Scientia Iranica E (2023) 30(2), 814–821

Sharif University of Technology
Scientia Iranica
Transactions E: Industrial Engineering
http://scientiairanica.sharif.edu

L-Moments and calibration-based variance estimators under double stratified random sampling scheme: Application of Covid-19 pandemic

U. Shahzad[a,b,*], I. Ahmad[a], I. Mufrah Almanjahie[c,d], M. Hanif [b], and N.H. Al-Noor[e]

a. *Department of Mathematics and Statistics, International Islamic University, Islamabad 46000, Pakistan.*
b. *Department of Mathematics and Statistics - PMAS-Arid Agriculture University, Rawalpindi 46300, Pakistan.*
c. *Department of Mathematics, College of Science, King Khalid University, Abha 62529, Saudi Arabia.*
d. *Statistical Research and Studies Support Unit, King Khalid University, Abha 62529, Saudi Arabia.*
e. *Department of Mathematics, College of Science, Mustansiriyah University, Baghdad 10011, Iraq.*

Received 22 September 2020; received in revised form 9 December 2020; accepted 5 July 2021

KEYWORDS
Extreme observations;
Variance estimation;
L-Moments;
Calibration;
Double stratified random sampling.

Abstract. Extreme events gives rise to outrageous results in terms of population-related parameters and their estimates are usually done using traditional moments. Traditional moments are usually affected by extreme observations. This study aims to propose some new calibration estimators considering the L-Moments scheme for variance, which is one of the most important population parameters. a number of suitable calibration constraints under double stratified random sampling were defined for these estimators. The proposed estimators, which were based on L-Moments, were relatively more robust despite extreme values. The empirical efficiency of the proposed estimators was also assessed through simulation. Covid-19 pandemic data from January 22, 2020 to August 23, 2020 was taken into account in the simulation study.

1. Introduction

Auxiliary information can be used in different stages to improve the efficiency of the variance estimator for a finite population case. There are numerous real-life examples where there is a roughly linear relationship between the study variable Y and auxiliary variable X. Take height and weight for example: taller people tend to be heavier. The same holds for the body mass index and total cholesterol given the direct positive relationship between them. This sort of linear relationship allows researchers to use auxiliary variable X for improved estimation of any parameter of study variable Y. For more discussion on the auxiliary information, refer to the studies carried out by Koyuncu [1], Al-Omari [2], Zaman [3,4], Naz et al. [5,6], and Shahzad et al. [7,8]. An alternative method for situations with further available auxiliary information is ranked set sampling due to McIntyre [9]. The proposed method is more cost-efficient than simple random sampling method (see [10–16]).

Double sampling is a technique in which the

*. *Corresponding author.*
E-mail addresses: usman.stat@yahoo.com (U. Shahzad); ishfaq.ahmad@iiu.edu.pk (I. Ahmad); imalmanjahi@kku.edu.sa (I. Mufrah Almanjahie); mhpuno@hotmail.com (M. Hanif); nadialnoor@uomustansiriyah.edu.iq (N.H. Al-Noor)

doi: 10.24200/sci.2021.56853.4942

information related to the auxiliary variable X is not available at the first phase, while the information related to the study variable Y is available on a smaller sub-sample chosen from the first-phase sampling. For instance, consider a situation where a physiologist needs to assess variation (variance) of leaf zone (area) for another strain of wheat. It might, in some cases, not be necessary to pluck all the leaves in the total population of 120 plants and get the weight X to construct a variance estimator of the leaf zone Y. Instead, it could be more proper to choose a substantially large first-stage sample of leaves and measure the weight X for the sample leaves. A sub-sample from this underlying sample of leaves could then be chosen to determine the leaf zone. As a result, an estimate of the variance of weight from all 120 plants was acquired from the perception made on the first-phase sample. This estimate of weight X' could be used in the variance estimation of the leaf zone Y instead of population weight X.

Assume that (Y, X) belongs to the population $\Omega = \{v_{11}, v_{12}, ..., v_n\}$ of size N. Here, Ω is classified into H'' strata. Further, $\phi_h = \frac{N_h}{N}$ is the stratum weight where N_h represents the size of h^{th} stratum for $h = 1, 2, \ldots, H''$. The overall size of the population containing all the strata is denoted by $\sum_{h=1}^{H''} N_h = N$. Now, the first-phase simple random sample (srs) of size $n_h^{'}$ is drawn without replacement from the h^{th} stratum such that $\sum_{h=1}^{H''} n_h^{'} = n^{'}$ and then, a second stage sample n_h $(n_h < n_h^{'})$ is selected. In light of this double stratified random sampling design, the traditional estimator of variance can be expressed as follows:

