A New Class of L-Moments Based Calibration Variance Estimators

Usman Shahzad1,2, *, Ishfaq Ahmad1, Ibrahim Mufrah Almanjahie3,4, Nadia H. Al Noor5 and Muhammad Hanif2

1Department of Mathematics and Statistics, International Islamic University, Islamabad, 46000, Pakistan
2Department of Mathematics and Statistics, PMAS-Arid Agriculture University, Rawalpindi, 46300, Pakistan
3Department of Mathematics, King Khalid University, Abha, 62529, Saudi Arabia
4Statistical Research and Studies Support Unit, King Khalid University, Abha, 62529, Saudi Arabia
5Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, 10011, Iraq

*Corresponding Author: Usman Shahzad. Email: usman.stat@yahoo.com

Received: 31 August 2020; Accepted: 19 October 2020

Abstract: Variance is one of the most important measures of descriptive statistics and commonly used for statistical analysis. The traditional second-order central moment based variance estimation is a widely utilized methodology. However, traditional variance estimator is highly affected in the presence of extreme values. So this paper initially, proposes two classes of calibration estimators based on an adaptation of the estimators recently proposed by Koyuncu and then presents a new class of L-Moments based calibration variance estimators utilizing L-Moments characteristics (L-location, L-scale, L-CV) and auxiliary information. It is demonstrated that the proposed L-Moments based calibration variance estimators are more efficient than adapted ones. Artificial data is considered for assessing the performance of the proposed estimators. We also demonstrated an application related to apple fruit for purposes of the article. Using artificial and real data sets, percentage relative efficiency (PRE) of the proposed class of estimators with respect to adapted ones are calculated. The PRE results indicate to the superiority of the proposed class over adapted ones in the presence of extreme values. In this manner, the proposed class of estimators could be applied over an expansive range of survey sampling whenever auxiliary information is available in the presence of extreme values.

Keywords: L-moments; variance estimation; calibration approach; stratified random sampling

1 Introduction

All around, it was considered that the utilization of auxiliary knowledge (information) in the test review (sampling survey) configuration brings about effective estimators of population parameters. The literature on test review portrays an incredible assortment of strategies for utilizing auxiliary information. Ratio technique is the acceptable delineation in this specific situation,
see [1–5]. In some cases when some population parameters, as the mean, coefficient of variation, the standard deviation of the auxiliary variable have been known many authors including [6,7] imagined an enormous number of improved estimators for the population mean of the variate of interest. This urge researchers to utilize the characteristics of the auxiliary variable (mean, coefficient of variation, standard deviation, median and quartiles) in order to improve the population variance estimate of the study variable.

Variations are available in our life. For example, a doctor needs a full comprehension of varieties in the level of human circulatory strain, internal heat level, and heartbeat rate for satisfactory medicine. A maker needs steady information on the degree of varieties in individuals’ response to his item to have the option to realize whether to decrease or increment his cost or improve the nature of his item. An agriculturist requests a sufficient comprehension of the varieties in climatic factors particularly all around (or from time to time) to have the option to anticipate when, where, and how to plant his harvest. A lot of circumstances can be experienced where the population variance estimate of the study variable assumes a huge significance. Consequently, different authors, for example, [8,9] have given their consideration to the improved estimation of population variance of the study variable with the utilization of auxiliary information. However, all these quoted estimators based on traditional second-order central moment based variance estimation which is highly affected in presence of extreme values. So, in the current article, we have characterized a class of estimators for the population variance of the study variable based on L-Moments and calibration approach that is less effective to extreme values.

An experimental investigation utilizing apple production data sets was led, and we got good outcomes, numerically. The analysis of apple fruit production has great importance in food sciences. In view of the applications here, this study well provide a significant premise for future food sciences, medical researches, and many engineering applications.

The rest of the article is arranged in the following major sections. In Section 2, we present some classes of calibration based variance estimators according to the traditional second-order central moment. In Section 3, we provide L-Moments characteristics (L-location, L-scale, L-CV) and, propose a new class of calibration based variance estimators under stratified random sampling scheme. Results discussion is documented in Section 4. Finally, some conclusions are given in Section 5.

2 Adapted Classes of Estimators

Consider a finite population $\Omega_a = \{v_{a1}, v_{a2}, \ldots, v_{am}\}$ of size $N$. Suppose $(Y, X)$ be the study and auxiliary variables associated with $\Omega_a$. Let $\Omega_a$ is stratified into $L$ strata with the $h$th stratum containing $N_h$ units, where $h = 1, 2, \ldots, L$, such that $\sum_{h=1}^{L} N_h = N$. A simple random sample (srs) of size $n_h$ is drawn without replacement from the $h$th stratum such that $\sum_{h=1}^{L} n_h = n$. Further, $(x_{hi}, y_{hi})$ are the observed values of $Y$ and $X$ in $h$th stratum, where $i = 1, 2, \ldots, N_h$ and $h = 1, 2, \ldots, L$. In light of this stratified sampling design, taking motivation from [1], we present the first calibrated class of variance estimators as given below

$$T_{(SH_a)} = \sum_{h=1}^{L} \Phi^SH_{h} y^2_{yh}$$ (1)
where \( s^2_{yh} \) is the sample variance of the study variable in \( h \)th stratum. Further, \( \Phi^SH_h \) are calibrated weights that are selected in such a way that the sum of the chi-square type distance measure

\[
\sum_{h=1}^{L} \frac{(\Phi^SH_h - W_h)^2}{\psi_h W_h}
\]

is minimum, subject to the calibration constraints

\[
\sum_{h=1}^{L} \Phi^SH_h = 1
\]

\[
\sum_{h=1}^{L} \Phi^SH_h \hat{\Delta}_{xh} = \sum_{h=1}^{L} W_h \Delta_{xh}
\]

Note that \( W_h = \frac{N_h}{N} \) denoting the traditional stratum weight, \((\hat{\Delta}_{xh}, \Delta_{xh})\) are the sample and population characteristics of the auxiliary variable in \( h \)th stratum and, \( \Psi_h \) are appropriately selected weights that express the form of the estimator. Some suitable choices for \( \Psi_h \) are provided in Tab. 1.

