Estimation of variance of the difference-cum-ratio-type exponential estimator in simple random sampling

Umer Daraz\textsuperscript{b} and Mursala Khan\textsuperscript{a,c}

\textsuperscript{a}School of Business and Economics Freie University Berlin, Germany; \textsuperscript{b}School of Mathematical Sciences, Soochow University, Suzhou Jiangsu China; \textsuperscript{c}Department of Mathematics and Statistics, Riphah International University, I-14 Islamabad, Pakistan

\textbf{ABSTRACT}

In this article, we have suggested a class of estimators for the estimation of the population variance of the variable of interest. The proposed estimators used some certain known information of the auxiliary variable, such as kurtosis, coefficient of variation, and the minimum and maximum values. The properties of the suggested class of estimators such as the bias and mean squared error (MSE) are obtained up to the first order of approximation. In order to check the performances of the estimators and to verify the theoretical results, we conducted a simulation study. The results of the simulation study show that the proposed class of estimators have lower MSE than other existing estimators. This holds for all simulation scenarios. In the application part, we used data from Statistical Bureau of Pakistan, and from the Textbook of Cochran, which also confirms that the suggested class of estimators is more efficient than the usual unbiased variance estimator, ratio estimator, traditional regression estimator, and other existing estimators in survey literature.

\textbf{1. Introduction}

The purpose of survey sampling is to get accurate information about the characteristics of the population for improving the efficiency of the estimators under study at the lowest costs, less time and human efforts (for more details, see Yang et al. (2020)). In several populations, there has been a few extreme values and to estimate the unknown population parameters without including this information is very sensitive. In which case, the results will be underestimated or overestimated. To solve this issue, it is important to use this information in estimating the population parameters. Isaki (1983), Bahl and Tuteja (1991), Upadhyaya and Singh (1999), Kadilar and Cingi (2006), Dubey and Sharma (2008), H. Singh and Chandra (2008), Shabbir and Gupta (2010), H. P. Singh and Solanki (2013), and Yadav et al. (2015) have all suggested some wider classes of estimators for estimating finite population variance. Consider a finite population $U = \{U_1, U_2, U_3, \ldots, U_N\}$ of size $N$ units. Let $y_i$ and $x_i$ be the values of the study variable $Y$ and the auxiliary variable $X$ for the $i$th units respectively. Let $\bar{Y} = (1/N) \sum_{i=1}^{N} Y_i$ and $\bar{X} = (1/N) \sum_{i=1}^{N} X_i$ be the population mean of the study and the auxiliary variable, respectively. It is further assumed that $S_y^2 = (1/N - 1) \sum_{i=1}^{N} (Y_i - \bar{Y})^2$ and $S_x^2 = (1/N - 1) \sum_{i=1}^{N} (X_i - \bar{X})^2$ be the population variances of the study as well as auxiliary variable, respectively.

To estimate the unknown population parameter $\bar{Y}$, we select a random sample of size $n$ units from the population by using simple random sampling without replacement (SRSWOR). Let $\bar{y} = (1/n) \sum_{i=1}^{n} y_i$ and $\bar{x} = (1/n) \sum_{i=1}^{n} x_i$ be the sample means of the study and the auxiliary variables, respectively, and their corresponding sample variances are $S_{y}^2 = (1/n - 1) \sum_{i=1}^{n} (y_i - \bar{y})^2$ and $S_{x}^2 = (1/n - 1) \sum_{i=1}^{n} (x_i - \bar{x})^2$, respectively.

To find the bias and MSE for different estimators, we define the following terms. Let $e_0 = \left(\frac{\bar{y} - S_y}{S_y}\right)$, $e_1 = \left(\frac{\bar{y} - S_y}{S_y}\right)$ and $e_2 = \left(\frac{\bar{x} - \bar{X}}{X}\right)$ such that $E(e_i) = 0$ for $i = 0, 1, 2$.