$$T_o = \sum_{h=1}^{H'} \phi_h s_{yh}^2, \tag{1}$$

where s_{yh}^2 denotes the traditional variance of study variable in the h^{th} stratum for $h = 1, 2, \ldots, H''$ and T_o the traditional unbiased variance estimator under double stratified random sampling. T_o is based on traditional moments and, thus, influenced by extreme values. In the literature, much of the development has done to tackle this issue regarding mean estimation. For instance, Zaman and Bulut [17,18] introduced robust regression techniques for controlling the effects of extreme values. Ali et al. [19] extended their idea for the mean estimation of a sensitive variable. Abid et al. [20] used some non-conventional descriptive measures of statistics for variance estimation. Of note, the estimates of both mean and variance used in all these described studies are based on the traditional moments. On the contrary, in this study, Linear Moments (L-Moments) were taken into consideration to construct some new estimators of the variance based on L-Moments characteristics of auxiliary and study variables rather than traditional moments. L-Moments were highly robust in the presence of extreme observations that would provide a suitable estimate of population variance under a double stratified sampling scheme.

Motivated by the abovementioned developments, this study aims to propose two new estimators to estimate the population variance by more meticulous applications of an auxiliary variable. The objective is achieved using the L-Moments characteristics such as L-scale, L-location, L-skewness, and L-kurtosis of auxiliary variable. The applicability of the proposition is further elaborated in double stratified random sampling scheme based on the dataset about Covid-19 taken from four continents. In this regard, a comparative study of the traditional unbiased variance estimator is conducted using numerical simulations. The simulation evaluation reveals the superior performance of the proposed estimators compared to the others.

The rest of the study is organized as follows. Section 2 presents the preliminaries with reference to L-Moments along with the proposed estimators. Section 3 highlights the simulation-based performance evaluation. Finally, Section 4 concludes the study.

2. L-Moments and proposed estimators

2.1. Extreme events and L-Moments

Human beings may witness a variety of intense events throughout their lives. For instance, the ongoing Covid-19 pandemic is still an obvious example of such events that have repercussions on human culture in the long run. In this regard, taking control over the negative effects of such events becomes a necessity to have a better estimate of the population parameters such as mean, variance, and quantiles. As mentioned earlier, the current study was conducted based on variance estimation, which is one of the most important population parameters. To the best of our knowledge, different variance estimators were developed based on the traditional moments, which are usually affected by extreme values. An elective procedure with a specific ability to settle this issue is L-Moments that are exceptionally affected by extreme values, compared to the traditional moments [21].

L-Moments are based on the linear combination of order statistics. It should be noted that "L" in L-Moments represents their linearity. Hence, these moments are free from higher powers and known as linear moments. This is also one of the major differences between the traditional (nonlinear) moments and L-Moments. For the auxiliary variable X, the hth stratum, one may define the population L-Moments as [21]:

$$L_{1x_\ell} = E(X_{1:1}),$$

$$L_{2x_\ell} = \frac{E(X_{2:2} - X_{1:2})}{2},$$

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

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

The sample L-Moments can be written by the equations shown in Box I, where $x_{h(k)}$ represents the kth order statistics with binomial coefficient (:). Further, we can write the mathematical expressions of L-Moments for the study variable by adapting the structure of the sample and population L-Moments related to the auxiliary variable. For a detailed study of the L-Moments see [22].

Some notations for the upcoming proposed work in light of L-Moments with respect to the hth stratum are given below:

- $\bar{X}_{h_\ell} = L_{1xh_\ell}$ and $\bar{x}_{h_\ell} = \hat{L}_{1xh_\ell}$ are the population and sample means (L-location) of the auxiliary variable, respectively, based on the L-Moments;
- $\bar{Y}_{h_\ell} = L_{1yh_\ell}$ and $\bar{y}_{h_\ell} = \hat{L}_{1yh_\ell}$ are the population and sample means (L-location) of the study variable, respectively, based on L-Moments;
- $S^2_{hx_\ell} = L^2_{2x_\ell}$, $s^2_{hx_\ell} = \hat{L}^2_{1x_\ell}$ are the population and sample variance (L-dispersion) of the auxiliary variable, respectively, based on L-Moments;
- $S^2_{hy_\ell} = L^2_{2y_\ell}$, $s^2_{hy_\ell} = \hat{L}^2_{1y_\ell}$ are the population and sample variance (L-dispersion) of the study variable, respectively, based on L-Moments.

2.2. Calibration approach and proposed variance estimators

Calibration is one of the general methods for parameter estimation based on which the original weights ϕ_h can be improved by minimizing chi-square value or any other suitable loss function. The improved weights are called the calibrated weights. However, the minimization of the loss function is based on some suitable calibration constraints. These constraints belong to the auxiliary variable. Deville and Srndal [23] initially developed the concept of calibration-based estimation of parameters. Tracy et al. [24] introduced the idea of calibration-based estimation in double stratified random sampling. Koyuncu [25] extended her idea by defining a new and unique constraint, i.e., combination of original and calibrated weights. Of note, descriptive statistics (mean, variance, etc.) used in these studies were based on the traditional moments. However, the L-Moments characteristics, which are substantially robust in the presence of extreme values, have not been effectively utilized yet. In this respect, the current study proposed L-Moments characteristics based on the calibration estimators of the population variance under the double stratified random sampling scheme as follows:

