Tech Science Press

L-Moments Based Calibrated Variance Estimators Using Double Stratified Sampling

Usman Shahzad[1,2,*], Ishfaq Ahmad[1], Ibrahim Mufrah Almanjahie[3,4] and Nadia H.Al –Noor[5]

[1]Department of Mathematics and Statistics, International Islamic University, Islamabad, 44000, Pakistan
[2]Department of Mathematics and Statistics, PMAS-Arid Agriculture University, Rawalpindi, 46300, Pakistan
[3]Department of Mathematics, College of Science, King Khalid University, Abha, 62529, Saudi Arabia
[4]Statistical Research and Studies Support Unit, King Khalid University, Abha, 62529, Saudi Arabia
[5]Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, 10011, Iraq
[*]Corresponding Author: Usman Shahzad. Email: usman.stat@yahoo.com
Received: 19 January 2021; Accepted: 04 March 2021

Abstract: Variance is one of the most vital measures of dispersion widely employed in practical aspects. A commonly used approach for variance estimation is the traditional method of moments that is strongly influenced by the presence of extreme values, and thus its results cannot be relied on. Finding momentum from Koyuncu's recent work, the present paper focuses first on proposing two classes of variance estimators based on linear moments (L-moments), and then employing them with auxiliary data under double stratified sampling to introduce a new class of calibration variance estimators using important properties of L-moments (L-location, L-cv, L-variance). Three populations are taken into account to assess the efficiency of the new estimators. The first and second populations are concerned with artificial data, and the third populations is concerned with real data. The percentage relative efficiency of the proposed estimators over existing ones is evaluated. In the presence of extreme values, our findings depict the superiority and high efficiency of the proposed classes over traditional classes. Hence, when auxiliary data is available along with extreme values, the proposed classes of estimators may be implemented in an extensive variety of sampling surveys.

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

1 Introduction

Planning is an integral part of the administrative process for the development of any field. Among the most important outputs of the planning process are the plans and programs that institutions seek to execute. One of the most important pillars of planning success is the availability of data and information that enables the decision-maker to conduct scientific analysis. In statistical literature, the additional information attached to each element is referred to as auxiliary (or ancillary, supplementary, supporting, concomitant) information. Whatever type of information is offered, it can be used to identify better sampling strategies. Auxiliary information has been

used with sampling techniques for many years. The authors of [1,2] were pioneers in the usage of auxiliary information regarding the development of estimation techniques with high estimation accuracy. Recently, there have been many interesting works using auxiliary information in different ways [3–12].

In all sample surveys, the major concern is the derivation of point estimators for various parameters of interest. Nevertheless, it is equally important to evaluate the performance of these estimators. The importance of variance estimators lies primarily in the fact that the estimated variance, of any estimator, is a major component of its quality. Reference [13] pointed out that the importance of variance estimation lies in the fact that it offers an indicator of the quality of estimators. It can be used in calculating confidence intervals, and drawing accurate conclusions, and can provide indicators of data quality. The sampling design that underlies a sample survey is one of the most important factors determining both the size of sample and the procedure needed to estimate the variances. More specifically, there are many components of sample designs related to the estimation of variances, including the number of sampling stages. In single or one-stage sample designs, the stage is very direct, and the closed formula can be derived for estimation of variance. In designs with more than one stage, the state becomes complicated since there is more than one source of variance. At each stage, unit sampling (primary, secondary, etc.) leads to an additional component of variance. In cases where all other components of sampling and estimation are rather simple, a closed formula can be obtained by calculating the variance at each stage. However, common practice is to roughly estimate the variance by estimating the variation among the initial sampling units, since this is the dominant component of the overall variance. For example, with double or two-stage sampling, there are two sources of variance such as variation resulting from the selection of the primary sampling units and the variation resulting from the selection of the secondary sampling units (for more details, see [14]). There are also many studies that have employed double sampling for real data [15–18]. In this paper, we consider double stratified random sampling. With stratified sampling, the population is split into subpopulations that are not overlapping; these are known as strata and typically describe homogeneous subpopulations, resulting in reduced overall variability. A random sample is chosen from each stratum, independently of the other stratum. A stratified sampling pattern may be the same or different from that of other stratum.

Consider X and Y as the auxiliary and study variables associated with a finite population of size N, and $\Omega = \{v_1, v_2, \ldots, v_n\}$, where Ω is stratified into R strata with the hth stratum including N_h units. $h = 1, 2, \ldots, R$, and $\sum_{h=1}^{R} N_h = N$. For the first stage, a simple random sample with size n_h^* is chosen from the stratum h without replacement such as $\sum_{h=1}^{R} n_h^* = n^*$. Then the sample n_h $\left(n_h < n_h^*\right)$ for the second stage is selected. $h = 1, 2, \ldots, R$, (x_{hi}, y_{hi}) represents the observed values of X and Y with $i = 1, 2, \ldots, N_h$, and (s_{xh}^{*2}, s_{xh}^2) and (s_{yh}^{*2}, s_{yh}^2) represent the variances of X and Y for the first and second stage samples, respectively. In view of this double stratified sampling design, the traditional variance estimator is

$$T_o = \sum_{h=1}^{R} W_h s_{yh}^2 \tag{1}$$

It is worth noting that s_{yh}^2 is based on traditional moments and hence is highly affected by the presence of extreme values. Note also that $W_h = \dfrac{N_h}{N}$ is the stratum's weight.

The analysis of sample data is complex. The complexity of the analysis increases when the data contains unusual points (outliers or extreme values) that affect the robustness of the variance estimation under traditional central moments. One of the solutions to tackle this issue is to use L-moments instead of traditional central moments. L-moments provide a robust statistical framework for the analysis. L-moments [19] are determined by linear combinations of the expected values of the order statistics (O.S.). Furthermore, calibration estimation is another common statistical approach that relies on the use of auxiliary information to adjust the original weights of the design and improve the accuracy of estimators. The authors of [20] were pioneers in the use of calibration estimation with survey data and several additional works on mean estimation have been published since (for example, see [21–23]).

In the present paper, our objective is to develop some new classes of variance estimators for a variable of interest, based on L-moments and the calibration approach under double stratified random sampling. The remainder of this article is organized as follows. In Section 2, the L-moments and proposed classes are presented in detail. Numerical illustrations of three populations are offered in Section 3 to evaluate the performance of the new estimators. Finally, Section 4 provides conclusions.