Minimization of Eq. (2), subject to the calibration constraints set out in Eqs. (3) and (4), the optimum weights shall be calculated by

\[
\Phi^SH_h = W_h + \left[ \frac{\left( W_h \psi_h \hat{\Delta}_{xh} \right) \left( \sum_{h=1}^{L} W_h \psi_h \right) - \left( W_h \psi_h \right) \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} \right)^2}{\left( \sum_{h=1}^{L} W_h \psi_h \right) \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh}^2 \right) - \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} \right)^2} \right] \times \left[ \sum_{h=1}^{L} W_h \left( \Delta_{xh} - \hat{\Delta}_{xh} \right) \right]
\]

and thus

\[
T_{(SH_a)} = \sum_{h=1}^{L} W_h s^2_{yh} + \hat{\beta}(SH_a) \left[ \sum_{h=1}^{L} W_h \left( \Delta_{xh} - \hat{\Delta}_{xh} \right) \right]
\]

where

\[
\hat{\beta}(SH_a) = \frac{\left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} s^2_{yh} \right) \left( \sum_{h=1}^{L} W_h \psi_h \right) - \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} \right) \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} s^2_{yh} \right)}{\left( \sum_{h=1}^{L} W_h \psi_h \right) \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh}^2 \right) - \left( \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} \right)^2}
\]

The whole class \( T_{(SH_a)} \) is provided in Tab. 1.

Taking motivation from [10], we adapt the following calibration estimator of the population variance given by

\[
T_{(SH_b)} = \sum_{h=1}^{L} \Phi^P_h s^2_{yh}
\]
Table 1: First adapted family of estimators

| $T_{(SH_a)}$ | $\psi_h$ | $\hat{\Delta}_{xh}$ | $\Delta_{xh}$ |
|-------------|-----------|----------------------|---------------|
| $T_{(SH_{a1})}$ | $1$ | $\bar{x}_{h}$ | $\bar{x}_{h}$ |
| $T_{(SH_{a2})}$ | $\frac{1}{x_{h}}$ | $\bar{x}_{h}$ | $\bar{x}_{h}$ |
| $T_{(SH_{a3})}$ | $\frac{1}{s_{xh}}$ | $\bar{x}_{h}$ | $\bar{x}_{h}$ |
| $T_{(SH_{a4})}$ | $\frac{1}{s_{xh}}$ | $\bar{s}_{xh}$ | $S_{xh}^2$ |
| $T_{(SH_{a5})}$ | $\frac{1}{\bar{x}_{h}}$ | $\bar{s}_{xh}$ | $S_{xh}^2$ |
| $T_{(SH_{a6})}$ | $1$ | $s_{xh}^2$ | $S_{xh}^2$ |
| $T_{(SH_{a7})}$ | $\frac{1}{x_{h}}$ | $s_{xh}^2$ | $S_{xh}^2$ |
| $T_{(SH_{a8})}$ | $\frac{1}{s_{xh}}$ | $s_{xh}^2$ | $S_{xh}^2$ |
| $T_{(SH_{a9})}$ | $\frac{1}{\bar{x}_{h}}$ | $s_{xh}^2$ | $S_{xh}^2$ |
| $T_{(SH_{a10})}$ | $\frac{1}{\bar{x}_{h}}$ | $s_{xh}^2$ | $S_{xh}^2$ |

subject to the following constraints

$$\sum_{h=1}^{L} \Phi_h S_{xh}^2 = \sum_{h=1}^{L} W_h S_{xh}^2$$  \hspace{1cm} (9)

Minimization of Eq. (8), subject to the calibration constraints set out in Eqs. (9) and (10), the optimum weights shall be calculated by

$$\Phi_h = W_h + W_h \psi_h \hat{\Delta}_{xh} H_1 + W_h \psi_h s_{xh}^2 H_2$$

where

$$H_1 = \frac{\left[ \sum_{h=1}^{L} W_h (\Delta_{xh} - \hat{\Delta}_{xh}) \right] \left[ \sum_{h=1}^{L} W_h \psi_h s_{xh}^4 \right] - \left[ \sum_{h=1}^{L} W_h (S_{xh}^2 - s_{xh}^2) \right] \left[ \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} s_{xh}^2 \right]}{\left[ \sum_{h=1}^{L} W_h \psi_h s_{xh}^4 \right] \left[ \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh}^2 \right] - \left[ \sum_{h=1}^{L} W_h \psi_h \bar{s}_{xh} s_{xh}^2 \right]^2}$$

$$H_2 = \frac{\left[ \sum_{h=1}^{L} W_h (S_{xh}^2 - s_{xh}^2) \right] \left[ \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh}^2 \right] - \left[ \sum_{h=1}^{L} W_h (\Delta_{xh} - \hat{\Delta}_{xh}) \right] \left[ \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh} s_{xh}^2 \right]}{\left[ \sum_{h=1}^{L} W_h \psi_h s_{xh}^4 \right] \left[ \sum_{h=1}^{L} W_h \psi_h \hat{\Delta}_{xh}^2 \right] - \left[ \sum_{h=1}^{L} W_h \psi_h \bar{s}_{xh} s_{xh}^2 \right]^2}$$

and thus

$$T_{(SH_b)} = \sum_{h=1}^{L} W_h \bar{s}_{xh}^2 + \hat{b}_{1(SH_b)} \sum_{h=1}^{L} W_h (\Delta_{xh} - \hat{\Delta}_{xh}) + \hat{b}_{2(SH_b)} \sum_{h=1}^{L} W_h (S_{xh}^2 - s_{xh}^2)$$  \hspace{1cm} (11)
where

$$
\hat{b}_{1(SH_b)} = \left[ \frac{\sum_{h=1}^{L} W_h \Psi_h \Delta x_{bh}^2}{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4} + \frac{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4}{\sum_{h=1}^{L} W_h \Psi_h \Delta x_{bh}^2} - \frac{\sum_{h=1}^{L} W_h \Psi_h \Delta x_{bh}^2}{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4} \right]^{2}
$$

$$
\hat{b}_{2(SH_b)} = \left[ \frac{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4}{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4} + \frac{\sum_{h=1}^{L} W_h \Psi_h \Delta x_{bh}^2}{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4} - \frac{\sum_{h=1}^{L} W_h \Psi_h \Delta x_{bh}^2}{\sum_{h=1}^{L} W_h \Psi_h s_{bh}^4} \right]^{2}
$$