$$G_{st(j)} = \sum_{h=1}^{H''} \vartheta'_h s^2_{hy_\ell} \quad \text{for } j = 1, 2, \tag{2.1}$$

where ϑ'_h stands for the calibration weights in the chi-square loss function:

$$L(\vartheta'_h, \phi_h) = \sum_{h=1}^{H''} \frac{(\vartheta'_h - \phi_h)^2}{\phi_h \Delta_h}, \tag{2.2}$$

and subject to the following calibration constraints:

$$\sum_{h=1}^{H''} \vartheta'_h \bar{x}_{h_\ell} = \sum_{h=1}^{H''} \phi_h \bar{X}_{h_\ell}, \tag{2.3}$$

$$\sum_{h=1}^{H''} \vartheta'_h s^2_{hx_\ell} = \sum_{h=1}^{H''} \phi_h S^2_{hx_\ell}, \tag{2.4}$$

$$\hat{L}_{1x\ell} = \begin{pmatrix} n_h \\ 1 \end{pmatrix}^{-1} \sum_{k=1}^{n_h} x_{h(k)},$$

$$\hat{L}_{2x\ell} = \frac{1}{2}\begin{pmatrix} n_h \\ 2 \end{pmatrix}^{-1} \sum_{k=1}^{n_h} \left\{ \begin{pmatrix} k-1 \\ 1 \end{pmatrix} - \begin{pmatrix} n-k \\ 1 \end{pmatrix} \right\} x_{h(k)},$$

$$\hat{L}_{3x\ell} = \frac{1}{3}\begin{pmatrix} n_h \\ 3 \end{pmatrix}^{-1} \sum_{k=1}^{n_h} \left\{ \begin{pmatrix} k-1 \\ 2 \end{pmatrix} - 2\begin{pmatrix} k-1 \\ 1 \end{pmatrix}\begin{pmatrix} n-k \\ 1 \end{pmatrix} + \begin{pmatrix} n-k \\ 2 \end{pmatrix} \right\} x_{h(k)},$$

$$\hat{L}_{4x\ell} = \frac{1}{4}\begin{pmatrix} n_h \\ 4 \end{pmatrix}^{-1} \sum_{k=1}^{n_h} \left\{ \begin{pmatrix} k-1 \\ 3 \end{pmatrix} - 3\begin{pmatrix} k-1 \\ 2 \end{pmatrix}\begin{pmatrix} n-k \\ 1 \end{pmatrix} + 3\begin{pmatrix} k-1 \\ 1 \end{pmatrix}\begin{pmatrix} n-k \\ 2 \end{pmatrix} - \begin{pmatrix} n-k \\ 3 \end{pmatrix} \right\} x_{h(k)},$$

Box I

$$\sum_{h=1}^{H''} \vartheta'_h \hat{\tau}_{hx_{\ell(j)}} = \sum_{h=1}^{H''} \phi_h \tau_{hx_{\ell(j)}}, \tag{2.5}$$

where $\tau_{hx_{\ell(j)}}$ (for $j = 1, 2$) are the population L-skewness and L-kurtosis of the auxiliary variable X, respectively. Similarly, $\hat{\tau}_{hx_{\ell(j)}}$ are the sample L-skewness and L-kurtosis of the auxiliary variable X, respectively. In addition, Δ_h sands for the suitably chosen weights used for determining different forms of estimators (see Koyuncu [25]). The Lagrange function is given below:

$$\Omega = \sum_{h=1}^{H''} \frac{(\vartheta'_h - \phi_h)^2}{\phi_h \Delta_h} - 2\lambda'_1 \left(\sum_{h=1}^{H''} \vartheta'_h \bar{x}_{h_\ell} - \sum_{h=1}^{H''} \phi_h \bar{X}_{h_\ell} \right)$$
$$- 2\lambda'_2 \left(\sum_{h=1}^{H''} \vartheta'_h s^2_{hx_\ell} - \sum_{h=1}^{H''} \phi_h S^2_{hx_\ell} \right)$$
$$- 2\lambda'_3 \left(\sum_{h=1}^{H''} \vartheta'_h \hat{\tau}_{hx_{\ell(j)}} - \sum_{h=1}^{H''} \phi_h \tau_{hx_{\ell(j)}} \right). \tag{2.6}$$

Minimizing the chi-square loss function (Eq. (2.2)) subject to the calibration Constraints (Eqs. (2.3), (2.4), and (2.5)) determines the calibration weights for stratified sampling as follows:

$$\vartheta'_h = \phi_h + \phi_h \Delta_h (\lambda'_1 \bar{x}_{hx} + \lambda'_2 s^2_{hx_\ell} + \lambda'_3 \hat{\tau}_{hx_{\ell(j)}}). \tag{2.7}$$