2 L-Moments and Proposed Classes

Reference [19] described the L-moments as expectations of the order statistics of certain linear combinations. L-moments can be specified for any random variable for which a mean exists. They are used to describe probability distributions and estimate parameters, and their estimates are used for summarizing and describing the samples of observed data. There are many advantages of L-moments over traditional moments: they are linear data functions, they suffer less from the effects of sample change-ability, they are more robust to outliers/extreme values in data, and they enable safer inferences made from small samples about any fundamental population parameter. The general population mathematical forms of first four L-moments for the auxiliary variable X in relation to the stratum h are defined as follows:

$$L_{1xl} = E(X_{1:1})$$

$$L_{2xl} = \frac{1}{2} E(X_{2:2} - X_{1:2})$$

$$L_{3xl} = \frac{1}{3} E(X_{3:3} - 2X_{2:3} + X_{1:3})$$

$$L_{4xl} = \frac{1}{4} E(X_{4:4} - 3X_{3:4} + 3X_{2:4} - X_{1:4}).$$

Similarly, we can write second-stage sample L-moments of the auxiliary variable as

$$l_{1xl} = \binom{n_h}{1}^{-1} \sum_{d=1}^{n_h} x_{h(d)}$$

$$l_{2xl} = \frac{1}{2}\binom{n_h}{2}^{-1}\sum_{d=1}^{n_h}\left\{\binom{d-1}{1} - \binom{n-d}{1}\right\}x_{h(d)}$$

$$l_{3xl} = \frac{1}{3}\binom{n_h}{3}^{-1}\sum_{d=1}^{n_h}\left\{\binom{d-1}{2} - 2\binom{d-1}{1}\binom{n-d}{1} + \binom{n-d}{2}\right\}x_{h(d)}$$

$$l_{4xl} = \frac{1}{4}\binom{n_h}{4}^{-1}\sum_{d=1}^{n_h}\left\{\binom{d-1}{3} - 3\binom{d-1}{2}\binom{n-d}{1} + 3\binom{d-1}{1}\binom{n-d}{2} - \binom{n-d}{3}\right\}x_{h(d)}.$$

where $x_{h(d)}$ represents the dth order statistics with binomial coefficient (:). Similarly, we can write the L-moments expression for the first-stage sample as l_{1xl}^*, l_{2xl}^*, l_{3xl}^*, and l_{4xl}^*. Furthermore, we can write the mathematical expressions of L-moments for the study variable Y by adapting the structure of auxiliary variable X.

2.1 First Proposed Class of Estimators

The authors of [9,10] used robust regression and robust co-variance matrices methodologies for improved estimation of the population's mean. Their use of robust regression and robust co-variance matrices allows us to utilize robust moments (L-moments) instead of traditional moments. Hence, taking motivation from [21], we propose the following class of L-moments based calibration estimators of variance under double stratified sampling:

$$V_{ai} = \sum_{h=1}^{R}\gamma_h s_{ymh}^2 \tag{2}$$

where the calibrated weights are selected to minimize the measure of chi-square distance

$$\sum_{h=1}^{R}\frac{(\gamma_h - W_h)^2}{W_h\theta_h} \tag{3}$$

is subject to the following calibration constraints

$$\sum_{h=1}^{R}\gamma_h = \sum_{h=1}^{R}W_h \tag{4}$$

$$\sum_{h=1}^{R}\gamma_h\Delta_{xmh} = \sum_{h=1}^{R}W_h\Delta_{xmh}^* \tag{5}$$

where $s_{ymh}^2 = l_{2yl}^2$ is the second-stage L-variance of Y; $\Delta_{xmh}^* = \left(\bar{x}_h^* = l_{1xl}^*, C_{xmh}^* = \frac{l_{2xl}^*}{l_{1xl}^*}, s_{xmh}^{*2} = l_{2xl}^{*2}\right)$ is the first-stage L-location, L-cv, and L-variance; and $\Delta_{xmh} = \left(\bar{x}_h = l_{1xl}, C_{xmh} = \frac{l_{2xl}}{l_{1xl}}, s_{xmh}^2 = l_{2xl}^2\right)$ is the second-stage L-location, L-cv, and L-variance of X.

The Lagrange function is given as

$$T = \sum_{h=1}^{R} \frac{(\gamma_h - W_h)^2}{W_h \theta_h} - 2\mu_{11} \left(\sum_{h=1}^{R} \gamma_h - \sum_{h=1}^{R} W_h \right) - 2\mu_{12} \left(\sum_{h=1}^{R} \gamma_h \Delta_{xmh} - \sum_{h=1}^{R} W_h \Delta_{xmh}^* \right) \quad (6)$$

where μ_{11} and μ_{12} are the Lagrange multiples. To obtain the optimum value for the calibration weight, we differentiate the Lagrange function with respect to γ_h and set it equal to zero. Thus the weight of calibration can be obtained in the form

$$\gamma_h = W_h + W_h \theta_h (\mu_{11} + \mu_{12} \Delta_{xmh}) \quad (7)$$

Now, μ_{11} and μ_{12} can be obtained by replacing γ_h in Eqs. (4) and (5) with its value given by Eq. (7). Thus, we obtain a weight of calibration of

$$\gamma_h = W_h + W_h \theta_h \left[\frac{-\left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh} \right) \right)}{\left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh}^2 \right) \left(\sum_{h=1}^{R} W_h \theta_h \right) - \left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh} \right)^2} \right]$$

$$+ W_h \theta_h \Delta_{xmh} \left[\frac{\left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \left(\sum_{h=1}^{R} W_h \theta_h \right) \right)}{\left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh}^2 \right) \left(\sum_{h=1}^{R} W_h \theta_h \right) - \left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh} \right)^2} \right] \quad (8)$$

By substituting the value of γ_h from Eq. (8) with that from Eq. (2), we can obtain the proposed estimator of the calibration as follows:

$$V_{ai} = \sum_{h=1}^{R} W_h s_{ymh}^2 + \widehat{\beta}_{cv} \sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \quad (9)$$

where

$$\widehat{\beta}_{cv} = \left[\frac{\left(\sum_{h=1}^{R} W_h \theta_h \right) \left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh} s_{ymh}^2 \right) - \left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh} \right) \left(\sum_{h=1}^{R} W_h \theta_h s_{ymh}^2 \right)}{\left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh}^2 \right) \left(\sum_{h=1}^{R} W_h \theta_h \right) - \left(\sum_{h=1}^{R} W_h \theta_h \Delta_{xmh} \right)^2} \right] \quad (10)$$

The members of the first proposed class are provided in Tab. 1.

