RESEARCH ARTICLE

A simulation study: Improved ratio-in-regression type variance estimator based on dual use of auxiliary variable under simple random sampling

Sohaib Ahmad1, Sardar Hussain2, Kalim Ullah3, Erum Zahid4, Muhammad Aamir1, Javid Shabbir2,5, Zubair Ahmad2,6, Huda M. Alshanbari6, Wejdan Alajlan6

1 Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan, 2 Department of Statistics, Quaid-i-Azam University Islamabad, Pakistan, 3 Foundation University Medical College, Foundation University School of Health Sciences, DHA-I, Islamabad, Pakistan, 4 Department of Applied Mathematics and Statistics, Institute of Space Technology, Islamabad, Pakistan, 5 Department of Statistics, University of Wah at Wah Cantt, Wah Cantt, Pakistan, 6 Department of Mathematical Sciences, College of Science, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

* z.ferry21@gmail.com

Abstract

In this article, we proposed an improved finite population variance estimator based on simple random sampling using dual auxiliary information. Mathematical expressions of the proposed and existing estimators are obtained up to the first order of approximation. Two real data sets are used to examine the performances of a new improved proposed estimator. A simulation study is also recognized to assess the robustness and generalizability of the proposed estimator. From the result of real data sets and simulation study, it is examining that the proposed estimator give minimum mean square error and percentage relative efficiency are higher than all existing counterparts, which shown the importance of new improved estimator. The theoretical and numerical result illustrated that the proposed variance estimator based on simple random sampling using dual auxiliary information has the best among all existing estimators.

1. Introduction

In survey sampling, it is well known fact that efficiency of population parameters estimates can be improved by the use of auxiliary information, providing it is highly associated with the study variable. Auxiliary information plays a vital role in parameter selection and estimation to achieve efficient estimates of the unknown population parameters. Many authors have suggested estimators based on original auxiliary information. In fact, appropriate use of auxiliary variable in survey sampling results in significant decrease in the variance of the estimator of unknown population parameter(s). Variance is one of the features that can be used to describe a statistical population. Brilliant efforts have been taken to estimate this characteristic as accurately as possible. The goal is to reduce the scattering of the calculated estimate when it is obtained from many available samples. Furthermore, the importance of using information
from one or more auxiliary variables is to improve estimator efficiency. Many attempts have previously been made in this direction, including use of the ratio and transformed ratio-product-type. By utilizing the association among the study variable and auxiliary variable, the survey statisticians have shaped effective ratio and product estimators for population variance. The estimation of population variance of the study variable is required in a variety of scenarios in application. In a variety of fields, including agriculture, medicine, biology, and business, where we encounter populations that are likely to be skewed, variance estimate for the population is important.

Variations can be found in our daily lives throughout the world. It is a common belief that no two things or people are exactly the same. For example, a farmer must have a good alertness of how climate geographies change over time in order to determine how to plant his lot. For proper treatment, a doctor must have a detailed realization of differences in human hypertension and fever. Many authors suggested different types of estimators for variance including Das [1], Prasad and Singh [2], Swain and Mishra [3], Ahmad and Shabbir [4], Kadilar and Cingi [5], Subramani and Kumarapandiyan [6], Singh and Solanki [7], Yadav et al. [8], Ahmad et al. [21], Ozturk [22], Ozturk and Demirel [23], Frey and Feeman [24], Zamanzade [34,35], Yadav and Kadilar [33], Zaman and Bulut [30,31], Yadav and Zaman [32], Zaman and Dun-der [34] and Singh and Khalid [35]. We requisite to estimate population parameters for numerous variables in multidisciplinary surveys, such as house lengths, income level, and the number of unemployed people, in socio-economic surveys. The goal of this study is to offer a strategy for estimating population variance using a dual auxiliary variable in simple random sampling. We found no contribution on variance estimation using dual auxiliary variables under simple random sampling in the literature.

2. Notations and symbols

Consider \( \psi = (\psi_1, \psi_2, \ldots, \psi_N) \) be a population comprise of size \( N \) elements, let a sample of size \( n \) is chosen from \( \psi \) by using simple random sampling without replacement (SRSWOR). Let \( Y, X \) and \( R_x \) represent the study variable, auxiliary variable and rank of the auxiliary variable respectively. Consider population variance of the study variable \( Y \), auxiliary variable \( X \) and rank of the auxiliary variable \( R_x \) is denoted by \( S_y^2, S_x^2 \) and \( S_{rx}^2 \). Let \( y_i, x_i \) and \( r_{xi} \) be the sample for the \( i^{th} \) units correspondingly. Let \( s_y^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1}, s_x^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}, s_{rx}^2 = \frac{\sum_{i=1}^{n} (r_{xi} - \bar{r}_x)^2}{n-1} \), be the sample variances. \( S_y^2 = \frac{\sum_{i=1}^{N} (y_i - \bar{y})^2}{N-1}, S_x^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}, S_{rx}^2 = \frac{\sum_{i=1}^{N} (r_{xi} - \bar{r}_x)^2}{N-1} \), of \( Y, X \) and \( R_x \) be the corresponding variances. To derive the properties of the suggested estimator, we have the following error terms.