By substituting Constraint (2.7) into Constraints (2.3) (2.4), and (2.5), the following system of equations can be derived:

$$\begin{bmatrix} P_{11} & P_{12} & P_{13} \\ P_{12} & P_{22} & P_{23} \\ P_{13} & P_{23} & P_{33} \end{bmatrix} \begin{bmatrix} \lambda'_1 \\ \lambda'_2 \\ \lambda'_3 \end{bmatrix} = \begin{bmatrix} P_{10} \\ P_{20} \\ P_{30} \end{bmatrix}. \tag{2.8}$$

Solving the system of equations in Eq. (2.8) for λ'_s yields the equations shown in Box II, where:

$$P_{11} = \sum_{h=1}^{H''} \phi_h \Delta_h \bar{x}^2_{h_\ell}, \quad P_{22} = \sum_{h=1}^{H''} \phi_h \Delta_h s^4_{hx_\ell},$$

$$P_{33} = \sum_{h=1}^{H''} \phi_h \Delta_h \hat{\tau}^2_{hx_{\ell(j)}}, \quad P_{12} = \sum_{h=1}^{H''} \phi_h \Delta_h \bar{x}_{h_\ell} s^2_{hx_\ell},$$

$$P_{13} = \sum_{h=1}^{H''} \phi_h \Delta_h \bar{x}_{h_\ell} \hat{\tau}_{hx_{\ell(j)}},$$

$$P_{23} = \sum_{h=1}^{H''} \phi_h \Delta_h s^2_{hx_\ell} \hat{\tau}_{hx_{\ell(j)}},$$

$$P_{10} = \sum_{h=1}^{H''} \phi_h (\bar{X}_{h_\ell} - \bar{x}_{h_\ell}), \quad P_{20} = \sum_{h=1}^{H''} \phi_h (S^2_{hx_\ell} - s^2_{hx_\ell}),$$

$$P_{30} = \sum_{h=1}^{H''} \phi_h (\tau_{hx_{\ell(j)}} - \hat{\tau}_{hx_{\ell(j)}}).$$

By substituting λ'_s into Eq. (2.7) and the resulting equation in Constraint (2.1) while setting $\Delta_h = 1$, the proposed estimator for population variance can be derived, as shown below:

$$G_{st(j)} = \sum_{h=1}^{H''} \phi_h s^2_{hy_\ell} + D_{1h(\alpha)} P_{10} + D_{2h(\alpha)} P_{20}$$
$$+ D_{3h(\alpha)} P_{30}, \tag{2.9}$$

$D_{1h(\alpha)}$, $D_{2h(\alpha)}$, and $D_{3h(\alpha)}$ are calculated by the equations shown in Box III and:

$$T_{11} = \sum_{h=1}^{H''} \phi_h \bar{x}^2_{h_\ell}, \quad T_{22} = \sum_{h=1}^{H''} \phi_h s^4_{hx_\ell},$$

$$T_{33} = \sum_{h=1}^{H''} \phi_h \hat{\tau}^2_{hx_{\ell(j)}}, \quad T_{12} = \sum_{h=1}^{H''} \phi_h \bar{x}_{h_\ell} s^2_{hx_\ell},$$

$$\lambda'_1 = \frac{(P_{13}P_{23} - P_{12}P_{33})(P_{12}P_{20} - P_{22}P_{10}) - (P_{13}P_{22} - P_{12}P_{23})(P_{12}P_{30} - P_{23}P_{10})}{(P^2_{12} - P_{11}P_{22})(P_{13}P_{23} - P_{12}P_{33}) - (P_{13}P_{22} - P_{12}P_{23})(P_{12}P_{13} - P_{11}P_{23})},$$

$$\lambda'_2 = \frac{(P_{13}P_{23} - P_{12}P_{33})(P_{12}P_{10} - P_{11}P_{20}) - (P_{12}P_{13} - P_{23}P_{11})(P_{13}P_{20} - P_{12}P_{30})}{(P^2_{12} - P_{11}P_{22})(P_{13}P_{23} - P_{12}P_{33}) - (P_{13}P_{22} - P_{12}P_{23})(P_{12}P_{13} - P_{11}P_{23})},$$

$$\lambda'_3 = \frac{(P^2_{12} - P_{11}P_{22})(P_{13}P_{20} - P_{12}P_{30}) - (P_{13}P_{22} - P_{12}P_{23})(P_{12}P_{10} - P_{11}P_{20})}{(P^2_{12} - P_{11}P_{22})(P_{13}P_{23} - P_{12}P_{33}) - (P_{13}P_{22} - P_{12}P_{23})(P_{12}P_{13} - P_{11}P_{23})}.$$