$E(e_0^2) = \theta k_{10}, E(e_1^2) = \theta k_{20}, E(e_0^2) = \theta k_{20}, E(e_0 e_1) = \theta k_{22}, E(e_0 e_2) = \theta k_{21}$

$E(e_1 e_2) = \theta L_{12},$

where $\lambda_{10} = (\lambda_{10} - 1), \lambda_{04} = (\lambda_{04} - 1), \lambda_{22} = (\lambda_{22} - 1), \theta = \left(\frac{1}{n} - \frac{1}{N}\right)$. Also $\lambda_{10} = \frac{\mu_y}{\mu_y^2 \mu_x^2}$, where $\mu_x = \sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X})$. Here $\lambda_{40} = \beta_{2(y)}$ and $\lambda_{04} = \beta_{2(x)}$ are the population coefficients of kurtosis.

The usual variance estimator of $\hat{S}_y^2 = \hat{S}_y^2$ [1] for population variance is given by...
\[ \text{Var}(\hat{S}^2_R) = \theta S^2_R \lambda^*_t. \quad (1.1) \]

Isaki (1983) suggested a ratio-type estimator for the variance of the study variable \( Y \), which is denoted by \( \hat{S}^2_R \) [2], and is given by
\[ \hat{S}^2_R = s^2 \left( \frac{S^2}{s^2} \right), \quad (1.2) \]

Expressions for bias and MSE of \( \hat{S}^2_R \), in sample random sampling (SRS) are given by
\[ \text{Bias}(\hat{S}^2_R) \cong \theta S^4 (\lambda^*_t - \lambda^*_2), \quad (1.3) \]
and
\[ \text{MSE}(\hat{S}^2_R) \cong \theta S^4 (\lambda^*_t + \lambda^*_t - 2 \lambda^*_2). \quad (1.4) \]

The classical regression estimator \( \hat{S}^2_{Ip} \) [3] in SRS is given by
\[ \hat{S}^2_{Ip} = s^2 + b_{(i',i)}(S^2 - s^2), \quad (1.5) \]
where \( b_{(i',i)} = \frac{s^2_{i'} - s^2_2}{s^2_{i'} - s^2} \) is the sample regression coefficient. The MSE of the estimator \( \hat{S}^2_{Ip} \), is given by
\[ \text{MSE}(\hat{S}^2_{Ip}) \cong \theta S^4 \lambda^*_t (1 - p^2), \quad (1.6) \]
where
\[ p^* = \frac{\lambda^*_2}{\sqrt{\lambda^*_t \lambda^*_4}}. \quad (1.7) \]

Bahl and Tuteja (1991) suggested an exponential ratio-type estimator for the population variance of the study variable \( Y \), which is denoted by \( \hat{S}^2_{Bt} \) [4] and is given by:
\[ \hat{S}^2_{Bt} = s^2 \exp \left( \frac{S^2 - s^2}{s^2 + s^2} \right), \quad (1.8) \]
Expressions for bias and MSE respectively of \( \hat{S}^2_{Bt} \), are given by
\[ \text{Bias}(\hat{S}^2_{Bt}) \cong \frac{1}{2} \theta S^4 \left( \frac{3 \lambda^*_t}{4} - \lambda^*_t \right), \quad (1.9) \]
and
\[ \text{MSE}(\hat{S}^2_{Bt}) \cong \theta S^4 \left( \lambda^*_t + \frac{\lambda^*_t}{4} - \lambda^*_2 \right). \quad (1.10) \]

Upadhayaya and Singh (1999) proposed a ratio-type estimator \( \hat{S}^2_{US} \) [5], that uses the kurtosis of an auxiliary variable in SRS, given by
\[ \hat{S}^2_{US} = s^2 \left( \frac{S^2 + \lambda^*_t}{s^2 + \lambda^*_t} \right), \quad (1.11) \]
Expressions for bias and MSE respectively of \( \hat{S}^2_{US} \), are given by
\[ \text{Bias}(\hat{S}^2_{US}) \cong \theta S^2 g_0 \left( \lambda^*_t - \lambda^*_t \right), \quad (1.12) \]
and
\[ \text{MSE}(\hat{S}^2_{US}) \cong \theta S^4 \left( \lambda^*_t + \frac{3 \lambda^*_t}{4} - 2 \lambda^*_2 \right), \quad (1.13) \]
where
\[ g_0 = \frac{s^2}{s^2 + \lambda^*_t}. \quad (1.14) \]

Kadilar and Cingi (2006) suggested a class of ratio estimators \( \hat{S}^2_{KC} \) [6–8] which are given by
\[ \hat{S}^2_{KC1} = s^2 \left( \frac{S^2 + C_x}{s^2 + C_x} \right), \quad (1.15) \]
\[ \hat{S}^2_{KC2} = s^2 \left( \frac{\lambda^*_t S^2 + C_x}{\lambda^*_t S^2 + C_x} \right), \quad (1.16) \]
\[ \hat{S}^2_{KC3} = s^2 \left( \frac{C_x S^2 + \lambda^*_t}{C_x S^2 + \lambda^*_t} \right), \quad (1.17) \]
where \( C_x = \frac{s^2}{\lambda^*_t} \) is the population coefficient of variation.