The whole class $T_{(SH_b)}$ is provided in Tab. 2.

| $T_{(SH_b)}$ | $\Psi_h$ | $\Delta x_h$ | $\Delta x_h$ |
|--------------|-----------|--------------|--------------|
| $T_{(SH_{b1})}$ | 1 | $\bar{x}_h$ | $\bar{x}_h$ |
| $T_{(SH_{b2})}$ | $\frac{1}{x_h}$ | $\bar{x}_h$ | $\bar{x}_h$ |
| $T_{(SH_{b3})}$ | $\frac{1}{s_{bh}}$ | $\bar{x}_h$ | $\bar{x}_h$ |
| $T_{(SH_{b4})}$ | $\frac{1}{x_{bh}}$ | $\bar{x}_h$ | $\bar{x}_h$ |
| $T_{(SH_{b5})}$ | $\frac{1}{C_{bh}}$ | $\bar{x}_h$ | $\bar{x}_h$ |
| $T_{(SH_{b6})}$ | 1 | $\bar{C}_{bh}$ | $C_{bh}$ |
| $T_{(SH_{b7})}$ | $\frac{1}{x_h}$ | $\bar{C}_{bh}$ | $C_{bh}$ |
| $T_{(SH_{b8})}$ | $\frac{1}{s_{bh}}$ | $\bar{C}_{bh}$ | $C_{bh}$ |
| $T_{(SH_{b9})}$ | $\frac{1}{x_{bh}}$ | $\bar{C}_{bh}$ | $C_{bh}$ |
| $T_{(SH_{b10})}$ | $\frac{1}{C_{bh}}$ | $\bar{C}_{bh}$ | $C_{bh}$ |

3 L-Moments and Propose Class of Estimators

The occurrence of extreme values, for example, in agriculture engineering, hydrology, food sciences. Meteorology and climatology, among others, observed nowadays in various parts of the world may influence adversely on human culture. Therefore, it is essential to appraise the accurate measurements of the numerical data in the presence of extreme values. As we have depicted before that variance is one of the significant measures of data description. Variance estimation dependent on traditional moments which is profoundly influenced by extreme values. Hence, a quantifiable procedure is necessitated that will catch the extreme values in the assessment, however less influenced by their quality. An elective strategy that has the vital ability to settle this issue is the L-Moments. Because L-Moments are more robust than traditional moments in the presence of extreme values [11].
The general form of the population L-Moments for the auxiliary variable with respect to \( h \)th stratum are as given below
\[
\lambda_{1xh} = E(X_{1:1}) \\
\lambda_{2xh} = (2)^{-1} E(X_{2:2} - X_{1:2}) \\
\lambda_{3xh} = (3)^{-1} E(X_{3:3} - 2X_{2:3} + X_{1:3}) \\
\lambda_{4xh} = (4)^{-1} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} + X_{1:4})
\]

Corresponding to these, we can write the sample based L-Moments such as \( \hat{\lambda}_{1xh}, \hat{\lambda}_{2xh}, \hat{\lambda}_{3xh}, \hat{\lambda}_{4xh} \), based on sample observations, respectively. Similarly, we can write the general form of the L-Moments for study variable by replacing \( x \) with \( y \). For more details regarding L-Moments, interested readers may refer to [12]. Some notations for upcoming proposed work in light of L-Moments with respect to \( h \)th stratum are
\[
\bar{X}_h = \lambda_{1xh}, \ \bar{X}_h = \hat{\lambda}_{1xh} \text{ are the population and sample means (L-location) of auxiliary variable based on L-Moments.}
\]
\[
\bar{Y}_h = \lambda_{1y}, \ \bar{Y}_h = \hat{\lambda}_{y} \text{ are the population and sample means (L-location) of study variable based on L-Moments.}
\]
\[
S_{xmh}^2 = \lambda_{2xh}, \ S_{xmh}^2 = \hat{\lambda}_{2xh} \text{ are the population and sample variance of auxiliary variable based on L-Moments.}
\]
\[
S_{ymh}^2 = \lambda_{2y}, \ S_{ymh}^2 = \hat{\lambda}_{2y} \text{ are the population and sample variance of study variable based on L-Moments.}
\]
\[
S_{xmh} = \lambda_{2xh}, \ S_{xmh} = \hat{\lambda}_{2xh} \text{ are the population and sample standard deviation (L-scale) of auxiliary variable based on L-Moments.}
\]
\[
S_{ymh} = \lambda_{2y}, \ S_{ymh} = \hat{\lambda}_{2y} \text{ are the population and sample standard deviation (L-scale) of study variable based on L-Moments.}
\]
\[
C_{xmh} = \frac{\lambda_{2xh}}{\lambda_{1xh}}, \ C_{xmh} = \frac{\hat{\lambda}_{2xh}}{\hat{\lambda}_{1xh}} \text{ are the population and sample coefficient of variation (L-CV) of the auxiliary variable based on L-Moments.}
\]
\[
C_{ymh} = \frac{\lambda_{2y}}{\lambda_{1y}}, \ C_{ymh} = \frac{\hat{\lambda}_{2y}}{\hat{\lambda}_{1y}} \text{ are the population and sample coefficient of variation (L-CV) of the study variable based on L-Moments.}
\]