Box II

$$D_{1h(\alpha)}=\frac{T_{12}[T_{14}(T_{22}T_{33}-T_{23}^2)+T_{24}(T_{13}T_{23}-T_{12}T_{23})+T_{34}(T_{12}T_{23}-T_{13}T_{22})]}{(T_{12}^2-T_{11}T_{22})(T_{13}T_{23}-T_{12}^2)-(T_{13}T_{22}-T_{12}T_{23})(T_{12}T_{13}-T_{11}T_{23})},$$

$$D_{2h(\alpha)}=\frac{T_{12}[T_{14}(T_{13}T_{23}-T_{12}T_{33})+T_{24}(T_{11}T_{33}-T_{13}^2)+T_{34}(T_{12}T_{13}-T_{11}T_{23})]}{(T_{12}^2-T_{11}T_{22})(T_{13}T_{23}-T_{12}^2)-(T_{13}T_{22}-T_{12}T_{23})(T_{12}T_{13}-T_{11}T_{23})},$$

$$D_{3h(\alpha)}=\frac{T_{12}[T_{14}(T_{13}T_{22}-T_{12}T_{23})+T_{24}(T_{12}T_{13}-T_{11}T_{23})+T_{34}(T_{11}T_{22}-T_{12}^2)]}{(T_{12}^2-T_{11}T_{22})(T_{13}T_{23}-T_{12}^2)-(T_{13}T_{22}-T_{12}T_{23})(T_{12}T_{13}-T_{11}T_{23})}.$$

Box III

$$T_{13}=\sum_{h=1}^{H''}\phi_h\bar{x}_{h_\ell}\hat{\tau}_{hx_{\ell(j)}},$$

$$T_{14}=\sum_{h=1}^{H''}\phi_h\bar{x}_{h_\ell}s^2_{hy_\ell},$$

$$T_{23}=\sum_{h=1}^{H''}\phi_h s^2_{hx_\ell}\hat{\tau}_{hx_{\ell(j)}},\quad T_{24}=\sum_{h=1}^{H''}\phi_h s^2_{hx_\ell}s^2_{hy_\ell},$$

$$T_{34}=\sum_{h=1}^{H''}\phi_h\hat{\tau}_{hx_{\ell(j)}}s^2_{hy_\ell}.$$

3. Simulation study

The simulation study used the available data on Covid-19 pandemic, considering the total number of recoveries as the study variable Y and total number of cases as the auxiliary variable X in four continents namely Africa, Asia, Europe, and North America from January 22, 2020 to August 23, 2020 (Source: https://www.worldometers.info/coronavirus).

Each continent represents a stratum. There are 49 countries in Asia, 57 countries in Africa, 48 countries in Europe, and 39 countries in North America. The number of countries shows the size of each stratum. The scatter plot (X,Y) for each continent is provided in Figures 1–4 according to which Covid-19 pandemic data confirm the presence of the extreme values issue. In this respect, as already mentioned in Section 2, the proposed L-Moments-based variance estimators can be suitable candidates in this regard. The design of the sampling 0 is formed through random selection of a large first-phase sample n'_h from each continent. Note that the size of n'_h is 60% for each hth stratum. Followed by the selection of preliminary sample n'_h from each continent, we select 1000 times the second-phase sample whose sizes are denoted by n_h. The size of n_h is 25% for each hth stratum. The detailed characteristics of the data are listed in Table 1. Of note, the traditional correlation coefficient between (X,Y) is also provided in Table 1, where the strength of the relationship justifies the application of the total number of cases as an auxiliary variable X while estimating the study variable, i.e., total number of recoveries Y. The empirical mean square error and percentage relative efficiency results are calculated via the following formulae:

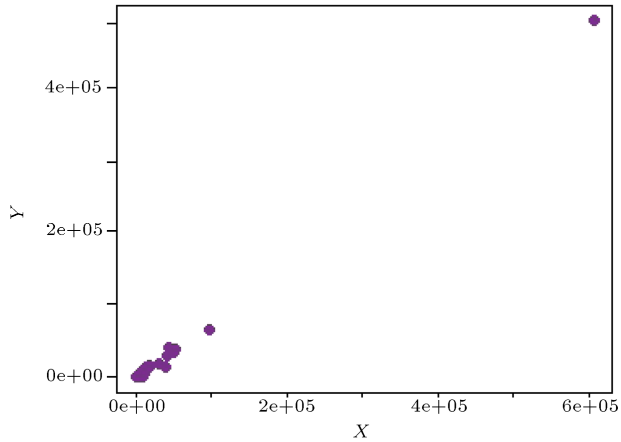

Figure 1. Scatter plot for the first stratum.

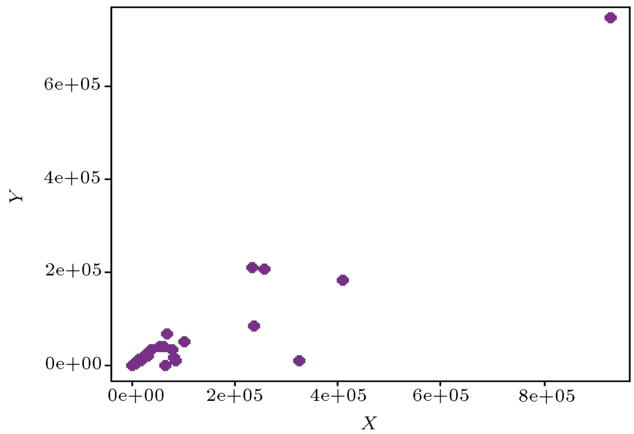

Figure 2. Scatter plot for the third stratum.