Let \( \zeta_{oi} = \frac{s_y^2 - S_y^2}{S_y^2}, \zeta_{1} = \frac{s_x^2 - S_x^2}{S_x^2}, \zeta_{2} = \frac{s_{rx}^2 - S_{rx}^2}{S_{rx}^2}, E(\zeta_i) = 0, \) for \( i = 0, 1, 2) \)

\[
E(\zeta^{2}_o) = \lambda \chi^*_{101}, E(\zeta^{2}_1) = \lambda \chi^*_{104}, E(\zeta^{2}_2) = \lambda \chi^*_{104}, E(\zeta_o \zeta_{1}) = \lambda \chi^*_{201}, E(\zeta_o \zeta_{2}) = \lambda \chi^*_{204}, E(\zeta_{1} \zeta_{2}) = \lambda \chi^*_{202}.
\]

\[
\lambda = \left( \frac{1}{n} - \frac{1}{N} \right).
\]
\[ \hat{X}_{rst} = \frac{u_{rst}}{u_{200} u_{020} u_{002}} = \frac{\sum(y_{xy} - \bar{Y})(x_{xy} - \bar{X})(r_{sy} - R_y)}{N - 1}. \]

Where \( r, s \) and \( t \) be the positive number and \( u_{200}, u_{020}, u_{002} \) are the second order moments about means of \( Y, X \) and rank of \( X \), and \( \hat{X}_{rst} \) is the moment ratio.

### 3. Existing estimators

In this section, we discussed several well-known estimators of population variance.

i. The usual estimator, is given by:

\[ \hat{Y}_0 = \frac{\sum_{i=1}^{n}(y_i - \bar{Y})^2}{n - 1} \]  \( \text{(1)} \)

The variance of \( \hat{Y}_0 \), is given by:

\[ \text{Var}(\hat{Y}_0) = \lambda S_y^4 \hat{X}_{400} \]  \( \text{(2)} \)

ii. Isaki \([36]\) suggested ratio estimator, is given by:

\[ \hat{Y}_1 = s_y^2 \left( \frac{S_s^2}{S_x^2} \right) \]  \( \text{(3)} \)

The properties of \( \hat{Y}_1 \), is given as:

\[ \text{Bias}(\hat{Y}_1) = \lambda S_y^2 \{ \hat{X}_{400} - \hat{X}_{220} \}, \]

and

\[ \text{MSE}(\hat{Y}_1) \cong \lambda S_y^2 \{ \hat{X}_{400} + \hat{X}_{400} - 2\hat{X}_{220} \}. \]  \( \text{(4)} \)

iii. The difference estimator, is given by:

\[ \hat{Y}_2 = s_y^2 + \Psi_{11} (S_x^2 - s_x^2). \]  \( \text{(5)} \)

Where

\[ \Psi_{11} = \begin{bmatrix} S_y^2 \hat{X}_{220} \\ S_x^2 \hat{X}_{400} \end{bmatrix}. \]

The minimum variance at the optimal value of \( \Psi_{11} \), is given by:

\[ \text{Var}(\hat{Y}_2) = \lambda S_y^4 \hat{X}_{400} (1 - \rho^2). \]  \( \text{(6)} \)
Where
\[ \rho = \frac{\hat{\chi}_{220}^*}{\sqrt{\hat{\chi}_{400}^* \hat{\chi}_{040}^*}}. \]

iv. Rao [37] difference type estimator, is given by:
\[ \hat{Y}_3 = \{\omega_1 \hat{s}_y^2 + \omega_2 (S_y^2 - \hat{s}_y^2)\}, \tag{7} \]
where
\[ \omega_1 = \frac{\hat{\chi}_{400}^* \hat{\chi}_{040}^*}{K(\hat{\chi}_{400}^* \hat{\chi}_{040}^* - \hat{\chi}_{220}^2) + \hat{\chi}_{040}^*}, \]
\[ \omega_2 = \frac{S_y^2 \hat{\chi}_{220}^*}{S_y^2 \{K(\hat{\chi}_{400}^* \hat{\chi}_{040}^* - \hat{\chi}_{220}^2) + \hat{\chi}_{040}^*\}}. \]

The properties at the optimal value of \(\omega_1\) and \(\omega_2\), are given by:
\[ \text{Bias}\left(\hat{Y}_3\right) = S_y^2 \left\{\frac{\hat{\chi}_{400}^*}{K(\hat{\chi}_{400}^* \hat{\chi}_{040}^* - \hat{\chi}_{220}^2) + \hat{\chi}_{040}^*} - 1\right\}, \tag{8} \]
and
\[ \text{MSE}\left(\hat{Y}_3\right) = \frac{K S_y^4 [\hat{\chi}_{400}^* \hat{\chi}_{040}^* - \hat{\chi}_{220}^2]}{K(\hat{\chi}_{400}^* \hat{\chi}_{040}^* - \hat{\chi}_{220}^2) + \hat{\chi}_{040}^*}. \tag{9} \]

v. Singh et al. [38] suggested the following estimators:
\[ \hat{Y}_4 = S_y^2 \exp\left\{S_y^2 - \hat{s}_2^2\right\} \tag{10} \]

The properties of \(\hat{Y}_4\), are given by:
\[ \text{Bias}\left(\hat{Y}_4\right) = K S_y^3 \left\{\frac{3}{8} \hat{\chi}_{400}^* + \frac{1}{4} \hat{\chi}_{040}^* - \frac{1}{2} \hat{\chi}_{220}^*\right\}, \tag{11} \]
\[ \text{MSE}\left(\hat{Y}_4\right) = K S_y^4 \left[\hat{\chi}_{400}^* + \frac{1}{4} \hat{\chi}_{040}^* - \hat{\chi}_{220}^*\right]. \tag{12} \]