2.2 Second Proposed Class of Estimators

By extending the idea of V_{ai}, we propose the second class of estimators of variance under double stratified sampling as given below:

$$V_{bi} = \sum_{h=1}^{R} \gamma_h s_{ymh}^2 \quad (11)$$

Through using the distance of chi-square,

$$\sum_{h=1}^{R} \frac{(\gamma_h - W_h)^2}{\theta_h W_h} \quad (12)$$

Table 1: First proposed class of estimators

V_{ai}	θ_h	Δ_{xmh}	Δ^*_{xmh}
V_{a1}	1	$\overline{x}_h$	$\overline{x}^*_h$
V_{a2}	$\dfrac{1}{\overline{x}_h}$	$\overline{x}_h$	$\overline{x}^*_h$
V_{a3}	$\dfrac{1}{s_{xmh}}$	$\overline{x}_h$	$\overline{x}^*_h$
V_{a4}	$\dfrac{1}{s^2_{xmh}}$	$\overline{x}_h$	$\overline{x}^*_h$
V_{a5}	$\dfrac{1}{C_{xmh}}$	$\overline{x}_h$	$\overline{x}^*_h$
V_{a6}	1	C_{xmh}	C^*_{xmh}
V_{a7}	$\dfrac{1}{\overline{x}_h}$	C_{xmh}	C^*_{xmh}
V_{a8}	$\dfrac{1}{s_{xmh}}$	C_{xmh}	C^*_{xmh}
V_{a9}	$\dfrac{1}{s^2_{xmh}}$	C_{xmh}	C^*_{xmh}
V_{a10}	$\dfrac{1}{C_{xmh}}$	C_{xmh}	C^*_{xmh}
V_{a11}	1	s^2_{xmh}	s^{*2}_{xmh}
V_{a12}	$\dfrac{1}{\overline{x}_h}$	s^2_{xmh}	s^{*2}_{xmh}
V_{a13}	$\dfrac{1}{s_{xmh}}$	s^2_{xmh}	s^{*2}_{xmh}
V_{a14}	$\dfrac{1}{s^2_{xmh}}$	s^2_{xmh}	s^{*2}_{xmh}
V_{a15}	$\dfrac{1}{C_{xmh}}$	s^2_{xmh}	s^{*2}_{xmh}

which is subject to the following three calibration constraints:

$$\sum_{h=1}^{R} \gamma_h \Delta_{xmh} = \sum_{h=1}^{R} W_h \Delta^*_{xmh} \tag{13}$$

$$\sum_{h=1}^{R} \gamma_h s_{xmh}^2 = \sum_{h=1}^{R} W_h s_{xmh}^{*2} \tag{14}$$

$$\sum_{h=1}^{R} \gamma_h = \sum_{h=1}^{R} W_h \tag{15}$$

The Lagrange function is given as

$$T = \sum_{h=1}^{R} \frac{(\gamma_h - W_h)^2}{\theta_h W_h} - 2\mu_{21}\left(\sum_{h=1}^{R} \gamma_h \Delta_{xmh} - \sum_{h=1}^{R} W_h \Delta_{xmh}^*\right) - 2\mu_{22}\left(\sum_{h=1}^{R} \gamma_h s_{xmh}^2 - \sum_{h=1}^{R} W_h s_{xmh}^{*2}\right)$$
$$- 2\mu_{23}\left(\sum_{h=1}^{R} \gamma_h - \sum_{h=1}^{R} W_h\right)$$

After taking the derivative of T with respect to γ and setting it equal to zero, we get

$$\gamma_h = W_h + \theta_h W_h \left(\mu_{21}\Delta_{xmh} + \mu_{22}s_{xmh}^2 + \mu_{23}\right). \tag{16}$$

The following equations system can be obtained by substituting Eq. (16) into Eqs. (13)–(15) respectively:

$$[P_a]_{3\times3}[P_b]_{3\times1} = [P_c]_{3\times1}$$

where

$$P_b = \begin{bmatrix} \mu_{21} \\ \mu_{22} \\ \mu_{23} \end{bmatrix}, \quad P_c = \begin{bmatrix} \displaystyle\sum_{h=1}^{R} W_h \Delta_{xmh}^* - \sum_{h=1}^{R} W_h \Delta_{xmh} \\ \displaystyle\sum_{h=1}^{R} W_h s_{xmh}^{*2} - \sum_{h=1}^{R} W_h s_{xmh}^2 \\ 0 \end{bmatrix}, \text{ and}$$

$$P_a = \begin{bmatrix} \displaystyle\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2 & \displaystyle\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{xmh}^2 & \displaystyle\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \\ \displaystyle\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{xmh}^2 & \displaystyle\sum_{h=1}^{R} \theta_h W_h s_{xmh}^4 & \displaystyle\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \\ \displaystyle\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} & \displaystyle\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 & \displaystyle\sum_{h=1}^{R} \theta_h W_h \end{bmatrix}$$