Expressions for bias and MSE’s respectively of \( \hat{S}^2_{KC}(i = 1, 2, 3) \), in SRS are given by
\[ \text{Bias}(\hat{S}^2_{KC1}) \cong \theta S^2 \left( g_1 \lambda^*_t - \lambda^*_t \right), \quad (1.18) \]
and
\[ \text{MSE}(\hat{S}^2_{KC1}) \cong \theta S^4 \left( g^1 + \frac{3 \lambda^*_t}{4} - 2 g_2 \lambda^*_2 \right), \quad (1.19) \]
where
\[ g^1 = \frac{s^2}{s^2 + C_x}, \quad g^2 = \frac{\lambda^*_t S^2}{\lambda^*_t S^2 + C_x}, \quad g^3 = \frac{C_x S^2}{C_x S^2 + \lambda^*_t}. \quad (1.20) \]

2. Proposed estimators

Motivated by Daraz et al. (2018), we proposed an improved class of estimators for estimating the finite population variance \( S^2 \) using certain known population parameters under simple random sampling scheme. The proposed estimator is given by
\[ \hat{S}^2_t = \left[ k_1 \left( \frac{S^2}{s^2} \right)^{a_3} + k_2 \left( \frac{S^2}{s^2} \right)^{a_3} \right] \exp \left( \frac{a_1 (S^2 - s^2)}{a_1 (S^2 + s^2) + 2 b_1} \right), \quad (2.1) \]
where \( k_1 \) and \( k_2 \) are the unknown constants whose values are to be determined such that the MSE’s are minimum, \( a_3 \) and \( b_1 \) are the parameters of the auxiliary variables. Also, \( a_1 \) and \( a_2 \) are the scalar quantities which contain the values (0, -1, 1) from (2.1) we can generate
the different classes of proposed estimator which are given in Table 1.

where

\[ L = \exp\left(\frac{a_1(s_x^2 - S_{x}^2)}{(a_1(s_x^2 + S_{x}^2) + 2b_1)}\right) \]  (2.2)

Properties of the proposed estimator

Rewriting (2.1) in term of errors, we have

\[ \hat{S}_D^2 = \left[k_1S_y^2 + k_2\bar{X}e_2(1 + e_1)^{-a_2} \right] \exp\left[\frac{-g_4e_1}{2}\left(1 + \frac{g_4}{2}e_1^2\right)^{-1}\right] \]  (2.2)

where

\[ g_4 = \frac{a_1S_x^2}{a_1S_x^2 + b_1} \]  (2.3)

By using Taylor series up to the first order of approximation, we have

\[ \hat{S}_D^2 - S_y^2 \approx -S_y^2 + k_1S_y^2 \left[1 + e_0 - e_1(a_1 + \frac{g_4}{2}) + e_1^2\left(\frac{a_1}{2} + \frac{3g_4}{8} + \frac{a_1(a_1 + 1)}{2}\right)\right] - k_2\bar{X} e_2 \left[ e_1 + e_2(2a_2 + \frac{g_4}{2}) \right] \]  (2.3)

Using (2.3), the bias of \( \hat{S}_D^2 \), is given by

\[ \text{Bias}(\hat{S}_D^2) \approx -\left[S_y^2 - k_1S_y^2 D - k_2G\right], \]  (2.4)

where \( D = \left[1 + \theta\left(\frac{a_1}{2} + 3g_4 + a_1(a_1 + 1) + 2a_1 + 1\right)\right] \), and \( G = \theta S_x^2_{\lambda_03}\left(\frac{a_1}{2} + 3g_4\right) \). By squaring and taking expectation on both sides of Equation (2.3), we get the mean squared error by using the first order of approximation, which is given by

