Enhancing Precision in Population Variance Vector Estimation: A Two-Phase Sampling Approach with Multi-Auxiliary Information
(Meningkatkan Ketepatan dalam Anggaran Vektor Varians Populasi: Pendekatan Persampelan Dua Fasa dengan Maklumat Berbilang Bantu)

AMBER ASGHAR¹, AAMIR SANAULLAH², MUHAMMAD HANIF³ & LAILA A. AL-ESSA⁴

¹Department of Statistics, Faculty of Science and Technology, Virtual University of Pakistan, Lahore, Pakistan
²Department of Statistics, COMSATS University Islamabad, Lahore Campus, Pakistan
³Department of Statistics, National College of Business Administration & Economics, Lahore, Pakistan
⁴Department of Mathematical Sciences, College of Sciences, Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia

Received: 11 January 2024/Accepted: 11 June 2024

ABSTRACT
To enhance precision in estimating unknown population parameters, an auxiliary variable is often used. However, in scenarios where required information on an auxiliary variable is partially or fully unavailable, two-phase sampling is commonly employed. The challenge of estimating the variance vector using multi-auxiliary variables is a less explored area in current literature. This paper addresses the estimation of vector of unknown population variances for multiple study variables by using an estimated vector of variances derived from multi-auxiliary information. This approach is particularly relevant when population variances for the multi-auxiliary variables are not known prior to the survey. The paper introduces a generalized variance and a vector of biases for the proposed multivariate estimator. Special cases of the proposed multivariate variance estimator are provided, accompanied by expressions for mean square errors. Theoretical mathematical conditions are discussed to guide the preference for the proposed estimator. Through the analysis of real-world application-based data, the applicability and efficiency of the proposed multivariate variance estimator are demonstrated, outperforming modified versions of multivariate variance estimators. Additionally, a simulation study validates the superior performance of the proposed estimator compared to its modified estimators.

Keywords: Generalized variance; multivariate estimator; regression-cum-exponential estimator; two-phase sampling; variance vector estimator

ABSTRAK

Kata kunci: Penganggar multivariat; penganggar regresi merangkap eksponen; penganggar vektor varians; pensampelan dua fasa; varians umum
INTRODUCTION

In recent years, survey sampling has become helpful in various sectors, including academia, healthcare, and both public and private industries. Surveys, employing both probability and non-probability sampling, are important in fields such as agriculture, industry, and healthcare. The crucial role of survey sampling in collecting data across diverse fields is indisputable.

As the general use of survey sampling grows, the demand for more advanced methods to interpret results becomes dominant. Multivariate estimation is also one of them. For example, in environmental monitoring, there are instances where auxiliary information (such as meteorological conditions) is not available for all monitoring stations or time periods. The use of a multivariate estimator (MV) without auxiliary information allows researchers to still derive estimates of air quality indicators, providing valuable insights into city-wide pollution levels. Among these methods, variance estimation emerges as the best choice for addressing the elaboration of complex survey designs. Neyman (1938) introduced a cost-effective two-phase sampling technique which is particularly valuable when collecting data on the variable of interest proves financially burdensome. For a deeper understanding of the application of two-phase sampling, refer to Breidt and Fuller (1993), Cochran (1977), Hussain et al. (2018), and Rao (1973).

Cebrián and García (1997) proposed an almost unbiased multivariate ratio-type estimator for population variance. Das and Tripathi (1978) addressed variance estimation by incorporating both population variance and the coefficient of variation (CV) of an auxiliary variable. Isaki (1983) developed ratio and regression estimators using the variance of the auxiliary variable for variance estimation.


Shahzad et al. (2021a) estimated variance using attributes of auxiliary variables, while Zaman and Bulut (2019) proposed a generalized variance approach instead of traditional ratio estimators. Additionally, Shahzad et al. (2021b) proposed estimators for population variance based on L-moments, such as L-mean, L-standard deviation, and L-coefficient of variation.

Various researchers have proposed different estimators for the estimation of population variance under the assumption that the population variance of auxiliary variables is known prior to a survey. Notable works include those by Ahmad, Hussain and Hanif (2016), Arcos and Rueda (1997), Asghar et al. (2023), Niaz et al. (2022), Sanaullah et al. (2020), and Zaman and Bulut (2019). However, real-life situations often arise where population variances are not available before a survey. In such cases, two-phase sampling becomes essential to estimate the unknown population variance of auxiliary variables.