Upon solving the equations system for μs, we get

$$\mu_{21} = \frac{A_1}{D_1}, \quad \mu_{22} = \frac{B_1}{D_1}, \quad \mu_{23} = \frac{C_1}{D_1},$$

where

$$A_1 = \left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^4 \right)$$

$$- \left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right)^2$$

$$+ \left(\sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2 \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right)$$

$$- \left(\sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2 \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh} \right)$$

$$B_1 = \left(\sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2 \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2 \right)$$

$$- \left(\sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2 \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right)^2$$

$$- \left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh} \right) \left(\sum_{h=1}^{R} \theta_h W_h \right)$$

$$+ \left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right)$$

$$C_1 = \left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh} \right)$$

$$- \left(\sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh} \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh} \right)$$

$$+ \left(\sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2 \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{xmh}^2 \right)$$

$$- \left(\sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2 \right) \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right)$$

$$D_1 = \left(\sum_{h=1}^{R} \theta_h W_h\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^4\right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2\right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}\right)^2 \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^4\right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh}\right)^2$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2\right)^2 \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2\right)$$

$$+ 2\left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{xmh}^2\right)$$

When substituting these μs, into Eq. (16) and then Eq. (11), we obtain the following:

$$V_{bi} = \sum_{h=1}^{R} W_h s_{ymh}^2 + \beta_{3(dnew)} \sum_{h=1}^{R} W_h \left(\Delta_{xmh}^* - \Delta_{xmh}\right) + \beta_{4(dnew)} \sum_{h=1}^{R} W_h \left(s_{xmh}^{*2} - s_{xmh}^2\right)$$

where $\beta_{3(dnew)} = \frac{A_1^*}{D_1}$, $\beta_{4(dnew)} = \frac{B_1^*}{D_1}$, and

$$A_1^* = \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{ymh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^4\right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{ymh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2\right)^2$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 s_{ymh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh}\right) \left(\sum_{h=1}^{R} \theta_h W_h\right)$$

$$+ \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 s_{ymh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2\right)$$

$$+ \left(\sum_{h=1}^{R} \theta_h W_h s_{ymh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh}\right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h s_{ymh}^2\right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}\right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^4\right)$$

$$B_1^* = \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{ymh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} s_{ymh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh} \right)$$

$$+ \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 s_{ymh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2 \right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 s_{ymh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right)^2$$

$$+ \left(\sum_{h=1}^{R} \theta_h W_h s_{ymh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh} \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \Delta_{xmh} \right)$$

$$- \left(\sum_{h=1}^{R} \theta_h W_h s_{ymh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h \Delta_{xmh}^2 \right) \left(\sum_{h=1}^{R} \theta_h W_h s_{xmh}^2 \right)$$

The members of the second proposed class are listed in Tab. 2.

Table 2: Second proposed class of estimators

V_{bi}	θ_h	Δ_{xmh}	Δ_{xmh}^*
V_{b1}	1	$\overline{x}_h$	$\overline{x}_h^*$
V_{b2}	$\dfrac{1}{\overline{x}_h}$	$\overline{x}_h$	$\overline{x}_h^*$
V_{b3}	$\dfrac{1}{s_{xmh}}$	$\overline{x}_h$	$\overline{x}_h^*$
V_{b4}	$\dfrac{1}{s_{xmh}^2}$	$\overline{x}_h$	$\overline{x}_h^*$
V_{b5}	$\dfrac{1}{C_{xmh}}$	$\overline{x}_h$	$\overline{x}_h^*$
V_{b6}	1	C_{xmh}	C_{xmh}^*
V_{b7}	$\dfrac{1}{\overline{x}_h}$	C_{xmh}	C_{xmh}^*
V_{b8}	$\dfrac{1}{s_{xmh}}$	C_{xmh}	C_{xmh}^*
V_{b9}	$\dfrac{1}{s_{xmh}^2}$	C_{xmh}	C_{xmh}^*
V_{b10}	$\dfrac{1}{C_{xmh}}$	C_{xmh}	C_{xmh}^*

3 Numerical Illustrations

Here, we evaluate the performance of the proposed estimators through three populations.

3.1 Simulation Design (Population-1)

In this article, we consider the population with size $N = 1000$. Utilizing an equal allocation of a sample with size 100 is selected from hth stratum, and the total sample size $n_h = 400$. Furthermore, for stratum h, random variables X_h and Y_h are defined as follows:

$$Y_h = \delta + KX_h + \varepsilon X_h^p, \quad \text{for the } h^{th} \text{ stratum} \tag{17}$$

where X_h for $h = 1, 2, 3, 4$ follows Gamma distributions with parameter values as given below:

$X_1 \sim G_h(2.6, 3.8), \quad for \ h = 1$

$X_2 \sim G_h(2.0, 3.1), \quad for \ h = 2$

$X_3 \sim G_h(1.5, 2.7), \quad for \ h = 3$

$X_4 \sim G_h(2.9, 3.1), \quad for \ h = 4.$

ε follows a standard normal distribution, and $\delta = 5$, $p = 1.6$, and $K = 2$.

Figs. 1–4 show the scatter plots for each stratum. The existence of extreme values is clearly demonstrated by these figures and are therefore fitting for evaluating our proposed estimators.
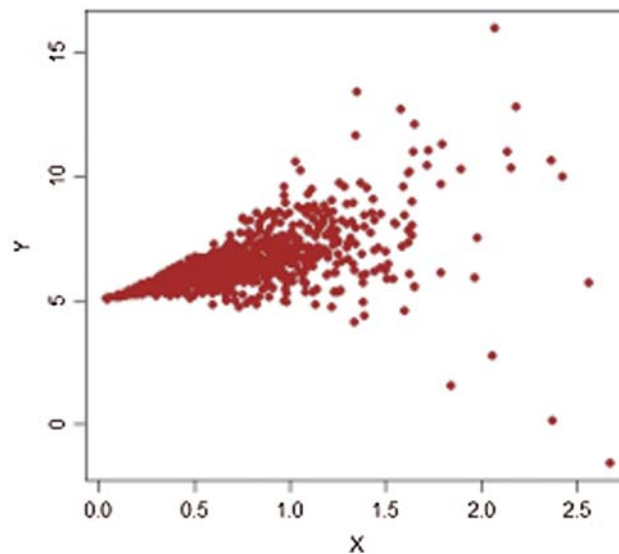
Figure 1: Population-1, $h = 1$

The simulation steps are as below:

Step 1: Select a random sample with size n_h through SRSWOR from stratum h.

Step 2: Find the value of variance estimate, say $\omega = V_{ai}, V_{bi}$ where $ai = 1, 2, \ldots, 15$ and $bi = 1, 2, \ldots, 10$.

Step 3: Repeat Steps 1 and 2 for $L = 5000$ times. Obtain $\omega_1, \omega_2, \ldots, \omega_L$.

