Scientia Iranica E (2021) 28(6), 3589–3601

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

An improved and robust class of variance estimator

M. Abid[a,*], R.A.Kh. Sherwani[b], M. Tahir[a], H.Z. Nazir[c], and M. Riaz[d]

a. *Department of Statistics, Government College University, Faisalabad, 38000, Pakistan.*
b. *College of Statistical and Actuarial Sciences, University of the Punjab Lahore, Pakistan.*
c. *Department of Statistics, University of Sargodha, Sargodha, Pakistan.*
d. *Department of Mathematics and Statistics, King Fahad University of Petroleum and Minerals, Dhahran, 31261, Saudi Arabia.*

Received 22 February 2019; received in revised form 11 December 2019; accepted 21 September 2020

KEYWORDS
Auxiliary variable;
Numerical methods;
Mean square error;
Monte Carlo;
Outliers;
Percentage relative efficiency;
Simulation;
Robust measures.

Abstract. The ratio, product, and regression estimators are commonly constructed based on conventional measures such as mean, median, quartiles, semi-interquartile range, semi-interquartile average, coefficient of skewness, and coefficient of kurtosis. In the case of the presence of outliers, these conventional measures lose their efficiency/performance ability and hence are less efficient as compared to those measures which performed efficiently in the presence of outliers. This study offers an improved class of estimators for estimating the population variance using robust dispersion measures such as probability-weighted moments, Gini, Downton and Bickel, and Lehmann measures of an auxiliary variable. Bias, mean square error and minimum mean square error of the suggested class of estimators have been derived. Application with two natural data sets is also provided to explain the proposal for practical considerations. In addition, a robustness study is also carried out to evaluate the performance of the proposed estimators in the presence of outliers by using environmental protection data. The results reveal that the proposed estimators perform better than their competitors and are robust, not only in simple conditions but also in the presence of outliers.

1. Introduction

Almost every sphere of life is prone to variation. For instance, a factory could produce five thousand piston rings of a specific size, and the exact size of a single- piston ring could vary slightly, even though they are all intended to be 'nominally' of the same size. It is because of the unavoidable variability in the manufacturing process, many more situations (such as manufacturing industry, pharmaceutical laboratories, medical and biological sciences and agriculture sectors having populations of different nature) can come across in practice where variation is present (cf. [1]). Therefore, the estimation of population variability is very important and has a major impact on obtaining accurate results.

In survey sampling, the use of auxiliary information can be adopted in the form of a ratio, product, and regression methods of estimation to get more precise estimators of the population variance of the study variable. The ratio and product methods of estimation are useful when there exists a high positive and negative correlation between the study variable, say y, and the auxiliary variable, say x, respectively (cf. [2]). The ratio method of estimation loses its efficiency when the regression line of y on x does not pass through the origin and in this situation regression method of estimation is a suitable choice (cf. [2]).

*. *Corresponding author.*
E-mail addresses: mabid@gcuf.edu.pk (M. Abid); rehan.stat@pu.edu.pk (R.A.Kh. Sherwani); tahirqaustat@yahoo.com (M. Tahir); hafizzafarnazir@yahoo.com (H.Z. Nazir); riaz76qau@yahoo.com (M. Riaz)

doi: 10.24200/sci.2020.52986.2986

Various types of estimators have been developed in the context of ratio, product, and regression estimators for estimating population variance by using the known auxiliary information based on different conventional measures such as mean, median, quartiles, semi-interquartile range, semi-interquartile average, coefficient of skewness, coefficient of kurtosis, etc. In this regard, the works [3–25].

A major disadvantage related with the estimators proposed by Isaki [3], Upadhyaya and Singh [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6–9], and Khan and Shabbir [10] is that these estimators lose their efficiency and performance ability in the presence of outliers (cf. [25–27]). For case with outliers, robust estimators are preferred and are of more practical use, because these estimators are not affected by the outliers. The development of a ratio estimator of variance that offers efficiency and robustness in the presence of outliers and measurement error in the data is a neglected area in survey sampling. So, the main concern of this work is to develop such kinds of estimators that perform well in the presence of outliers. In this study, a general class of estimators using robust measures of dispersion such as Probability-Weighted Moments (PWM), Gini, Downton, and Bickel and Lehmann measures for estimating population variance of the study variable. A major advantage of assuming these measures is their ability to be stable in the presence of outliers. The theoretical properties of these robust measures have been thoroughly investigated by Muhammad and Riaz [28], Gini [29], Downton [30], and Bickel and Lehmann [31], and these robust measures are considered to enhance the efficiency of the ratio estimator of variance in Simple Random Sampling Without Replacement (SRSWOR) scheme.

The rest of the article is planned as follows: Section 2 offers a comprehensive explanation of the existing variance estimators. The formation and performance assessment of the suggested estimators with existing estimators are given in Section 3. Section 4 includes a practical and robustness study of suggested estimators. In the end, concluding remarks are presented in Section 5.

2. Existing estimators

Isaki [3] proposed a usual ratio estimator to estimate the population variance (S_Y^2) of the study variable by using the known auxiliary information on the population variance (S_X^2) of the auxiliary variable and it is defined as:

$$\hat{S}_{E1}^2 = s_y^2\left(\frac{S_X^2}{s_x^2}\right), \qquad s_x^2 \neq 0, \tag{1}$$

where s_y^2 and s_x^2 are the sample mean of the study and auxiliary variables.

The approximate bias and Mean Square Error (MSE) of the Isaki [3] is given as:

$$Bias\left(\hat{S}_{E1}^2\right) \cong \psi S_y^2\left[(\beta_{2(x)}-1)-(\lambda_{22}-1)\right], \tag{2}$$

$$MSE\left(\hat{S}_{E1}^2\right) \cong \psi S_y^4\left[(\beta_{2(y)}-1)+(\beta_{2(x)}-1)-2(\lambda_{22}-1)\right]. \tag{3}$$

Upadhyaya and Singh [4] suggested an estimator to investigate population variance and showed that their suggested estimator is more efficient in comparison to the Isaki [3]. Upadhyaya and Singh [4] estimator is written as:

$$\hat{S}_{E2}^2 = s_y^2\left(\frac{S_X^2+\beta_{2(x)}}{s_x^2+\beta_{2(x)}}\right),$$

where $\beta_{2(x)}$ is the coefficient of kurtosis.

Based on the known auxiliary information on the coefficient of variation (C_x) and $\beta_{2(x)}$, Kadilar and Cingi [5] suggested some new estimators as follows:

$$\hat{S}_{E3}^2 = s_y^2\left(\frac{S_X^2+C_x}{s_x^2+C_x}\right), \quad \hat{S}_{E4}^2 = s_y^2\left(\frac{\beta_{2(x)}S_X^2+C_x}{\beta_{2(x)}s_x^2+C_x}\right),$$

$$\hat{S}_{E5}^2 = s_y^2\left(\frac{C_xS_X^2+\beta_{2(x)}}{C_xs_x^2+\beta_{2(x)}}\right).$$

Subramani and Kumarapandiyan [6] proposed an estimator by utilizing the information on the median (M_d) of an auxiliary variable and it is defined below:

$$\hat{S}_{E6}^2 = s_y^2\left(\frac{S_X^2+M_d}{s_x^2+M_d}\right).$$

Based on quartiles and their function of an auxiliary variable, Subramani and Kumarapandiyan [7] developed various ratio estimators which are shown below:

$$\hat{S}_{E7}^2 = s_y^2\left(\frac{S_X^2+Q_1}{s_x^2+Q_1}\right), \qquad \hat{S}_{E8}^2 = s_y^2\left(\frac{S_X^2+Q_3}{s_x^2+Q_3}\right),$$

$$\hat{S}_{E9}^2 = s_y^2\left(\frac{S_X^2+Q_r}{s_x^2+Q_r}\right), \qquad \hat{S}_{E10}^2 = s_y^2\left(\frac{S_X^2+Q_d}{s_x^2+Q_d}\right),$$

$$\hat{S}_{E11}^2 = s_y^2\left(\frac{S_X^2+Q_a}{s_x^2+Q_a}\right),$$

where Q_1, Q_3, Q_r, Q_d, and Q_a are the first quartile, third quartile, inter-quartile range, semi-quartile range, and semi-quartile average, respectively.