\[ \text{MSE}(\hat{S}_D^2) \approx \left[S_y^2 + k_1S_y^2 A + k_1^2B - 2k_1S_y^2 D - 2k_1S_y^2 G + 2k_1k_2S_y^2 F\right]. \]  (2.5)

where

\[ A = \left[1 + \theta\left(\frac{a_1}{2} + 3g_4 \right) + \left(\frac{a_1}{2} + \frac{3g_4}{4} + \frac{a_1(a_1 + 1)}{2}\right)\right] - 2\lambda_1(2a_1 + g_4) \]

\[ B = \theta S_x^2_{\lambda_1}, \quad F = \theta S_{\lambda_03}(a_1 + a_2 + g_4) - \lambda_{21}. \]

The optimum values of \( k_1 \) and \( k_2 \) obtained by minimizing (2.5) are \( k_1(\text{opt}) = \frac{BD - FG}{AB - F^2} \), and \( k_2(\text{opt}) = \frac{S_x^2(AG - DF)}{AB - F^2} \). By substituting the optimum values of \( k_1 \) and \( k_2 \) in (2.5), we get the minimum MSE of \( \hat{S}_D^2 \), which is given below:

\[ \text{MSE}(\hat{S}_D^2)_{\text{min}} \approx \left[1 - \frac{(AG^2 + BD^2 - 2DFG)}{AB - F^2}\right]. \]  (2.6)

3. Mathematical comparison

In this section, we compare the suggested class of estimator \( \hat{S}_D^2 \) with the existing estimators \( \hat{S}_y^2, \hat{S}_R^2, \hat{S}_F, \hat{S}_{BT}^2, \hat{S}_{US}^2, \) and \( \hat{S}_{KC^2}^2 \).

Condition (i): By (1.1) and (2.6), \( \text{Var} (\hat{S}_y^2) > \text{MSE}(\hat{S}_D^2)_{\text{min}} \) if

\[ \theta \lambda_{40}^+ + \left(\frac{AG^2 + BD^2 - 2DFG}{AB - F^2}\right) > 1. \]

Condition (ii): By (1.4) and (2.6), \( \text{MSE}(\hat{S}_R^2) > \text{MSE}(\hat{S}_D^2)_{\text{min}} \) if

\[ \theta \left(\lambda_{40}^+ + \lambda_{64}^+ - 2\lambda_{22}^+\right) + \left(\frac{AG^2 + BD^2 - 2DFG}{AB - F^2}\right) > 1. \]
Condition (iii): By (1.6) and (2.6), 

$$\theta_{40}^* (1 - \rho^2) + \left( \frac{AG^2 + BD^2 - 2DFG}{AB - F^2} \right) > 1.$$ 

Condition (iv): By (1.9) and (2.6), 

$$\theta \left( \frac{\lambda_{40}^* + \lambda_{04}^*}{4} - \lambda_{22}^* \right) + \left( \frac{AG^2 + BD^2 - 2DFG}{AB - F^2} \right) > 1.$$ 

Condition (v): By (1.12) and (2.6), 

$$\theta \left( \frac{\lambda_{40}^* + \lambda_{04}^* - 2\alpha_{12}}{2} \right) + \left( \frac{AG^2 + BD^2 - 2DFG}{AB - F^2} \right) > 1.$$ 

Condition (vi): By (1.7) and (2.6), 

$$\theta \left( \frac{\lambda_{40}^* + \lambda_{04}^* - 2\alpha_{12}}{2} \right) + \left( \frac{AG^2 + BD^2 - 2DFG}{AB - F^2} \right) > 1.$$ 

4. Simulation study

In order to verify the theoretical results in Section 3, we have conducted a simulation study by using the idea from Agarwal et al. (2012). We generated six different artificial populations of the auxiliary variable X by using the following probability distributions.

- X ~ Exponential (λ = 3) 
- X ~ Exponential (λ = 7) 
- X ~ Uniform (b_3 = 0, b_4 = 1) and X ~ Uniform (b_3 = 3, b_4 = 5) 
- X ~ Gamma (α_3 = 4, α_4 = 6) and X ~ Gamma (α_3 = 8, α_4 = 10).

After that, the study variable Y is computed as 

$$Y = r_{xy} \times X + e,$$

taking \( r_{xy} = 0.80 \), where \( r_{xy} \) is the correlation coefficient between the study and the auxiliary variables and \( e \sim N(0, 1) \) is the error term.