Step 4: Compute the mean square error (MSE) as

$$MSE\,(\omega) = \frac{1}{L}\sum_{i=1}^{L}(\omega - \overline{\omega})^2$$

Step 5: Compute the percentage relative efficiency (PRE) as

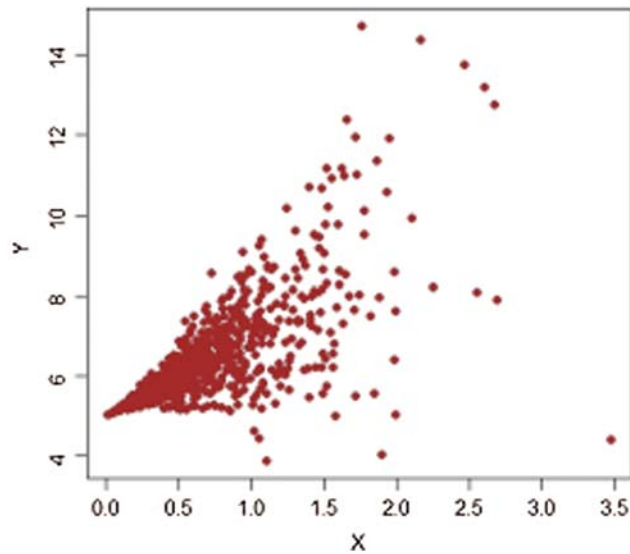$$PRE\,(\omega) = \frac{MSE\,(T_0)}{MSE\,(\omega)} \times 100$$
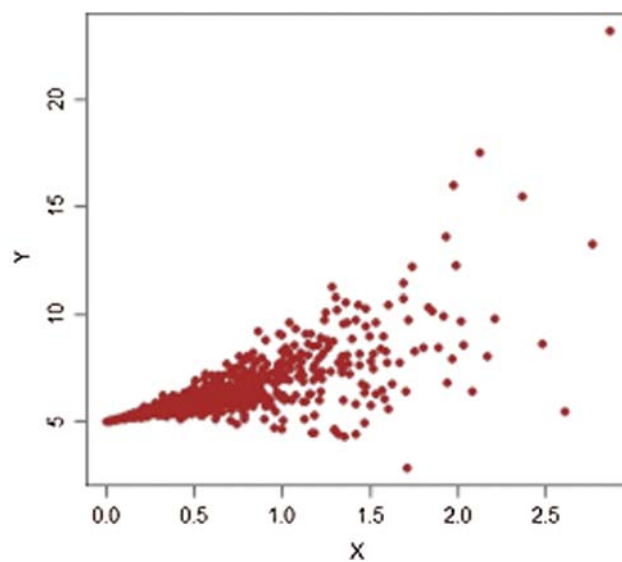

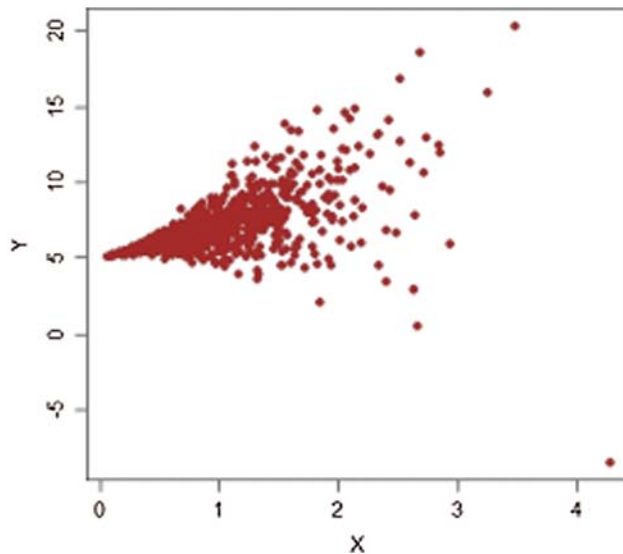**Figure 2:** Population-1, $h = 2$

Figure 3: Population-1, $h = 3$

Figure 4: Population-1, h $= 4$

The estimators' PRE obtained from the above five steps are provided in Tab. 3.

Table 3: PRE for Population-1

$V_{a1} - V_{a5}$	$V_{a6} - V_{a10}$	$V_{a11} - V_{a15}$	$V_{b1} - V_{b5}$	$V_{b6} - V_{b10}$
$V_{a1} = 422.00$	$V_{a6} = 188.39$	$V_{a11} = 478.67$	$V_{b1} = 140.80$	$V_{b6} = 134.25$
$V_{a2} = 412.72$	$V_{a7} = 186.00$	$V_{a12} = 475.95$	$V_{b2} = 140.67$	$V_{b7} = 134.86$
$V_{a3} = 417.20$	$V_{a8} = 187.29$	$V_{a13} = 478.54$	$V_{b3} = 140.86$	$V_{b8} = 134.44$
$V_{a4} = 410.80$	$V_{a9} = 185.05$	$V_{a14} = 475.64$	$V_{b4} = 140.79$	$V_{b9} = 134.44$
$V_{a5} = 422.86$	$V_{a10} = 187.10$	$V_{a15} = 476.58$	$V_{b5} = 140.72$	$V_{b10} = 133.43$

3.2 Real Life Data

The apple fruit is one of the most common types of fruits. It is native to Central Asia, but today it grows worldwide with different colors and sizes. The apple fruit is rich in fiber, vitamins, and antioxidants and has many health benefits.

In the present article, we use collected apple fruit data used by [24], where

Population-2: X $=$ number of apple trees in 1999, Y $=$ level of apple production in 1999.

Population-3: X $=$ level of apple production in 1998, Y $=$ level of apple production in 1999.