Researchers, such as Abu-Dayyeh and Ahmed (2005), have addressed this issue in the context of estimating the variance of a single study variable. Additionally, Sanaullah, Hanif and Asghar (2016) introduced generalized exponential-type ratio and product estimators for estimating variance when the mean auxiliary variable is unknown prior to the survey under two-phase sampling. More recently, Abid et al. (2020) proposed a ratio estimator for robust measures of population variance and compared it with competing estimators using quantiles of auxiliary variables, concluding its robust performance even in the presence of outliers.

Motivated by the studies discussed in the earlier text, this paper introduces a novel multivariate regression-cum-exponential (MRCE) type estimator for scenarios where information about the parameter(s) of auxiliary variables is not available. While usual regression and multiple regression estimators perform well in symmetric and linear situations, practical scenarios often exhibit non-symmetric and skewed distributions. In such cases, regression-cum-exponential type estimators are expected to yield superior results compared to simple regression estimators.
For estimation of vector of population variances, let us consider the notations and expectations given by,

\[
E \left( \varepsilon_{ij} \right) = E \left( \varepsilon_{ij} \right) = E \left( \varepsilon_{ij} \right) = 0
\]

\[
E \left( \varepsilon_{ij}^2 \right) = \gamma_i A_{ij}, \quad E \left( \varepsilon_{ij}^2 \right) = \gamma_j A_{ij}, \quad E \left( \varepsilon_{ij} \varepsilon_{jk} \right) = \gamma_{ij} A_{ij}, \quad E \left( \varepsilon_{ij} \varepsilon_{jk} \varepsilon_{kl} \right) = \gamma_{ij} \gamma_{kl}, \quad E \left( \varepsilon_{ij} \varepsilon_{jk} \varepsilon_{kl} \varepsilon_{mn} \right) = \gamma_{ij} \gamma_{kl} \gamma_{mn}, \quad E \left( \varepsilon_{ij} \varepsilon_{jk} \varepsilon_{kl} \varepsilon_{mn} \varepsilon_{op} \right) = \gamma_{ij} \gamma_{kl} \gamma_{mn} \gamma_{op}, \quad \text{and} \quad \gamma_{ij} \gamma_{kl} \gamma_{mn} \gamma_{op} \gamma_{qr}.
\]

\[
E \left( \varepsilon_{ij} \varepsilon_{jk} \varepsilon_{kl} \varepsilon_{mn} \varepsilon_{op} \varepsilon_{qr} \right) = \gamma_{ij} \gamma_{kl} \gamma_{mn} \gamma_{op} \gamma_{qr} \gamma_{st}, \quad \text{and} \quad \gamma_{ij} \gamma \gamma_{kl} \gamma \gamma_{mn} \gamma \gamma_{op} \gamma \gamma_{qr} \gamma \gamma_{st}.
\]

Let \( S^2 \) be the population variance, and its usual unbiased sample variance estimator vector, is given by

\[
t_i = \left[ t_{ij} \right]_{j=1}^J, \quad \text{where,} \quad t_{ij} = s_{ij}^2, \quad j = 1, 2, \ldots, m.
\]

The variance of the modified unbiased sample variance vector is defined as,

\[
\Sigma t_0 = \gamma_2 \sum_{j=1}^J t_{ij}(j),
\]

Let \( \Sigma \) be the sample variance estimator vector using sample means of auxiliary information is modified, and it is given by,

\[
t_{reg} = \left[ t_{reg} \right]_{j=1}^J, \quad \text{where,} \quad t_{reg}
\]

\[
= s_{reg}^2 + \sum_{k=1}^J \sum_{j=1}^J \left( \bar{x}_{1j} - \bar{x}_{2k} \right)^2.
\]

The expression of Generalized variance of \( t_{reg} \), is given by,

\[
\Sigma_{reg} = S \sum \left( \bar{x}_{2j} - \bar{x}_{2k} \right)^2. \sum_{j=1}^J \sum_{k=1}^J \left( \bar{x}_{1j} - \bar{x}_{1k} \right)^2.
\]