To incorporate the information on the deciles of an auxiliary variable, Subramani and Kumarapandiyan [8] introduced the following estimators:

$$\hat{S}_{E12}^2 = s_y^2\left(\frac{S_X^2+D_1}{s_x^2+D_1}\right), \qquad \hat{S}_{E13}^2 = s_y^2\left(\frac{S_X^2+D_2}{s_x^2+D_2}\right),$$

$$\hat{S}^2_{E14}=s^2_y\left(\frac{S^2_X+D_3}{s^2_x+D_3}\right),\qquad \hat{S}^2_{E15}=s^2_y\left(\frac{S^2_X+D_4}{s^2_x+D_4}\right),$$

$$\hat{S}^2_{E16}=s^2_y\left(\frac{S^2_X+D_5}{s^2_x+D_5}\right),\qquad \hat{S}^2_{E17}=s^2_y\left(\frac{S^2_X+D_6}{s^2_x+D_6}\right),$$

$$\hat{S}^2_{E18}=s^2_y\left(\frac{S^2_X+D_7}{s^2_x+D_7}\right),\qquad \hat{S}^2_{E19}=s^2_y\left(\frac{S^2_X+D_8}{s^2_x+D_8}\right),$$

$$\hat{S}^2_{E20}=s^2_y\left(\frac{S^2_X+D_9}{s^2_x+D_9}\right),\qquad \hat{S}^2_{E21}=s^2_y\left(\frac{S^2_X+D_{10}}{s^2_x+D_{10}}\right).$$

Making use of the information of C_X and M_d, Subramani and Kumarapandiyan [9] introduced an estimator which is given below:

$$\hat{S}^2_{E22}=s^2_y\left(\frac{C_X S^2_X+M_d}{C_X s^2_x+M_d}\right).$$

Khan and Shabbir [10] have suggested an estimator based on the coefficient of correlation (ρ) and Q_3 which is given below:

$$\hat{S}^2_{E23}=s^2_y\left(\frac{\rho S^2_X+Q_3}{s^2_x\rho+Q_3}\right).$$

The bias and MSE of the existing estimators suggested by Upadhyaya and Singh [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6–9], and Khan and Shabbir [10] i.e., $\hat{S}^2_{Ei}$, $i=2,3,\cdots,23$ up to the first degree of approximation are listed below:

$$Bias\left(\hat{S}^2_{Ei}\right)\cong\eta S^2_y\left[R^2_i(\lambda_{04}-1)-R_i(\lambda_{22}-1)\right],\quad (4)$$

$$MSE\left(\hat{S}^2_{Ei}\right)\cong\eta S^4_y\left[(\lambda_{40}-1)+R^2_i(\lambda_{04}-1)-2R_i(\lambda_{22}-1)\right],\quad (5)$$

where:

$$R_2=\frac{S^2_X}{S^2_X+\beta_{2(X)}},\qquad R_3=\frac{S^2_X}{S^2_X+C_X},$$

$$R_4=\frac{C_X S^2_X}{C_X S^2_X+\beta_{2(X)}},\qquad R_5=\frac{\beta_{2(X)}S^2_X}{\beta_{2(X)}S^2_X+C_X},$$

$$R_6=\frac{S^2_X}{S^2_X+M_d},\qquad R_7=\frac{S^2_X}{S^2_X+Q_1},$$

$$R_8=\frac{S^2_X}{S^2_X+Q_3},\qquad R_9=\frac{S^2_X}{S^2_X+Q_r},$$

$$R_{10}=\frac{S^2_X}{S^2_X+Q_d},\qquad R_{11}=\frac{S^2_X}{S^2_X+Q_a},$$

$$R_{12}=\frac{S^2_X}{S^2_X+D_1},\qquad R_{13}=\frac{S^2_X}{S^2_X+D_2},$$

$$R_{14}=\frac{S^2_X}{S^2_X+D_3},\qquad R_{15}=\frac{S^2_X}{S^2_X+D_4},$$

$$R_{16}=\frac{S^2_X}{S^2_X+D_5},\qquad R_{17}=\frac{S^2_X}{S^2_X+D_6},$$

$$R_{18}=\frac{S^2_X}{S^2_X+D_7},\qquad R_{19}=\frac{S^2_X}{S^2_X+D_8},$$

$$R_{20}=\frac{S^2_X}{S^2_X+D_9},\qquad R_{21}=\frac{S^2_X}{S^2_X+D_{10}},$$

$$R_{22}=\frac{C_X S^2_X}{C_X S^2_X+M_d},\qquad R_{23}=\frac{\rho S^2_X}{\rho S^2_X+Q_3}.$$

3. The proposed class of estimators

In this section, we propose a general class of ratio estimator of variance to evaluate S^2_Y by adopting the information on robust measures of dispersion such as PWM, Gini, Downton and Bickel, and Lehmann measures of an auxiliary variable. The PWM are investigated by Muhammad and Riaz [28] and they are defined as:

$$PWM=\frac{\sqrt{\pi}}{N^2}\sum_{i=1}^{N}(2i-N-1)X_{(i)},$$

where $X_{(i)}$ is ith order statistic of the population. The PWM is more efficient than conventional measures against outliers and therefore, provides more effective estimators than conventional measures (cf. [26]). The next measure included in this study is the Gini (G) measure suggested by Gini [29] and it is stated as:

$$G_X=\frac{4}{N-1}\sum_{i=1}^{N}\left(\frac{2i-N-1}{2N}\right)X_{(i)}.$$

The G measure is also more robust than the conventional measures in the presence of outliers (cf. [32]). The Downton (D) measure is introduced by Downton [30] and is written as:

$$D=\frac{2\sqrt{\pi}}{N(N-1)}\sum_{i=1}^{N}\left(i-\frac{N+1}{2}\right)X_{(i)}.$$

The main purpose to include D in this study is that it remains stable to outliers (cf. [26]). The last measure included in this study was suggested by Bickel and Lehmann [31] and hereafter named as B_n. The B_n measure is obtained by replacing pairwise averages with pairwise distances and it is defined as:

$$B_n=1.0483\,(\text{median}\,|x_i-x_l|\,;i<l).$$

This robust measure also has an efficiency to perform better against outliers in the data (cf. [33]).

The proposed class of estimators to estimate S^2_y is defined as:

$$\hat{S}_P^2 = \lambda \left[s_y^2 \left(\frac{\psi S_X^2 + \delta}{\psi s_x^2 + \delta} \right)^{\frac{\psi S_X^2}{\psi S_X^2 + \delta}} \right], \tag{6}$$

where λ is the constant lies between $0 \leq \lambda \leq 1$, $\psi(\neq 0)$ and δ are functions of the known parameters of auxiliary variable such as PWM, G, D, B_n, C_x, ρ, etc.

To find the bias, the MSE, and the minimum MSE of the suggested class of estimators, $\hat{S}_p^2$, we used the following conventions:

$$\varepsilon_0 = \frac{s_y^2 - S_y^2}{S_y^2}, \qquad \varepsilon_1 = \frac{s_x^2 - S_x^2}{S_x^2}.$$

Further, we can write:

$$s_y^2 = S_y^2(1+\varepsilon_0) \quad \text{and} \quad s_x^2 = S_x^2(1+\varepsilon_1),$$

where, ε_0 and ε_1 are the relative error of study and auxiliary variables, respectively.