We considered the following steps in R-Software to obtain the MSE’s of the proposed class of estimators:

**Step 1:** In the first step, we generated a population of size 1000 using a certain type of probability distributions.

**Step 2:** We obtained population total, minimum and maximum values of the auxiliary variable from Step 1. We also computed the optimum values of the unknown constants of the proposed estimator.

**Step 3:** We considered different sample sizes for each population to generate the samples using SRSWOR.

**Step 4:** For each sample size, the values of bias’s and MSE’s are computed for all the estimators considered in this paper.

**Step 5:** The process in Step 3 and Step 4 is repeated 50,000 times and the results for artificial populations are reported in Table 2, whereas the results of the real data sets are summarized in Table 3.

Finally, the MSE’s of the estimators over all replications are obtained by using the following formula.

$$\text{MSE} \left( S^2_k \right)_{\text{min}} = \frac{\sum_{i=1}^{5000} (S_i - S^*_i)^2}{5000},$$

for \( k = R, Ir, BT, US, KCi, D1, D2, \ldots, D8 \).

5. Numerical examples

To check the performances of the suggested class of estimators, we used three real data sets to compare the MSE’s of different estimators. The description and summary statistics are given by

**Data 1.** (Bureau of Statistics (2013), p. 135)

Y: Total number of students enrolls in 2012,

### Table 2. Mean squared error (MSE) of the estimators using the artificial populations

<table>
<thead>
<tr>
<th>Estimator</th>
<th>X ~ Exp(3)</th>
<th>X ~ Exp(7)</th>
<th>X ~ Uni(0,1)</th>
<th>X ~ Uni(3,5)</th>
<th>X ~ Gam(4,6)</th>
<th>X ~ Gam(8,10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] S^2</td>
<td>1.10e-3</td>
<td>9.70e-4</td>
<td>9.00e-4</td>
<td>1.40e-3</td>
<td>2.66e-2</td>
<td>1.10e-3</td>
</tr>
<tr>
<td>[2] S^2</td>
<td>8.00e-4</td>
<td>8.80e-4</td>
<td>8.00e-4</td>
<td>1.20e-3</td>
<td>2.19e-2</td>
<td>1.00e-3</td>
</tr>
<tr>
<td>[3] S^2</td>
<td>8.00e-4</td>
<td>7.70e-4</td>
<td>7.00e-4</td>
<td>1.10e-3</td>
<td>2.16e-2</td>
<td>1.00e-3</td>
</tr>
<tr>
<td>[4] S^2</td>
<td>7.00e-4</td>
<td>7.40e-4</td>
<td>7.00e-4</td>
<td>1.10e-3</td>
<td>2.16e-2</td>
<td>1.00e-3</td>
</tr>
<tr>
<td>[5] S^2</td>
<td>7.00e-4</td>
<td>7.50e-4</td>
<td>7.00e-4</td>
<td>1.10e-3</td>
<td>2.20e-2</td>
<td>1.00e-3</td>
</tr>
<tr>
<td>[6] S^2</td>
<td>7.00e-4</td>
<td>7.10e-4</td>
<td>7.00e-4</td>
<td>1.00e-3</td>
<td>2.12e-2</td>
<td>1.00e-3</td>
</tr>
<tr>
<td>[7] S^2</td>
<td>7.00e-4</td>
<td>7.10e-4</td>
<td>7.00e-4</td>
<td>1.10e-3</td>
<td>2.14e-2</td>
<td>1.00e-3</td>
</tr>
<tr>
<td>[8] S^2</td>
<td>7.00e-4</td>
<td>7.20e-4</td>
<td>7.00e-4</td>
<td>1.00e-3</td>
<td>2.17e-2</td>
<td>9.00e-4</td>
</tr>
<tr>
<td>[9] S^2</td>
<td>2.00e-4</td>
<td>1.40e-4</td>
<td>2.00e-4</td>
<td>2.60e-4</td>
<td>3.80e-3</td>
<td>1.00e-4</td>
</tr>
<tr>
<td>[10] S^2</td>
<td>1.00e-4</td>
<td>1.00e-4</td>
<td>1.00e-4</td>
<td>1.00e-4</td>
<td>1.30e-3</td>
<td>5.00e-5</td>
</tr>
<tr>
<td>[11] S^2</td>
<td>2.40e-4</td>
<td>1.50e-4</td>
<td>1.20e-4</td>
<td>1.50e-4</td>
<td>3.50e-3</td>
<td>1.00e-4</td>
</tr>
<tr>
<td>[12] S^2</td>
<td>2.00e-4</td>
<td>1.60e-4</td>
<td>1.70e-4</td>
<td>2.00e-4</td>
<td>3.70e-3</td>
<td>1.00e-4</td>
</tr>
<tr>
<td>[13] S^2</td>
<td>2.50e-4</td>
<td>2.00e-4</td>
<td>2.00e-4</td>
<td>3.00e-4</td>
<td>4.20e-3</td>
<td>1.00e-4</td>
</tr>
<tr>
<td>[14] S^2</td>
<td>2.50e-4</td>
<td>2.00e-4</td>
<td>2.00e-4</td>
<td>2.70e-4</td>
<td>5.30e-3</td>
<td>1.60e-4</td>
</tr>
<tr>
<td>[15] S^2</td>
<td>1.20e-4</td>
<td>1.00e-4</td>
<td>1.40e-4</td>
<td>2.00e-4</td>
<td>2.70e-3</td>
<td>1.00e-4</td>
</tr>
<tr>
<td>[16] S^2</td>
<td>2.00e-5</td>
<td>2.00e-5</td>
<td>3.00e-5</td>
<td>4.00e-5</td>
<td>6.00e-4</td>
<td>2.00e-5</td>
</tr>
</tbody>
</table>
Table 3. Mean squared error (MSE) of the estimators using empirical data sets