The calibration estimation is a method of adjusting the original design weights to expand the accuracy of estimates by utilizing auxiliary information. Calibration based estimators utilize calibrated/adjusted weights that are planned to minimize a given measure of distance to the original weights while fulfilling a group of constraints with the auxiliary information. The authors in [13] are pioneer of calibration estimation. After that much of the work has been done in this filed such as [1,10,14]. These results motivated us to define the following calibrated estimator of variance as
\[
T(P_{ai}) = \sum_{h=1}^{L} \Phi_h P_{ai} \frac{s_{ymh}^2}{\hat{\lambda}_{1y}} \quad (12)
\]
where $s_{ym}^2$ is the sample variance of the study variable in the $h$th stratum. Further, $\Phi_{h}^{Pai}$ are calibrated weights that are selected in such a way that the sum of the chi-square type distance measure

$$\sum_{h=1}^{L} \left( \frac{\Phi_{h}^{Pai} - W_{h}}{\Psi_{h} W_{h}} \right)^2$$

is minimum, subject to the calibration constraints

$$\sum_{h=1}^{L} \Phi_{h}^{Pai} = 1$$

$$\sum_{h=1}^{L} \Phi_{h}^{Pai} \hat{\Delta}_{xmh} = \sum_{h=1}^{L} W_{h} \Delta_{xmh}$$

Note that $W_{h} = \frac{N_{h}}{N}$ denotes the traditional stratum weight, $(\hat{\Delta}_{xmh}, \Delta_{xmh})$ are the sample and population characteristics of the auxiliary variable in the $h$th stratum and, $\Psi_{h}$ are appropriately selected weights that express the form of the estimator. In most situations, $\Psi_{h} = 1$. Minimization of Eq. (13), subject to the calibration constraints set out in Eqs. (14) and (15), the optimum weights shall be calculated by

$$\Phi_{h}^{Pai} = W_{h} + \left[ \left( \frac{W_{h} \Psi_{h} \hat{\Delta}_{xmh}}{\sum_{h=1}^{L} W_{h} \Psi_{h}} \right) - \left( \frac{W_{h} \Psi_{h} \hat{\Delta}_{xmh}^2}{\sum_{h=1}^{L} W_{h} \Psi_{h} \hat{\Delta}_{xmh}} \right) \right] \left( \sum_{h=1}^{L} W_{h} \Delta_{xmh} - \hat{\Delta}_{xmh} \right)$$

and thus

$$T_{(Pai)} = \sum_{h=1}^{L} W_{h} s_{ymh}^2 + \hat{\beta}_{(Pai)} \left[ \sum_{h=1}^{L} W_{h} \left( \Delta_{xmh} - \hat{\Delta}_{xmh} \right) \right]$$

where

$$\hat{\beta}_{(Pai)} = \frac{\left( \sum_{h=1}^{L} W_{h} \Psi_{h} \hat{\Delta}_{xmh} s_{ymh}^2 \right) \left( \sum_{h=1}^{L} W_{h} \Psi_{h} \right) - \left( \sum_{h=1}^{L} W_{h} \Psi_{h} s_{ymh}^2 \right) \left( \sum_{h=1}^{L} W_{h} \Psi_{h} \hat{\Delta}_{xmh} \right)}{\left( \sum_{h=1}^{L} W_{h} \Psi_{h} \right) \left( \sum_{h=1}^{L} W_{h} \Psi_{h} \hat{\Delta}_{xmh}^2 \right) - \left( \sum_{h=1}^{L} W_{h} \Psi_{h} \hat{\Delta}_{xmh} \right)^2}$$

The whole class $T_{(Pai)}$ is provided in Tab. 3.
Table 3: Propose class of estimators

| $T(P_{al})$ | $\psi_h$ | $\hat{\Delta}_{xmh}$ | $\Delta_{xmh}$ |
|-----------|---------|-----------------|-------------|
| $T(P_{al})$ | 1       | $\bar{x}_h$     | $\bar{X}_h$ |
| $T(P_{a2})$ | $\frac{1}{\bar{x}_h}$ | $\bar{x}_h$     | $\bar{X}_h$ |
| $T(P_{a3})$ | $\frac{1}{s_{xmh}}$ | $\bar{x}_h$     | $\bar{X}_h$ |
| $T(P_{a4})$ | $\frac{1}{S_{xmh}}$ | $\bar{x}_h$     | $\bar{X}_h$ |
| $T(P_{a5})$ | $\frac{1}{c_{xmh}}$ | $\bar{x}_h$     | $\bar{X}_h$ |
| $T(P_{a6})$ | 1       | $s_{xmh}^2$     | $S_{xmh}^2$ |
| $T(P_{a7})$ | $\frac{1}{s_{xmh}}$ | $s_{xmh}^2$     | $S_{xmh}^2$ |
| $T(P_{a8})$ | $\frac{1}{s_{xmh}}$ | $s_{xmh}^2$     | $S_{xmh}^2$ |
| $T(P_{a9})$ | $\frac{1}{s_{xmh}}$ | $s_{xmh}^2$     | $S_{xmh}^2$ |
| $T(P_{a10})$ | $\frac{1}{c_{xmh}}$ | $s_{xmh}^2$     | $S_{xmh}^2$ |

4 Results and Discussion

In this section, the performance of the suggested estimators is investigated through a simulation study.