Using the definition of ε_0 and ε_1, it can be shown that $E(\varepsilon_0) = E(\varepsilon_1) = 0$. Also $E(\varepsilon_0^2) = \eta(\lambda_{40} - 1)$, $E(\varepsilon_1^2) = \eta(\lambda_{04} - 1)$, and $E(\varepsilon_0\varepsilon_1) = \eta(\lambda_{22} - 1)$, where:

$$\lambda_{40} = \mu_{40}/\mu_{20}^2, \qquad \lambda_{04} = \mu_{04}/\mu_{02}^2,$$

$$\lambda_{22} = \mu_{22}/\mu_{20}\mu_{02},$$

$$\lambda_{rs} = (N-1)^{-1} \sum_{i=1}^{N} (Y_i - \bar{Y})^r \left(X_i - \bar{X}\right)^s,$$

$$S_Y^2 = (N-1)^{-1} \sum_{i=1}^{N} \left(Y_i - \bar{Y}\right)^2,$$

$$S_X^2 = (N-1)^{-1} \sum_{i=1}^{N} \left(X_i - \bar{X}\right)^2,$$

$$s_y^2 = (n-1)^{-1} \sum_{i=1}^{n} (y_i - \bar{y})^2,$$

and:

$$s_x^2 = (n-1)^{-1} \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

So, Eq. (6) can be written in terms of ε_0 and ε_1 as follows:

$$\hat{S}_P^2 = \lambda \left[S_y^2 (1+\varepsilon_0)(1+\Delta_j \varepsilon_1)^{-\Delta_j} \right],$$
$$j = 1, 2, \cdots, 12, \tag{7}$$

where:

$$\Delta_1 = \frac{S_X^2}{S_X^2 + PWM}, \qquad \Delta_2 = \frac{\rho S_X^2}{\rho S_X^2 + PWM},$$

$$\Delta_3 = \frac{C_x S_X^2}{C_x S_X^2 + PWM}, \qquad \Delta_4 = \frac{S_X^2}{S_X^2 + G},$$

$$\Delta_5 = \frac{\rho S_X^2}{\rho S_X^2 + G}, \qquad \Delta_6 = \frac{C_x S_X^2}{C_x S_X^2 + G},$$

$$\Delta_7 = \frac{S_X^2}{S_X^2 + D}, \qquad \Delta_8 = \frac{\rho S_X^2}{\rho S_X^2 + D},$$

$$\Delta_9 = \frac{C_x S_X^2}{C_x S_X^2 + D}, \qquad \Delta_{10} = \frac{S_X^2}{S_X^2 + B_n},$$

$$\Delta_{11} = \frac{\rho S_X^2}{\rho S_X^2 + B_n}, \qquad \Delta_{12} = \frac{C_x S_X^2}{C_x S_X^2 + B_n}.$$

Expanding Eq. (7) up to the first degree of approximation provides:

$$\hat{S}_P^2 = \lambda \left[S_y^2 \left\{ 1 + \varepsilon_0 - \Delta_j^2 \varepsilon_1 + \frac{1}{2}\Delta_j^3 \left(1+\Delta_j^2\right) \varepsilon_1^2 - \Delta_j^2 \varepsilon_0 \varepsilon_1 \right\} \right],$$

$$\hat{S}_P^2 - S_y^2 = \lambda S_y^2 + \lambda \left[S_y^2 \left\{ \varepsilon_0 - \Delta_j^2 \varepsilon_1 + \frac{1}{2}\Delta_j^3 (1+\Delta_j^2) \varepsilon_1^2 - \Delta_j^2 \varepsilon_0 \varepsilon_1 \right\} \right] - S_y^2. \tag{8}$$

The bias of an estimator $\hat{S}_P^2$ is defined as:

$$Bias\left(\hat{S}_P^2\right) = E\left(\hat{S}_P^2 - S_y^2\right).$$

Applying expectation on both sides of Eq. (8), we get:

$$E\left(\hat{S}_P^2 - S_y^2\right) = (\lambda - 1) S_y^2 + \lambda \left[S_y^2 \left\{ E(\varepsilon_0) - \Delta_j^2 E(\varepsilon_1) + \frac{1}{2}\Delta_j^3 (1+\Delta_j^2) E(\varepsilon_1^2) - \Delta_j^2 E(\varepsilon_0\varepsilon_1) \right\} \right],$$

$$Bias\left(\hat{S}_P^2\right) \cong (\lambda - 1) S_y^2 + \eta \lambda S_y^2 \left[\Delta_j^2 \left\{ \frac{1}{2}\Delta_j (1+\Delta_j^2)(\lambda_{04} - 1) - (\lambda_{22} - 1) \right\} \right]. \tag{9}$$

The MSE of an estimator $\hat{S}_P^2$ is defined as:

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

Thus, squaring and then applying expectation on both sides of Eq. (8), we get:

$$E\left(\hat{S}_P^2 - S_y^2\right)^2 = [(\lambda-1)^2 S_y^4 + \left\{\eta \lambda^2 S_y^4\right\} E(\varepsilon_0^2) + \eta S_y^4 \left\{ \lambda^2 \Delta_j^3 (1 + 2\Delta_j) - \lambda \Delta_j^3 (1+\Delta_j) \right\} E(\varepsilon_0^2) + \eta S_y^4 \left\{ \lambda^2 (4\Delta_j^2) - \lambda (2\Delta_j^2) E(\varepsilon_0 \varepsilon_1) \right\}].$$

After substituting the values of $E(\varepsilon_0^2)$, $E(\varepsilon_0^2)$ and $E(\varepsilon_0\varepsilon_1)$ in the above equation, we obtain:

$$MSE\left(\hat{S}_P^2\right) \cong (\lambda-1)^2 S_y^4 + \Gamma_1(\lambda_{40}-1) + \Gamma_2(\lambda_{04}-1) - \Gamma_3(\lambda_{22}-1), \tag{10}$$

where

$$\Gamma_1 = \eta\lambda^2 S_y^4,$$

$$\Gamma_2 = \eta S_y^4\{\lambda^2\Delta_j^3(1+2\Delta_j) - \lambda\Delta_j^3(1+\Delta_j)\},$$

$$\Gamma_3 = \eta S_y^4\{\lambda^2(4\Delta_j^2) - \lambda(2\Delta_j^2)\}.$$

Differentiating Eq. (10) with respect to λ and then setting it equal to zero, we get the equation shown in Box I, where:

$$A = 2S_y^4 + \eta S_y^4\{\Delta_j^3(1+\Delta_j)(\lambda_{04}-1) - 2\Delta_j^2(\lambda_{22}-1)\},$$

$$B = 2S_y^4 + 2\eta S_y^4[(\lambda_{40}-1) + \Delta_j^3(1+2\Delta_j)(\lambda_{04}-1) - 4\Delta_j^2(\lambda_{22}-1)].$$

Putting λ_{opt} in Eq. (10), we obtain the minimum MSE of $\hat{S}_P^2$:

$$MSE\left(\hat{S}_P^2\right)_{\min} \cong S_y^2\left[S_y^2 - \frac{A^2}{2BS_y^2}\right]. \tag{11}$$

3.1. Some members of the proposed class of estimators

In this subsection, we will propose some new estimators which belong to the proposed class of estimators given in Eq. (6) by substituting different values of λ, ψ, and δ.

Remark 3.1. If we put $(\lambda,\psi,\delta) = (1,0,1)$ in Eq. (6) then the proposed class reduces to the traditional simple random sampling estimator.

$$\hat{S}_{E0}^2 = s_y^2.$$

Remark 3.2. If we substitute $(\lambda,\psi,\delta) = (1,1,0)$ in Eq. (6) then proposed class belongs to the Isaki [3] estimator as follows:

$$\hat{S}_{E1}^2 = s_y^2\left(\frac{S_x^2}{s_x^2}\right), \qquad s_x^2 \neq 0.$$

Remark 3.3. If we put different choices of $\psi(1,\rho,C_x)$ and $\delta(PWM,G,D,B_n)$ in Eq. (6) then some new members belong to the proposed class $\hat{S}_P^2$ as below:

$$\hat{S}_{P1}^2 = \lambda s_y^2\left[\frac{S_X^2+PWM}{s_x^2+PWM}\right]^{\frac{S_X^2}{S_X^2+PWM}},$$

$$\hat{S}_{P2}^2 = \lambda s_y^2\left[\frac{\rho S_X^2+PWM}{\rho s_x^2+PWM}\right]^{\frac{\rho S_X^2}{\rho S_X^2+PWM}},$$

$$\hat{S}_{P3}^2 = \lambda s_y^2\left[\frac{C_X S_X^2+PWM}{C_X s_x^2+PWM}\right]^{\frac{C_X S_X^2}{C_X S_X^2+PWM}},$$

