat{S}_{p3}^2$	0.5387	0.7894	0.9768	0.6072	3.3579	869.3280
$\hat{S}_{p4}^2$	0.5092	0.7596	0.9807	0.0795	2.5186	887.3120
$\hat{S}_{p5}^2$	0.5476	0.7494	0.9781	0.8296	2.2434	875.4970
$\hat{S}_{p6}^2$	0.5182	0.7159	0.9818	0.1213	1.3900	892.4940

Based on Figures 1 and 2, it is clearly observed that there exist outliers in the data; therefore, the recommended estimators are expected to outperform the usual estimator and the existing estimators.

The characteristics of two outlier datasets are reported in Table 7, and the MSE values of the existing and the proposed estimators are given in Table 8. Based on Table 8, it is revealed that the recommended estimators have the least values of MSE in comparison to the usual and existing estimators, proving that the

Table 5. The mean squared error values of the existing ratio estimators of variance.

Estimators	Population 1	Population 2	Population 3
$\hat{S}_r^2$	6816.634	3924.679	17428438
$\hat{S}_1^2$	3706.764	3657.936	17412179
$\hat{S}_2^2$	9269.550	4003.635	17431277
$\hat{S}_3^2$	33998.990	4249.233	17444749
$\hat{S}_4^2$	7373.862	3951.765	17428912
$\hat{S}_5^2$	38737.400	4371.851	17444068
$\hat{S}_6^2$	4828.674	3319.838	17257349
$\hat{S}_7^2$	5470.456	3480.095	17312838
$\hat{S}_8^2$	3511.960	2908.550	17112429
$\hat{S}_9^2$	3704.843	3097.985	17223314
$\hat{S}_{10}^2$	4572.937	3426.733	17324810
$\hat{S}_{11}^2$	4051.555	3132.900	17211588
$\hat{S}_{12}^2$	5755.732	3589.974	17334676
$\hat{S}_{13}^2$	5581.849	3520.089	17320816
$\hat{S}_{14}^2$	5359.549	3422.823	17306464
$\hat{S}_{15}^2$	5060.764	3370.162	17291096
$\hat{S}_{16}^2$	4828.674	3319.838	17257349
$\hat{S}_{17}^2$	4404.946	3265.482	17226968
$\hat{S}_{18}^2$	3704.416	2977.819	17180200
$\hat{S}_{19}^2$	3433.220	2876.527	17069806
$\hat{S}_{20}^2$	3338.948	2737.079	16803036
$\hat{S}_{21}^2$	3223.476	2660.747	16224064
$\hat{S}_{22}^2$	3451.181	2878.338	17106745
$\hat{S}_{23}^2$	4759.064	3180.344	17264364

Table 6. The mean squared error values of the proposed ratio estimators of variance.

Estimators	Population 1	Population 2	Population 3
$\hat{S}_{p1}^2$	3372.184	2880.487	16781910
$\hat{S}_{p2}^2$	3343.609	2800.603	16886640
$\hat{S}_{p3}^2$	3348.396	2851.133	16770718
$\hat{S}_{p4}^2$	3338.740	2775.057	16877142
$\hat{S}_{p5}^2$	3356.334	2753.287	16807105
$\hat{S}_{p6}^2$	3338.975	2697.643	16908006

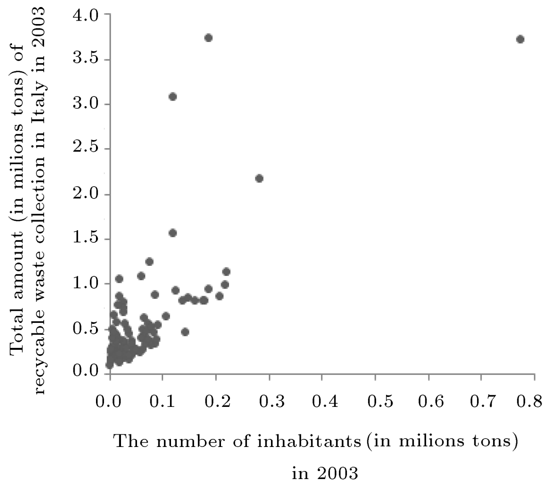

Figure 1. Scatter graph of the first auxiliary and study variables.

