



OPEN

## An enhanced estimator of finite population variance using two auxiliary variables under simple random sampling

Sohaib Ahmad<sup>1</sup>, Nitesh Kumar Adichwal<sup>2✉</sup>, Muhammad Aamir<sup>1</sup>, Javid Shabbir<sup>3</sup>, Najwan Alsadat<sup>4</sup>, Mohammed Elgarhy<sup>5</sup> & Hijaz Ahmad<sup>6,7,8,9</sup>

In this article, we have suggested a new improved estimator for estimation of finite population variance under simple random sampling. We use two auxiliary variables to improve the efficiency of estimator. The numerical expressions for the bias and mean square error are derived up to the first order approximation. To evaluate the efficiency of the new estimator, we conduct a numerical study using four real data sets and a simulation study. The result shows that the suggested estimator has a minimum mean square error and higher percentage relative efficiency as compared to all the existing estimators. These findings demonstrate the significance of our suggested estimator and highlight its potential applications in various fields. Theoretical and numerical analyses show that our suggested estimator outperforms all existing estimators in terms of efficiency. This demonstrates the practical value of incorporating auxiliary variables into the estimation process and the potential for future research in this area.

The use of auxiliary information can increase the precision of population parameter estimates in survey sampling, assuming that it is strongly associated with the study variable. To get precise estimates of the unknown population characteristics, auxiliary information is essential in parameter selection and at estimation stage. Estimating population variance is specifically important in populations that are likely to be skewed, and the accuracy of the estimation has been the focus of many research studies. By reducing the dispersion of the estimate obtained from various samples, we can obtain a more accurate representation of the population. Variance is a feature that can be used to describe a statistical population, and brilliant efforts have been made to estimate this characteristic as accurately as possible. The goal is to reduce the scattering of the calculated values, and appropriate use of auxiliary variables in survey sampling results in significant decrease in the variance of the estimator of unknown population parameter(s). In many situations, it is necessary to estimate the population variance of the study variable. The variance estimate for the population is crucial in a number of disciplines, such as agriculture, health, biology, and business, where we confront populations that are likely to be skewed. Everywhere in the world, there are differences in the things we do every day. The idea that no two things or persons are exactly alike is one that is widely held. For instance, a farmer must be well aware of how the geographical climate varies over time in order to decide how to sow his land. A clinician must have a thorough understanding of the variations in human hypertension and fever in order to administer the appropriate care.

The estimation of population variance has been a matter of interest for several researchers over the years. Garcia and Cebrian<sup>1</sup> suggested the ratio estimator for the population variance. Upadhyaya et al.<sup>2</sup> recommended a broad class of estimators for variance of the ratio estimator. Chandra and Singh<sup>3</sup> discussed a family of estimators for population variance. Arcos et al.<sup>4</sup> suggested incorporating the auxiliary information available in variance estimation. Kadilar and Cingi<sup>5</sup> discussed improvement in variance estimation in simple random sampling. Grover<sup>6</sup> discussed a correction note on improvement in variance using auxiliary information. Sharma and

<sup>1</sup>Department of Statistics, Abdul Wali Khan University Mardan, Mardan, Pakistan. <sup>2</sup>School of Business, University of Petroleum and Energy Studies, Dehradun, India. <sup>3</sup>Department of Statistics, University of Wah at Wah Cant, Wah, Pakistan. <sup>4</sup>Department of Quantitative Analysis, College of Business Administration, King Saud University, P.O.Box 71115, 11587 Riyadh, Saudi Arabia. <sup>5</sup>Mathematics and Computer Science Department, Faculty of Science, Beni-Suef University, Beni-Suef 62521, Egypt. <sup>6</sup>Section of Mathematics, International Telematic University Uninettuno, Corso Vittorio Emanuele II, 39, 00186 Rome, Italy. <sup>7</sup>Operational Research Center in Healthcare, Near East University, TRNC Mersin 10, 99138 Nicosia, Turkey. <sup>8</sup>Center for Applied Mathematics and Bioinformatics, Gulf University for Science and Technology, Mishref, Kuwait. <sup>9</sup>Department of Computer Science and Mathematics, Lebanese American University, Beirut, Lebanon. ✉email: nnitesh139@gmail.com

Singh<sup>7</sup> discussed a generalized class of estimators for finite population variance. Singh and Solanki<sup>8</sup> suggested improved estimation of finite population variance using auxiliary information. Yadav and Kadilar<sup>9</sup> recommended a two parameter variance estimator using auxiliary information. Adichwal et al.<sup>10</sup> proposed generalized class of estimators for population variance using auxiliary attribute. Adichwal et al.<sup>11</sup> recommended generalized class of estimators for population variance using information on two auxiliary variables. Lone and Tailor<sup>12</sup> suggested estimation of population variance in simple random sampling. Singh and Khalid<sup>13</sup> discussed effective estimation strategy of population variance in two-phase successive sampling under random non-response. Audu et al.<sup>14</sup> proposed difference-cum-ratio estimators for estimating finite population coefficient of variation in simple random sampling. Shahzad et al.<sup>15</sup> suggested a new variance estimators for calibration approach under stratified random sampling. Singh and Khalid<sup>16</sup> suggested a composite class of estimators to deal with the issue of variance estimation under the situations of random non-response in two-occasion successive sampling. Zaman and Bulut<sup>17</sup> suggested a new class of robust ratio estimators for finite population variance. Shahzad et al.<sup>18</sup> discussed use of calibration constraints and linear moments for variance estimation under stratified adaptive cluster sampling. Ashutosh et al.<sup>19</sup> suggested calibration Approach for Variance Estimation of Small Domain. Ahmad<sup>20</sup> discussed improved estimation of population variance under stratified random sampling. Ahmad<sup>21</sup> discussed improved variance estimator using dual auxiliary variable under simple random sampling. Ahmad et al.<sup>22</sup> suggested enhanced generalized class of estimators under simple random sampling.