Following Shabbir and Gupta (2015), a MRCE type estimator is modified for estimating a vector of population variances, and it is given with a form given by, where,

\[
t_s = \left[ t_{s} \right]_{j=1}^J, \quad \text{where,} \quad t_{sj} = s_{sj}^2 + \sum_{k=1}^J \sum_{j=1}^J \left( \bar{x}_{1j} - \bar{x}_{1k} \right)^2.
\]
The Generalized variance of \( t_j \) is given by,
\[
\sum \varepsilon = SS \left[ \gamma_2 \sum y_{\text{obs}} - (\gamma_2 - \gamma_1) \sum y_{\text{obs}} \right].
\]

PROPOSED METHODOLOGY FOR ESTIMATING THE VECTOR OF VARIANCES

In this section, a generalized MRCE estimator under two-phase sampling is proposed for estimating a vector of population variances. Form of the proposed MV estimator is given by,
\[
t_p = \left[ t_{\theta (j)} \right]_{j=1}^{m},
\]
where
\[
t_{\theta (j)} = \left( s_{(j)}^2 + \sum_{k=1}^{\frac{1}{2}} b_j \left( s_{(j)k}^2 - s_{(k)j}^2 \right) \right) \exp \left( \sum_{k=1}^{\frac{1}{2}} b_j \left( s_{(j)k}^2 - s_{(k)j}^2 \right) \right) \left( \sum_{k=1}^{\frac{1}{2}} b_j \left( s_{(j)k}^2 - s_{(k)j}^2 \right) \right)
\]
(5)

It is clear that positive values of \( b_j \) produce different families of multivariate exponential (ME) ratio estimators, and negative values of \( b_j \) will give different families of ME product estimators. \( \delta_j \) is assumed to be unknown and needs to be estimated and emphasized its optimal value.

DERIVATION OF THE VECTOR OF THE BIAS, AND THE GENERALIZED VARIANCE

To determine the generalized variance and bias vector, the proposed estimator, as defined in Equation (5), incorporates sampling errors. At the start, focus on the \( j \)-th estimator of this proposed estimator, expressed as follows,
\[
t_{\theta (j)} = \left( s_{(j)}^2 + \sum_{k=1}^{\frac{1}{2}} b_j \left( s_{(j)k}^2 - s_{(k)j}^2 \right) \right) \exp \left( \sum_{k=1}^{\frac{1}{2}} b_j \left( s_{(j)k}^2 - s_{(k)j}^2 \right) \right)
\]
(6)

Retaining the terms up to the order \( O(n^{-1}) \), the proposed estimator is given as,
\[
t_{\theta (j)} = \left( s_{(j)}^2 + \sum_{k=1}^{\frac{1}{2}} b_j \left( e_{(j)k} - e_{(k)j} \right) \right) \exp \left( \sum_{k=1}^{\frac{1}{2}} b_j \left( e_{(j)k} - e_{(k)j} \right) \right)
\]
(7)

After some simplification, the equation is obtained as,
\[
t_{\theta (j)} = S_{(j)}^2
\]
(8)

Applying expectations, a vector of the bias expression for \( j \)-th estimator is given by,
\[
\text{Bias} \left( t_{\theta (j)} \right) = S_{(j)}^2
\]
(9)

Consider Equation (8) to proceed as,
\[
\Sigma_{\text{var}} = E \left( \left( t_{\text{var}} - S_{\text{var}} \right) \left( t_{\text{var}} - S_{\text{var}} \right) ^T \right) = S_{(j)}^2 E \left( \left( t_{\text{var}} - S_{\text{var}} \right) \left( t_{\text{var}} - S_{\text{var}} \right) ^T \right)
\]
(10)

where, \( \Psi_{\text{var}} = \left( 2 \delta_j + b_k \right)_{(k=1)} \), \( k = 1, 2, ..., l, j = 1, 2, ..., m \).

Alternatively, Equation (10) can take another form, given by:
\[
\Sigma_{\text{var}} = S \left[ \gamma_2 \Sigma_{\text{var}} + \frac{1}{2} \left( \gamma_1 - \gamma_2 \right) \Sigma_{\text{var}} \right] \Psi_{\text{var}}
\]
(11)
The optimum value of $\Psi$ are obtained by differentiating Equation (11) w.r.t $\Psi$, and equating the first derivative to zero. The optimum value is given by,

$$
\Psi_{opt/\varphi} = 2\sum_{l(i,m)}^{-1} \sum_{\eta(i,m)} \varphi(i,m).
$$

Put the optimum value of $\Psi$ in Equation (11), the minimum Generalized variance of $t_p$ is obtained, and it is given by

$$
\min \sum_{j(m,n)} = S'(\gamma_2 \sum_{j(m,n)} - \gamma_1 \sum_{\eta(i,m)} \sum_{\eta(i,m)} \delta(i,m)) (12)
$$