4.1 Simulation Design

The simulation design is organized as follows: A random variable $X_h$ and random variable $Y_h$ are defined as follows:

$$Y_h = v + RX_h + \epsilon X_h^p,$$  \hspace{1cm} for $h$th stratum  \hspace{1cm} (19)

Figure 1: Pop-1, $h = 1$
where

\[
X_1 \sim \text{Gamma}(2.6, 3.8), \quad \text{for } h = 1 \\
X_2 \sim \text{Gamma}(2.0, 3.1), \quad \text{for } h = 2 \\
X_3 \sim \text{Gamma}(1.5, 2.7), \quad \text{for } h = 3 \\
X_4 \sim \text{Gamma}(2.9, 3.1), \quad \text{for } h = 4 
\]

Further, we assume that \( p = 1.6, v = 5, R = 2 \) and \( \varepsilon \) has a standard normal distribution in Eq. (19). We consider the population of size \( N = 1000 \) for \( h \)th stratum. The overall sample of size \( n_h = 400 \) is selected for purposes of the article. Using equal allocation of a sample of size 100 is
selected from hth stratum. A scatter plot for each stratum is provided (see Figs. 1–4). These figures clearly show the presence of extreme values and hence suitable for our proposed estimators.

![Figure 4: Pop-1, h = 4](image)

**Table 4:** PRE of proposed family with respect to $T_{(SH_a)}$ for artificial data

| $\phi$ | $T_{(SH_{a1})}$ | $T_{(SH_{a2})}$ | $T_{(SH_{a3})}$ | $T_{(SH_{a4})}$ | $T_{(SH_{a5})}$ |
|-------|----------------|----------------|----------------|----------------|----------------|
| $T_{(P_{a1})}$ | 4029.244 | 4023.545 | 4053.027 | 4040.295 | 32026.127 |
| $T_{(P_{a2})}$ | 3908.058 | 3902.530 | 3931.126 | 3918.776 | 31062.389 |
| $T_{(P_{a3})}$ | 3947.085 | 3941.502 | 3970.384 | 3957.911 | 31373.455 |
| $T_{(P_{a4})}$ | 3861.196 | 3855.735 | 3883.988 | 3871.787 | 30690.034 |
| $T_{(P_{a5})}$ | 4045.904 | 4040.181 | 4069.786 | 4057.001 | 32158.017 |
| $T_{(P_{a6})}$ | 2715.207 | 2711.366 | 2731.234 | 2722.654 | 21581.018 |
| $T_{(P_{a7})}$ | 2679.445 | 2675.656 | 2695.261 | 2686.794 | 21297.875 |
| $T_{(P_{a8})}$ | 2704.947 | 2701.122 | 2720.914 | 2712.366 | 21500.474 |
| $T_{(P_{a9})}$ | 2678.691 | 2674.902 | 2694.502 | 2686.038 | 21291.040 |
| $T_{(P_{a10})}$ | 2704.851 | 2701.025 | 2720.817 | 2712.270 | 21499.319 |

| $\phi$ | $T_{(SH_{b1})}$ | $T_{(SH_{b2})}$ | $T_{(SH_{b3})}$ | $T_{(SH_{b4})}$ | $T_{(SH_{b5})}$ |
|-------|----------------|----------------|----------------|----------------|----------------|
| $T_{(P_{b1})}$ | 6016.945 | 6035.710 | 5932.726 | 5970.067 | 5953.911 |
| $T_{(P_{b2})}$ | 5835.975 | 5854.176 | 5754.290 | 5790.507 | 5774.838 |
| $T_{(P_{b3})}$ | 5894.255 | 5912.638 | 5811.754 | 5848.333 | 5832.507 |
| $T_{(P_{b4})}$ | 5765.996 | 5783.979 | 5685.291 | 5721.073 | 5705.592 |
| $T_{(P_{b5})}$ | 6041.823 | 6060.666 | 5957.257 | 5994.752 | 5978.530 |
| $T_{(P_{b6})}$ | 4054.668 | 4067.314 | 3997.916 | 4023.079 | 4012.192 |
| $T_{(P_{b7})}$ | 4001.265 | 4013.744 | 3945.260 | 3970.092 | 3959.348 |
| $T_{(P_{b8})}$ | 4039.348 | 4051.946 | 3982.810 | 4007.878 | 3997.032 |
| $T_{(P_{b9})}$ | 4000.139 | 4012.614 | 3944.150 | 3968.974 | 3958.234 |
| $T_{(P_{b10})}$ | 4039.204 | 4051.801 | 3982.668 | 4007.735 | 3996.889 |
The steps of simulation study can be carried out as:

Step-1: A sample of size $n_h$ is selected through SRSWOR from the $h$th stratum.

Step-2: Using Step-1, we find the value of variance estimate (say) $\hat{\phi} = \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}$, where

\[
\hat{\alpha}_1 = T_{(S_{H_1})}, T_{(S_{H_2})}, T_{(S_{H_3})}, T_{(S_{H_4})}, T_{(S_{H_5})}, T_{(S_{H_6})}, T_{(S_{H_7})}, T_{(S_{H_8})}, T_{(S_{H_9})}, T_{(S_{H_{10}})},
\]

\[
\hat{\alpha}_2 = T_{(S_{H_{12}})}, T_{(S_{H_{13}})}, T_{(S_{H_{14}})}, T_{(S_{H_{15}})}, T_{(S_{H_{16}})}, T_{(S_{H_{17}})}, T_{(S_{H_{18}})}, T_{(S_{H_{19}})}, T_{(S_{H_{20}})},
\]

\[
\hat{\beta} = T_{(P_{a_1})}, T_{(P_{a_2})}, T_{(P_{a_3})}, T_{(P_{a_4})}, T_{(P_{a_5})}, T_{(P_{a_6})}, T_{(P_{a_7})}, T_{(P_{a_8})}, T_{(P_{a_9})}, T_{(P_{a_{10}})}.
\]

Step-3: The Step-1 and Step-2 are replicated with $K = 5000$ times. Thus, $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_k$ are obtained.

Step-4: The mean square error (MSE) of the estimators is computed as $MSE(\hat{\phi}) = \frac{1}{K}\sum_{i=1}^K (\hat{\phi}_i - \bar{\phi})^2$.

Step-5: The PRE is computed as $PRE(\hat{\phi}) = \frac{MSE(\hat{\alpha}_1, \hat{\alpha}_2)}{MSE(\hat{\beta})} \times 100$ and the results are provided in Tabs. 4 and 5.