Numerous authors have contributed to the estimation of population variance as mentioned above. In this article, we have suggested an enhanced estimator, which has valuable contribution and highlighted the importance of auxiliary variables in estimating of population variance. By using actual data and a simulation study, the suggested estimator has minimum mean square error and higher percentage relative efficiency as compared to recent well known existing estimators.

The primary goal of the current work is given by:

1. To propose a new estimator for finite population variance using two auxiliary variables under simple random sampling.
2. The properties i.e. bias and mean squared error of the propose estimator is derived up to the first order of approximation.
3. Through the use of the actual data and a simulation study, the application of the propose estimator is highlighted.
4. A comparison of estimators is done with existing estimators in terms of minimum mean square error.
5. Our proposed estimator provides a novel and valuable contribution to the field of population variance estimation, and we believe it has the potential to be useful in a wide range of applications.

### Notations and symbols

Consider  $\omega = (\omega_1, \omega_2, \dots, \omega_N)$  is a population comprised of size  $N$ , consider a sample of size  $n$  is chosen from  $\omega$  by using simple random sampling without replacement (SRSWOR). Let  $Y$  be the population study variable, and  $X_1$  and  $X_2$  denote the two auxiliary variables respectively. Consider population variance of  $Y$ ,  $X_1$ , and  $X_2$  are denoted by  $S_y^2$ ,  $S_{x_1}^2$ , and  $S_{x_2}^2$ . Let  $s_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1}$ ,  $s_{x_1}^2 = \frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}{n-1}$ ,  $s_{x_2}^2 = \frac{\sum_{i=1}^n (x_{2i} - \bar{x}_2)^2}{n-1}$ , be the sample variances.

$S_y^2 = \frac{\sum_{i=1}^N (y_i - \bar{Y})^2}{N-1}$ ,  $S_{x_1}^2 = \frac{\sum_{i=1}^N (x_{1i} - \bar{X}_1)^2}{N-1}$ ,  $S_{x_2}^2 = \frac{\sum_{i=1}^N (x_{2i} - \bar{X}_2)^2}{N-1}$ , be the population variances. Let  $\bar{y}$ ,  $\bar{x}_1$ , and  $\bar{x}_2$  be the sample mean.

To derive the bias and mean square error, we consider the following error terms:

$$\xi_0 = \frac{s_y^2 - S_y^2}{S_y^2}, \xi_1 = \frac{s_{x_1}^2 - S_{x_1}^2}{S_{x_1}^2}, \xi_2 = \frac{s_{x_2}^2 - S_{x_2}^2}{S_{x_2}^2}, E(\xi_i) = 0, \text{ for } (i = 0, 1, 2).$$

$$E(\xi_0^2) = \hat{\lambda}_{400}^*, E(\xi_1^2) = \hat{\lambda}_{040}^*, E(\xi_2^2) = \hat{\lambda}_{004}^*, E(\xi_0 \xi_1) = \hat{\lambda}_{220}^*, E(\xi_0 \xi_2) = \hat{\lambda}_{202}^*, E(\xi_1 \xi_2) = \hat{\lambda}_{022}^*,$$

$$\hat{\lambda}_{lmn} = \frac{v_{lmn}}{v_{200}^{1/2} v_{020}^{m/2} v_{002}^{n/2}}, v_{lmn} = \frac{\sum (y_i - \bar{Y})^l (x_{1i} - \bar{X}_1)^m (x_{2i} - \bar{X}_2)^n}{N-1}.$$

where  $\hat{\lambda}_{lmn}^* = (\hat{\lambda}_{lmn} - 1)$ , and  $\hat{\lambda} = (\frac{1}{n} - \frac{1}{N})$ , where  $l$ ,  $m$ , and  $n$  be the positive numerals and  $v_{200}$ ,  $v_{020}$ , and  $v_{002}$  are the second-order moments about means of  $Y$ ,  $X_1$ , and  $X_2$ , and  $\hat{\lambda}_{lmn}$  is the moment ratio.

The article has been structured as follows:

In Section "Introduction", we have provided a brief overview of the topic and highlighted the importance of estimating population variance. In Section "Notations and symbols", we review some of the existing estimators for estimation of population variance. In Section "Existing estimators", we proposed a new estimator for population variance under simple random sampling. In Section "Proposed estimator", we will compare the proposed estimator with the existing estimators in terms of MSE. In Section "Efficiency assessment", we will present a numerical study to illustrate the performance of the proposed estimator and compare it with the existing estimators. In Section "Numerical study", we will perform simulations to validate the results of the numerical study and provide additional insights into the performance of the estimators. In Section "Simulation study", we will discuss the results of our study, the limitations of the proposed estimator, and the scope for future research in this field. In Section "Discussion", we will summarize the main findings of our study and provide concluding remarks on the estimation of population variance under simple random sampling.

## Existing estimators

In this section, we study different variance estimators based on simple random sampling that are available in the literature.

- (i) The usual estimator  $T_0$ , is given by:

$$T_0 = \frac{\sum_{i=1}^n (y_i - \bar{Y})^2}{n-1} \quad (1)$$

The variance of  $T_0$ , is given by:

$$\text{Var}(T_0) = \hat{\lambda} S_y^4 \lambda_{400}^* \quad (2)$$

- (ii) Isaki<sup>23</sup> suggested a ratio estimator, which is given by:

$$T_1 = s_y^2 \left( \frac{S_x^2}{s_x^2} \right) \quad (3)$$

The bias and mean square error of  $T_1$  are given as:

$$\text{Bias}(T_1) = \hat{\lambda} S_y^2 \{ \lambda_{040}^* - \lambda_{220}^* \},$$

and

$$\text{MSE}(T_1) \cong \hat{\lambda} S_y^4 \{ \lambda_{400}^* + \lambda_{040}^* - 2\lambda_{220}^* \}. \quad (4)$$

- (iii) The difference type estimator is given by:

$$T_2 = s_y^2 + k(S_x^2 - s_x^2). \quad (5)$$

where  $k$  is the constant, i.e.  $k = \left[ \frac{S_y^2 \lambda_{220}^*}{S_x^2 \lambda_{040}^*} \right]$ .