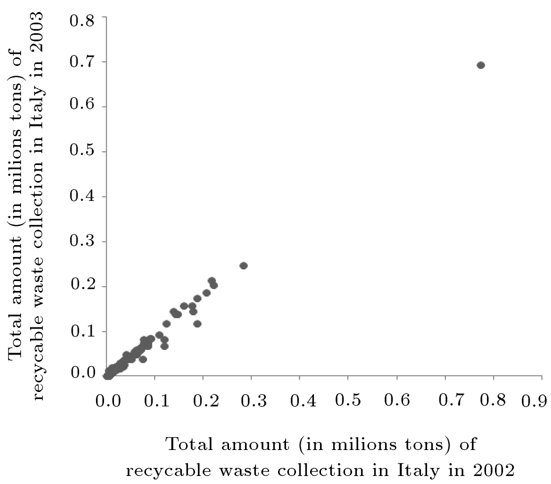

Figure 2. Scatter graph of the second auxiliary and study variables.

proposed estimators are also superior in the presence of outliers to the estimators used in this study.

Therefore, it can be concluded that the performance of the proposed estimators is relatively better than those of the usual and existing ratio estimators of the variance in the case of the datasets with and without outliers.

6. Summary and conclusions

To enhance the precision of estimators, the use of auxiliary information is very important in both the designing and estimation stages. This study developed some new ratio estimators of variance based on the information obtained from the midrange and an interdecile range of auxiliary variables. The proposed estimators were compared with the estimators introduced by Isaki [3], Upadhyaya and Singh [19], Kadilar and Cingi [20], Subramani and Kumarapandiyan [21-24], and Khan and Shabbir [25]. Based on the results of this study, it was observed that the proposed estimators showed better performance than their competitors in terms of bias and mean squared error. The suggested estimators also performed more efficiently in the presence of extreme values than the usual and existing estimators considered in this study. Hence, this study strongly recommends using the proposed estimators over the estimators considered in this study when usual and unusual observations are present in the auxiliary variables.

Table 7. Characteristics of the populations of extreme values.

Parameters	Population 4	Population 5
N	103	103
n	40	40
$\bar{Y}$	626.2123	62.621
$\bar{X}$	557.1909	556.5541
ρ	0.9936	0.7298
S_y	913.5498	91.3550
C_y	1.4588	1.4588
S_x	818.1117	610.1643
C_x	1.4683	1.0963
$\beta_{2(x)}$	37.3216	17.8738
$\beta_{2(y)}$	37.1279	37.1279
λ_{22}	37.2055	17.2220
M_d	308.05	373.82
Q_1	142.9950	259.3830
Q_3	665.6250	628.0235
Q_r	522.63	368.6405
Q_d	261.3150	184.3203
Q_a	404.31	443.7033
MR	3469.657	1906.840
IDR	1344.654	757.087

Acknowledgement

The authors are thankful to the anonymous reviewers and the corresponding editor for the constructive comments that helped improve the last version of the paper. The author, Muhammad Riaz, is indebted to King Fahd University of Petroleum and Minerals

Table 8. The mean squared error values of the existing and proposed ratio estimators of variance in case of extreme values populations.