vi. Grover and Kaur [39], suggested the following estimator, is given by:
\[ \hat{Y}_5 = \left[\omega_3 \hat{s}_y^2 + \omega_4 (S_y^2 - \hat{s}_y^2)\right] \exp\left\{\frac{S_y^2 - \hat{s}_2^2}{S_y^2 + \hat{s}_2^2}\right\} \tag{13} \]
\[ \text{Bias}(\hat{Y}_5) = \left[-S_y^2 + \omega_3 S_y^2 \left\{1 + \frac{3}{8} \hat{\chi}_{400}^* - \frac{1}{2} \hat{\chi}_{220}^*\right\} + \frac{1}{2} \omega_4 S_y^2 \hat{\chi}_{040}^*\right]. \tag{14} \]
Where \( \omega_3 \) and \( \omega_4 \) are the unknown constant

\[
\omega_3 = \frac{\{8\kappa_{4040} - \kappa_{220}^2\}}{8[\kappa_0^2(\kappa_{4040} + \kappa_{220}^2) + \kappa_{4040}^2]},
\]

\[
\omega_4 = \frac{S_2(\kappa_{040}^2 - \kappa_{220}^2) - 8\kappa_{4040} + 4\{\kappa_{4040}^2 - \kappa_{220}^2\}}{8S_4[\kappa_{4040}^2(\kappa_{040}^2 - \kappa_{220}^2) + \kappa_{040}^2]}.\]

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

\[
\text{MSE}(\hat{\gamma}_5) = \frac{\kappa_0^2S_4 S_040^2 - \kappa_{220}^2 - 16\kappa_2^{4040} \kappa_{4040}^2 - \kappa_{220}^2]}{64\{\kappa_4040^2 - \kappa_{220}^2 + \kappa_{040}^2\}} \tag{13}
\]

vii. Ahmad et al. [40], suggested the following estimator, given by:

\[
\hat{\gamma}_6 = [\omega_5 S_5^2 + \omega_6 S_5^2 - S_5^2] \exp\left( \frac{S_2^2 - S_5^2}{S_5^2 + S_6^2} \right) \tag{14}
\]

After expanding the above equation, we have

\[
\left( \hat{\gamma}_6 - S_5^2 \right) = -S_5^2 + \omega_5 S_5^2 + \omega_6 S_5^2 - 1/2 \omega_5 S_5^2 - \omega_6 S_5^2 - 1/2 \omega_5 S_5^2 + 3/8 \omega_6 S_5^2 + 1/2 \omega_6 S_5^2
\]

The bias and MSE of \( \hat{\gamma}_6 \), are given by:

\[
\text{Bias}(\hat{\gamma}_6) \approx -S_5^2 + \omega_5 S_5^2 (1 + 3/8 \kappa_{040}^2 - 1/2 \kappa_{220}^2) + 1/2 \kappa (\omega_6 S_5^2 \kappa_{040}^2 + \omega_7 S_5^2 \kappa_{022}^2),
\]

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

\[
\omega_5(\text{opt}) = \frac{8 - \kappa_{040}}{\kappa_{040}^2 \left( \frac{1}{\kappa_{040}^2} + \kappa (\gamma^2 + 1) \right)},
\]

\[
\omega_6(\text{opt}) = \frac{\kappa_{040}^2 (\kappa_{040}^2 \kappa_{004}^2 - \kappa_{220}^2) + (\kappa_{020}^2 \kappa_{004}^2 - \kappa_{220}^2) (8 - \kappa_{040}^2) + 4\kappa_{040}^2 (\kappa_{4040}^2 - \kappa_{002}^2)}{8S_4^2 \kappa_{040}^2 \left( \frac{1}{\kappa_{040}^2} + \kappa (\gamma^2 + 1) \right)} \tag{15}
\]

and

\[
\omega_7(\text{opt}) = \frac{S_5^2 (8 - \kappa_{040}^2) \left( \kappa_{220}^2 \kappa_{002}^2 - \kappa_{004}^2 \kappa_{202}^2 \right)}{8S_4^2 \kappa_{040}^2 \left( \frac{1}{\kappa_{040}^2} + \kappa (\gamma^2 + 1) \right)}. \tag{16}
\]
The minimum mean square error at the optimal values of \( \omega_{\gamma_{\text{opt}}}, \omega_{\delta_{\text{opt}}} \) and \( \omega_{\tau_{\text{opt}}} \) are given by:

\[
\text{MSE}(\hat{\gamma}_{\text{opt}})_{\text{min}} = \frac{\kappa S_{\gamma}^2 \left[ 64(\gamma^* + 1) - \kappa \left( \frac{\gamma^*}{\gamma^* + 1} \right)^2 - 16\lambda \chi_{404}^* (\gamma^* + 1) \right]}{64 \left( \frac{1}{\gamma^* + 1} + \lambda (\gamma^* + 1) \right)},
\]

where

\[
\gamma^* = \frac{2\lambda \chi_{202}^* \chi_{722}^* - \chi_{404}^* \chi_{202}^* - \chi_{404}^* \chi_{220}^*}{\chi_{404}^* (\chi_{404}^* \chi_{202}^* - \chi_{220}^*)}
\]