The minimum variance at the optimal value of  $k$  is given by:

$$\text{Var}(T_2) = S_y^4 \hat{\lambda}_{400}^* (1 - \rho^2), \quad (6)$$

where  $\rho = \frac{\lambda_{220}^*}{(\lambda_{400}^*)^{\frac{1}{2}} (\lambda_{040}^*)^{\frac{1}{2}}}$ .

- (iv) The difference type estimator given by:

$$T_3 = \left\{ \omega_1 s_y^2 + \omega_2 (S_x^2 - s_x^2) \right\}, \quad (7)$$

where  $\omega_1$  and  $\omega_2$  is the unidentified constants:

$$\omega_1 = \frac{\lambda_{040}^*}{\hat{\lambda} \{ \lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2} \} + \lambda_{040}^*},$$

and

$$\omega_2 = \frac{S_y^2 \lambda_{220}^*}{S_x^2 \{ \hat{\lambda} (\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}) + \lambda_{040}^* \}}.$$

The bias and minimum MSE at the optimal value of  $\omega_1$  and  $\omega_2$ , are given by:

$$\text{Bias}(T_3) = S_y^2 \left\{ \frac{\lambda_{040}^*}{\hat{\lambda} \{ \lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2} \} + \lambda_{040}^*} - 1 \right\},$$

and

$$\text{MSE}(T_3) = \frac{\hat{\lambda} S_y^4 [\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}]}{\hat{\lambda} [\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}] + \lambda_{040}^*} \quad (8)$$

- (v) The exponential ratio and product type estimators by Singh et al.<sup>24</sup>:

$$T_4 = s_y^2 \exp \left\{ \frac{S_x^2 - s_x^2}{S_x^2 + s_x^2} \right\} \quad (9)$$

$$T_5 = s_y^2 \exp\left\{\frac{s_x^2 - S_x^2}{s_x^2 + S_x^2}\right\} \tag{10}$$

The bias and MSE of  $T_4$  and  $T_5$  are given by:

$$\text{Bias}(T_4) = \lambda S_y^2 \left\{ \frac{3}{8} \lambda_{040}^* - \frac{1}{2} \lambda_{220}^* \right\}, \text{MSE}(T_4) = \lambda S_y^4 \left[ \lambda_{400}^* + \frac{1}{4} \lambda_{040}^* - \lambda_{220}^* + \frac{1}{4} \right] \tag{11}$$

and

$$\text{Bias}(T_5) = \lambda S_y^2 \left\{ \frac{1}{2} \lambda_{220}^* - \frac{1}{8} \lambda_{040}^* \right\}, \text{MSE}(T_5) = \lambda S_y^4 \left[ \lambda_{400}^* + \frac{1}{4} \lambda_{040}^* - \lambda_{220}^* + \frac{9}{4} \right] \tag{12}$$

(vi) Grover and Kaur<sup>25</sup>, suggested the following estimator, which is given by:

$$T_6 = \left[ \omega_3 s_y^2 + \omega_4 (S_x^2 - s_x^2) \right] \exp\left(\frac{S_x^2 - s_x^2}{S_x^2 + s_x^2}\right) \tag{13}$$

The bias of  $T_6$ , is given by:

$$\text{Bias}(T_6) = \left[ -S_y^2 + \omega_3 S_y^2 \left\{ 1 + \frac{3}{8} \lambda_{040}^* - \frac{1}{2} \lambda_{220}^* \right\} + \frac{1}{2} \omega_4 S_x^2 \lambda_{040}^* \right],$$

The minimum values of  $\omega_3$  and  $\omega_4$  are given by:

$$\omega_3 = \frac{\{8\lambda_{040}^* - \lambda_{220}^{*2}\}}{8[\lambda(\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}) + \lambda_{040}^*]},$$

and

$$\omega_4 = \frac{S_y^2 (\lambda_{040}^{*2} - \lambda_{040}^* \lambda_{220}^* + 8\lambda_{220}^* - 4\{\lambda_{040}^* - \lambda(\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2})\})}{8S_x^2 [\lambda\{\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}\} + \lambda_{040}^*]}.$$

The minimum MSE at the optimum values of  $\omega_3$  and  $\omega_4$ , are given by:

$$\text{MSE}(T_6) = \frac{S_y^4 [64\{\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}\} - \lambda_{040}^{*3} - 16\lambda_{040}^* \{\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}\}]}{64\{\lambda(\lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}) + \lambda_{040}^*\}}. \tag{14}$$

### Proposed estimator

The use of auxiliary variables can improve the precision and efficiency of an estimator during both the design and estimation procedures. Taking motivation from the work of Ahmad et al.<sup>26</sup>, we propose an enhanced estimator for estimation of finite population variance using two auxiliary variables under simple random sampling. Our suggested variance estimator based on simple random sampling is more efficient and flexible as compared to all the existing estimators considered in this study, which is given by:

$$T_{prop} = \left[ \omega_5 s_y^2 + \omega_6 (S_{x1}^2 - s_{x1}^2) + \omega_7 (S_{x2}^2 - s_{x2}^2) \right] \exp\left(\frac{S_{x1}^2 - s_{x1}^2}{S_{x1}^2 + s_{x1}^2}\right), \tag{15}$$

where  $\omega_5$ ,  $\omega_6$  and  $\omega_7$  are unknown constants.

$$T_{prop} = \left[ \omega_5 S_y^2 (1 + \xi_0) - \omega_6 S_{x1}^2 \xi_1 - \omega_7 S_{x1}^2 \xi_2 \right] \left( 1 - \frac{\xi_1}{2} + \frac{3\xi_1^2}{8} + \dots \right).$$