Estimators	Population 4	Population 5
$\hat{S}^2_r$	670393403	35796744
$\hat{S}^2_1$	670169924	35796635
$\hat{S}^2_2$	670402276	35796751
$\hat{S}^2_3$	670620841	35796853
$\hat{S}^2_4$	670393641	35796744
$\hat{S}^2_5$	670547872	35796844
$\hat{S}^2_6$	668667194	35794497
$\hat{S}^2_7$	669558617	35795178
$\hat{S}^2_8$	667000664	35793005
$\hat{S}^2_9$	667623709	35794527
$\hat{S}^2_{10}$	668911758	35795628
$\hat{S}^2_{11}$	668182966	35794084
$\hat{S}^2_{12}$	670096346	35795656
$\hat{S}^2_{13}$	669691427	35795383
$\hat{S}^2_{14}$	669396238	35795075
$\hat{S}^2_{15}$	669118092	35794793
$\hat{S}^2_{16}$	668667194	35794497
$\hat{S}^2_{17}$	668111552	35794131
$\hat{S}^2_{18}$	667307609	35793572
$\hat{S}^2_{19}$	666714281	35792130
$\hat{S}^2_{20}$	664721078	35791247
$\hat{S}^2_{21}$	663322785	35777147
$\hat{S}^2_{22}$	666982742	35791666
$\hat{S}^2_{23}$	669188509	35794692
$\hat{S}^2_{p1}$	**666362968**	**35785940**
$\hat{S}^2_{p2}$	**664829191**	**35792258**
$\hat{S}^2_{p3}$	**666446248**	**35782352**
$\hat{S}^2_{p4}$	**664809727**	**35790664**
$\hat{S}^2_{p5}$	**663996842**	**35786824**
$\hat{S}^2_{p6}$	**666049543**	**35792642**

(KFUPM), Dhahran, Saudi Arabia, for providing excellent research facilities.

Nomenclature

N	Population size
n	Sample size
Y	Study variable
X	Auxiliary variable
$\bar{X}, \bar{Y}$	Population means
$\bar{x}, \bar{y}$	Sample means
S_y^2, S_x^2	Population variance of Y and X
s_y^2, s_x^2	Sample variance of Y and X
S_{xy}	Population covariance between X and Y
C_x, C_y	Coefficient of variation
M_d	Median of X
Q_1	First (lower) quartile
Q_3	Third (upper) quartile
$\gamma = 1/n$	
$Q_r = Q_3 - Q_1$	Inter quartile range
$Q_d = (Q_3 - Q_1)/2$	Semi quartile range
$Q_a = (Q_3 + Q_1)/2$	Semi quartile average
$MR = \frac{X_{(1)} + X_{(N)}}{2}$	Mid-range
$IDR = D_9 - D_1$	Interdecile range
$B(.)$	Bias of the Estimator
$MSE(.)$	Mean squared error of the estimator
$\hat{S}_i^2$	Existing modified ratio type variance estimator of S_y^2
$\hat{S}_{pj}^2$	Proposed modified ratio type variance estimator of S_y^2

Subscript

i	For existing estimators
j	For proposed estimators

Greek

ρ	Coefficient of correlation between X and Y

References

1. Neyman, J. "Contribution to the theory of sampling human populations", *Journal of American Statistical Association*, **33**, pp. 101-116 (1938).

2. Cochran, W.G. "The estimation of the yields of the cereal experiments by sampling for the ratio of grain to total produce", *Journal of Agricultural Sciences*, **30**, pp. 262-275 (1940).

3. Isaki, C.T. "Variance estimation using auxiliary information", *Journal of the American Statistical Association*, **78**, pp. 117-123 (1983).

4. Prasad, B. and Singh, H.P. "Some improved ratio type estimators of finite population variance in sample surveys", *Communications in Statistics-Theory and Methods*, **19**, pp. 1127-1139 (1990).

5. Arcos, A., Rueda, M., Martinez, M.D., Gonzalez, S., and Roman, Y. "Incorporating the auxiliary information available in variance estimation", *Applied Mathematics and Computation*, **160**, pp. 387-399 (2005).

6. Agarwal, M.C. and Sithapit, A.B. "Unbiased ratio type estimation", *Statistics and Probability Letters*, **25**, pp. 361-364 (1995).