<table>
<thead>
<tr>
<th>Estimator</th>
<th>MSE: Data 1</th>
<th>MSE: Data 2</th>
<th>MSE: Data 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1] $S_1^2$</td>
<td>1.21e+21</td>
<td>1.08e+21</td>
<td>8140.74</td>
</tr>
<tr>
<td>[2] $S_2^2$</td>
<td>1.42e+21</td>
<td>1.40e+21</td>
<td>5702.73</td>
</tr>
<tr>
<td>[3] $S_3^2$</td>
<td>1.07e+21</td>
<td>1.13e+21</td>
<td>5667.55</td>
</tr>
<tr>
<td>[4] $S_4^2$</td>
<td>1.28e+21</td>
<td>1.19e+21</td>
<td>6442.13</td>
</tr>
<tr>
<td>[5] $S_5^2$</td>
<td>1.40e+21</td>
<td>1.40e+21</td>
<td>5712.85</td>
</tr>
<tr>
<td>[6] $S_6^2$</td>
<td>1.46e+21</td>
<td>1.40e+21</td>
<td>5703.41</td>
</tr>
<tr>
<td>[7] $S_7^2$</td>
<td>1.42e+21</td>
<td>1.04e+21</td>
<td>5703.06</td>
</tr>
<tr>
<td>[8] $S_8^2$</td>
<td>1.42e+21</td>
<td>1.40e+21</td>
<td>5785.88</td>
</tr>
<tr>
<td>[9] $S_9^2$</td>
<td>3.19e+20</td>
<td>2.49e+20</td>
<td>5401.90</td>
</tr>
<tr>
<td>[10] $S_{10}^2$</td>
<td>3.74e+20</td>
<td>5.12e+20</td>
<td>5084.03</td>
</tr>
<tr>
<td>[11] $S_{11}^2$</td>
<td>5.94e+20</td>
<td>5.60e+20</td>
<td>4759.47</td>
</tr>
<tr>
<td>[12] $S_{12}^2$</td>
<td>5.52e+20</td>
<td>3.11e+20</td>
<td>4340.05</td>
</tr>
<tr>
<td>[13] $S_{13}^2$</td>
<td>5.42e+20</td>
<td>4.74e+20</td>
<td>4233.81</td>
</tr>
<tr>
<td>[14] $S_{14}^2$</td>
<td>5.01e+20</td>
<td>3.01e+20</td>
<td>3594.18</td>
</tr>
<tr>
<td>[15] $S_{15}^2$</td>
<td>4.57e+20</td>
<td>7.34e+19</td>
<td>3068.23</td>
</tr>
<tr>
<td>[16] $S_{16}^2$</td>
<td>1.99e+19</td>
<td>2.85e+19</td>
<td>510.23</td>
</tr>
</tbody>
</table>

(a) Artificial data: $X \sim \text{Exponential(}\lambda = 3\text{)}$. 
(b) Artificial data: $X \sim \text{Exponential(}\lambda = 7\text{)}$.
(c) Artificial data: $X \sim \text{Uniform(b}_3 = 0, b_4 = 1\text{)}$.
(d) Artificial data: $X \sim \text{Uniform(b}_3 = 3, b_4 = 5\text{)}$.