### 4. Proposed improved ratio-in-regression type variance estimator

The need for auxiliary variable can boost estimator efficiency during the design and estimation stage. When the association among the study variable and auxiliary variables exist, then rank of the auxiliary variable is also associated with the study variable. Therefore, the rank of the auxiliary variable (that includes rank of the auxiliary variable) can be considered as a novel auxiliary variable, and this evidence may help us in increasing the precision of an estimator. Based on these ideas, we are encouraged to investigate dual use of the auxiliary variable for population variance because it has been examined that there has been no work before. The suggested estimator includes the auxiliary information in the form of an auxiliary variable and in the form of a rank auxiliary variable. There are several estimators for estimating population variance based on original auxiliary variables. The key advantage of our proposed estimator in simple random sampling is that it is more flexible and best than existing estimators. By taking motivation from Ahmad et al. [40] we proposed the following estimator:

\[
\hat{Y}_{\text{prop}} = \omega_{\delta} S_{\delta}^2 + \omega_{\tau} (S_{\tau}^2 - \bar{s}_x^2) \exp \left( \frac{S_{\delta}^2 - S_{\tau}^2}{S_{\delta}^2 + S_{\tau}^2} \right) + \omega_{\tau_1} (S_{\tau}^2 - \bar{s}_x^2) \exp \left( \frac{S_{\tau}^2 - S_{\tau_1}^2}{S_{\tau}^2 + S_{\tau_1}^2} \right)
\]

Simplifying (16), we have

\[
\hat{Y}_{\text{prop}} = \omega_{\delta} S_{\delta}^2 (1 + \bar{\xi}_\delta) - \omega_{\delta} S_{\delta}^2 \bar{\xi}_\delta \left( 1 - \frac{1}{2} \bar{\xi}_1 + \frac{3}{8} \bar{\xi}_2 \right) - \omega_{\tau_1} S_{\tau_1}^2 \bar{\xi}_\tau \left( 1 - \frac{1}{2} \bar{\xi}_2 + \frac{3}{8} \bar{\xi}_2 \right)
\]

\[
\hat{Y}_{\text{prop}} - S_{\gamma}^2 = (\omega_{\delta} - 1) S_{\delta}^2 + \omega_{\delta} S_{\delta}^2 \bar{\xi}_\delta - \omega_{\tau_1} S_{\tau_1}^2 \bar{\xi}_\tau \left( \bar{\xi}_1 - \frac{1}{2} \bar{\xi}_2 + \frac{3}{8} \bar{\xi}_2 \right) - \omega_{\tau_2} S_{\tau_2}^2 \bar{\xi}_\tau \left( \bar{\xi}_2 - \frac{1}{2} \bar{\xi}_2 \right)
\]

The bias of \( \hat{\gamma}_{\text{prop}} \) is:

\[
\text{Bias}(\hat{\gamma}_{\text{prop}}) = (\omega_{\delta} - 1) S_{\delta}^2 + \frac{1}{2} \omega_{\delta} S_{\delta}^2 \chi_{404}^* + \frac{1}{2} \omega_{\tau_2} S_{\tau_2}^2 \lambda \chi_{404}^*
\]

Simplify (17), we have

\[
\text{MSE}(\hat{\gamma}_{\text{prop}}) = (\omega_{\delta} - 1) S_{\delta}^4 + \omega_{\delta} S_{\delta}^2 S_{\gamma}^2 + \omega_{\tau_1} S_{\tau_1}^2 S_{\gamma}^2 + \omega_{\tau_2} S_{\tau_2}^2 S_{\gamma}^2
\]

\[
+ 2 \omega_{\delta} (\omega_{\delta} - 1) S_{\delta}^2 \bar{\xi}_\delta - 2 (\omega_{\delta} - 1) S_{\delta}^2 \bar{\xi}_\delta \left( \bar{\xi}_1 - \frac{1}{2} \bar{\xi}_2 \right)
\]

\[
- 2 (\omega_{\delta} - 1) \omega_{\delta} S_{\delta}^2 \bar{\xi}_\delta \left( \bar{\xi}_2 - \frac{1}{2} \bar{\xi}_2 \right) - 2 \omega_{\delta} \omega_{\tau_1} S_{\tau_1}^2 \bar{\xi}_\tau \left( \bar{\xi}_1 - \bar{\xi}_1 \right) - 2 \omega_{\delta} \omega_{\tau_2} S_{\tau_2}^2 \bar{\xi}_\tau \left( \bar{\xi}_2 - \bar{\xi}_2 \right)
\]

\[
+ 2 \omega_{\delta} \omega_{\tau_1} S_{\tau_1}^2 \bar{\xi}_\tau \left( \bar{\xi}_1 - \bar{\xi}_1 \right)
\]
From (2) and (19), we have:

\[
\text{MSE} \left( \hat{Y}_{\text{prop}} \right) = (\omega_s - 1)^2 S_i^4 + \omega_s^2 S_i^4 \bar{x}_{400} + \omega_s^2 S_i^4 \bar{x}_{004}^* + \omega_s^2 S_i^4 \bar{x}_{004}^* + \frac{2(\omega_s \omega_9 - \omega_{10}) S_i^4 \bar{x}_{004}^*}{2} + 2(\omega_s \omega_9 - \omega_{10}) S_i^4 \bar{x}_{004}^* - 2(\omega_s \omega_9 - \omega_{10}) S_i^4 \bar{x}_{220} - 2\omega_s \omega_{10} S_i^4 \bar{x}_{004}^* + 2\omega_s \omega_{10} S_i^4 \bar{x}_{220}^*.
\]