$$\hat{S}_{P4}^2 = \lambda s_y^2\left[\frac{S_X^2+G}{s_x^2+G}\right]^{\frac{S_X^2}{S_X^2+G}},$$

$$\hat{S}_{P5}^2 = \lambda s_y^2\left[\frac{\rho S_X^2+G}{\rho s_x^2+G}\right]^{\frac{\rho S_X^2}{\rho S_X^2+G}},$$

$$\hat{S}_{P6}^2 = \lambda s_y^2\left[\frac{C_X S_X^2+G}{C_X s_x^2+G}\right]^{\frac{C_X S_X^2}{C_X S_X^2+G}},$$

$$\hat{S}_{P7}^2 = \lambda s_y^2\left[\frac{S_X^2+D}{s_x^2+D}\right]^{\frac{S_X^2}{S_X^2+D}},$$

$$\hat{S}_{P8}^2 = \lambda s_y^2\left[\frac{\rho S_X^2+D}{\rho s_x^2+D}\right]^{\frac{\rho S_X^2}{\rho S_X^2+D}},$$

$$\hat{S}_{P9}^2 = \lambda s_y^2\left[\frac{C_X S_X^2+D}{C_X s_x^2+D}\right]^{\frac{C_X S_X^2}{C_X S_X^2+D}},$$

$$\hat{S}_{P10}^2 = \lambda s_y^2\left[\frac{S_X^2+B_n}{s_x^2+B_n}\right]^{\frac{S_X^2}{S_X^2+B_n}},$$

$$\hat{S}_{P11}^2 = \lambda s_y^2\left[\frac{\rho S_X^2+B_n}{\rho s_x^2+B_n}\right]^{\frac{\rho S_X^2}{\rho S_X^2+B_n}},$$

$$\hat{S}_{P12}^2 = \lambda s_y^2\left[\frac{C_X S_X^2+B_n}{C_X s_x^2+B_n}\right]^{\frac{C_X S_X^2}{C_X S_X^2+B_n}}.$$

Remark 3.4. We can also get some new members which belong to the proposed class by replacing C_x and ρ by $\beta_{1(x)}$ and $\beta_{2(x)}$.

The theoretical and numerical comparisons are usually performed between the proposed and the existing estimators. In the next sections, we present

$$\lambda_{opt} = \frac{2S_y^4 + \eta S_y^4\{\Delta_j^3(1+\Delta_j)(\lambda_{04}-1) - 2\Delta_j^2(\lambda_{22}-1)\}}{2S_y^4 + 2\eta S_y^4[(\lambda_{40}-1) + \Delta_j^3(1+2\Delta_j)(\lambda_{04}-1) - 4\Delta_j^2(\lambda_{22}-1)]} = \frac{A}{B}.$$

Box I

these comparisons between the proposed and existing estimators. The performance comparison is carried out on the basis of MSE. The proposed family of estimators is compared with the estimators suggested by Upadhyaya and Singh [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6–9], and Khan and Shabbir [10].

3.2. Efficiency comparisons

In this section, we find the conditions in which the proposed estimators perform more efficiently in comparison to the traditional and the existing ratio estimators of variance.

3.2.1. Comparison with traditional ratio estimator

The proposed estimators are more efficient than the traditional ratio estimator suggested by Isaki [3], if they have smaller values of MSE against Isaki [3] estimator. Mathematically, it is defined as:

$$MSE\left(\hat{S}_P^2\right)_{\min} < MSE\left(\hat{S}_{\epsilon 1}^2\right).$$

By Eqs. (3) and (11):

$$S_y^2\left[S_y^2-\frac{A^2}{2BS_y^2}\right] < \eta S_y^4[(\lambda_{40}-1)+(\lambda_{04}-1) - 2(\lambda_{22}-1)].$$

After some simplification, we get:

$$A^2 > 2BS_y^4\{1-\eta(\lambda_{40}+\lambda_{04}-2\lambda_{22})\}. \tag{12}$$

3.2.2. Comparison with existing estimators

The proposed estimators perform better if the values of MSE of the suggested estimators are lesser than the values of MSE of existing estimators proposed by Upadhyaya and Singh [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6–9], and Khan and Shabbir [10] and in algebraic form, it is expressed as:

$$MSE\left(\hat{S}_P^2\right)_{\min} < MSE\left(\hat{S}_{\epsilon i}^2\right), \qquad i=2,3,...,23.$$

By Eqs. (5) and (11):

$$S_y^2\left[S_y^2-\frac{A^2}{2BS_y^2}\right] < \eta S_y^4[(\lambda_{40}-1)+R_i^2(\lambda_{04}-1) - 2R_i(\lambda_{22}-1)].$$

After solving the above equation, we obtain:

$$A^2 > 2BS_y^4\{1-\eta(\lambda_{40}-1)+R_i^2(\lambda_{04}-1) - 2R_i(\lambda_{22}-1)\}. \tag{13}$$

If the conditions given in Eqs. (12) and (13) are held then it indicates the supremacy of the proposed estimators against the existing estimators discussed in Section 2.

In the next section, we present the empirical and outliers study comparisons between the proposed and the existing estimators.

4. Empirical study

To assess the performance of the suggested estimators against their competing estimators, we use two real populations. The descriptions of two populations are given below:

Population I (Source: Murthy [34], page 228):

$Y=$ Output, $X=$ The number of workers,

$N=80$, $n=20$, $\bar{Y}=51.826$,

$\bar{X}=2.851$, $\rho=0.915$, $S_y=18.357$,

$C_y=0.354$, $S_x=2.704$, $C_x=0.948$,

$\lambda_{04}=3.581$, $\lambda_{40}=2.267$, $\lambda_{22}=2.323$,

$M_d=1.480$, $Q_1=0.865$, $Q_3=4.453$,

$Q_r=3.588$, $Q_d=1.794$, $Q_a=2.659$,

$PWM=2.448$, $G=2.797$, $D=2.479$,

$B_n=2.101$.

Population II (Source: Murthy [34], page 228):

$Y=$ Output, $X=$ The fixed capital,

$N=80$, $n=20$, $\bar{Y}=51.826$,

$\bar{X}=11.265$, $\rho=0.941$, $S_y=18.357$,

$C_y=0.354$, $S_x=8.456$, $C_x=0.751$,

$\lambda_{04}=2.866$, $\lambda_{40}=2.267$, $\lambda_{22}=2.221$,

$M_d=7.575$, $Q_1=5.150$, $Q_3=16.975$,

$Q_r=11.825$, $Q_d=5.913$, $Q_a=11.063$,

$PWM=7.913$, $G=9.040$, $D=8.013$,

$B_n=6.714$.

The validity and utility of an estimator are generally assessed by the metric called MSE (cf. [4,5,25]). The MSE measures the divergence of the estimator $(\hat{\theta})$ values from the true parameter (θ) value and mathematically it is defined as:

$$MSE=E\left(\hat{\theta}-\theta\right)^2. \tag{14}$$

We also find the Percentage Relative Efficiencies (PREs) of proposed and existing estimators with respect to the traditional ratio estimator of variance.

The mathematical expressions of PREs are given below:

$$PRE(E_i;E_1)=\frac{MSE\left(\hat{S}^2_{E1}\right)}{MSE\left(\hat{S}^2_{Ei}\right)}, \tag{15}$$

$$PRE(P_j;E_1)=\frac{MSE\left(\hat{S}^2_{E1}\right)}{MSE\left(\hat{S}^2_{P_j}\right)}, \tag{16}$$

where, $i=2,3,\cdots,23$, $j=1,2,\cdots,9$. The values of MSE and PREs are given in Table 1.