**Remark 1**

It is observed that considering $b_{i} = 0$ in Equation (5) the MR estimator may be obtained for the no information case using MAVs. The vector of the bias and the generalized variance may be obtained directly from Equations (9) and (11).

$$
t_{o} = \{s_{y(i)} + \sum_{i=1}^{l} \delta(i,m)(s_{y(i)} - s_{y(i,m)})\}
$$

where, $t_{o} = \{s_{y(i)} + \sum_{i=1}^{l} \delta(i,m)(s_{y(i)} - s_{y(i,m)})\}$.

**Remark 2**

Similarly, it is also noted that the ME estimator may be obtained by putting $\delta(i,m) = 0$ directly in Equation (5) and its bias and generalized variance may be obtained from Equations (9) and (11) for the situation when the required information about the parameters or the function of parameters, such as the population means, variances, correlation, coefficient of variation etcetera of the MAVs is not available.

$$
t_{o} = \{s_{y(i)} + \sum_{i=1}^{l} \delta(i,m)(s_{y(i)} - s_{y(i,m)})\}
$$

where, $t_{o} = \exp\{\sum_{i=1}^{l} \delta(i,m)(s_{y(i)} - s_{y(i,m)})\}$.

**Remark 3**

Observing that $j = 1, k = 1, 2, 3, \ldots$, is considered in Equation (5), a generalized univariate regression-cum-exponential estimator can be derived. This setting refers to situations where no information is available, and the resulting estimator can be expressed as,

$$
t_{o} = \left[t_{o} \right]^{(1,0)}, j = 1, 2, 3, \ldots, m,
$$

where, $t_{o} = (s_{y(i)} + \sum_{k=1}^{l} \delta(i,m)(s_{y(i)} - s_{y(i,m)})\exp\left\{\sum_{k=1}^{l} \delta(i,m)(s_{y(i)} - s_{y(i,m)})\right\}$.

Under the same conditions, one may obtain the bias and generalized variance for the generalized univariate estimator from Equations (9) (11).

**RELATIVE PERFORMANCE OF THE PROPOSED ESTIMATOR**

To assess the effectiveness of the suggested MREV estimator, a theoretical comparison is conducted, evaluating its performance in a univariate context against various alternative modified estimators. The proposed estimator $t_{o}$ will perform more efficiently than usual sample variance estimator $t_{o}$ iff,

$$
\frac{A_{o}}{A_{o}} > \frac{\delta^{2} + 1}{4}.
$$

The proposed estimator $t_{o}$ performs better than Isaki’s (1983) regression estimator $t_{o}$ when

$$
\frac{A_{o}}{A_{o}} > \frac{\delta^{2} + 1}{4}.
$$

The proposed estimator $t_{o}$ performs better than Shabbir and Gupta (2015) regression estimator $t_{o}$ when following condition is met,

$$
\sqrt{A_{o}} \Psi - 2A_{o} > (A_{o})^{2} < c,
$$

$$
\frac{A_{o}}{A_{o}} > \frac{\delta^{2} + 1}{4} \left(1 - \frac{\sqrt{c}}{\sqrt{A_{o}}} \right) \text{ or } \frac{A_{o}}{A_{o}} > \frac{\delta^{2} + 1}{4} \left(1 + \frac{\sqrt{c}}{\sqrt{A_{o}}} \right),
$$

where, $c = \left(\frac{\Psi}{\Psi} A_{o} - 8A_{o} \Psi \delta^{2} + \frac{2A_{o} \Psi^{2}}{A_{o}} \right)^{2}$.

The proposed estimator $t_{o}$ performs better than Sanaullah, Hanif and Asghar (2016) regression-cum-exponential estimator $t_{o}$ when,
\[
\frac{A_{yx}}{A_x} > -\phi \delta - \frac{1}{2} \left[ \frac{\sqrt{\rho}}{\sqrt{A_x}} - 1 \right], \quad \text{or} \quad \frac{A_{yx}}{A_x} > \phi \delta - \frac{1}{2} \left[ \frac{\sqrt{\rho}}{\sqrt{A_x}} + 1 \right],
\]
where, \( d = \beta A_{xy} \left( \beta - 4 \frac{A_{yx}}{A_{xx}} \right) + \left( \frac{2A_{yx}}{\sqrt{A_x}} \right)^2. \)

### RESULTS AND DISCUSSION

In this study, the MRCE is utilized, demonstrating its application with real-life data extracted from Canadian climate records, as published by the National Oceanic and Atmospheric Administration (NOAA). The dataset focuses on daily weather records for the month of May 2017, collected from 37,247 different weather stations across Canada.