The optimum values of \(\omega_s, \omega_9\) and \(\omega_{10}\) are given by:

\[
\omega_s = \frac{\bar{k} \left( N_i / 400^* - 2 (y_{11} + N_{11}) \right) - 1 / 400^*}{\bar{k} \left( 4 \left( \gamma^* + 1 \right) + (y_{11} + N_{11}) \right) - N_i / 400^* + 1 / 400^*},
\]

\[
\omega_9 = \frac{S_i^2 \left( 2 \bar{k} x_{400}^* \bar{x}_{004}^* (x_{404}^* - x_{022}^*) + \bar{k} x_{004}^* x_{220}^* (x_{004}^* - 2 x_{220}^*) + \bar{k} x_{004}^* x_{220}^* (x_{004}^* - x_{220}^*) + 4 (x_{004}^* x_{220}^* - x_{220}^* x_{022}^*) \right)}{S_i^2 \left( x_{004}^* x_{004}^* - x_{022}^* x_{022}^* \right) \left[ \bar{k} \left( 4 \left( \gamma^* + 1 \right) + (y_{11} + N_{11}) \right) - N_i / 400^* + 1 / 400^* \right]},
\]

\[
\omega_{10} = \frac{S_i^2 \left( 2 \bar{k} x_{400}^* x_{004}^* (x_{404}^* - x_{022}^*) + \bar{k} x_{004}^* x_{220}^* (x_{004}^* - x_{220}^*) + \bar{k} x_{004}^* x_{220}^* (x_{004}^* - x_{220}^*) + 4 (x_{004}^* x_{220}^* - x_{220}^* x_{022}^*) \right)}{S_i^2 \left( x_{004}^* x_{004}^* - x_{022}^* x_{022}^* \right) \left[ \bar{k} \left( 4 \left( \gamma^* + 1 \right) + (y_{11} + N_{11}) \right) - N_i / 400^* + 1 / 400^* \right]}. \]

Where,

\[
y_1 = \frac{x_{220}^2 x_{022}^2 - 2 x_{220}^* x_{022}^* x_{004}^* x_{220}^*}{x_{004}^* x_{004}^* - x_{022}^* x_{022}^*}, \quad N_1 = \frac{x_{004}^* x_{004}^* (x_{004}^* + x_{004}^* - 2 x_{220}^*)}{x_{004}^* x_{004}^* - x_{022}^* x_{022}^*},
\]

\[
y_{11} = \frac{x_{220}^* x_{220}^* (x_{004}^* - x_{022}^*)}{x_{004}^* x_{004}^* - x_{022}^* x_{022}^*}, \quad N_{11} = \frac{x_{220}^* x_{220}^* (x_{004}^* - x_{022}^*)}{x_{220}^* x_{220}^* - x_{004}^* x_{220}^* - x_{004}^* x_{220}^* - x_{004}^* x_{220}^*}. \]

The minimum MSE at the optimum values of \(\omega_s, \omega_9\) and \(\omega_{10}\) in (18), we have:

\[
\text{MSE} \left( \hat{Y}_{\text{prop}} \right)_{\text{min}} = \frac{\bar{k} S_i^2 \left( (y_1 - N_1) \bar{k} + (\gamma^* + 1) \right)}{\bar{k} \left( 4 \left( \gamma^* + 1 \right) + (y_{11} + N_{11}) \right) - N_i / 400^* + 1 / 400^*}. \]

**5. Efficiency comparison**

In this section, the theoretical comparison of the suggested estimator with existing estimators is considered:

i. From (2) and (19),

\[
\text{MSE} (\hat{Y}_{\text{prop}}) < \text{Var} (\hat{Y}_0) \text{ if } \bar{k} S_i^2 (x_{400} - M) > 0
\]

\[
\text{Var} (\hat{Y}_0) - \text{MSE} (\hat{Y}_{\text{prop}}) > 0.
\]
where

\[ M_j = \frac{[(y_{i1} - N_{ij})l + (y^* + 1)]}{\frac{4[(y^* + 1) + (y_{i11} + N_{i11})]}{n_i/400^*} + 1/400^*}} \]

ii. From (4) and (19),

\[ \text{MSE}(\hat{Y}_{\text{prop}}) < \text{MSE}(\hat{Y}_1) \text{ if } \]

\[ \text{MSE}(\hat{Y}_1) - \text{MSE}(\hat{Y}_{\text{prop}}) > 0 \]

\[ \Delta S_j^4(\{l_{400}^* + l_{140}^* - 2l_{220}^*\} - M_j) > 0 \]

iii. From (6) and (19),

\[ \text{MSE}(\hat{Y}_{\text{prop}}) < \text{Var}(\hat{Y}_2) \text{ if } \]

\[ \text{Var}(\hat{Y}_2) - \text{MSE}(\hat{Y}_{\text{prop}}) > 0 \]

\[ \Delta S_j^4(l_{400}^* (1 - \rho^2) - M_j) > 0 \]

iv. From (8) and (19),

\[ \text{MSE}(\hat{Y}_{\text{prop}}) < \text{MSE}(\hat{Y}_3) \text{ if } \]

\[ \text{MSE}(\hat{Y}_3) - \text{MSE}(\hat{Y}_{\text{prop}}) > 0 \]

\[ \Delta S_j^4\left(\frac{l_{400}^* l_{140}^* - l_{220}^*}{l_{400}^* l_{140}^* - l_{220}^* + l_{340}^*} - M_j\right) > 0 \]

v. From (11) and (19),

\[ \text{MSE}(\hat{Y}_{\text{prop}}) < \text{MSE}(\hat{Y}_4) \text{ if } \]

\[ \text{MSE}(\hat{Y}_4) - \text{MSE}(\hat{Y}_{\text{prop}}) > 0 \]

\[ \Delta S_j^4\left(\{l_{300}^* + \frac{1}{4} l_{340}^* - l_{220}^*\} - M_j\right) > 0 \]
vi. From (13) and (19),

\[ \text{MSE}(\hat{Y}_{\text{prop}}) < \text{MSE}(\hat{Y}_5) \text{ if } \]

\[ \text{MSE}(\hat{Y}_5) - \text{MSE}(\hat{Y}_{\text{prop}}) > 0 \]