After expanding the above equation, we have

$$\begin{aligned} (T_{prop} - S_y^2) &= -S_y^2 + \omega_5 S_y^2 + \omega_5 S_y^2 \xi_0 - \frac{1}{2} \omega_5 S_y^2 \xi_1 - \omega_6 S_{x1}^2 \xi_1 - \omega_7 S_{x2}^2 \xi_2 + \frac{1}{2} \omega_5 S_y^2 \xi_0 \xi_1 \\ &+ \frac{3}{8} \omega_5 S_y^2 \xi_1^2 + \frac{1}{2} \omega_6 S_{x1}^2 \xi_1^2 + \frac{1}{2} \omega_7 S_{x2}^2 \xi_1 \xi_2 \end{aligned} \tag{16}$$

The bias and mean square error of  $T_{prop}$ , is given by:

$$\begin{aligned}
 \text{Bias}(T_{prop}) &\cong S_y^2(\omega_5 - 1) + \frac{3}{8}S_y^2\omega_5\lambda_{040}^* + \frac{1}{2}S_x^2\omega_6\lambda_{040}^* - \frac{1}{2}S_y^2\omega_5\lambda_{220}^* + \frac{1}{2}S_{x2}^2\lambda_{022}^* \\
 \text{MSE}(T_{prop}) &\cong S_y^4(\omega_5 - 1)^2 + S_y^4\lambda_{400}^*\omega_5^2 + S_{x1}^4\lambda_{040}^*\omega_6 + S_{x2}^2\lambda_{004}^*\omega_7 + S_y^4\omega_5^2\lambda_{040}^* - S_{y^2}^2S_{x1}^2\omega_6\lambda_{040}^* \\
 &\quad + 2S_y^2S_{x1}^2\omega_5\omega_6\lambda_{040}^* - \frac{3}{4}S_y^4\omega_5\lambda_{040}^* + S_y^4\omega_5\lambda_{220}^* - 2S_y^4\omega_5^2\lambda_{220}^* - 2S_{y^2}^2S_{x1}^2\omega_5\omega_6\lambda_{220}^* \\
 &\quad - 2S_y^2S_{x2}^2\omega_5\omega_7\lambda_{202}^* - S_{y^2}^2S_{x2}^2\omega_7\lambda_{022}^* + 2S_y^2S_{x2}^2\omega_5\omega_7\lambda_{022}^* - 2S_{y^2}^2S_{x2}^2\omega_6\omega_7\lambda_{022}^*
 \end{aligned} \tag{17}$$

The optimal values of  $\omega_5$ ,  $\omega_6$ , and  $\omega_7$  are obtained by minimizing (17), and are given by:

$$\begin{aligned}
 \omega_{5(Opt)} &= \frac{8 - \lambda_{040}^*}{\lambda_{400}^* \left[ 8 \left\{ \frac{1}{\lambda_{400}^*} + \hat{\lambda}(\gamma^* + 1) \right\} \right]}, \\
 \omega_{6(Opt)} &= \frac{S_y^2 \left[ \lambda_{040}^* (\lambda_{040}^*\lambda_{004}^* - \lambda_{220}^*) + (\lambda_{004}^*\lambda_{220}^* - \lambda_{202}^*\lambda_{022}^*) (8 - \lambda_{040}^*) + 4\lambda_{400}^* (\lambda_{040}^*\lambda_{004}^* - \lambda_{220}^*) \right]}{8S_{x^2}^2\lambda_{400}^* (\lambda_{040}^*\lambda_{004}^* - \lambda_{220}^*) \left\{ \left( \frac{1}{\lambda_{400}^*} \right) + \hat{\lambda}(\gamma^* + 1) \right\}},
 \end{aligned}$$

and

$$\omega_{7(Opt)} = \frac{S_y^2 [8 - \lambda_{040}^*] \{ \lambda_{220}^*\lambda_{002}^* - \lambda_{040}^*\lambda_{202}^* \}}{8S_{x^2}^2\lambda_{400}^* (\lambda_{040}^*\lambda_{004}^* - \lambda_{220}^*) \left\{ \left( \frac{1}{\lambda_{400}^*} \right) + \hat{\lambda}(\gamma^* + 1) \right\}}.$$

The minimal MSE at the optimum values of  $\omega_{5(Opt)}$ ,  $\omega_{6(Opt)}$  and  $\omega_{7(Opt)}$ , are given by:

$$\text{MSE}(T_{prop}) = \frac{\hat{\lambda}S_y^4 \left[ 64(\gamma^* + 1) - \hat{\lambda} \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\hat{\lambda}\lambda_{040}^*(\gamma^* + 1) \right]}{64 \left\{ \frac{1}{\lambda_{400}^*} + \hat{\lambda}(\gamma^* + 1) \right\}}, \tag{18}$$

where

$$\gamma^* = \frac{2\lambda_{220}^*\lambda_{202}^*\lambda_{022}^* - \lambda_{040}^*\lambda_{202}^* - \lambda_{004}^*\lambda_{220}^*}{\lambda_{400}^* (\lambda_{040}^*\lambda_{004}^* - \lambda_{220}^*)}.$$

### Efficiency assessment

In this section, we compare the proposed estimator with existing estimators in terms of mean square error.