To contour the analysis, transform the data into weekly summaries, specifically for the weeks of May 08-14, 2017; May 15-21, 2017; and May 22-28, 2017. To ensure data wholeness, stations with incomplete records for a full week are excluded from the analysis. Only weeks with recorded data for the entire week are retained for further investigation.

For the auxiliary variables, monthly temperature data for the last three years (2016, 2015, and 2013) is considered. Stations lacking temperature records for the entire week in these years are excluded. As a result, 704 stations are retained with complete temperature data for the full week for our analysis. To address the missing values and facilitate variable transformations, the tidyverse Package (2016) is used. The study variable is defined as the Temperature Averages (TAVG) for the three weeks of May 2017, denoted as \( i = 1, 2, 3 \). The auxiliary variables for the previous three years (2016, 2015, and 2013) consist of monthly TAVGs, also denoted as \( i = 1, 2, 3 \).

A finite population approach is used to model the variance and covariance of weekly temperatures across Canada. A comprehensive description of population characteristics is presented in Table A1 in the appendix. In this study, a two-phase sampling strategy is employed for estimating the population variance. Initially, a sample of size \( n_1 = 211 \) is selected from the population, and relevant calculations are performed based on this first-phase sample. Subsequently, using a simple random sampling without replacement, a second-phase sample of size \( n_2 = 105 \) is drawn from the previously selected first-phase sample. For both the first-phase and second-phase samples, key statistics are computed. These statistics are then integrated into the proposed multivariate (MV) estimator, as well as the existing MV estimators, to derive estimates for the population variance. This entire process is simulated 1000 times to ensure robustness in the estimation.

Finally, the determinants for the generalized variance of each MV estimator are calculated. The results are presented in Table 1, setting the corresponding values for both the proposed and modified existing estimators. To address the missing-ness in the data where it is present, mice R statistical package (2011) is used. The Broom (2016) package for replicating two-phase sampling procedure is adopted, so as to take first-phase and second phase samples in a faster and more efficient way.

Table 2 shows the determinants of the proposed, and existing estimators. From Table 2, the proposed MV estimator is attaining the smaller value of determinant of its generalized variance, where the determinants for each of the mentioned existing estimators are larger. The percent relative efficiency (PRE) value for each of the mentioned estimators including the proposed estimator is also computed. Here, PRE values also indicate that the proposed MV estimator is better in performance as compared to the performance of mentioned existing estimators. Tables 1 and 2 clearly indicate that the empirical results are in favor of the proposed MV estimator.

### SIMULATION RESULTS AND DISCUSSION

A simulation is performed to assess the performance of the proposed MV, along with other mentioned estimators, for estimating population variance under a two-phase sampling design. Consider the three study variables \((Y_1, Y_2, Y_3)\) and the three auxiliary variables \((X_1, X_2, X_3)\) for multivariate estimation. The auxiliary variables are generated following the mechanism defined as follow:

\[
X_1 \sim N(12, 3.3); X_i \sim X_i^* \rho + \sqrt{1-\rho^2} \cdot N(12, 2.5);
\]

\[
X_i \sim X_i^* \rho + \sqrt{1-\rho^2} \cdot N(15, 2.6),
\]

where, \( \rho = 0.9 \).

The study variables are simulated using the mechanism given as:

\[
Y_{ij} = \sum_{i=1}^{k} \alpha_i X_i^2 + \epsilon, \quad \text{where}, \quad \epsilon \sim N(0, 1).
\]

The three study variables with three auxiliary variables under the discussed model are distinct by the equations:
The statistics computed from both the first-phase and second-phase samples are then incorporated into the proposed estimator and the specified existing estimators to estimate the population variance. This simulation process is repeated 10000 times to compute the generalized variance for the proposed estimator and the mentioned existing estimators. The simulation results, including the generalized variance and the PRE values, are presented in Tables 3 and 4, respectively.