**Table 5: PRE of proposed family with respect to $T_{(S_{H_i})}$ for artificial data**

| $\phi$  | $T_{(S_{H_{12}})}$ | $T_{(S_{H_{13}})}$ | $T_{(S_{H_{14}})}$ | $T_{(S_{H_{15}})}$ | $T_{(S_{H_{16}})}$ |
|---------|---------------------|---------------------|---------------------|---------------------|---------------------|
| $T_{(P_{a_1})}$ | 3904.298           | 3867.100            | 3861.923            | 3848.535            | 3964.176            |
| $T_{(P_{a_2})}$ | 3786.870           | 3750.791            | 3745.770            | 3732.784            | 3844.947            |
| $T_{(P_{a_3})}$ | 3824.687           | 3788.248            | 3783.176            | 3770.061            | 3883.344            |
| $T_{(P_{a_4})}$ | 3741.462           | 3705.815            | 3700.855            | 3688.025            | 3798.842            |
| $T_{(P_{a_5})}$ | 3920.442           | 3883.090            | 3877.892            | 3864.448            | 3980.567            |
| $T_{(P_{a_6})}$ | 2631.009           | 2605.942            | 2602.454            | 2593.432            | 2671.359            |
| $T_{(P_{a_7})}$ | 2596.357           | 2571.620            | 2568.177            | 2559.274            | 2636.175            |
| $T_{(P_{a_8})}$ | 2621.068           | 2596.096            | 2592.620            | 2583.633            | 2661.266            |
| $T_{(P_{a_9})}$ | 2595.626           | 2570.896            | 2567.454            | 2558.554            | 2635.433            |
| $T_{(P_{a_{10}})}$ | 2620.974          | 2596.003            | 2592.528            | 2583.540            | 2661.171            |

| $\phi$  | $T_{(S_{H_{19}})}$ | $T_{(S_{H_{20}})}$ | $T_{(S_{H_{21}})}$ | $T_{(S_{H_{22}})}$ | $T_{(S_{H_{23}})}$ |
|---------|---------------------|---------------------|---------------------|---------------------|---------------------|
| $T_{(P_{a_1})}$ | 2267.264           | 2312.984            | 2282.897            | 2314.902            | 2265.450            |
| $T_{(P_{a_2})}$ | 2199.072           | 2243.417            | 2214.235            | 2245.278            | 2197.313            |
| $T_{(P_{a_3})}$ | 2221.033           | 2265.820            | 2236.347            | 2267.700            | 2219.256            |
| $T_{(P_{a_4})}$ | 2172.703           | 2216.516            | 2187.684            | 2218.355            | 2170.965            |
| $T_{(P_{a_5})}$ | 2276.638           | 2322.547            | 2292.336            | 2324.474            | 2274.817            |
| $T_{(P_{a_6})}$ | 1527.852           | 1558.662            | 1538.387            | 1559.955            | 1526.630            |
| $T_{(P_{a_7})}$ | 1507.729           | 1538.133            | 1518.125            | 1539.409            | 1506.523            |
| $T_{(P_{a_8})}$ | 1522.079           | 1552.773            | 1532.574            | 1554.061            | 1520.862            |
| $T_{(P_{a_9})}$ | 1507.305           | 1537.700            | 1517.698            | 1538.976            | 1506.099            |
| $T_{(P_{a_{10}})}$ | 1522.025          | 1552.717            | 1532.520            | 1554.005            | 1520.808            |
4.2 Real Life Data

To illustrate the behavior of the proposed estimators in this article, we consider a data set of apple fruit, used by [15], where

\[ X = \text{number of apple trees (1 unit = 100 trees)}, \]
\[ Y = \text{level of apple production (1 unit = 100 tonnes)}. \]

It is worth mentioning that we consider 477 villages in 4 strata, respectively (like 1: Marmarian, 2: Agean, 3: Mediterranean, and 4: Central Anatolia) in 1999. We draw a scatter plot for each stratum (see Figs. 5–8). The scatter plot of each stratum representing extreme values. Some important characteristics of the data are provided in Tab. 6. The PRE of the estimators is calculated using defined five steps and available in Tabs. 7 and 8.

![Figure 5: Pop-2, h = 1](image)

![Figure 6: Pop-2, h = 2](image)
However, this investigation can assist with revealing insight into the value of the proposed class. Actually, it gives a reasonable sign that more mind-boggling options than the traditional moments based variance estimators can be helpful when the data is contaminated with extreme values. Because each PRE result of the proposed estimator is greater than 100. Which means that the proposed class beats the adapted ones thoroughly. Based on real and artificial numerical illustration, we feel confident that the same could hold in other contexts of food sciences and agricultural engineering.
### Table 6: Characteristics of apple data

| Stratum-I | Stratum-II | Stratum-III | Stratum-IV |
|-----------|------------|-------------|------------|
| \( N_1 = 106 \) | \( N_2 = 106 \) | \( N_3 = 94 \) | \( N_4 = 171 \) |
| \( \bar{X}_1 = 24375.59 \) | \( \bar{X}_2 = 27421.7 \) | \( \bar{X}_3 = 72409.95 \) | \( \bar{X}_4 = 74364.68 \) |
| \( \bar{Y}_1 = 1536.774 \) | \( \bar{Y}_2 = 2212.594 \) | \( \bar{Y}_3 = 9384.309 \) | \( \bar{Y}_4 = 5588.012 \) |
| \( S_{x1} = 49189.08 \) | \( S_{x2} = 57460.61 \) | \( S_{x3} = 160757.3 \) | \( S_{x4} = 285603.1 \) |
| \( S_{y1} = 6425.087 \) | \( S_{y2} = 11551.53 \) | \( S_{y3} = 29907.48 \) | \( S_{y4} = 28643.42 \) |
| \( C_{x1} = 2.017964 \) | \( C_{x2} = 2.095443 \) | \( C_{x3} = 2.2201 \) | \( C_{x4} = 3.840575 \) |
| \( C_{y1} = 4.180894 \) | \( C_{y2} = 5.220807 \) | \( C_{y3} = 3.186967 \) | \( C_{y4} = 5.12587 \) |
| \( n_1 = 29 \) | \( n_2 = 29 \) | \( n_3 = 26 \) | \( n_4 = 47 \) |
| \( S_{xm1} = 16920.78 \) | \( S_{xm2} = 20100.85 \) | \( S_{xm3} = 54805.62 \) | \( S_{xm4} = 60429.63 \) |
| \( S_{ym1} = 1289.806 \) | \( S_{ym2} = 1923.234 \) | \( S_{ym3} = 8149.858 \) | \( S_{ym4} = 5003.631 \) |
| \( C_{xm1} = 0.694196 \) | \( C_{xm2} = 0.733027 \) | \( C_{xm3} = 0.7568798 \) | \( C_{xm4} = 0 : 812612 \) |
| \( C_{ym1} = 0.8392945 \) | \( C_{ym2} = 0.8692213 \) | \( C_{ym3} = 0.8684559 \) | \( C_{ym4} = 0.8954225 \) |