7. Ahmed, M.S., Rahman, M.S., and Hossain, M.I. "Some competitive estimators of finite population variance using multivariate auxiliary information", *Information and Management Sciences*, **11**(1), pp. 49-54 (2000).

8. Arcos, A. and Rueda, M. "Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator", *Metrika*, **45**, pp. 171-178 (1997).

9. Al-Jararha, J. and Al-Haj Ebrahem, M. "A ratio estimator under general sampling design", *Australian Journal of Statistics*, **41**(2), pp. 105-115 (2012).

10. Bhushan, S. "Some efficient sampling strategies based on ratio type estimator", *Electronic Journal of Applied Statistical Analysis*, **5**(1), pp. 74-88 (2012).

11. Das, A.K. and Tripathi, T.P. "Use of auxiliary information in estimating the finite population variance", *Sankhyã*, **40**, pp. 139-148 (1978).

12. Cebrián, A. and Garcıa, M. "Variance estimation using auxiliary information: An almost unbiased multivariate ratio estimator", *Metrika*, **45**(1), pp. 171-178 (1997).

13. Gupta, S. and Shabbir, J. "Variance estimation in simple random sampling using auxiliary information", *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics*, **37**(1), pp. 57-67 (2008).

14. Reddy, V.N. "On Transformed ratio method of estimation", *Sankhyã*, **36**, pp. 59-70 (1974).

15. Singh, H.P., Upadhyaya, U.D., and Namjoshi, U.D. "Estimation of finite population variance", *Current Science*, **57**(24), pp. 1331-1334 (1988).

16. Singh, H.P., Chandra, P., and Singh, S. "Variance estimation using multi-auxiliary information for random non-response in survey sampling", *Statistica*, **63**, pp. 23-40 (2003).

17. Upadhyaya, L.N. and Singh, H.P. "Almost unbiased ratio and product-type estimators of finite population variance in sample surveys", *Statistics in Transition*, **7**(5), pp. 1087-1096 (2006).

18. Wolter, K.M., *Introduction to Variance Estimation*, Springer-Verlag, Berlin (1985).

19. Upadhyaya, L.N. and Singh, H.P. "An estimator of population variance that utilizes the kurtosis of an auxiliary variable in sample surveys", *Vikram Mathematical Journal*, **19**, pp. 14-17 (1999).

20. Kadilar, C. and Cingi, H. "Improvement in variance estimation using auxiliary information", *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics*, **35**(1), pp. 111-115 (2006).

21. Subramani, J. and Kumarapandiyan, G. "Variance estimation using median of the auxiliary variable", *International Journal of Probability and Statistics*, **1**(3), pp. 36-40 (2012).

22. Subramani, J. and Kumarapandiyan, G. "Variance estimation using quartiles and their functions of an auxiliary variable", *International Journal of Statistics and Applications*, **2**(5), pp. 67-72 (2012).

23. Subramani, J. and Kumarapandiyan, G. "Estimation of variance using deciles of an auxiliary variable", *Proceedings of International Conference on Frontiers of Statistics and Its Applications*, **33**, pp. 143-149 (2012).

24. Subramani, J. and Kumarapandiyan, G. "Estimation of variance using known coefficient of variation and median of an auxiliary variable", *Journal of Modern Applied Statistical Methods*, **12**(1), pp. 58-64 (2013).

25. Khan, M. and Shabbir, J. "A ratio type estimators for the estimation of population variance using quartiles of an auxiliary variable", *Journal of Statistics Applications and Probability*, **2**(3), pp. 319-325 (2013).

26. Ferrell, E.B. "Control charts using midranges and medians", *Industrial Quality Control*, **9**, pp. 30-34 (1953).

27. Rana, S., Siraj-Ud-Doulah, Md., Midi, H., and Imon, A.H.M.R. "Decile mean: A new robust measure of central tendency", *Chiang Mai Journal of Science*, **39**(3), pp. 478-485 (2012).