\[ \Lambda^S_{ij} \left( \frac{64\{\lambda_{120}^* \lambda_{040}^* - \lambda_{220}^*\} - \Lambda \lambda_{040} - 16 \Lambda \lambda_{040} \{\lambda_{120}^* \lambda_{040}^* - \lambda_{220}^*\}}{64\{\Lambda (\lambda_{120}^* \lambda_{040}^* - \lambda_{220}^* + \lambda_{040}^*) - \Lambda\} - \Lambda} \right) > 0 \]

vii. From (15) and (19),

\[ \text{MSE}(\hat{Y}_{\text{prop}}) < \text{MSE}(\hat{Y}_6) \text{ if } \]

\[ \text{MSE}(\hat{Y}_6) - \text{MSE}(\hat{Y}_{\text{prop}}) > 0 \]

\[ \Lambda^S_{ij} \left( \frac{64(\gamma^* + 1) - \Lambda \left( \frac{\lambda_{120}^*}{\lambda_{040}^*} \right) - 16 \Lambda \lambda_{040} (\gamma^* + 1)}{64\{\gamma^* + 1\} - \Lambda} \right) > 0 \]

6. Data description

In this part, we carry out a mathematical study to see the performances of proposed estimator in comparison of existing estimator. We considered two real data sets under SRS. The data descriptions of these populations are given in Table 1. We compare the production of our proposed estimator with existing counterparts in terms of PRE. To obtain the PRE, we used the following expression.

\[ \text{PRE}(u) = \frac{\text{Var}(\hat{Y}_u)}{\text{MSE}(\hat{Y}_u)} \times 100 \]

Where \((u) = (0, 1, 2, 3, 4, 5, 6)\).

**Population 1:** [Source: Singh [41]]
- \(Y\): Estimated number of fish in year 1995.
- \(X\): Estimated number of fish in year 1994,
- \(R_x\) = Rank of \(X\) variable

**Population 2:** [Source: Singh [41]]
- \(Y\): Estimated number of fish in year 1995.
- \(X\): Estimated number of fish in year 1993,
- \(R_x\) = Rank of \(X\) variable

7. Simulation study

We have generated two populations of size 5,000 from normal distribution with different parameters by using R language. The populations have different correlations, i.e. population I is positively correlated and population-II has strong positive correlation between \(X\) and \(Y\) variables. The population details are given below:
The percentage relative efficiency (PRE) is calculated as follows:

$$\text{PRE} = \frac{\text{Var}(\hat{\theta}_j)}{\text{MSE}(\hat{\theta}_j)} \times 100,$$

where \( j = 0,1, \ldots, 6 \).

The results are given below:

8. Discussion

To show the performances of our suggested improved variance estimator, we used two real populations. Data description of these mentioned populations are given in Table 1. The bias result using real data sets are given in Table 2. The MSE and PRE are presented in Table 3.

From the numerical results, the proposed estimator is found to be more effective than its existing estimators. The gain in efficiency in population 2 is more as compared to population 1. As we increase the sample size, the MSE decreases and PRE increases, yielding the expected results. We also evaluated the efficiency of our proposed improved variance estimator in simple random sampling using simulation study. The bias result using simulated data sets are
Table 2. Bias result using Populations 1–2 using two real data sets.

| Estimators | Population-I | Population-II |
|------------|--------------|---------------|
| \( \hat{\gamma}_0 \) | - | - |
| \( \hat{\gamma}_1 \) | 1.187433e+14 | 1.089019e+14 |
| \( \hat{\gamma}_2 \) | - | - |
| \( \hat{\gamma}_3 \) | 1.655189e+18 | 1.205829e+13 |
| \( \hat{\gamma}_4 \) | -1.912397e+13 | -2.405279e+13 |
| \( \hat{\gamma}_5 \) | 6.540944e+16 | 5.91082e+16 |
| \( \hat{\gamma}_6 \) | -1.269339e+46 | -5.184275e+45 |
| \( \hat{\gamma}_{prop} \) | 5.168886e+16 | 5.564613e+16 |

Table 3. MSE and Percentage relative efficiency using Populations 1–2.

| Estimators | Population-I | Population-II |
|------------|--------------|---------------|
| \( \hat{\gamma}_0 \) | 4.725399e+14 | 4.725399e+14 |
| \( \hat{\gamma}_1 \) | 8.206155e+13 | 6.231377e+13 |
| \( \hat{\gamma}_2 \) | 5.960812e+13 | 4.342992e+13 |
| \( \hat{\gamma}_3 \) | 5.714651e+14 | 4.210837e+14 |
| \( \hat{\gamma}_4 \) | 1.203095e+13 | 1.104193e+14 |
| \( \hat{\gamma}_5 \) | 4.639459e+13 | 3.301273e+13 |
| \( \hat{\gamma}_6 \) | 4.518437e+13 | 3.231554e+13 |
| \( \hat{\gamma}_{prop} \) | 2.928899e+13 | 2.600185e+13 |

From Table 5, it is concluded that our suggested improved variance estimator performs better than the standard estimator and other existing estimators for a number of correlation coefficient values. The outcome of the simulation study clearly determines that for the simulated data sets 1–2, the PRE of the proposed estimator is greater than the PRE of the existing estimate. Overall, the gain in efficiency of our proposed estimator is the best as compared to all existing counterparts.