Table 1. L-Moments characteristics of Covid-19 data set.

Stratum-I	Stratum-II	Stratum-III	Stratum-IV
$N_1 = 57$	$N_2 = 49$	$N_3 = 48$	$N_4 = 39$
$n_1^{'} = 34$	$n_2^{'} = 29$	$n_3^{'} = 29$	$n_4^{'} = 23$
$n_1 = 9$	$n_2 = 7$	$n_3 = 7$	$n_4 = 6$
$\rho_{xy_1} = 0.9985$	$\rho_{xy_2} = 0.9987$	$\rho_{xy_3} = 0.9352$	$\rho_{xy_4} = 0.9995$
$\bar{X}_{1_\ell} = 20760.74$	$\bar{X}_{2_\ell} = 127006.30$	$\bar{X}_{3_\ell} = 69487.58$	$\bar{X}_{4_\ell} = 176996.7$
$\bar{Y}_{1_\ell} = 15840.37$	$\bar{Y}_{2_\ell} = 99818.78$	$\bar{Y}_{3_\ell} = 41214.02$	$\bar{Y}_{4_\ell} = 99015.46$
$S_{1x_\ell} = 17345.22$	$S_{2x_\ell} = 104999.5$	$S_{3x_\ell} = 55359.53$	$S_{4x_\ell} = 173204.70$
$S_{1y_\ell} = 13624.12$	$S_{2y_\ell} = 81958.28$	$S_{3y_\ell} = 33999.61$	$S_{4y_\ell} = 96747.00$
$\tau_{1x_{\ell(1)}} = 0.81740$	$\tau_{1x_{\ell(2)}} = 0.75474$	$\tau_{1x_{\ell(3)}} = 0.69328$	$\tau_{1x_{\ell(4)}} = 0.96329$
$\tau_{2x_{\ell(1)}} = 0.70339$	$\tau_{2x_{\ell(2)}} = 0.60642$	$\tau_{2x_{\ell(3)}} = 0.47242$	$\tau_{2x_{\ell(4)}} = 0.92148$

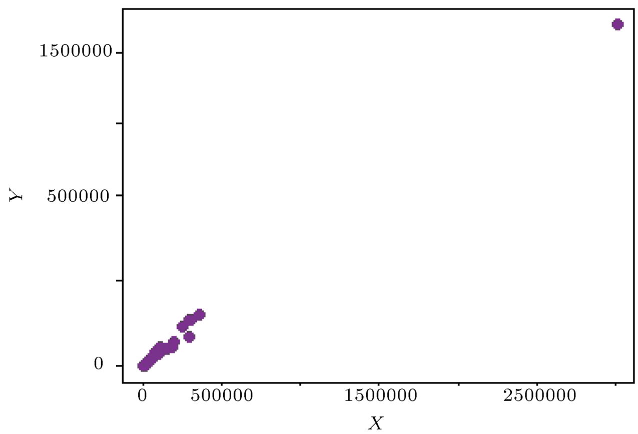

Figure 3. Scatter plot for the second stratum.

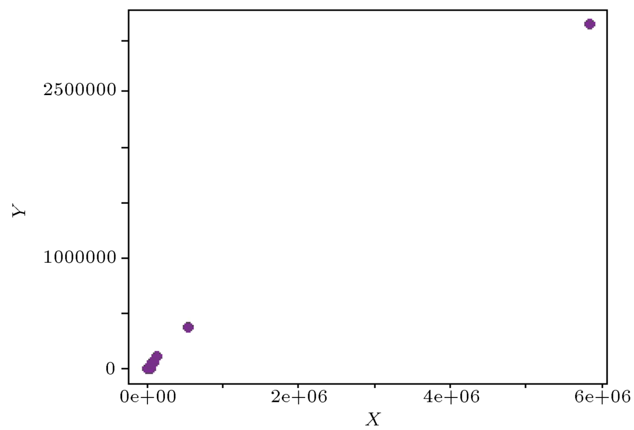

Figure 4. Scatter plot for the fourth stratum.

$$MSE(G_{st(j)}) = \frac{1}{\binom{N}{n}} \sum_{i=1}^{\binom{N}{n}} (G_{st(j)} - \sigma_y^2)^2,$$

$$PRE(G_{st(j)}) = \frac{MSE(T_o)}{MSE(G_{st(j)})} \times 100.$$

Table 2. PREs of estimators

T_o	$G_{st(1)}$	$G_{st(2)}$
100	8471.74	8469.912

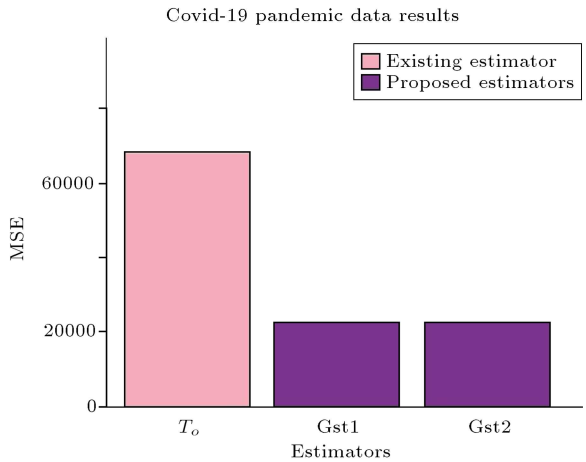

Figure 5. MSE of estimators.