(1) By taking (2) and (18)

$$\text{Var}(T_0) - \text{MSE}(T_{prop}) > 0$$

$$S_y^4\lambda_{400}^* - \frac{\hat{\lambda}S_y^2 \left[ 64(\gamma^* + 1) - \hat{\lambda} \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\hat{\lambda}\lambda_{040}^*(\gamma^* + 1) \right]}{64 \left\{ \frac{1}{\lambda_{400}^*} + \hat{\lambda}(\gamma^* + 1) \right\}} > 0 \tag{19}$$

$$\frac{\hat{\lambda}S_y^2 \left[ 64S_{y^2}^2\lambda_{400}^* \{ \Psi_1 \} + \left[ 16\hat{\lambda}\lambda_{040}^*(\gamma^* + 1) - 64\gamma^* + \hat{\lambda} \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 64 \right] \right]}{64\Psi_1} > 0 \tag{20}$$

where  $\Psi_1 = \frac{1}{\lambda_{400}^*} + \hat{\lambda}(\gamma^* + 1)$ ,

$$\Psi_2 = \left[ 64\gamma^* + 64 - \hat{\lambda} \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\hat{\lambda}\lambda_{040}^*(\gamma^* + 1) \right],$$

$$\Psi_3 = \frac{1}{4}\hat{\lambda}\lambda_{040}^*(\gamma^* + 1) - \gamma^* + \frac{1}{64}\hat{\lambda} \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 1$$

(2) By taking (4) and (18)

$$\text{MSE}(T_1) - \text{MSE}(T_{prop}) > 0$$

$$\lambda S_y^4 \{ \lambda_{400}^* + \lambda_{040}^* - 2\lambda_{220}^* \} - \frac{\lambda S_y^2 \left[ 64(\gamma^* + 1) - \lambda \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\lambda_{040}^* (\gamma^* + 1) \right]}{64\Psi_1} > 0 \tag{21}$$

$$\frac{\lambda S_y^2 (S_y^2 \mathcal{Y} \Psi_1 + \Psi_3)}{\Psi_1} > 0 \tag{22}$$

(3) By taking (6) and (18)

$$\text{Var}(T_2) - \text{MSE}(T_{prop}) > 0$$

$$S_y^4 \lambda_{400}^* (1 - \rho^2) - \frac{\lambda S_y^2 \left[ 64(\gamma^* + 1) - \lambda \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\lambda_{040}^* (\gamma^* + 1) \right]}{64 \left\{ \frac{1}{\lambda_{400}^*} + \lambda(\gamma^* + 1) \right\}} > 0 \tag{23}$$

$$\frac{\lambda S_y^2 (S_y^2 \lambda_{400}^* (1 - \rho^2) \Psi_1 + \Psi_3)}{\Psi_1} > 0 \tag{24}$$

(4) By taking (8) and (18)

$$\text{MSE}(T_3) - \text{MSE}(T_{prop}) > 0$$

$$\frac{\lambda S_y^4 [\mathcal{Y}]}{\lambda[\mathcal{Y}] + \lambda_{040}^*} - \frac{\lambda S_y^2 \left[ 64(\gamma^* + 1) - \lambda \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\lambda_{040}^* (\gamma^* + 1) \right]}{64 \left\{ \frac{1}{\lambda_{400}^*} + \lambda(\gamma^* + 1) \right\}} > 0 \tag{25}$$

$$\frac{\frac{1}{64} \left[ \lambda S_y^2 (\lambda_{040}^* \Psi_2) + \Psi_1 S_y^2 + \Psi_2 (\mathcal{Y}) \right]}{\lambda([\mathcal{Y}] + \lambda_{040}^*) \Psi_1} > 0. \tag{26}$$

where  $\mathcal{Y} = \lambda_{400}^* \lambda_{040}^* - \lambda_{220}^{*2}$   
 (5) By taking (11) and (18)

$$\text{MSE}(T_4) - \text{MSE}(T_{prop}) > 0$$

$$\lambda S_y^4 \left[ \lambda_{400}^* + \frac{1}{4} \lambda_{040}^* - \lambda_{220}^* + \frac{1}{4} \right] - \frac{\lambda S_y^2 \left[ 64(\gamma^* + 1) - \lambda \left( \frac{\lambda_{040}^{*2}}{\lambda_{400}^*} \right) - 16\lambda_{040}^* (\gamma^* + 1) \right]}{64 \left\{ \frac{1}{\lambda_{400}^*} + \lambda(\gamma^* + 1) \right\}} > 0 \tag{27}$$

$$\frac{\lambda S_y^2 \left[ S_y^2 \left\{ \lambda_{400}^* + \frac{1}{4} \lambda_{040}^* - \lambda_{220}^* + \frac{1}{4} \right\} \Psi_1 + \Psi_3 \right]}{\Psi_1} > 0 \tag{28}$$

(6) By taking (12) and (18)

$$\text{MSE}(T_5) - \text{MSE}(T_{prop}) > 0$$

$$\lambda S_y^4 \left[ \lambda_{400}^* + \frac{1}{4} \lambda_{040}^* - \lambda_{220}^* + \frac{9}{4} \right] - \frac{\lambda S_y^2 \left[ 64(\gamma^* + 1) - \lambda \left( \frac{\lambda_{040h}^{*2}}{\lambda_{400h}^*} \right) - 16\lambda_{040h}^* (\gamma^* + 1) \right]}{64 \left\{ \frac{1}{\lambda_{400h}^*} + \lambda(\gamma^* + 1) \right\}} > 0 \tag{29}$$

$$\frac{\lambda S_y^2 \left[ S_y^2 \left\{ \lambda_{400}^* + \frac{1}{4} \lambda_{040}^* - \lambda_{220}^* + \frac{9}{4} \right\} \Psi_1 + \Psi_3 \right]}{\Psi_1} > 0 \tag{30}$$

(7) By taking (14) and (18)