It should be noted that we consider 477 villages in four strata in 1999, termed (1: Marmarian), (2: Agean), (3: Mediterranean), and (4: Central Anatolia). The scatter plots of extreme values for each stratum are shown in Figs. 5–12. The estimators' PREs are computed as defined in Subsection 3.1, and are presented in Tabs. 4 and 5. The first-stage samples with sizes n_1^*, n_2^*, n_3^*

and n_4^* are selected, and then from these samples the second-stage samples with sizes n_1, n_2, n_3, and n_4 are selected:

$N_1 = 106$, $N_2 = 106$, $N_3 = 94$, $N_4 = 171$,

$n_1^* = 58$, $n_2^* = 58$, $n_3^* = 52$, $n_4^* = 94$,

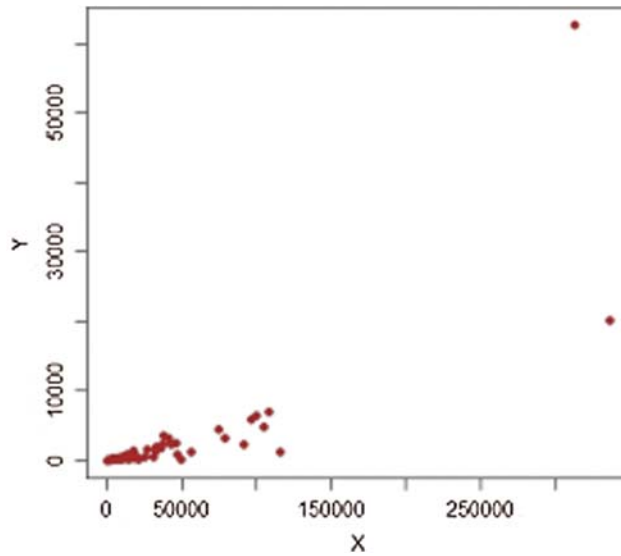$n_1 = 29$, $n_2 = 29$, $n_3 = 26$, $n_4 = 47$.

Figure 5: Population-2, $h = 1$
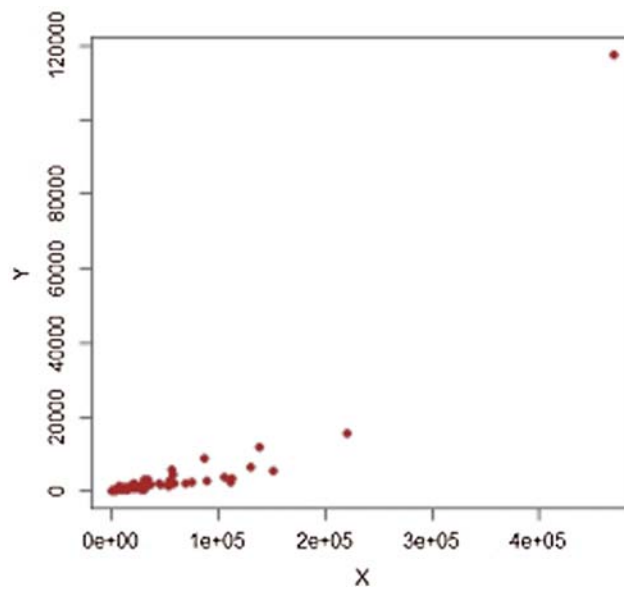
Figure 6: Population-2, $h = 2$
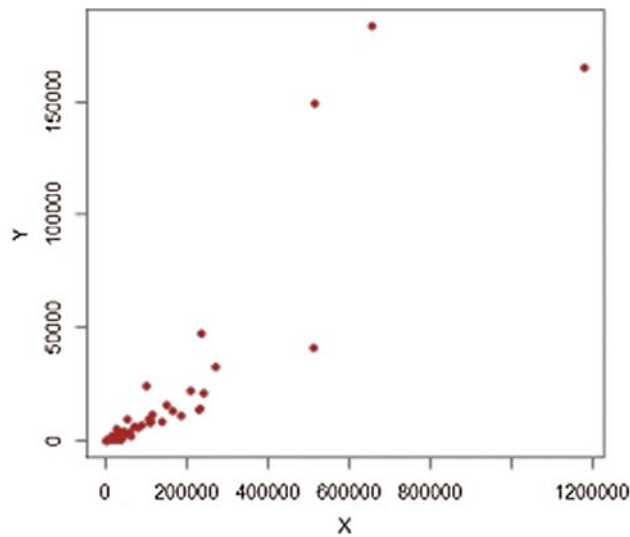
Figure 7: Population-2, $h = 3$
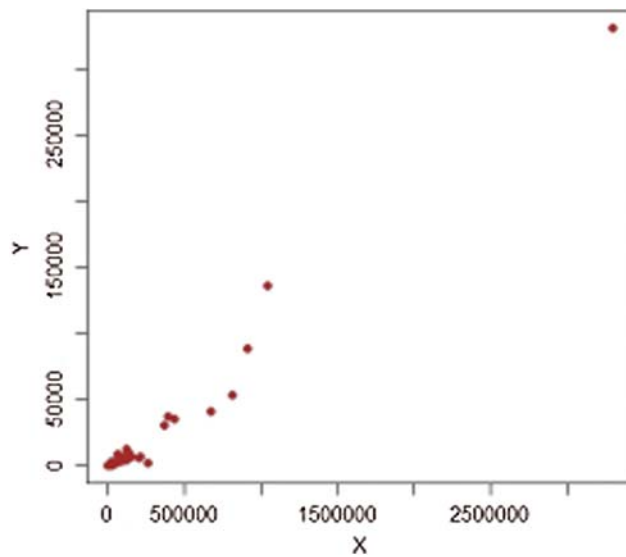
Figure 8: Population-2, $h = 4$

3.3 Findings

1: From Tab. 3, we can see that the results (V_{ai}, V_{bi}) of Population-1 indicates that

$PRE(V_{a11-15}) > PRE(V_{a1-5}) > PRE(V_{a6-10})$, w.r.t. V_{ai}

$PRE(V_{b1-5}) > PRE(V_{b6-10})$, w.r.t. V_{bi}.

The proposed estimators V_{a11} and V_{b3} record the highest efficiency compared to other competitor estimators.
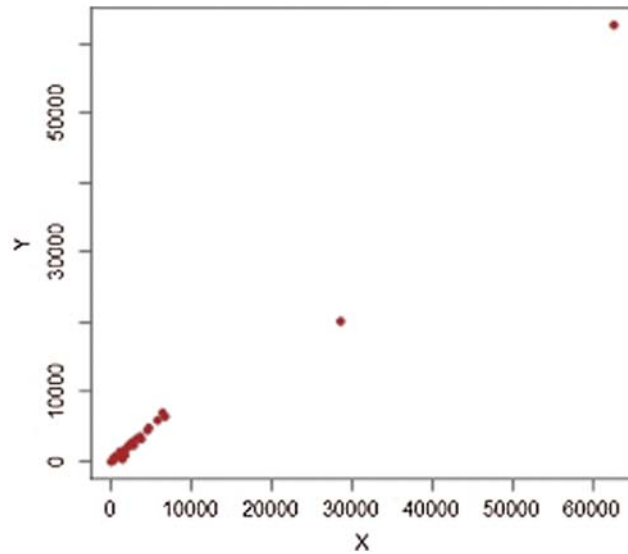
Figure 9: Population-3, $h = 1$