It appears that in Table 2, the PRE values associated with the proposed estimators are greater than 100, meaning that the proposed estimators $(G_{st(1)}, G_{st(2)})$ actually outperform the traditional estimator T_o. Figure 5 confirms this finding mainly because the highest values of the MSE are clearly associated with the traditional estimator and consequently, its performance is quite poor, compared to that of the proposed estimators when the data includes extreme values.

4. Conclusion

The current research proposed a new way of estimating population variance using L-Moments and calibration approach under double stratified random sampling. To this end, a simulation study was carried out using the Covid-19 pandemic dataset as a real-life application of the proposed estimators. Table 2 presents the simulation-based percentage relative efficiency results, and Figure 5 graphically shows the mean square error results. The obtained results confirmed that the proposed estimators, compared to traditional unbiased variance estimator, were characterized by high efficiency with small mean square error under the double stratified sampling scheme. Hence, it is recommended that the proposed estimators be used in the presence of extreme observations. It should be noted that some other estimators could be derived in the forthcoming studies by adding the suitable calibration constraints based on L-Moments characteristics of the auxiliary information, such as L-Moments-based coefficient of variation or skewness of the auxiliary variable, to the proposed estimators, as in the studies of Shahzad et al. [8].

Acknowledgment

The authors would like to express their gratitude to the Editor-In-Chief, Professor S.T.A. Niaki and three anonymous referees for constructive comments which helped improve the previous version of this paper. 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.1/189/41. Ibrahim Mufrah Almanjahie received the grant.

References

1. Koyuncu, N. "Efficient estimators of population mean using auxiliary attributes", *Applied Mathematics and Computation*, **218**, pp. 10900–10905 (2012).
2. Al-Omari, A.I. "Ratio estimation of the population mean using auxiliary information in simple random sampling and median ranked set sampling", *Statistics and Probability Letters*, **82**(11), pp. 1883–1890 (2012).
3. Zaman, T. "Improvement of modified ratio estimators using robust regression methods", *Applied Mathematics and Computation*, **348**, pp. 627–631 (2019).
4. Zaman, T. "An efficient exponential estimator of the mean under stratified random sampling", *Mathematical Population Studies*, **28**, pp. 104–121 (2021).
5. Naz, F., Abid, M., Nawaz, T., et al. "Enhancing efficiency of ratio-type estimators of population variance by a combination of information on robust location measures", *Scientia Iranica*, **27**(4), pp. 2040–2056 (2020a).
6. Naz, F., Nawaz, T., Pang, T., et al. "Use of nonconventional dispersion measures to improve the efficiency of ratio-type estimators of variance in the presence of outliers", *Symmetry*, **12**(16), pp. 1–26 (2020b).
7. Shahzad, U., Hanif, M., Sajjad, I., et al. "Quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information", *Scientia Iranica*, **29**, pp. 1705–1715 (2022).
8. Shahzad, U., Ahmad, I., Almanjahie, I., et al. "A new class of L-Moments based calibration variance Estimators", *Computers Materials and Continua*, **66**(3), pp. 3013–3028 (2021).
9. McIntyre, G.A. "A method of unbiased selective sampling using ranked sets", *Australian Journal of Agricultural Research*, **3**, pp. 358–390 (1952).
10. Adel Rastkhiz, S.E., MobiniDehkordi, A., Yadollahi Farsi, J., et al. "A new approach to evaluating entrepreneurial opportunities", *Journal of Small Business and Enterprise Development*, **26**, pp. 67–84 (2019).
11. Zamanzade, E. and Vock, M. "Variance estimation in ranked set sampling using a concomitant variable", *Statistics and Probability Letters*, **105**, pp. 1–5 (2015).
12. Zamanzade, E. and Wang, X. "Estimation of population proportion for judgment post-stratification", *Computational Statistics and Data Analysis*, **112**, pp. 257–269 (2017).
13. Zamanzade, E. and Mahdizadeh, M. "A more efficient proportion estimator in ranked set sampling", *Statistics and Probability Letters*, **129**, pp. 28–33 (2017).
14. Zamanzade, E. and Wang, X. "Proportion estimation in ranked set sampling in the presence of tie information", *Computational Statistics*, **33**(3), pp. 1349–1366 (2018).
15. Mahdizadeh, M. and Zamanzade, E. "Efficient body fat estimation using multistage pair ranked set sampling", *Statistical Methods in Medical Research*, **28**(1), pp. 223–234 (2019).
16. Shahzad, U., Ahmad, I., Oral, E., et al. "Estimation of the population mean by successive use of an auxiliary variable in median ranked set sampling", *Mathematical Population Studies*, **28**, pp. 176–199 (2021).
17. Zaman, T. and Bulut, H. "Modified ratio estimators using robust regression methods", *Communications in Statistics - Theory and Methods*, **48**(8), pp. 2039–2048 (2019).
18. Zaman, T. and Bulut, H. "Modified regression estimators using robust regression methods and covariance matrices in stratified random sampling", *Communications in Statistics - Theory and Methods*, **49(14)**, pp. 3407–3420 (2020).
19. Ali, N., Ahmad, I., Hanif, M., et al. "Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information", *Communications in Statistics - Theory and Methods*, **50**, pp. 979–992 (2021).
20. Abid, M., Ahmed, S., Tahir, M., et al. "Improved ratio estimators of variance based on robust measures", *Scientia Iranica*, **26**(4), pp. 2848–2494 (2019).
21. Hosking, J.R. "L-moments: Analysis and estimation of distributions using linear combinations of order statistics", *Journal of the Royal Statistical Society: Series B (Methodological)*, **52**(1), pp. 105–124 (1990).