$$\text{MSE}(T_6) - \text{MSE}(T_{prop}) > 0$$

Parameters	Populations			
	I	II	III	IV
$N$	80	33	69	34
$N$	24	8	25	10
$\hat{\lambda}$	0.02916667	0.09469697	0.02550725	0.7058824
$\bar{Y}$	5182.637	27.49091	4514.899	765.3529
$\bar{X}_1$	285.125	3.727273	4954.435	218.4118
$\bar{X}_2$	1126.463	72.54545	4591.072	201.4118
$S_y^2$	3,369,642	102.6327	37,199,578	222,931.4
$S_{x1}^2$	73,132.09	2.329545	49,829,270	28,123.22
$S_{x2}^2$	715,055.8	111.8807	39,881,874	23,154.86
$C_y$	0.3541939	0.3685139	1.350893	0.6169129
$C_{x1}$	0.9484593	0.4094911	1.424781	0.7678148
$C_{x2}$	0.7506772	0.1458033	1.375541	0.755503
$\rho_{yx1}$	0.9149811	0.432738	0.9601401	0.8307546
$\rho_{yx2}$	0.9413055	0.2521603	0.956907	0.8992855
$\rho_{x1x2}$	0.9884207	- 0.06598959	0.972902	0.929922
$\hat{\lambda}_{400}^*$	1.210342	4.382758	6.544991	1.583558
$\hat{\lambda}_{040}^*$	2.491817	1.245254	8.69773	2.091393
$\hat{\lambda}_{004}^*$	1.79522	1.015036	8.698631	2.341373
$\hat{\lambda}_{220}^*$	1.294326	0.388943	7.053057	1.004473
$\hat{\lambda}_{202}^*$	1.193045	1.218625	7.190267	1.607816
$\hat{\lambda}_{022}^*$	2.093142	0.4465917	8.523088	1.433162

**Table 1.** Summary statistics of population’s I–IV.

$$\frac{S_y^4 [64\{Y\} - \hat{\lambda}_{040}^{*3} - 16\hat{\lambda}_{040}^*\{Y\}]}{64\{\hat{\lambda}(Y) + \hat{\lambda}_{040}^*\}} - \frac{\hat{\lambda}S_y^4 [64(\gamma^* + 1) - \hat{\lambda}\left(\frac{\lambda_{040}^{*2}}{\lambda_{400}^*}\right) - 16\hat{\lambda}_{040}^*(\gamma^* + 1)]}{64\{\Psi_1\}} > 0. \tag{31}$$

### Numerical study

In this section, we conduct a mathematical exploration to compare the performance of our suggested estimator with existing estimators. We consider four actual datasets that are sampled using simple random sampling, and their descriptions are provided in Table 1. To compare the performance of the estimators, we use the concept of percentage relative efficiency (PRE), which is a widely used measure in statistics.

To calculate the PRE, we use the following expression:

$$PRE(.) = \frac{Var(T_u)}{MSE(T_{prop})} \times 100$$

where  $(u) = (0, 1, 2, 3, 4, 5, 6)$ .

The expression of mean square error (MSE) using four actual data sets are given in Table 2. The conditional values of estimators are given in Table 3. The expression of percentage relative efficiency (PRE) using four actual data sets are given in Table 4.

**Population 1:** [Source: Murthy<sup>27</sup>]

$Y$ : Output of the workshop,  
 $X_1$ =Fixed capital,  
 $X_2$ = Number of worker.

**Population 2:** [Source: Cochran<sup>28</sup>]

$Y$ : Food cost,  
 $X_1$ : size of the family,  
 and  $X_2$ = Income.

**Population 3:** [Source: Singh<sup>29</sup>]

$Y$ : Estimated fish caught in the year 1995,  
 $X_1$ : Estimated fish caught in the year 1994,  
 $X_2$ = Estimated fish caught in the year 1993.

**Population 4:** [Source: Sukhathme<sup>30</sup>]

$Y$ : Area under wheat in 1937,  
 $X_1$ = Area under wheat in 1936,  
 $X_2$ = Total cultivated area in 1931.

Estimators	Data-I	Data-II	Data-III	Data-IV
T <sub>0</sub>	400,831,931,440	4371.75	231,019,500,000,000	5,555,316,462
T <sub>1</sub>	368,762,720,252	4837.944	40,118,900,000,000	5,844,551,874
T <sub>2</sub>	178,180,420,058	4250.572	29,141,750,000,000	3,862,867,615
T <sub>3</sub>	175,427,522,009	3028.487	28,540,710,000,000	3,584,275,528
T <sub>4</sub>	261,285,129,365	4543.687	67,642,260,000,000	4,742,748,835
T <sub>5</sub>	923,630,298,554	6538.664	138,236,600,000,000	11,758,993,996
T <sub>6</sub>	171,317,454,680	2937.575	25,915,450,000,000	3,436,287,618
T <sub>prop</sub>	75,030,307,291	2200.201	18,697,750,000,000	1,549,496,322

**Table 2.** Mean square error using population's I-IV.

### Simulation study

This portion shows the algebraic method for the evaluation of T<sub>prop</sub> with the existing counterparts T<sub>0</sub>, T<sub>1</sub>, T<sub>2</sub>, T<sub>3</sub>, T<sub>4</sub>, T<sub>5</sub>, and T<sub>6</sub>. The algorithm for simulation study to demonstrate the performance of numerous estimators of T<sub>0</sub> is as follows:

1. Produce two independent random variables X from N(μ, σ<sup>2</sup>) and Z from N(μ<sub>1</sub>, σ<sub>1</sub><sup>2</sup>) using the box-Muller method.
2. Set Y = ρX + √(1 - ρ<sup>2</sup>), where 0 < ρ = 0.5, 0.6, 0.7, 0.8 > 1.
3. Return the pair (Y, X, Z).
4. Let the population-I with the parameters μ = 3, σ = 2, μ<sub>1</sub> = 5 and σ<sub>1</sub> = 3 and in step-1 and repeat steps 1-3 for 500 times. This population will have different variances for X and Z.
5. Likewise, produce the population-II with the parameters μ = 2, σ = 3, μ<sub>1</sub> = 4 and σ<sub>1</sub> = 3 in step-1 and repeat steps 1-3 for 500 times.
6. From the population of size N = 50, draw 500 SRSWOR (y<sub>i</sub>, x<sub>i</sub>) (i = 1, 2, 3, ..., n) of size n = 10, 15 and 20.
7. The Average MSE of the estimators is defined by:

$$AverageMSE(T) = \frac{1}{500} \sum_{k=1}^{500} E(T_k - T_0)^2$$

8. The percentage relative efficiency of estimators as compared to usual estimator T<sub>0</sub> is defined by:

$$PRE(T) = \frac{Var(T_0) \times 100}{MSE(T)}$$

9. The Average PRE of the estimators is defined by

$$AveragePRE(T) = \frac{1}{500} \sum_{k=1}^{500} PRE(T_k)$$

The results of the simulation study are presented in Tables 5 and 6.