To get more insight into the study, we also used a new measure called the Percentage Decrease (PD) in MSE which hereafter is named as MSE_{PD}. The values of MSE_{PD} between existing and proposed estimators are given in Table 2. An estimator which has a larger MSE_{PD} is considered to be better as compared to other estimators. The MSE_{PD} can be computed as:

$$MSE_{PD}=\frac{(MSE_E-MSE_p)}{MSE_E}\times 100.$$

From the analysis of Tables 1 and 2, the key observations about the study could be summarized as follows:

i. By comparing all existing estimators, it is found that the performance of the estimators $\hat{S}^2_{E19}$ and $\hat{S}^2_{E20}$ is relatively better for Populations I and II, respectively;

ii. By comparing all proposed estimators with the usual ratio estimator and existing estimators, it is observed that proposed estimators perform more efficiently due to smaller MSE values and higher PREs values;

iii. It is seen that proposed estimators have lesser values of MSE ranges from (2393, 2421) and (1950, 2102) against existing estimators for Populations I and II, respectively;

iv. It is also observed that the proposed estimators have higher values of PREs ranges from (209.81, 213.62) and (140.05, 150.97) against their competing estimators for Populations I and II, respectively;

v. The estimators $\hat{S}^2_{P8}$ and $\hat{S}^2_{P6}$ have smaller values of MSE (2393, 1950) and higher values of PREs (213.62,150.97) by comparing all other proposed estimators for Populations I and II, respectively;

vi. The estimator $\hat{S}^2_{P8}$ has a larger MSE_{PD} against other proposed estimators for Population I in comparison to existing estimators (cf. Table 2).

A graphical comparison between the proposed and the existing estimators is also made. Most efficient estimators from the existing estimators suggested by Isaki [3], Upadhyaya and Singh [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6–9], and Khan and Shabbir [10] and from the proposed estimators are chosen and then compared. Figure 1(a) and (b), depict that the proposed estimators have smaller values of MSE in comparison to the existing

Table 1. Values of Mean Square Error (MSE) and Percentage Relative Efficiency (PRE) for empirical study.

Estimators	**Population I**		**Population II**	
Usual	**MSE**	**PRE**	**MSE**	**PRE**
$\hat{S}^2_R$	5112	100	2944	100
Existing:				
$\hat{S}^2_{E1}$	2780	183.90	2744	107.29
$\hat{S}^2_{E2}$	4028	126.92	2888	101.95
$\hat{S}^2_{E3}$	4752	107.59	2924	100.68
$\hat{S}^2_{E4}$	2741	186.55	2686	109.62
$\hat{S}^2_{E5}$	3622	141.17	2491	118.22
$\hat{S}^2_{E6}$	4103	124.61	2609	112.83
$\hat{S}^2_{E7}$	2634	194.10	2183	134.88
$\hat{S}^2_{E8}$	2779	183.99	2326	126.57
$\hat{S}^2_{E9}$	3430	149.06	2572	114.46
$\hat{S}^2_{E10}$	3039	168.25	2351	125.26
$\hat{S}^2_{E11}$	4317	118.43	2693	109.32
$\hat{S}^2_{E12}$	4186	122.12	2641	111.49
$\hat{S}^2_{E13}$	4020	127.19	2568	114.66
$\hat{S}^2_{E14}$	3796	134.70	2528	116.45
$\hat{S}^2_{E15}$	3622	141.17	2491	118.22
$\hat{S}^2_{E16}$	3304	154.75	2450	120.19
$\hat{S}^2_{E17}$	2778	184.01	2234	131.80
$\hat{S}^2_{E18}$	2575	198.55	2158	136.44
$\hat{S}^2_{E19}$	2504	204.16	2053	143.39
$\hat{S}^2_{E20}$	2643	193.46	1996	147.51
$\hat{S}^2_{E21}$	3569	143.23	2386	123.40
$\hat{S}^2_{E22}$	2588	197.52	2160	136.30
Proposed:				
$\hat{S}^2_{P1}$	2400	212.99	2053	143.40
$\hat{S}^2_{P2}$	2393	213.62	2036	144.63
$\hat{S}^2_{P3}$	2395	213.49	1977	148.92
$\hat{S}^2_{P4}$	2394	213.55	2015	146.08
$\hat{S}^2_{P5}$	2405	212.56	2000	147.25
$\hat{S}^2_{P6}$	2399	213.10	1950	150.97
$\hat{S}^2_{P7}$	2399	213.13	2049	143.66
$\hat{S}^2_{P8}$	2393	213.62	2032	144.88
$\hat{S}^2_{P9}$	2394	213.55	1974	149.13
$\hat{S}^2_{P10}$	2437	209.81	2102	140.05
$\hat{S}^2_{P11}$	2412	211.99	2084	141.28
$\hat{S}^2_{P12}$	2421	211.20	2019	145.86

Table 2. Percentage decrease in Mean Square Error (MSE) for Population I.

Existing	Proposed											
	$\hat{S}^2_{P1}$	$\hat{S}^2_{P2}$	$\hat{S}^2_{P3}$	$\hat{S}^2_{P4}$	$\hat{S}^2_{P5}$	$\hat{S}^2_{P6}$	$\hat{S}^2_{P7}$	$\hat{S}^2_{P8}$	$\hat{S}^2_{P9}$	$\hat{S}^2_{P10}$	$\hat{S}^2_{P11}$	$\hat{S}^2_{P12}$
$\hat{S}^2_{R}$	53.052	53.189	53.149	53.169	52.954	53.071	53.071	53.189	53.169	52.328	52.817	52.641
$\hat{S}^2_{E1}$	13.669	13.921	13.849	13.885	13.489	13.705	13.705	13.921	13.885	12.338	13.237	12.914
$\hat{S}^2_{E2}$	40.417	40.591	40.541	40.566	40.293	40.442	40.442	40.591	40.566	3Z9.499	40.119	39.896
$\hat{S}^2_{E3}$	49.495	49.642	49.600	49.621	49.390	49.516	49.516	49.642	49.621	48.716	49.242	49.053
$\hat{S}^2_{E4}$	12.441	12.696	12.623	12.660	12.258	12.477	12.477	12.696	12.660	11.091	12.003	11.675
$\hat{S}^2_{E5}$	33.738	33.932	33.876	33.904	33.600	33.766	33.766	33.932	33.904	32.717	33.407	33.158
$\hat{S}^2_{E6}$	41.506	41.677	41.628	41.652	41.384	41.531	41.531	41.677	41.652	40.604	41.214	40.994
$\hat{S}^2_{E7}$	8.884	9.150	9.074	9.112	8.694	8.922	8.922	9.150	9.112	7.479	8.428	8.087
$\hat{S}^2_{E8}$	13.638	13.890	13.818	13.854	13.458	13.674	13.674	13.890	13.854	12.307	13.206	12.882
$\hat{S}^2_{E9}$	30.029	30.233	30.175	30.204	29.883	30.058	30.058	30.233	30.204	28.950	29.679	29.417
$\hat{S}^2_{E10}$	21.027	21.257	21.191	21.224	20.862	21.060	21.060	21.257	21.224	19.809	20.632	20.336
$\hat{S}^2_{E11}$	44.406	44.568	44.522	44.545	44.290	44.429	44.429	44.568	44.545	43.549	44.128	43.919
$\hat{S}^2_{E12}$	42.666	42.833	42.785	42.809	42.547	42.690	42.690	42.833	42.809	41.782	42.379	42.164
$\hat{S}^2_{E13}$	40.299	40.473	40.423	40.448	40.174	40.323	40.323	40.473	40.448	39.378	40.000	39.776
$\hat{S}^2_{E14}$	36.776	36.960	36.907	36.934	36.644	36.802	36.802	36.960	36.934	35.801	36.459	36.222
$\hat{S}^2_{E15}$	33.738	33.932	33.876	33.904	33.600	33.766	33.766	33.932	33.904	32.717	33.407	33.158
$\hat{S}^2_{E16}$	27.361	27.573	27.512	27.542	27.209	27.391	27.391	27.573	27.542	26.241	26.998	26.725
$\hat{S}^2_{E17}$	13.607	13.859	13.787	13.823	13.427	13.643	13.643	13.859	13.823	12.275	13.175	12.851
$\hat{S}^2_{E18}$	6.796	7.068	6.990	7.029	6.602	6.835	6.835	7.068	7.029	5.359	6.330	5.981
$\hat{S}^2_{E19}$	4.153	4.433	4.353	4.393	3.954	4.193	4.193	4.433	4.393	2.676	3.674	3.315
$\hat{S}^2_{E20}$	9.194	9.459	9.383	9.421	9.005	9.232	9.232	9.459	9.421	7.794	8.740	8.400
$\hat{S}^2_{E21}$	32.754	32.950	32.894	32.922	32.614	32.782	32.782	32.950	32.922	31.718	32.418	32.166
$\hat{S}^2_{E22}$	7.264	7.535	7.457	7.496	7.071	7.303	7.303	7.535	7.496	5.835	6.801	6.453

estimators which shows that proposed estimators are more efficient than the existing estimators.