9. Conclusion

In this article, we proposed an improved ratio-in-regression type exponential estimator for the finite population variance using dual auxiliary variables under simple random sampling. Expressions for mean square error of the proposed estimator are derived up to first order of approximation. The proposed estimator is compared with several existing counterparts to judge their uniqueness and superiority using two real data sets. Moreover, simulation study is also conducted to check the robustness and generalizability of the proposed variance estimator. A numerical study is carried out to support the theoretical results. Based on the numerical findings, it is observed from the result of real data sets and simulation study, showing when compared to its existing counterparts; the suggested estimator produces well results. Therefore we recommend the use of our proposed estimators for efficiently estimating the finite population variance under simple random sampling using dual auxiliary variable. The current work can be extended to develop an improved class of estimators under measurement error.
probability proportional to size (PPS), two-stage sampling using dual auxiliary variable for estimating the population variance.

**Acknowledgments**

Princess Nourah bint Abdulrahman University Researchers Supporting Project number (PNURSP2022R299), Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

**Author Contributions**

**Conceptualization:** Sohaib Ahmad, Sardar Hussain, Kalim Ullah, Erum Zahid, Muhammad Aamir, Javid Shabbir, Zubair Ahmad.

**Data curation:** Sohaib Ahmad, Sardar Hussain, Kalim Ullah, Huda M. Alshanbari.

**Formal analysis:** Sohaib Ahmad, Sardar Hussain, Erum Zahid, Zubair Ahmad, Huda M. Alshanbari, Wejdan Alajlan.

**Methodology:** Sardar Hussain.

**Software:** Muhammad Aamir, Huda M. Alshanbari, Wejdan Alajlan.

**Supervision:** Muhammad Aamir, Javid Shabbir.

---

**Table 4. Bias result using Populations 1–2 using simulated data sets.**

| Estimators | Population-I | Population-II |
|------------|--------------|---------------|
| $\hat{Y}_a$ | -            | -             |
| $\hat{Y}_1$ | 199.1479     | 9.756393      |
| $\hat{Y}_2$ | -            | -             |
| $\hat{Y}_3$ | 12722895     | 1619653       |
| $\hat{Y}_4$ | 39.8726      | -21.2332      |
| $\hat{Y}_5$ | 465181.2     | 136970.7      |
| $\hat{Y}_6$ | -1.156381e+19 | -3.046733e+20 |
| $\hat{Y}_{prop}$ | -128030.3 | -5891.692 |

https://doi.org/10.1371/journal.pone.0276540.t004

**Table 5. MSE and PRE of $\hat{Y}_{prop}$ w.r.t the existing counterparts using two simulated data.**

| Estimators | Population-1 | Population-II |
|------------|--------------|---------------|
| $\hat{Y}_a$ | 185.8621     | 172.0718      |
| $\hat{Y}_1$ | 135.8605     | 136.8036      |
| $\hat{Y}_2$ | 109.091      | 170.3734      |
| $\hat{Y}_3$ | 107.9588     | 172.1601      |
| $\hat{Y}_4$ | 112.3095     | 165.491       |
| $\hat{Y}_5$ | 107.3989     | 173.0577      |
| $\hat{Y}_6$ | 107.3748     | 173.0966      |
| $\hat{Y}_{prop}$ | 70.44824 | 263.8278 |

https://doi.org/10.1371/journal.pone.0276540.t005
Validation: Javid Shabbir.

Writing – original draft: Sohaib Ahmad, Sardar Hussain, Kalim Ullah, Erum Zahid, Zubair Ahmad, Wejdan Alajlan.

Writing – review & editing: Zubair Ahmad, Huda M. Alshanbari.