28. Abid, M., Abbas, N., Nazir, H.Z., and Lin, Z. "Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters", *Revista Colombiana de Estadistica*, **39**(1), pp. 63-79 (2016).

29. Abid, M., Abbas, N., and Riaz, M. "Improved modified ratio estimators of population mean based on deciles", *Chiang Mai Journal of Science*, **43**(1), pp. 1311-1323 (2016).

30. Abid, M., Abbas, N., Sherwani, R.A.K., and Nazir, H.Z. "Improved ratio estimators for the population mean using non-conventional measures of dispersion", *Pakistan Journal of Statistics and Operation Research*, **12**(2), pp. 353-367 (2016).

31. Abid, M., Sherwani, R.A.K., Abbas, N., and Raza, M.A. "Some Improved Modified Ratio Estimators Based on Decile Mean of an Auxiliary Variable", *Pakistan Journal of Statistics and Operation Research*, **12**(4), pp. 787-797 (2016).

32. Murthy, M.N., *Sampling Theory and Methods*, Statistical Publishing Society, Calcutta, India (1967).

33. Cochran, W.G., *Sampling Techniques*, John Wiley and Sons, New York (1977).

34. IBEP (2004). http://www.os servatorio nazionalerifiuti.it/elencodocpub.asp?A_ tipodoc=6

Biographies

Muhammad Abid obtained his MSc and MPhil degrees in Statistics from Quaid-i-Azam University, Islamabad, Pakistan in 2008 and 2010, respectively. He did his PhD in Statistics from the Institute of Statistics, Zhejiang University, Hangzhou, China in 2017. He served as a Statistical Officer in National Accounts Wing, Pakistan Bureau of Statistics (PBS) during 2010-2011. He is now serving as an Assistant Professor at the Department of Statistics, Government College University, Faisalabad, Pakistan from 2017 to present. He has published more than 15 research papers in research journals. His research interests include statistical quality control, Bayesian statistics, non-parametric techniques, and survey sampling.

Shabbir Ahmad obtained his MSc degree in Statistics from the PMAS-Arid Agriculture University Rawalpindi, Pakistan in 2002 and his MPhil degree in Statistics from Quaid-i-Azam University, Islamabad, Pakistan in 2005. He served as a Statistical Officer in the National Accounts Wing, Pakistan Bureau of Statistics during 2006 to 2007. He obtained his PhD degree in Statistics at the Department of Mathematics, Institute of Statistics, Zhejiang University, Hangzhou, China in 2013. He is serving as an Assistant Professor at the Department of Mathematics, COMSATS Institute of Information Technology, Wah Cantonment, Pakistan. His current research interests include statistical process control and application of sampling techniques.

Muhammad Tahir obtained his PhD in Statistics from Quaid-i-Azam University, Islamabad, Pakistan in 2017. Currently, he is an Assistant Professor of Statistics in Government College University, Faisalabad, Pakistan. He has published more than 25 research papers in national and international reputed journals. His research interests include Bayesian inference, reliability analysis, and mixture distributions.

Hafiz Zafar Nazir obtained his MSc and MPhil degrees in Statistics from the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan in 2006 and 2008, respectively. He did his PhD in Statistics from the Institute of Business and Industrial Statistics University of Amsterdam, The Netherlands in 2014. He served as a Lecturer at the Department of Statistics, University of Sargodha, Pakistan during 2009-2014. He is now serving as an Assistant Professor at the Department of Statistics, University of Sargodha, Pakistan from 2015 until now. His current research interests include statistical process control, non-parametric techniques, and robust methods.

Muhammad Riaz obtained his PhD in Statistics from the Institute for Business and Industrial Statistics, University of Amsterdam, The Netherlands in 2008. He is a Professor at the Department of Mathematics and Statistics, King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia. His current research interests include statistical process control, non-parametric techniques, and experimental design.