### Table 7: PRE of proposed family with respect to \( T_{(SH_a)} \) for real life data

| \( \phi \) | \( T_{(SH_{a1})} \) | \( T_{(SH_{a2})} \) | \( T_{(SH_{a3})} \) | \( T_{(SH_{a4})} \) | \( T_{(SH_{a5})} \) |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( T_{(P_{a1})} \) | 60745.99 | 58011.08 | 46175.25 | 52506.23 | 344276122.52 |
| \( T_{(P_{a2})} \) | 60323.30 | 57607.42 | 45853.95 | 52140.87 | 341880544.34 |
| \( T_{(P_{a3})} \) | 59891.91 | 57195.45 | 45526.03 | 51768.00 | 339435644.33 |
| \( T_{(P_{a4})} \) | 47586.12 | 45443.70 | 36171.95 | 41131.40 | 269692941.07 |
| \( T_{(P_{a5})} \) | 61314.87 | 58554.36 | 46607.68 | 52997.95 | 347500261.41 |
| \( T_{(P_{a6})} \) | 20766.74 | 19831.78 | 15785.56 | 17949.88 | 117694922.83 |
| \( T_{(P_{a7})} \) | 20393.02 | 19474.88 | 15501.48 | 17626.85 | 115576839.25 |
| \( T_{(P_{a8})} \) | 20312.28 | 19397.87 | 15440.11 | 17557.09 | 115119246.35 |
| \( T_{(P_{a9})} \) | 16618.67 | 15870.46 | 12632.46 | 14364.46 | 94185815.80 |
| \( T_{(P_{a10})} \) | 20947.46 | 20004.37 | 15922.94 | 18106.09 | 118791415.99 |

| \( \phi \) | \( T_{(SH_{a6})} \) | \( T_{(SH_{a7})} \) | \( T_{(SH_{a8})} \) | \( T_{(SH_{a9})} \) | \( T_{(SH_{a10})} \) |
|----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| \( T_{(P_{a1})} \) | 696700.06 | 677047.85 | 606140.97 | 640044.69 | 655240.63 |
| \( T_{(P_{a2})} \) | 691852.21 | 672336.75 | 601923.25 | 635591.07 | 650681.26 |
| \( T_{(P_{a3})} \) | 686904.55 | 667528.65 | 597618.70 | 631045.75 | 646028.03 |
| \( T_{(P_{a4})} \) | 545768.58 | 530373.77 | 474827.99 | 501386.89 | 513290.82 |
| \( T_{(P_{a5})} \) | 703224.65 | 683388.39 | 611817.47 | 646038.70 | 661376.94 |
| \( T_{(P_{a6})} \) | 238175.28 | 231456.93 | 207216.56 | 218806.96 | 224001.87 |
| \( T_{(P_{a7})} \) | 233888.98 | 227291.54 | 203487.41 | 214869.22 | 219970.65 |
| \( T_{(P_{a8})} \) | 232962.97 | 226391.65 | 202681.76 | 214018.51 | 219099.74 |
| \( T_{(P_{a9})} \) | 190600.68 | 185224.30 | 165825.85 | 175101.11 | 179258.36 |
| \( T_{(P_{a10})} \) | 240247.96 | 233471.15 | 209019.83 | 220711.09 | 225951.21 |
Table 8: PRE of proposed family with respect to $T_{(SH_b)}$ for real life data

| $\phi$ | $T_{(SH_b1)}$ | $T_{(SH_b2)}$ | $T_{(SH_b3)}$ | $T_{(SH_b4)}$ | $T_{(SH_b5)}$ |
|--------|---------------|---------------|---------------|---------------|---------------|
| $T_{(P_{a1})}$ | 221326.6      | 244744.1      | 247714.1      | 371784.6      | 220241.8      |
| $T_{(P_{a2})}$ | 219786.6      | 243041.1      | 245990.4      | 369197.6      | 218709.3      |
| $T_{(P_{a3})}$ | 218214.8      | 241303.0      | 244231.3      | 366557.3      | 217145.3      |
| $T_{(P_{a4})}$ | 173378.9      | 191723.3      | 194049.9      | 291242.0      | 172529.2      |
| $T_{(P_{a5})}$ | 223399.4      | 247036.1      | 250033.9      | 375266.3      | 222304.4      |
| $T_{(P_{a6})}$ | 75663.17      | 83668.69      | 84684.04      | 127099.02     | 75292.32      |
| $T_{(P_{a7})}$ | 74301.50      | 82162.96      | 83160.03      | 124811.70     | 73937.33      |
| $T_{(P_{a8})}$ | 74007.33      | 81837.66      | 82830.78      | 124317.54     | 73644.59      |
| $T_{(P_{a9})}$ | 60549.74      | 66956.20      | 67768.73      | 101711.48     | 60252.97      |
| $T_{(P_{a10})}$ | 76321.61      | 84396.81      | 85420.99      | 128205.08     | 75947.54      |