References
1. Das AK. Use of auxiliary information in estimating the finite population variance. Sankhya, c. 1978; 40:139–48.
2. Prasad B, Singh HP. Some improved ratio-type estimators of finite population variance in sample surveys. Communications in Statistics-Theory and Methods. 1990 Jan 1; 19(3):1127–39.
3. Swain AK, Mishra G. Estimation of finite population variance under unequal probability sampling. Sankhyā: The Indian Journal of Statistics, Series B. 1994 Dec 1:374–88.
4. Ahmad S, Shabbir J. Use of extreme values to estimate finite population mean under pps sampling scheme. Journal of Reliability and Statistical Studies. 2018 Oct 5:99–112.
5. Kadilar C, Cingi H. Improvement in variance estimation in simple random sampling. Communications in Statistics—Theory and Methods. 2007 Aug 9; 36(11):2075–81.
6. Subramani J, Kumarapandyan G. Variance estimation using quartiles and their functions as an auxiliary variable. Ratio. 2012; 1(1).
7. Singh HP, Solanki RS. Improved estimation of finite population variance using auxiliary information. Communications in Statistics-Theory and Methods. 2013 Aug 3; 42(15):2718–30.
8. Yadav R., Upadhyaya L. N., Singh H. P., & Chatterjee S. (2013). A generalized family of transformed ratio-product estimators for variance in sample surveys. Communications in Statistics-Theory and Methods, 42(10), 1839–1850.
9. Singh GN, Priyanka P, Prasad S, Singh S, Kim JM. A class of estimators for population variance in two occasion rotation patterns. Communications for Statistical Applications and Methods. 2013; 20(4):247–57.
10. Yadav SK, Kadilar C. A two parameter variance estimator using auxiliary information. Applied Mathematics and Computation. 2014 Jan 1; 226:117–22.
11. Adichwal NK, Sharma P, Verma HK, Singh R. Generalized class of estimators for population variance using auxiliary attribute. International journal of applied and computational mathematics. 2016 Dec; 2(4):499–508.
12. Adichwal NK, Sharma P, Singh R. Generalized class of estimators for population variance using information on two auxiliary variables. International Journal of Applied and Computational Mathematics. 2017 Jun; 3(2):651–61.
13. Yasmeen U, Noor-ul-Amin M, Hanif M. Exponential estimators of finite population variance using transformed auxiliary variables. Proceedings of the National Academy of Sciences, India Section A: Physical Sciences. 2019 Mar; 89(1):185–91.
14. Zaman T. An efficient exponential estimator of the mean under stratified random sampling. Mathematical Population Studies. 2021 Apr 3; 28(2):104–21.
15. Zaman T. Generalized exponential estimators for the finite population mean. Statistics in Transition new series. 2020; 21(1):159–68.
16. Zaman T, Bulut H. An efficient family of robust-type estimators for the population variance in simple and stratified random sampling. Communications in Statistics-Theory and Methods. 2021 Jul 15:1–5.
17. Zaman T, Sagir M, Şahin M. A new exponential estimators for analysis of COVID-19 risk. Concurrency and Computation: Practice and Experience. 2022 May 1; 34(10):e6806.
18. Zaman T, Dünér E, Adu A, Allah DA, Shahzad U, Hanif M. Robust regression-ratio-type estimators of the mean utilizing two auxiliary variables: A simulation study. Mathematical Problems in Engineering. 2021 Sep 6; 2021.
19. Bhusan S, Kumar A, Kumar S. Efficient classes of estimators for population variance using attribute. International Journal of Mathematics in Operational Research. 2022; 22(1):74–92.
20. Shahzad U, Ahmad I, Almanjahie IM, Koyuncu N, Hanif M. Variance estimation based on L-moments and auxiliary information. Mathematical Population Studies. 2022 Jan 2; 29(1):31–46.
21. Ahmad S, Aamir M, Hussain S, Shabbir J, Zahid E, Subkrajàng K, Jirawattanapinit A. A New Generalized Class of Exponential Factor-Type Estimators for Population Distribution Function Using Two Auxiliary Variables. Mathematical Problems in Engineering. 2022 Aug 8; 2022.
22. Ozturk O. (2014). Statistical inference for population quantiles and variance in judgment post-stratified samples. Computational Statistics & Data Analysis, 77, 188–205.

23. Ozturk O, Demirel N. Estimation of population variance from multi-ranker ranked set sampling designs. Communications in Statistics-Simulation and Computation. 2016 Nov 25; 45(10):3568–83.

24. Frey J., & Feeman T. G. (2013). Variance estimation using judgment post-stratification. Annals of the Institute of Statistical Mathematics, 65(3), 551–569.

25. Zamanzade E, Vock M. Variance estimation in ranked set sampling using a concomitant variable. Statistics & Probability Letters. 2015 Oct 1; 105:1–5.

26. Mahdizadeh M, Zamanzade E. Estimating asymptotic variance of M-estimators in ranked set sampling. Computational Statistics. 2020 Dec; 35(4):1785–803.

27. Mahdizadeh M, Zamanzade E. New estimator for the variances of strata in ranked set sampling. Soft Computing. 2021 Jul; 25(13):8007–13.

28. Bhushan S, Kumar A, Kumar S, Singh S. SOME MODIFIED CLASSES OF ESTIMATORS FOR POPULATION VARIANCE USING AUXILIARY ATTRIBUTE. Pakistan Journal of Statistics. 2022 Apr 1; 38(2).

29. Bhushan S, Kumar A, Singh S, Kumar S. An improved class of estimators of population mean under simple random sampling. Philippine Statist. 2021; 70(1):33–47.

30. Zaman T, Bulut H. A new class of robust ratio estimators for finite population variance. Scientia Iranica. 2022 Jan 4.

31. Zaman T, Bulut H. A simulation study: Robust ratio double sampling estimator of finite population mean in the presence of outliers. Scientia Iranica. 2021 Oct 5.

32. Yadav SK, Zaman T. Use of some conventional and non-conventional parameters for improving the efficiency of ratio-type estimators. Journal of Statistics and Management Systems. 2021 Jul 4; 24(5):1077–100.

33. Zaman T, Kadilar C. Novel family of exponential estimators using information of auxiliary attribute. Journal of Statistics and Management Systems. 2019 Nov 17; 22(8):1499–509.

34. Zaman T, Dünder E. Proposing Novel Modified Ratio Estimators by Adding an Exponential Parameter. Lobachevskii Journal of Mathematics. 2020 Mar; 41(3):451–8.

35. Singh GN, 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. Communications in Statistics-Simulation and Computation. 2022 Apr 3; 51(4):1454–73.

36. Isaki CT. Variance estimation using auxiliary information. Journal of the American Statistical Association. 1983 Mar 1; 78(381):117–23.

37. Rao TJ. On certain methods of improving ration and regression estimators. Communications in Statistics-Theoory and Methods. 1991 Jan 1; 20(10):3325–40.

38. Singh R, Chauhan P, Sawan N, Smarandache F. Improved exponential estimator for population variance using two auxiliary variables. Infinite Study; 2009 Feb 1.

39. Grover LK, Kaur P. Ratio type exponential estimators of population mean under linear transformation of auxiliary variable: theory and methods. South African Statistical Journal. 2011 Jan 1; 45(2):205–30.

40. Ahmad S, Hussain S, Aamir M, Yasmeen U, Shabbir J, Ahmad Z. Dual use of auxiliary information for estimating the finite population mean under the stratified random sampling scheme. Journal of Mathematics. 2021 Nov 16; 2021.

41. Singh S. Advanced Sampling Theory with Applications: How Michael”” Selected”” Amy. Springer Science & Business Media; 2003.