### Discussion

As stated before we take four data sets and simulation, to notice the performance of the proposed enhanced variance estimator based on simple random sampling using two auxiliary variables. The proposed estimator is compared with existing estimators concerning MSE and PRE. The summary statistic of actual data sets is presented in Table 1. The MSE of estimators are given in Table 2 and the conditional values of estimators are given

Conditional values	Data-I	Data-II	Data-III	Data-IV
1	325,801,624,149	2171.549	212,321,800,000,000	4,005,820,140
2	293,732,412,961	2637.743	21,421,150,000,000	4,295,055,552
3	103,150,112,767	2050.371	10,444,000,000,000	2,313,371,293
4	100,397,214,718	828.286	9,842,960,000,000	2,034,779,206
5	186,254,822,074	2343.486	48,944,510,000,000	3,193,252,513
6	848,600,000,000	4338.463	119,538,900,000,000	10,209,497,674
7	96,287,147,389	737.374	7,217,700,000,000	1,886,791,296

**Table 3.** Conditional values using actual data sets.



Estimators	Data-I	Data-II	Data-III	Data-IV
T <sub>0</sub>	100	100	100	100
T <sub>1</sub>	108.6964	90.36379	575.8359	95.0512
T <sub>2</sub>	224.9585	102.8508	792.7441	143.8133
T <sub>3</sub>	228.4886	144.3542	809.4386	154.9913
T <sub>4</sub>	153.4079	96.2159	341.5313	117.1328
T <sub>5</sub>	43.39744	66.85998	167.119	47.24313
T <sub>6</sub>	233.9703	148.8217	891.4353	161.6662
T <sub>prop</sub>	534.2267	198.6978	1235.547	358.524

**Table 4.** Percentage relative efficiency w.r.t T<sub>0</sub> of population's I–IV.

$\rho$	N	Average PRE's								
		T <sub>0</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>prop</sub>	
0.5	10	100.00	47.42	103.86	118.43	68.90	37.88	123.56	124.09	
	15	100.00	44.45	101.50	109.32	67.12	36.44	112.12	112.48	
	20	100.00	45.74	102.28	106.54	66.64	33.71	107.94	108.64	
0.6	10	100.00	42.94	102.18	115.53	65.66	35.89	120.97	121.87	
	15	100.00	53.24	102.08	109.84	71.89	37.77	111.77	112.09	
	20	100.00	33.66	103.00	106.09	56.56	26.63	107.91	108.02	
0.7	10	100.00	18.50	103.24	113.27	42.77	24.99	132.30	133.85	
	15	100.00	39.90	101.76	107.87	62.42	31.34	110.52	111.60	
	20	100.00	48.88	102.37	106.02	67.43	31.61	107.10	109.01	
0.8	10	100.00	29.70	102.52	113.65	55.26	30.24	122.89	123.07	
	15	100.00	36.96	101.31	107.49	60.68	31.13	110.59	111.08	
	20	100.00	49.21	101.69	105.95	68.80	34.50	107.15	107.55	

**Table 5.** Evaluation of T<sub>prop</sub> with other existing counterparts for Population-I.

$\rho$	N	Average PRE's								
		T <sub>0</sub>	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>prop</sub>	
0.5	10	100.00	33.49	103.55	112.82	56.57	27.87	118.74	119.25	
	15	100.00	38.63	101.48	105.76	58.40	25.35	107.66	110.68	
	20	100.00	38.80	102.53	105.45	59.93	26.57	106.80	107.81	
0.6	10	100.00	34.59	101.65	110.67	57.33	27.82	115.92	116.16	
	15	100.00	47.10	102.33	111.01	68.85	38.40	113.77	114.45	
	20	100.00	42.67	103.10	106.61	63.67	30.06	107.98	108.54	
0.7	10	100.00	43.50	102.70	112.68	63.78	30.70	116.54	117.09	
	15	100.00	41.27	102.24	107.71	62.42	29.77	109.94	110.04	
	20	100.00	49.73	103.13	107.11	68.95	33.36	108.30	109.67	
0.8	10	100.00	40.37	101.55	110.84	60.96	29.07	114.79	115.00	
	15	100.00	41.47	103.11	109.56	63.87	32.42	112.27	113.43	
	20	100.00	55.97	103.38	107.88	72.28	36.05	108.90	111.05	

**Table 6.** Evaluation of T<sub>prop</sub> with other existing counterparts for Population-II.

in Table 3. From the mathematical results using actual data, and PRE are given in Table 4, it is observed that the proposed estimator is efficient in terms of efficiency. We evaluated the efficiency of our proposed variance estimator in simple random sampling using a simulation study. From Tables 5 and 6, it can be concluded that for different values of the correlation coefficient  $\rho=0.5, 0.6, 0.7, 0.8$  and the sample size  $n=10, 15, 20$  the performance of proposed estimator  $T_{prop}$  is more efficient in comparison with the usual estimator and other existing estimators. The result of the simulation study clearly shows that the average PRE of the proposed estimator  $T_{prop}$

is higher as compared to the average PRE of the existing estimator for both the simulated data sets I and dataset II. Hence it can be recommended to use it further.

## Conclusion

In this article, we have suggested a novel variance estimator using two auxiliary variables under simple random sampling. We conducted a comprehensive comparison of our suggested estimator with several existing counterparts using four real data sets. We also obtained the properties of the proposed estimator under the first order of approximation and carried out a simulation study to test its robustness and generalizability. The results showed that the suggested estimator outperformed to the existing estimators in terms of efficiency. Future research could focus on extending our proposed estimator to two-phase sampling designs, and to situations with non-response and measurement error, where information from auxiliary variables can be utilized to estimate population variance more accurately. Additionally, it would be interesting to investigate the performance of our proposed estimator in more complex survey settings, such as clustered and stratified sampling.