4.1. Simulation study

A simulation study is also carried out to assess the performance of the existing and proposed estimators. The statistical programming language R is used to carry out the simulation study. The following procedure is adopted to compute the MSE of proposed and existing estimators:

i. Generating a random sample of size $n = 30$ and 40 from the bivariate normal distribution;

ii. Computing the MSE of the proposed and existing estimators by using the expressions given in Eqs. (3), (5), and (11) of the random samples generated in Step (i);

iii. Repeating Steps (i) and (ii) 20000 times to obtain MSEs;

iv. Averaging these MSEs to obtain the value of MSE of proposed and existing estimators.

The results of the simulation study in terms of MSEs are reported in Table 3 and Figure 2(a) and (b). Following are the main findings based on these results:

i. The proposed estimators are more efficient in comparison to the usual and existing ratio estimators of variance;

ii. As the value of n increases, the value of MSE decreases, and vice versa;

iii. From all the proposed estimators, the estimator $\hat{S}^2_{p9}$ has the smaller MSE values for each choice of n.

4.2. Performance of the proposed estimators in case of outliers

As mentioned in the previous section, the estimators suggested in this study are robust and efficient for case with outliers in the data. So, the efficiency of the proposed estimators against outliers is evaluated in this section. For the said purpose, two natural populations taken from the Italian national institute for environment protection and research (cf. [35]) are considered. The comparison between proposed and

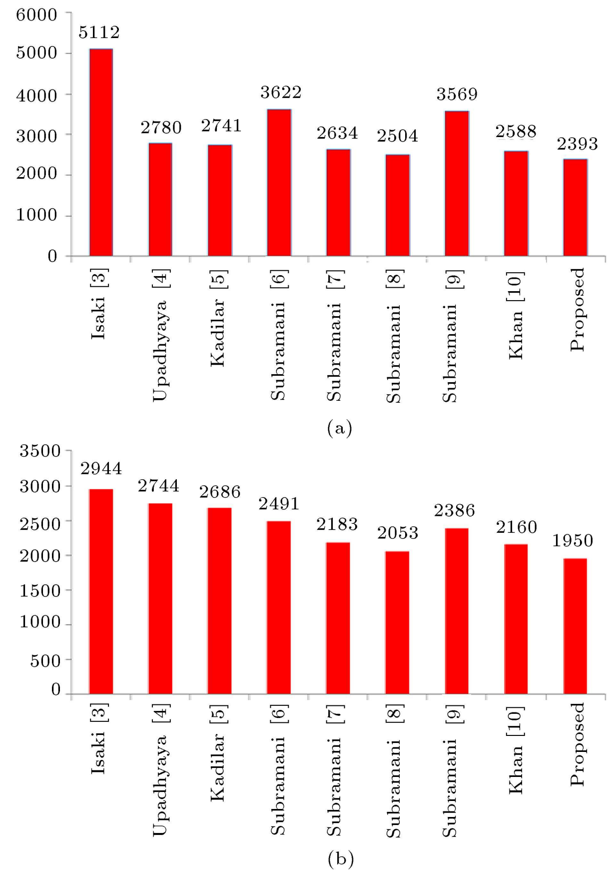

Figure 1. Mean Square Error (MSE) values of the proposed and existing estimators for (a) Population I and (b) Population II.

existing estimators for case with outliers is also done on the basis of MSE and PREs.

The characteristics of the two populations are listed below:

Population III (Source: ISPRA [35]):

$Y =$ Total amount of recycled waste collection in Italy in 2003,

$X =$ The number of inhabitants in 2003,

$N = 103$, $n = 40$, $\bar{Y} = 626.212$,

$\bar{X} = 557.191$, $\rho = 0.994$, $S_y = 913.541$,

$C_y = 1.457$, $S_x = 818.112$, $C_x = 1.468$,

$\lambda_{04} = 37.322$, $\lambda_{40} = 37.128$, $\lambda_{22} = 37.206$,

$M_d = 308.050$, $Q_1 = 142.995$, $Q_3 = 665.625$,

$Q_r = 522.630$ $Q_d = 261.315$, $Q_a = 404.310$,

$PWM = 556.424$, $G = 633.885$, $D = 561.879$,

Table 3. Values of Mean Square Error (MSE) based on the simulation study.

Estimators	$n = 30$	$n = 40$
Usual	**MSE**	**MSE**
$\hat{S}^2_R$	2840.26	1704.16
Existing:		
$\hat{S}^2_{E1}$	1544.49	926.69
$\hat{S}^2_{E2}$	2237.78	1342.67
$\hat{S}^2_{E3}$	2639.79	1583.88
$\hat{S}^2_{E4}$	1522.53	913.52
$\hat{S}^2_{E5}$	2011.95	1207.17
$\hat{S}^2_{E6}$	2279.36	1367.61
$\hat{S}^2_{E7}$	1463.32	877.99
$\hat{S}^2_{E8}$	1543.69	926.21
$\hat{S}^2_{E9}$	1905.39	1143.23
$\hat{S}^2_{E10}$	1688.15	1012.89
$\hat{S}^2_{E11}$	2398.22	1438.93
$\hat{S}^2_{E12}$	2325.77	1395.46
$\hat{S}^2_{E13}$	2233.15	1339.89
$\hat{S}^2_{E14}$	2108.65	1265.19
$\hat{S}^2_{E15}$	2011.95	1207.17
$\hat{S}^2_{E16}$	1835.39	1101.24
$\hat{S}^2_{E17}$	1543.51	926.10
$\hat{S}^2_{E18}$	1430.51	858.30
$\hat{S}^2_{E19}$	1391.23	834.74
$\hat{S}^2_{E20}$	1468.12	880.87
$\hat{S}^2_{E21}$	1982.94	1189.77
$\hat{S}^2_{E22}$	1437.99	862.80
Proposed:		
$\hat{S}^2_{P1}$	1368.40	827.85
$\hat{S}^2_{P2}$	1366.42	826.39
$\hat{S}^2_{P3}$	1365.74	825.84
$\hat{S}^2_{P4}$	1365.69	825.45
$\hat{S}^2_{P5}$	1369.00	827.05
$\hat{S}^2_{P6}$	1366.61	825.83
$\hat{S}^2_{P7}$	1366.24	826.25
$\hat{S}^2_{P8}$	1366.42	825.74
$\hat{S}^2_{P9}$	1365.44	825.39
$\hat{S}^2_{P10}$	1388.25	841.11
$\hat{S}^2_{P11}$	1373.94	831.66
$\hat{S}^2_{P12}$	1378.94	835.00

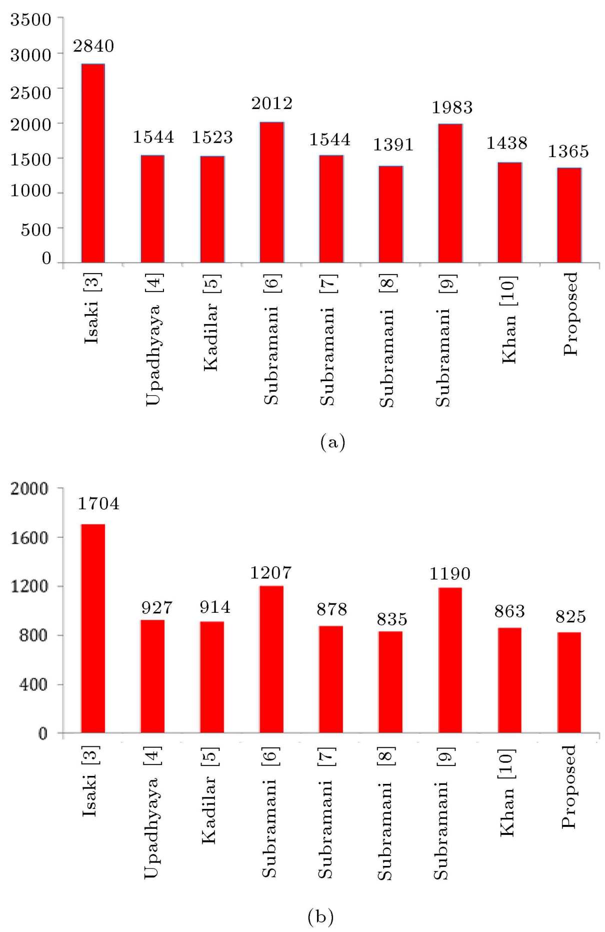

Figure 2. Mean Square Error (MSE) values of the proposed and existing estimators for simulation study: (a) $n = 30$ and (b) $n = 40$.