Figure 1. Graphical display of the MSE’s results of the estimators using the artificial data. Note: The vertical line of the figures shows the MSE’s of the estimators, while the horizontal line indicates the corresponding estimators. For easiness, we denote the estimators by different numbers starting from 1 to 16. For more details, see Table 2. Source: Own computations.
6. Conclusion

In this paper, we proposed a class of estimators for estimating the population variance of the study variable using some known information of the auxiliary variable. The properties of the proposed class of estimators are compared with other existing estimators. For this purpose, we reported some theoretical conditions in Section 3 under which the proposed estimators are more efficient than the existing estimators. These theoretical conditions are verified through the help of a simulation study and some empirical data sets. MSE’s results of various estimators over the simulation setup are demonstrated in Table 2. In comparing the MSE’s of the estimators, it is clear from the table that the proposed class of estimators performs the best over the cited existing estimators. The MSE’s results of various estimators in Table 2 are plotted in Figures 1 and 2 which demonstrates that the MSE’s of the proposed class of estimators are significantly smaller than the MSE’s of other estimators. Similar results are obtained from the empirical data, which also confirms the theoretical results in Section 3. The empirical results are displayed in Table 3, which is then graphically shown in Figure 3. Hence, based on our simulation results as well as through empirical results, we observed that the proposed class of estimators \( \hat{S}^2_i \) \((i = 9, 10, 11, \ldots, 16)\) are more efficient than the other considered estimators. Among the suggested class of estimators, \( \hat{S}^2_{16} \) is preferable because of its least MSE.

Figure 2. Graphical display of the MSE’s results of the estimators using the artificial data. Note: The vertical line of the figures shows the MSE’s of the estimators, while the horizontal line indicates the corresponding estimators. For easiness, we denote the estimators by different numbers starting from 1 to 16. For more details, see Table 2. Source: Own computations.
In this article, we have suggested a class of estimators for the estimation of the population variance by using the maximum and minimum values of independent variable. In order to check the performance of the estimators and to verify the theoretical results we conducted a simulation study from different distribution and also used the data sets from real life application and display it graphically which confirmed that the suggested class of estimator is more efficient than the existent estimators because it's least mean squared errors.

**Figure 3.** Graphical display of the MSE’s results of the estimators using the empirical data. Note: The vertical line of the Figures shows the MSE’s of the estimators, while the horizontal line indicates the corresponding estimators. For easiness, we denote the estimators by different numbers starting from 1 to 16. For more details, see Table 3. Source: Own computations.

**PUBLIC INTEREST STATEMENT**

In this article, we have suggested a class of estimators for the estimation of the population variance by using the maximum and minimum values of independent variable. In order to check the performance of the estimators and to verify the theoretical results we conducted a simulation study from different distribution and also used the data sets from real life application and display it graphically which confirmed that the suggested class of estimator is more efficient than the existent estimators because it’s least mean squared errors.

**Acknowledgment**

This work was supported by NSFC of China with grant [12071329]. We are very thankful to the two unknown referees, and the editor for their insightful comments and suggestions which greatly improved this paper.

**Disclosure statement**

The authors declare that there is no conflict of interests regarding the publication of this article.
Notes on contributors

Umer Daraz received his M.Phil degree in Survey Sampling from Quaid-i-Azam University, Islamabad, Pakistan in 2016. He is currently pursuing his PhD degree under the supervision of Prof. Tang Yu at Soochow University, Suzhou, China. His research interests lie in the survey sampling, design experiment and combination design.

Mursala Khan got his Doctorate degree from Free University Berlin, Germany. His field of specialization is survey sampling. Currently, he is working as an assistant professor in the Department of Mathematics and Statistics, Riphah International University, Islamabad, Pakistan.

References