| $\phi$ | $T_{(SH_b6)}$ | $T_{(SH_b7)}$ | $T_{(SH_b8)}$ | $T_{(SH_b9)}$ | $T_{(SH_b10)}$ |
|--------|---------------|---------------|---------------|---------------|---------------|
| $T_{(P_{a1})}$ | 191172.6      | 214522.1      | 215987.1      | 280450.4      | 190268.8      |
| $T_{(P_{a2})}$ | 189842.4      | 213029.4      | 214484.2      | 278498.9      | 188944.8      |
| $T_{(P_{a3})}$ | 188484.8      | 211505.9      | 212950.4      | 276507.3      | 187593.6      |
| $T_{(P_{a4})}$ | 149757.4      | 168048.5      | 169196.2      | 219694.3      | 149049.4      |
| $T_{(P_{a5})}$ | 192963.0      | 216531.1      | 218009.8      | 283076.8      | 192050.6      |
| $T_{(P_{a6})}$ | 65354.66      | 73363.95      | 73837.79      | 95875.33      | 65045.66      |
| $T_{(P_{a7})}$ | 64178.51      | 72017.15      | 72508.97      | 94149.92      | 63875.08      |
| $T_{(P_{a8})}$ | 63924.42      | 71732.02      | 72221.89      | 93777.16      | 63622.18      |
| $T_{(P_{a9})}$ | 52300.32      | 58688.18      | 59088.97      | 76724.60      | 52053.04      |
| $T_{(P_{a10})}$ | 65923.40      | 73975.16      | 74480.35      | 96709.67      | 65611.71      |

5 Conclusions

Usually, in test reviews, it uses traditional moments-based techniques to obtain improved designs and more accurate estimators. However, in the presence of extreme values, traditional moments-based techniques may provide misleading results. In this paper, a calibration approach is utilized and developed two adapted families of estimators. Then a new class of L-Moments based variance estimators has been proposed which provides a better estimate of variance in the presence of extreme values. The performance of adapted and proposed estimators has been assessed through artificial and apple fruit data set. Based on satisfactory results, we recommend survey practitioners utilize the proposed class of variance estimators in various environmental researches.

Acknowledgement: The authors are thankful to the reviewers for their valuable suggestions.

Funding Statement: The authors are grateful to the Deanship of Scientific Research at King Khalid University, Kingdom of Saudi Arabia for funding this study through the research groups program under project number R.G.P.2/67/41. Ibrahim Mufrah Almanjahie received the grant.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.
References

[1] N. Koyuncu, “Calibration estimator of population mean under stratified ranked set sampling design,” *Communications in Statistics-Theory and Methods*, vol. 47, no. 23, pp. 5845–5853, 2018.

[2] U. Shahzad, P. F. Perri and M. Hanif, “A new class of ratio-type estimators for improving mean estimation of nonsensitive and sensitive variables by using supplementary information,” *Communications in Statistics-Simulation and Computation*, vol. 48, no. 9, pp. 2566–2585, 2019.

[3] T. Zaman and H. Bulut, “Modified ratio estimators using robust regression methods,” *Communications in Statistics-Theory and Methods*, vol. 48, no. 8, pp. 2039–2048, 2019.

[4] U. Shahzad, N. H. Al-Noor, M. Hanif, I. Sajjad and A. M. Muhammad, “Imputation based mean estimators in case of missing data utilizing robust regression and variance-covariance matrices,” *Communications in Statistics-Simulation and Computation*, vol. 3, no. 2, pp. 1–20, 2020.

[5] M. Subzar, A. I. Al-Omari and A. R. A. Alanzi, “The robust regression methods for estimating of finite population mean based on SRSWOR in case of outliers,” *Computers, Materials & Continua*, vol. 65, no. 1, pp. 125–138, 2020.

[6] T. Zaman and E. Toksoy, “Improvement in estimating the population mean in simple random sampling using information on two auxiliary attributes and numerical application in agricultural engineering,” *Fresenius Environmental Bulletin*, vol. 28, no. 6, pp. 4584–4590, 2019.

[7] T. Zaman, “Improvement of modified ratio estimators using robust regression methods,” *Applied Mathematics and Computation*, vol. 348, pp. 627–631, 2019.

[8] M. Abid, S. Ahmed, M. Tahir, N. Zafar and M. Riaz, “Improved ratio estimators of variance based on robust measures,” *Scientia Iranica*, vol. 26, no. 4, pp. 2484–2494, 2019.

[9] F. Naz, M. Abid, T. Nawaz and T. Pang, “Enhancing the efficiency of the ratio-type estimators of population variance with a blend of information on robust location measures,” *Scientia Iranica*, 2019.

[10] D. S. Tracy, S. Singh and R. Arnab, “Note on calibration in stratified and double sampling,” *Survey Methodology*, vol. 29, no. 1, pp. 99–104, 2003.

[11] J. R. Hosking, “L-moments: Analysis and estimation of distributions using linear combinations of order statistics,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 52, no. 1, pp. 105–124, 1990.

[12] J. R. M. Hosking and J. R. Wallis, “Regional Frequency Analysis: An Approach Based on L-Moments,” New York: Cambridge University Press, 2005.

[13] J. C. Deville and C. E. Srndal, “Calibration estimators in survey sampling,” *Journal of the American Statistical Association*, vol. 87, no. 418, pp. 376–382, 1992.

[14] E. P. Clement, “Calibration approach separate ratio estimator for population mean in stratified sampling,” *International Journal of Modern Mathematical Sciences*, vol. 13, no. 4, pp. 377–384, 2015.

[15] N. Koyuncu and C. Kadilar, “Ratio and product estimators in stratified random sampling,” *Journal of Statistical Planning and Inference*, vol. 139, no. 8, pp. 2552–2558, 2009.