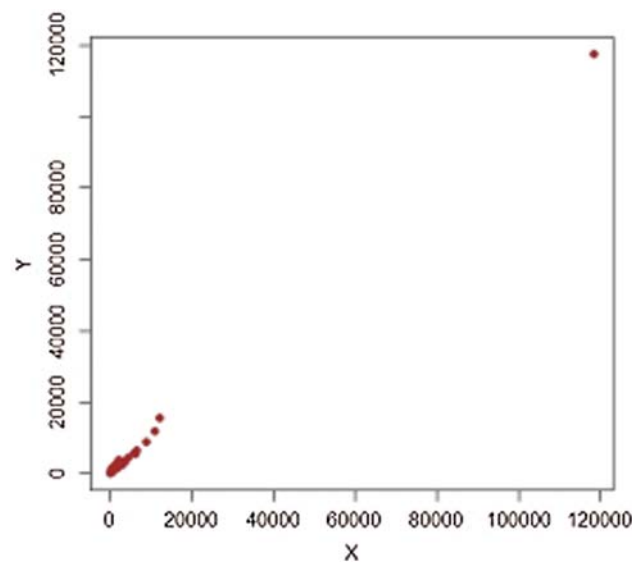**Figure 10:** Population-3, $h = 2$

2: Meanwhile, the results (V_{ai}, V_{bi}) of Population-2 in Tab. 4 indicate that

$PRE(V_{a11-13}, V_{a15}) > PRE(V_{a1-5}, V_{a14}) > PRE(V_{a6-10})$, w.r.t. V_{ai}

$PRE(V_{b1-5}) > PRE(V_{b6-10})$, w.r.t. V_{bi}.

The proposed estimators V_{a11} and V_{b5} record the highest efficiency compared to other competitor estimators.
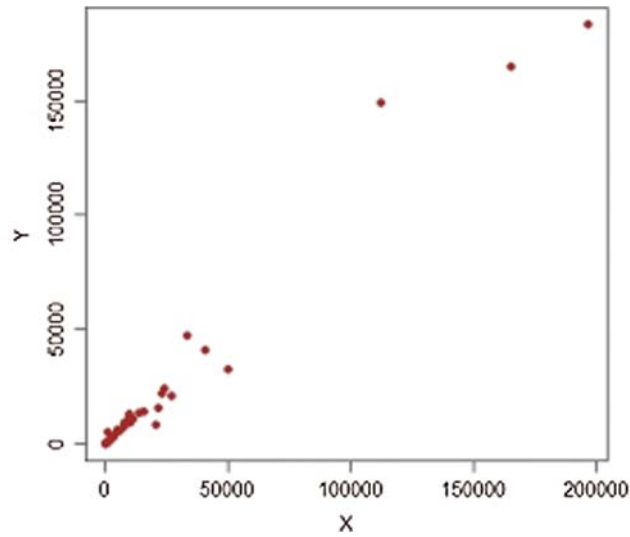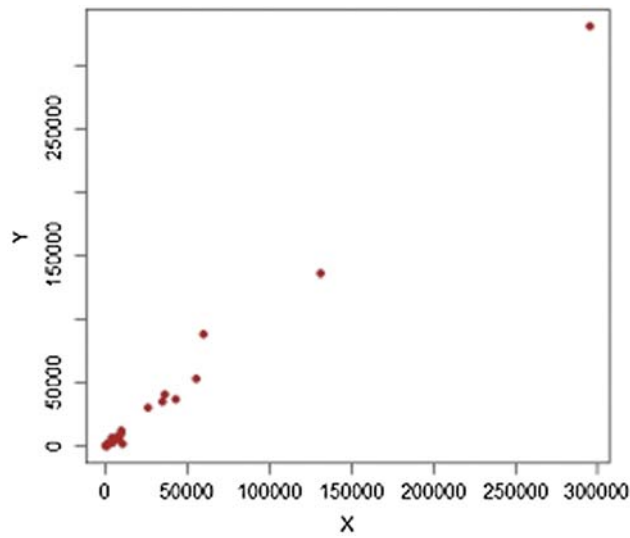
Figure 11: Population-3, h $= 3$

Figure 12: Population-3, h $= 4$

3: The results (V_{ai}, V_{bi}) of Population-3 (see Tab. 5) reveal that

$PRE(V_{a11-15}) > PRE(V_{a1-5}) > PRE(V_{a6-10})$, w.r.t. V_{ai}

$PRE(V_{b1}, V_{b5-8}, V_{b10}) > PRE(V_{b2-4}, V_{b9})$, w.r.t. V_{bi} .

Hence, the proposed estimators V_{a11} and V_{b1} record the highest efficiency of all compared estimators.

4: Comparing the two proposed classes for each population, leads us to the following findings:

Population-1: $PRE\left(V_{a1-15}\right) > PRE\left(V_{b1-10}\right)$;

Population-2: $PRE\left(V_{a1-15}\right) > PRE\left(V_{b1-10}\right)$;

Population-3: $PRE\left(V_{a1-5}, V_{a11-15}\right) > PRE\left(V_{b1-10}\right)$.

5: Overall, all the members of new classes have $PRE > 100$ with respect to T_o, and this clearly indicates that the performance of the proposed estimators is better than that of traditional estimators.

6: Furthermore, the proposed variance estimator V_{a11} is the best estimator among all proposed estimators, having PREs of $478.67, 28051.41$, and 77307.88 for populations 1–3, respectively.