$B_n = 370.155.$

Population IV (Source: ISPRA [35]):

Y = Total amount of recycled waste collection in Italy in 2003,

X = Total amount of recycled waste collection in Italy in 2002,

$N = 103,\quad n = 40,\quad \bar{Y} = 62.621,$

$\bar{X} = 556.554,\quad \rho = 0.730,\quad S_y = 91.354,$

$C_y = 1.459,\quad S_x = 610.164,\quad C_x = 1.096,$

$\lambda_{04} = 17.874,\quad \lambda_{40} = 37.128,\quad \lambda_{22} = 17.222,$

$M_d = 373.820,\quad Q_1 = 259.383,\quad Q_3 = 628.023,$

$Q_r = 388.641\quad Q_d = 183.320,\quad Q_a = 443.703,$

$PWM = 423.086,\quad G = 481.985,\quad D = 427.234,$

$B_n = 250.776.$

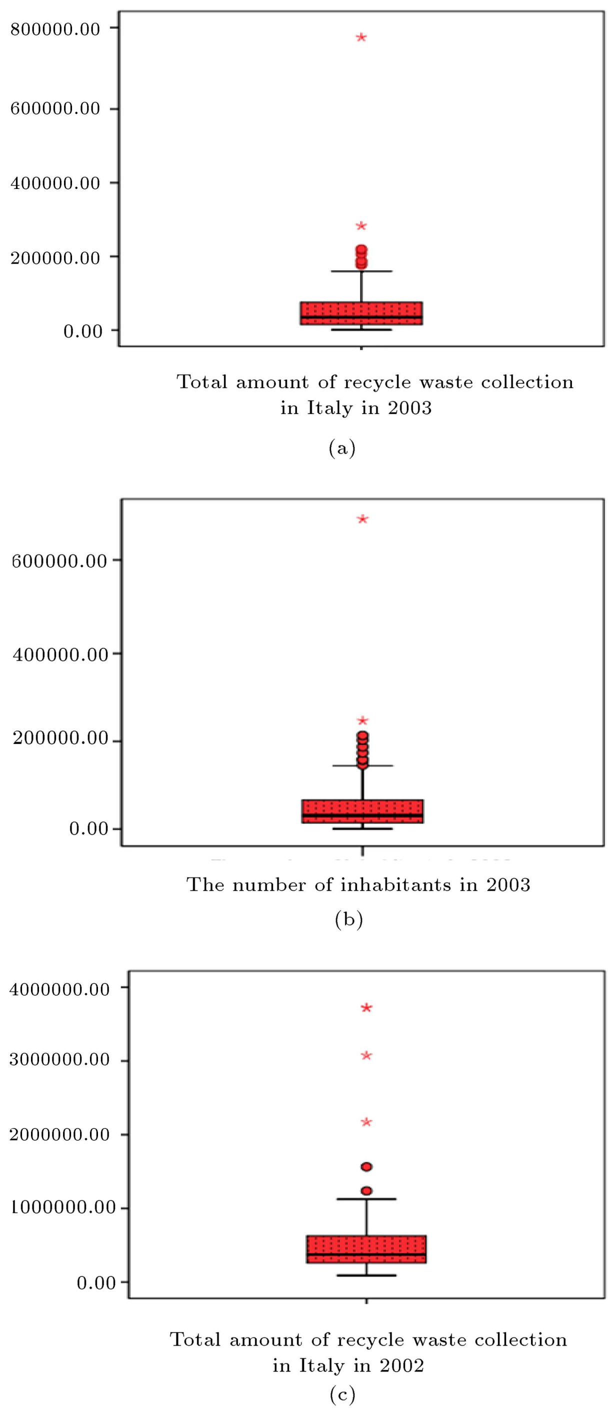

Figure 3. Boxplot of (a) total amount of recycled waste collection in Italy in 2003, (b) the number of inhabitants in 2003, and (c) total amount of recycled waste collection in Italy in 2002.

In addition, boxplots of Populations III and IV are provided to display that these populations contain outliers. It is obvious from Figure 3(a)–(c), that there

Table 4. Values of Mean Square Error (MSE) and Percentage Relative Efficiency (PRE) in the presence of outliers.

Estimators	Population III		Population IV	
Usual	**MSE**	**PRE**	**MSE**	**PRE**
$\hat{S}^2_R$	410046451	100	21895096	100
Existing:				
$\hat{S}^2_{E1}$	409909759	100.03	21895029	100.00
$\hat{S}^2_{E2}$	410041027	100.00	21895092	100.00
$\hat{S}^2_{E3}$	410046305	100.00	21895096	100.00
$\hat{S}^2_{E4}$	409953093	100.02	21895035	100.00
$\hat{S}^2_{E5}$	408990614	100.26	21893721	100.01
$\hat{S}^2_{E6}$	409535853	100.12	21894138	100.00
$\hat{S}^2_{E7}$	407971280	100.51	21892809	100.01
$\hat{S}^2_{E8}$	408352366	100.41	21893740	100.01
$\hat{S}^2_{E9}$	409140202	100.22	21894413	100.00
$\hat{S}^2_{E10}$	408694436	100.33	21893469	100.01
$\hat{S}^2_{E11}$	409864755	100.04	21894430	100.00
$\hat{S}^2_{E12}$	409617086	100.10	21894267	100.00
$\hat{S}^2_{E13}$	409436534	100.15	21894075	100.00
$\hat{S}^2_{E14}$	409266406	100.19	21893902	100.01
$\hat{S}^2_{E15}$	408990614	100.26	21893721	100.01
$\hat{S}^2_{E16}$	408650755	100.34	21893498	100.01
$\hat{S}^2_{E17}$	408157840	100.46	21893156	100.01
$\hat{S}^2_{E18}$	407796113	100.55	21892274	100.01
$\hat{S}^2_{E19}$	406576970	100.85	21891734	100.02
$\hat{S}^2_{E20}$	425355684	96.40	21883109	100.05
$\hat{S}^2_{E21}$	409309476	100.18	21893840	100.01
$\hat{S}^2_{E22}$	407960318	100.51	21891990	100.01
Proposed:				
$\hat{S}^2_{P1}$	403711178	101.57	14171434	154.50
$\hat{S}^2_{P2}$	403715386	101.57	14176295	154.45
$\hat{S}^2_{P3}$	403557550	101.61	14170280	154.51
$\hat{S}^2_{P4}$	403810934	101.54	14173262	154.48
$\hat{S}^2_{P5}$	403816787	101.54	14178797	154.42
$\hat{S}^2_{P6}$	403592339	101.60	14171948	154.50
$\hat{S}^2_{P7}$	403717552	101.57	14171563	154.50
$\hat{S}^2_{P8}$	403721868	101.57	14176471	154.45
$\hat{S}^2_{P9}$	403559699	101.61	14170397	154.51
$\hat{S}^2_{P10}$	403552638	101.61	14166083	154.56
$\hat{S}^2_{P11}$	403553952	101.61	14168966	154.53
$\hat{S}^2_{P12}$	403511522	101.62	14165398	154.57

exist outliers in Populations III and IV, so we believe that the proposed class of estimators perform well against their competitor estimators.