In Table 3, a value on the main diagonal of a generalized variance is a MSE of the estimator corresponding to the row and the column. Tables 3 and 4 show that the proposed estimator is more efficient than the considered estimators because a lower MSE indicates that the estimator tends to be closer to the true value of the parameter being estimated. The conclusion based on the simulation indicates that the proposed estimator remains consistent with the conclusion drawn through the real-life data.

\[
Y_1 = 1.7X_1^2 + 1.3X_2^2 + 1.6X_3^2 + \varepsilon, \\
Y_2 = 1.2X_1^2 + 1.5X_2^2 + 1.8X_3^2 + \varepsilon, \\
Y_3 = 1.4X_1^2 + 1.5X_2^2 + 1.2X_3^2 + \varepsilon.
\]

The conclusion based on the simulation indicates that the proposed estimator remains consistent with the conclusion drawn through the real-life data.

**TABLE 1. Generalized variance of the proposed and existing estimators**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Generalized variance of ( t_p )</th>
<th>Generalized variance of ( t_o )</th>
<th>Generalized variance of ( t_{reg} )</th>
<th>Generalized variance of ( t_r )</th>
<th>Generalized variance of ( t_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_p )</td>
<td>( \begin{bmatrix} 354.46678 &amp; 52.91568 &amp; 50.98563 \ 52.91568 &amp; 51.92468 &amp; 37.58348 \ 50.98563 &amp; 37.58348 &amp; 71.66995 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 573.5104 &amp; 157.2954 &amp; 141.02297 \ 157.2955 &amp; 92.34649 &amp; 74.25925 \ 141.0230 &amp; 74.25925 &amp; 96.01242 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 364.71483 &amp; 29.39202 &amp; 24.90685 \ 29.39202 &amp; 62.93120 &amp; 49.83716 \ 24.90685 &amp; 49.83716 &amp; 75.40119 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 618.0714 &amp; 172.21395 &amp; 156.89180 \ 172.2140 &amp; 98.04573 &amp; 80.48775 \ 156.8918 &amp; 80.48775 &amp; 102.89652 \end{bmatrix} )</td>
<td>( \begin{bmatrix} 597.2441 &amp; 168.76185 &amp; 154.09151 \ 168.7618 &amp; 97.50908 &amp; 80.23343 \ 154.0915 &amp; 80.23343 &amp; 102.76829 \end{bmatrix} )</td>
</tr>
</tbody>
</table>

**TABLE 2. Determinants of the generalized variance of the proposed and existing estimators**

<table>
<thead>
<tr>
<th>Estimator</th>
<th>( t_p )</th>
<th>( t_o )</th>
<th>( t_{reg} )</th>
<th>( t_r )</th>
<th>( t_s )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determinant</td>
<td>1004805</td>
<td>685570.9</td>
<td>793535.9</td>
<td>1115737</td>
<td>1070918</td>
</tr>
<tr>
<td>PRE</td>
<td>146.5647</td>
<td>100</td>
<td>126.6238</td>
<td>90.0575</td>
<td>93.82655</td>
</tr>
</tbody>
</table>
TABLE 3. Simulation based on generalized variance of the proposed and existing estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>Generalized variance of $t_p$</th>
<th>Generalized variance of $t_0$</th>
<th>Generalized variance of $t_{reg}$</th>
<th>Generalized variance of $t_{rg}$</th>
<th>Generalized variance of $t_{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$(78278828, 79774707, 61232339)$</td>
<td>$(160429771, 163784762, 125783555)$</td>
<td>$(160420103, 163775008, 125774928)$</td>
<td>$(150722369, 154107242, 118140599)$</td>
<td>$(160435404, 163790189, 125788487)$</td>
</tr>
<tr>
<td></td>
<td>$(79774707, 81814920, 62339927)$</td>
<td>$(163784762, 167725477, 128352965)$</td>
<td>$(163775008, 167715614, 128344245)$</td>
<td>$(154107242, 158089118, 120732127)$</td>
<td>$(163790189, 167730726, 128357699)$</td>
</tr>
<tr>
<td></td>
<td>$(61232339, 62339927, 48054423)$</td>
<td>$(125783555, 128352965, 98774953)$</td>
<td>$(125774928, 128344245, 98767358)$</td>
<td>$(118140599, 120732127, 92757288)$</td>
<td>$(125788487, 128357699, 98779173)$</td>
</tr>
</tbody>
</table>