Table 4: PRE for Population-2

$V_{a1} - V_{a5}$	$V_{a6} - V_{a10}$	$V_{a11} - V_{a15}$	$V_{b1} - V_{b5}$	$V_{b6} - V_{b10}$
$V_{a1} = 22054.22$	$V_{a6} = 10802.69$	$V_{a11} = 28051.41$	$V_{b1} = 4679.23$	$V_{b6} = 2920.96$
$V_{a2} = 19909.06$	$V_{a7} = 10657.67$	$V_{a12} = 25287.50$	$V_{b2} = 4568.57$	$V_{b7} = 2712.07$
$V_{a3} = 19482.79$	$V_{a8} = 10556.63$	$V_{a13} = 24738.15$	$V_{b3} = 4555.36$	$V_{b8} = 2679.75$
$V_{a4} = 14317.23$	$V_{a9} = 8598.21$	$V_{a14} = 18935.43$	$V_{b4} = 4054.82$	$V_{b9} = 2316.42$
$V_{a5} = 21880.95$	$V_{a10} = 10848.33$	$V_{a15} = 27712.04$	$V_{b5} = 4689.26$	$V_{b10} = 2901.67$

Table 5: PRE for Population-3

$V_{a1} - V_{a5}$	$V_{a6} - V_{a10}$	$V_{a11} - V_{a15}$	$V_{b1} - V_{b5}$	$V_{b6} - V_{b10}$
$V_{a1} = 38630.57$	$V_{a6} = 4710.165$	$V_{a11} = 77307.88$	$V_{b1} = 29995.12$	$V_{b6} = 28542.51$
$V_{a2} = 39188.94$	$V_{a7} = 5415.428$	$V_{a12} = 76371.62$	$V_{b2} = 24700.47$	$V_{b7} = 27754.20$
$V_{a3} = 39183.04$	$V_{a8} = 5430.796$	$V_{a13} = 76342.05$	$V_{b3} = 24669.71$	$V_{b8} = 27732.49$
$V_{a4} = 39160.33$	$V_{a9} = 4463.861$	$V_{a14} = 73964.38$	$V_{b4} = 18395.29$	$V_{b9} = 24642.28$
$V_{a5} = 38630.45$	$V_{a10} = 4745.627$	$V_{a15} = 77297.10$	$V_{b5} = 29953.95$	$V_{b10} = 28544.35$

4 Conclusion

The difficulty of data analysis arises from the presence of extreme values that adversely impact the variance estimation based on central moments. One of the ways to solve this issue is to use L-moments that provide a robust statistical structure for analysis. Calibration estimation is a common statistical approach that relies on the use of auxiliary information to adjust the original weights of design and to improve the accuracy of estimators. Motivation by [21], we propose new classes of estimators to estimate the population variance based on L-moments and present a calibration approach for double stratified random sampling. The percentage relative efficiency is adopted to compare the performance of the proposed estimators through three populations and through a simulation as well as application to real-life data. Our numerical results show that the proposed estimators are always superior and more efficient to existing estimators.

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

References

[1] D. J. Watson, "The estimation of leaf area in field crops," *Journal of Agricultural Science*, vol. 27, no. 3, pp. 474–483, 1937.

[2] W. G. Cochran, "The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce," *Journal of Agricultural Science*, vol. 30, no. 2, pp. 262–275, 1940.

[3] K. Ibrahim, M. Syam and A. I. Al-Omari, "Estimating the population mean using stratified median ranked set sampling," *Applied Mathematical Sciences*, vol. 4, no. 47, pp. 2341–2354, 2010.

[4] M. Syam, K. Ibrahim and A. I. Al-Omari, "The efficiency of stratified quartile ranked set sample in estimating the population mean," *Tamsui Oxford Journal of Information and Mathematical Sciences*, vol. 28, no. 2, pp. 175–190, 2012.

[5] M. Abid, N. Abbas and M. Riaz, "Improved modified ratio estimators of population mean based on deciles," *Chiang Mai Journal of Science*, vol. 43, no. 1, pp. 1311–1323, 2016.

[6] 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.

[7] N. Ali, I. Ahmad, M. Hanif and U. Shahzad, "Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information," *Communications in Statistics—Theory and Methods*, vol. 50, no. 4, pp. 979–992, 2019.

[8] M. Hanif and U. Shahzad, "Estimatıon of population variance using kernel matrix," *Journal of Statistics and Management Systems*, vol. 22, no. 3, pp. 563–586, 2019.

[9] 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.

[10] T. Zaman and H. Bulut, "Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling," *Communications in Statistics-Theory and Methods*, vol. 49, no. 14, pp. 3407–3420, 2020.

[11] U. Shahzad, N. H. Al-Noor, M. Hanif and I. Sajjad, "An exponential family of median based estimators for mean estimation with simple random sampling scheme," *Communications in Statistics-Theory and Methods*, vol. 43, no. 1, pp. 1–10, 2020.

[12] 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.

[13] F. Gagnon, H. Lee, E. Rancourt and C. E. Särndal, "Estimating the variance of the generalised regression estimator in the presence of imputation for the generalised estimation system," in *Proc. of the Survey Methods Section, Statistical Society of Canada*, pp. 151–156, 1996.

[14] C. E. Särndal, B. Swensson and J. Wretman, *Model Assisted Survey Sampling*. New York: Springer-Verlang, 1992.

[15] A. A. Jemain, A. I. Al-Omari and K. Ibrahim, "Modified ratio estimator for the population mean using double median ranked set sampling," *Pakistan Journal of Statistics*, vol. 24, no. 3, pp. 217–226, 2008.

[16] A. I. Al-Omari, "Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling," *Statistics and Probability Letters*, vol. 82, no. 11, pp. 1883–1890, 2012.

[17] M. Syam, K. Ibrahim and A. I. Al-Omari, "The efficiency of stratified double percentile ranked set sample for estimating the population mean," *Far East Journal of Mathematical Sciences*, vol. 73, no. 1, pp. 157–177, 2013.

[18] M. Syam, A. I. Al-Omari and K. Ibrahim, "On the mean estimation using stratified double median ranked set sampling," *American Journal of Applied Sciences*, vol. 14, no. 4, pp. 496–502, 2017.

[19] 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.

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

[21] 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.

[22] U. Shahzad, I. Ahmad, I. Almanjahie, N. H. Al-Noor and M. Hanif, "A new class of L-moments based calibration variance estimators," *Computers Materials and Continua*, vol. 66, no. 3, pp. 3013–3028, 2021.

[23] U. Shahzad, I. Ahmad, I. Almanjahie, M. Hanif and N. H. Al-Noor, "L-moments and calibration based variance estimators under double stratified random sampling scheme: An application of covid-19 pandemic," *Scientia Iranica*, pp. 1–17, 2021.

[24] 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.