The MSE and PREs values of the existing and the proposed estimators in the presence of outliers are reported in Table 4. It is noted that when outliers are present, the suggested estimators based on PWM, G, D, and B_n show good resistance and perform efficiently against the existing estimators (cf. Table 4). On the other hand, the existing estimators perform badly in the presence of outliers in comparison to the proposed estimators (larger MSE values and smaller PREs values, (cf. Table 4)). For case with outliers, the estimator $\hat{S}^2_{P12}$ performs well followed by the other proposed estimators for Populations III and IV (cf. Table 4). So, we can say that proposed estimators offer higher resistance against outliers as compared to the existing estimators considered in this study.

5. Concluding remarks

This study proposes a class of estimators for estimating population variance using robust measures of dispersion including probability-weighted moments, Gini, Downton, and B_n measures. The performance of the proposed estimators is compared with the competing estimators suggested by Isaki [3], Upadhyaya and Singh [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6–9], and Khan and Shabbir [10]. The mean square error and the percentage relative efficiency are used as performance metrics. It is found that the suggested estimators offer higher efficiency than existing estimators due to smaller values of mean square error and higher percentage relative efficiency. When there are outliers in the data, the proposed estimators based on robust dispersion measures offer quite robust behavior relative to the existing estimators. In brief, the proposed estimators based on PWM, G, D and B_n behave very well under empirical and outliers studies. The scope of the current work can also be extended to other sampling techniques such as stratified sampling, ranked set sampling, and systematic sampling.

References

1. Solanki, R.S., Singh, H.P., and Pal, S.K. "Improved ratio-type estimators of finite population variance using quartiles", *Hacettepe Journal of Mathematics and Statistics*, **44**(3), pp. 747–754 (2015).
2. Cochran, W.G., *Sampling Techniques*, 3rd Edn., John Wiley and Sons, New York (1977).
3. Isaki, C.T. "Variance estimation using auxiliary information", *Journal of the American Statistical Association*, **78**, pp. 117–123 (1983).
4. Upadhyaya, L.N. and Singh., H.P. "An estimator for population variance that utilizes the kurtosis of an auxiliary variable in sample surveys", *Vikram Mathematical Journal*, **19**, pp. 14–17 (1999).
5. Kadilar, C. and Cingi, H. "Ratio estimators for population variance in simple and stratified sampling", *Applied Mathematics and Computation*, **173**, pp. 1047–1058 (2006).

6. 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).

7. 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).

8. 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).

9. 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).

10. 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).

11. Hussain, Z. and Shabbir, J. "Estimation of the mean of a socially undesirable characteristics", *Scientia Iranica*, **20**(3), pp. 839–845 (2013).

12. Zamanzade, E. and Vock, M. "Variance estimation in ranked set sampling using a concomitant variable", *Statistics and Probability Letters*, **105**, pp. 1–5 (2015).

13. Yaqub, M. and Shabbir, J. "An improved class of estimators for finite population variance", *Hacettepe Journal of Mathematics and Statistics*, **45**(5), pp. 1641–1660 (2016).

14. 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. 11311–1323 (2016).

15. Maqbool, S. and Javaid, S. "Variance estimation using linear combination of tri-mean and quartile average", *American Journal of Biological and Environmental Statistics*, **3**(1), p. 5 (2017).

16. Adichwal, N.K., Sharma, P., and Singh, R. "Generalized class of estimators for population variance using information on two auxiliary variables", *International Journal of Applied and Computational Mathematics*, **3**(2), pp. 651–661 (2017).

17. Maji, R., Singh, G.N., and Bandyopadhyay, A. "Effective estimation strategy of finite population variance using multi-auxiliary variables in double sampling", *Journal of Modern Applied Statistical Methods*, **16**(1), pp. 158–178 (2017).

18. Zamanzade, E. and Wang, X. "Estimation of population proportion for judgment post-stratification", *Computational Statistics and Data Analysis*, **112**, pp. 257–269 (2017).

19. Zamanzade, E. and Mahdizadeh, M. "A more efficient proportion estimator in ranked set sampling", *Statistics and Probability Letters*, **129**, pp. 28–33 (2017).

20. Singh, H.P., Pal, S.K., and Yadav, A. "A study on the chain ratio-ratio-type exponential estimator for finite population variance", *Communications in Statistics-Theory and Methods*, **47**(6), pp. 1442–1458 (2018).

21. Zamanzade, E. and Wang, X. "Proportion estimation in ranked set sampling in the presence of tie information", *Computational Statistics*, **33**(3), pp. 1349–1366 (2018).

22. Hussain, S., Song, L., Ahmad, S., and Riaz, M. "On auxiliary information based improved EWMA median control charts", *Scientia Iranica*, **25**(2), pp. 954–982 (2018).

23. Muneer, S., Khalil, A., Shabbir, J., and Narjis, G. "A new improved ratio-product type exponential estimator of finite population variance using auxiliary information", *Journal of Statistical Computation and Simulation*, **88**(16), pp. 3179–3192 (2018).

24. Mahdizadeh, M. and Zamanzade, E. "Efficient body fat estimation using multistage pair ranked set sampling", *Statistical Methods in Medical Research*, **28**(1), pp. 223–234 (2019).

25. Abid, M., Ahmed, S., Tahir, M., Nazir, H.Z., and Riaz, M. "Improved ratio estimators of variance based on robust measures", *Scientia Iranica*, **26**(4), pp. 2484–2494 (2019).

26. Abid, M., Abbas, N., Sherwani, R.A.K., and Nazir, Z.A. "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).

27. Naz, F., Abid, M., Nawaz, T., and Pang, T. "Enhancing the efficiency of the ratio-type estimators of population variance with a blend of information on robust location measures", *Scientia Iranic*, **27**, pp. 2040–2056 (2020).

28. Muhammad, F. and Riaz, M. "Probability weighted moments approach to quality control charts", *Economic Quality Control*, **21**, pp. 251–260 (2006).

29. Gini, C. "Variabilitá e mutabilitá, contributo allo studio delle distribuzioni e delle relazioni statistiche", Studi Economico-Giuridici della R. Universitá di Cagliari, **3**, pp. 153–159 (1912).

30. Downton, F. "Linear estimates with polynomial coefficients", *Biometrika*, **53**, pp. 129–141 (1966).

31. Bickel, P.J. and Lehmann, E.L. "Descriptive statistics for nonparametric models III: dispersion", In *Contributions to Statistics*, Hájek Memorial Volume, J. Jurekova, Ed., Academia, Prague, pp. 33–40 (1979).

32. David, H.A. "Gini's mean difference rediscovered", *Biometrika*, **55**, pp. 573–575 (1968).

33. Shamos, M.I. "Geometry and statistics: problems at the interface", In *New York Directions and Recent Results in Algorithms and Complexity*, J.F. Traub, Ed., Academic Press, New York, pp. 251–280 (1976).

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

35. ISPRA. (2004). http://www.isprambiente.gov.it/en/databases

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 the National Accounts Wing, Pakistan Bureau of Statistics (PBS) during 2010–2011. He is now serving as an Assistant Professor in the Department of Statistics, Government College University, Faisalabad, Pakistan, from 2017 to the present. He has published more than 60 research papers in research journals. His research interests include statistical process control, Bayesian statistics, non-parametric techniques, and survey sampling.

Rehan Ahmed Khan Sherwani is working as an Associate Professor in the College of Statistical and Actuarial Sciences, University of the Punjab, Lahore, Pakistan. His fields of interest are multilevel models, statistical process control, and computational statistics.

Muhammad Tahir obtained his PhD in Statistics from Quaid-i-Azam University, Islamabad, Pakistan, in 2017. Currently, he is an Assistant Professor of Statistics at 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 M.Phil degrees in Statistics from the Department of Statistics, Quaid-i-Azam University, Islamabad, Pakistan, in 2006 and 2008, respectively. He received his PhD in Statistics from the Institute of Business and in Industrial Statistics University of Amsterdam, The Netherlands, in 2014. He served as a Lecturer in the Department of Statistics from, University of Sargodha, Pakistan, during 2009–2014. He is now serving as an Assistant Professor in the Department of Statistics, University of Sargodha, Pakistan, from 2015 until the present. 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 holds the position of Professor in 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.