TABLE 4. Determinants of the proposed and existing estimators

<table>
<thead>
<tr>
<th>Estimator</th>
<th>$t_p$</th>
<th>$t_0$</th>
<th>$t_{reg}$</th>
<th>$t_{rg}$</th>
<th>$t_{a}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Determinant</td>
<td>$6.009 \times 10^{18}$</td>
<td>$1.226 \times 10^{19}$</td>
<td>$1.2258 \times 10^{19}$</td>
<td>$1.161 \times 10^{19}$</td>
<td>$1.2256 \times 10^{19}$</td>
</tr>
<tr>
<td>PRE</td>
<td>204.0389</td>
<td>100</td>
<td>100.0186</td>
<td>105.5396</td>
<td>100.0306</td>
</tr>
</tbody>
</table>

CONCLUSIONS

In conclusion, this paper contributes to the field of multivariate variance estimation by introducing a novel approach that addresses the challenges of utilizing auxiliary information when the required parameters of the auxiliary variables are not readily available. In this study, the MSEs of each estimator, including the proposed estimator and competitor estimators, are computed using real-life application data (Tables 1 & 2) and simulated data (Tables 3 & 4). From these tables, it is evident that the proposed estimators have smaller MSEs than those of the competitor estimators. Therefore, it can be concluded that the proposed estimator is more efficient based on achieving smaller MSEs. The proposed MV estimator shows superior performance, as demonstrated by both real-life data analysis and simulation studies. This work lays the foundation for further advancements in precision-enhancing techniques for population parameter estimation.
ACKNOWLEDGMENTS
Princess Nourah bint Abdulrahman University Researchers
Supporting Project number (PNURSP2024R443), Princess
Nourah bint Abdulrahman University, Riyadh, Saudi
Arabia.

REFERENCES
Abid, M., Sherwani, K.A.R., Tahir, M., Nazir, Z.H. & Riaz, M.
2020. An improved and robust class of variance estimator.
Scien
ta Arania 28(6): 3589-3601.
estimators for the variance under two-phase sampling.
population variance in successive sampling using
multi-auxiliary variables. Communication in Statistics-
Theor
ey and Methods 45(3): 553-565.
regression-cum exponential estimator for population
variance vector in two phase sampling. Journal of King
Saud University-Science 30: 223-228.
Asghar, A., Sanaullah, A., Abbasi, A.M. & Hanif, M.
2023. Advancing sampling techniques: Multivariate
ratio estimation for variance vector in two-phase
473-484.
Breidt, F.J. & Fuller, W.A. 1993. Regression weighting for
auxiliary information an almost unbiased multivariate ratio
Wiley & Sons.
Das, A.K. & Tripathi, T.P. 1978. Use of auxiliary information
in estimating the finite population variance. Sank
ya 40:
139-148.
information based improved EWMA median control charts.
Scientia Iranica 25(2): 954-982.
Journal of the American Statistical Association 78:
117-123.
of population variance with the use of supplementary
information in survey sampling. Mathematical Problems
in Engineering 2021: 9931217.
Muneer, S., Khalila, A., Shabbir, J. & Narjisb, G. 2018. A
new improved ratio-product type exponential estimator
of finite population variance using auxiliary information.
Journal of Statistical Computation and Simulation 88(16):
3179-3192.
Neyman, J. 1938. Contribution to the theory of sampling
Niaz, I., Sanaullah, A., Saleem, I. & Shabbir, J. 2022. An
improved efficient class of estimators for the population
variance. Concurren
cy and Computation: Practice and Experience 34(4):
e6620.
Rao, J. 1973. On double sampling for stratification and
analytical surveys. Biometrika 60: 125-133.
exponential estimators for population variance under
two-phase sampling. International Journal Applied and
Computational Mathematics 2: 75-84.
Sanaullah, A., Niaz, I., Shabbir, J. & Elsan, I. 2020. A class of
hybrid type estimators for variance of a finite population
in simple random sampling. Communications in Statistics
- Simulation and Computation 51(10): 5609-5619.
Shabbir, J. & Gupta, S. 2015. A note on generalized
exponential type estimator for population variance in
survey sampling. Revista Colombiana de Estadística 38(2):
385-397.
Shahzad, U., Ahmad, I., Almanjahie, I.M., Al-Noor, N.MH.
& Hanif, M. 2021a. A novel family of variance estimators
based on L-moments and calibration approach under
stratified random sampling. Communications in Statistics
- Simulation and Computation 52(8): 3782-3795.
Shahzad, U., Ahmad, I., Almanjahie, I.M., Koyuncu, N. &
Hanif, M. 2021b. Variance estimation based on L-moments
and auxiliary information. Mathematical Population Studies
29(1): 31-46.
using multi-auxiliary information for random non-response
auxiliary information for estimating finite population
Zaman, T. & Bulut, H. 2019. Modified regression estimators
using robust regression methods and covariance matrices in
stratified random sampling. Communications in Statistics
for estimating the population mean and variance. Hacettepe