Received: 31 March 2023; Accepted: 4 October 2023

Published online: 05 December 2023

## References

- Garcia, M. R. & Cebrian, A. A. Repeated substitution method: The ratio estimator for the population variance. *Metrika* **43**(1), 101–105 (1996).
- Upadhyaya, L. N., Singh, H. P. & Singh, S. A class of estimators for estimating the variance of the ratio estimator. *J. Japan Stat. Soc.* **34**(1), 47–63 (2004).
- Chandra, P. & Singh, H. P. A family of estimators for population variance using knowledge of kurtosis of an auxiliary variable in sample survey. *Stat. Transit.* **7**(1), 27–34 (2005).
- Arcos, A., Rueda, M., Martinez, M. D., González, S. & Roman, Y. Incorporating the auxiliary information available in variance estimation. *Appl. Math. Comput.* **160**(2), 387–399 (2005).
- Kadilar, C. & Cingi, H. Improvement in variance estimation in simple random sampling. *Commun. Stat. Theory Methods* **36**(11), 2075–2081 (2007).
- Grover, L. K. A correction note on improvement in variance estimation using auxiliary information. *Commun. Stat. Theory Methods* **39**(5), 753–764 (2010).
- Sharma, P. & Singh, R. A generalized class of estimators for finite population variance in presence of measurement errors. *J. Mod. Appl. Stat. Methods* **12**(2), 13 (2013).
- Singh, H. P. & Solanki, R. S. Improved estimation of finite population variance using auxiliary information. *Commun. Stat. Theory Methods* **42**(15), 2718–2730 (2013).
- Yadav, S. K. & Kadilar, C. A two parameter variance estimator using auxiliary information. *Appl. Math. Comput.* **226**, 117–122 (2014).
- Adichwal, N. K., Sharma, P., Verma, H. K. & Singh, R. Generalized class of estimators for population variance using auxiliary attribute. *Int. J. Appl. Comput. Math.* **2**(4), 499–508 (2016).
- Adichwal, N. K., Sharma, P. & Singh, R. Generalized class of estimators for population variance using information on two auxiliary variables. *Int. J. Appl. Comput. Math.* **3**(2), 651–661 (2017).
- Lone, H. A. & Tailor, R. Estimation of population variance in simple random sampling. *J. Stat. Manag. Syst.* **20**(1), 17–38 (2017).
- Singh, G. N. & Khalid, M. Effective estimation strategy of population variance in two-phase successive sampling under random non-response. *J. Stat. Theor. Pract.* **13**(1), 1–28 (2019).
- Audu, A. *et al.* Difference-cum-ratio estimators for estimating finite population coefficient of variation in simple random sampling. *Asian J. Probab. Stat.* **13**(3), 13–29 (2021).
- Shahzad, U., Ahmad, I., Almanjahie, I. M., Al-Noor, N. H. & Hanif, M. A novel family of variance estimators based on L-moments and calibration approach under stratified random sampling. *Commun. Stat. Simul. Comput.* **1**, 1–14 (2021).
- Singh, G. N. & Khalid, M. A composite class of estimators to deal with the issue of variance estimation under the situations of random non-response in two-occasion successive sampling. *Commun. Stat. Simul. Comput.* **51**(4), 1454–1473 (2022).
- Zaman, T. & Bulut, H. A new class of robust ratio estimators for finite population variance. *Sci. Iran.* **1**, 1 (2022).
- Shahzad, U., Ahmad, I., Al-Noor, N. H. & Benedict, T. J. Use of calibration constraints and linear moments for variance estimation under stratified adaptive cluster sampling. *Soft Comput.* **26**(21), 11185–11196 (2022).
- Ashutosh, A., Shahzad, U., Ahmad, I., Raza, M. A. & Benedict, T. J. Calibration approach for variance estimation of small domain. *Math. Probl. Eng.* **2022**, 1–8 (2022).
- Ahmad, S. *et al.* Improved estimation of finite population variance using dual supplementary information under stratified random sampling. *Math. Probl. Eng.* **2022**(1), 1 (2022).
- Ahmad, S. *et al.* A simulation study: Improved ratio-in-regression type variance estimator based on dual use of auxiliary variable under simple random sampling. *Plos One* **17**(11), e0276540 (2022).
- Ahmad, S. *et al.* A new improved generalized class of estimators for population distribution function using auxiliary variable under simple random sampling. *Sci. Rep.* **13**(1), 5415 (2023).
- Isaki, C. T. Variance estimation using auxiliary information. *J. Am. Stat. Assoc.* **78**(381), 117–123 (1983).
- Singh, R., Chauhan, P., Sawan, N. & Smarandache, F. Improvement in estimating the population mean using exponential estimator in simple random sampling. *Int. J. Stat. Econ.* **3**(A09), 13–18 (2009).
- Grover, L. K. & Kaur, A. Ratio type exponential estimators of population mean under linear transformation of auxiliary variable: Theory and methods. *S. Afr. Stat. J.* **45**(2), 205–230 (2011).
- Ahmad, S. *et al.* Dual use of helping information for estimating the finite population mean under the stratified random sampling scheme. *J. Math.* **1**, 1 (2021).
- Murthy, M. N. (1967). *Sampling: Theory and methods*. Statistical Pub. Society.
- Cochran, W. G. *Sampling techniques* (Wiley, 1977).
- Singh, S. *Advanced sampling theory with applications: How michael “Selected” Amy* (Vol. 2) (Springer, 2003).
- Sukhathme, P. V. *Sampling theory of survey with application* (The Indian Society of Agricultural Statistics, 1970).

## Acknowledgements

Researchers Supporting Project number (RSPD2023R458), King Saud University, Riyadh, Saudi Arabia.

### Author contributions

The entire manuscript is being written by S.A. and N.K.A. M.A. and J.S. help us to improve the language of the manuscript and supervise the overall manuscript. N.A., M.E. and H.A. help us to revise the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Correspondence** and requests for materials should be addressed to N.K.A.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2024