22. Hosking, J.R.M. and Wallis, J.R. "Regional frequency analysis: an approach based on L-moments", Cambridge University Press (2005).

23. Deville, J.C. and Srndal, C.E. "Calibration estimators in survey sampling", *Journal of the American statistical Association*, **87**(418), pp. 376–382 (1992).

24. Tracy, D.S., Singh, S., and Arnab, R. "Note on calibration in stratified and double sampling", *Survey Methodology*, **29**, pp. 99–104 (2003).

25. Koyuncu, N. "Calibration estimator of population mean under stratified ranked set sampling design", *Communications in Statistics-Theory and Methods*, **47**(23), pp. 5845–5853 (2018).

Biographies

Usman Shahzad obtained his MSc degree in Statistics from International Islamic University, Islamabad, Pakistan. He completed his MPhil in Statistics from Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Pakistan. Currently, he is studying for his PhD in the same major from International Islamic University, Islamabad, Pakistan. He served as a lecturer at Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Pakistan. He has published more than 35 research papers in research journals. His research interests include survey sampling, extreme value theory, stochastic process, probability, and non-parametric statistics.

Ishfaq Ahmad was born in Pakistan on March 26th, 1981. He received his BSc degree in Statistics, Physics and Mathematics from Bahudin Zakaria University, Multan Pakistan in 1999. Afterward, he received MSc and MPhil degrees in Statistics from Quaid-e-Azam University, Islamabad, Pakistan in 2003 and 2005, respectively. Finally, he received a PhD degree in Probability and Mathematical Statistics from the Institute of Applied Mathematics, University of Chinese Academy of Science (UCAS) Beijing, China in 2010. Currently, He is working as an Associate Professor of Statistics at the Department of Mathematics and Statistics, Faculty of Basic and Applied Sciences, International Islamic University, Islamabad, Pakistan. Before this, he was working as an Assistant Professor at King Khalid University, Saudi Arabia. His main areas of research interest are extreme value theory, statistical inference, survey sampling, optimization, Bayesian analysis, and functional analysis. He has published more than 47 articles in journals of international repute, such as the International Journal of Climatology, Scientific reports, and many others.

Ibrahim Mufrah Almanjahie was born in Saudi Arabia in 1979. He received his BSc degree in Mathematics from King Khalid University, Saudi Arabia in 2002. He received the MSc degree in Mathematical and Statistical Sciences and PhD in Probability and Statistical Modeling from University of Western Australia, Australia in 2008 and 2015, respectively. Currently, He is an Associate Professor at the Department of Mathematics, College of Sciences king Khalid University. He is also the president of the Saudi Association for Statistical Sciences. His main areas of research interest are modeling and analysis of ion channel data, hidden Markov models, M algorithm, finite mixture models MCMC, computational methods in statistics, time series modeling, applied statistics, sampling and functional statistics.

Muhammad Hanif obtained his MSc degree in Statistics from the University of Agriculture, Faisalabad, Pakistan. He completed his MPhil in statistics from Government College University, Lahore, Pakistan. He completed his PhD degree in Statistics from Zhejiang University, Hangzhou, China. Currently, he is an Associate Professor and serving as a Chairman of the Department of Mathematics and Statistics at Pir Mehr Ali Shah Arid Agriculture University Rawalpindi, Pakistan. He has published more than 50 research papers in research journals. His main research interests include stochastic process, probability and mathematical statistics, survey sampling, and non-parametric estimation.

Nadia H. Al-Noor completed her PhD degree in Mathematical Statistics. Currently, she is serving as a Professor of Mathematical Statistics at the Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, Iraq. She has published more than 40 research papers in research journals. She supervised more than 10 postgraduate students. Her main research interest includes probability and mathematical statistics, survey sampling, statistical inference, reliability theory, robust regression, and non-parametric estimation methods.