*Corresponding author; email: chaamirsanaullah@yahoo.com
### DESCRIPTION OF THE REAL-LIFE DATA

#### TABLE A1. Details of variables for population

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>16.4344</td>
<td>11.7119</td>
<td>11.2634</td>
<td>5.0995</td>
<td>2.0234</td>
<td>0.5231</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>11.7119</td>
<td>10.0879</td>
<td>9.8238</td>
<td>1.3889</td>
<td>2.50897</td>
<td>0.7491</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>11.2634</td>
<td>9.8238</td>
<td>9.7610</td>
<td>1.3933</td>
<td>3.0050</td>
<td>0.9370</td>
</tr>
<tr>
<td>$X_1$</td>
<td>5.0996</td>
<td>1.3889</td>
<td>1.3933</td>
<td>7.4839</td>
<td>0.0406</td>
<td>-0.0538</td>
</tr>
<tr>
<td>$X_2$</td>
<td>2.0234</td>
<td>2.50897</td>
<td>3.0050</td>
<td>0.0406</td>
<td>4.6808</td>
<td>0.0080</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.5231</td>
<td>0.7491</td>
<td>0.9370</td>
<td>-0.0538</td>
<td>0.0080</td>
<td>2.0420</td>
</tr>
</tbody>
</table>

#### TABLE A2. Variance-covariance matrix

<table>
<thead>
<tr>
<th>Population</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>1.0000</td>
<td>0.8869</td>
<td>0.7654</td>
<td>0.9026</td>
<td>0.8528</td>
<td>0.9122</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>0.8869</td>
<td>1.0000</td>
<td>0.7535</td>
<td>0.8397</td>
<td>0.7789</td>
<td>0.8513</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>0.7654</td>
<td>0.7535</td>
<td>1.0000</td>
<td>0.8335</td>
<td>0.8118</td>
<td>0.8193</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.9026</td>
<td>0.8397</td>
<td>0.8335</td>
<td>1.0000</td>
<td>0.9421</td>
<td>-0.9453</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.8528</td>
<td>0.7789</td>
<td>0.8118</td>
<td>0.9421</td>
<td>1.0000</td>
<td>0.9208</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.9122</td>
<td>0.8513</td>
<td>0.8193</td>
<td>-0.9453</td>
<td>0.9208</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

#### TABLE A3. Correlation matrix

<table>
<thead>
<tr>
<th>Population</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$X_1$</th>
<th>$X_2$</th>
<th>$X_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y_1$</td>
<td>1.0000</td>
<td>0.8869</td>
<td>0.7654</td>
<td>0.9026</td>
<td>0.8528</td>
<td>0.9122</td>
</tr>
<tr>
<td>$Y_2$</td>
<td>0.8869</td>
<td>1.0000</td>
<td>0.7535</td>
<td>0.8397</td>
<td>0.7789</td>
<td>0.8513</td>
</tr>
<tr>
<td>$Y_3$</td>
<td>0.7654</td>
<td>0.7535</td>
<td>1.0000</td>
<td>0.8335</td>
<td>0.8118</td>
<td>0.8193</td>
</tr>
<tr>
<td>$X_1$</td>
<td>0.9026</td>
<td>0.8397</td>
<td>0.8335</td>
<td>1.0000</td>
<td>0.9421</td>
<td>-0.9453</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.8528</td>
<td>0.7789</td>
<td>0.8118</td>
<td>0.9421</td>
<td>1.0000</td>
<td>0.9208</td>
</tr>
<tr>
<td>$X_3$</td>
<td>0.9122</td>
<td>0.8513</td>
<td>0.8193</td>
<td>-0.9453</td>
<td>0.9208</td>
<td>1.0000</td>
</tr>
</tbody>
</table>