Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 5866085, 9 pages
https://doi.org/10.1155/2022/5866085

Research Article

Efficient Estimators of Finite Population Mean Based on Extreme Values in Simple Random Sampling

Anum Iftikhar,[1] Hongbo Shi,[1] Saddam Hussain ,[2] Mohsin Abbas,[3] and Kalim Ullah[4]

[1]*School of Statistics, Shanxi University of Finance and Economics, Taiyuan, China*
[2]*Department of Mathematics and Statistics, Institute of Southern Punjab, Multan, Pakistan*
[3]*Department of Physiotherapy, KAIMS International Institute, Multan, Pakistan*
[4]*Foundation University Medical College, Foundation University Islamabad, DHA-I, Islamabad 44000, Pakistan*

Correspondence should be addressed to Saddam Hussain; saddamhussain.stat885@gmail.com

Received 25 January 2022; Accepted 5 March 2022; Published 11 May 2022

Academic Editor: Tahir Mehmood

Copyright © 2022 Anum Iftikhar et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The use of extreme values of the auxiliary variable is sometimes more beneficial to get the high efficiency of the estimators, and the study variable can have a correlation with the rank of the decently correlated auxiliary variable. As a result, it can be regarded as additional data for the study variable that can be used to improve the estimators' efficiency. When the knowledge of the minimum and maximum values, as well as the rankings of the auxiliary variable, is known, various better estimators for calculating the finite population mean of the research variable based on extreme values under simple random sampling are proposed in this paper. The suggested estimators' bias and mean squared error expressions are derived using first-order approximation. The recommended estimators have been compared mathematically to the current estimators. The suggested estimators are more exact in terms of relative efficiency than the other estimators addressed here, as shown by simulation and real datasets used to demonstrate the estimation of a limited population mean based on extreme values.

1. Introduction

The purpose of survey sampling is to utilize the maximum amount of information about the characteristics of interest. Many fields of study require estimation of the finite population mean for a variable of interest. For example, average wheat production per acre, average income of households, mean weight of meat producing animals, etc. The mean per unit estimator is a base line estimator to estimate the finite population mean.

Unusual observations can occur in sample survey data. The mean estimator is sensitive to very large and/or small values if included in the sample. It can provide biased results and, ultimately, tempted to delete the sample data. However, generally, when there are extreme values in the data, the efficiency of classical estimators declines. (for more details, see [1] and the reference cited therein).

The use of supplementary information to enhance the precision of an estimator is a typical strategy in survey sampling. To improve their relative efficiency, the ratio, regression, and product-type estimators all require supplemental information on one or more auxiliary variables in addition to the information on the study variable. For example, when estimating the total household income, the household members and total expenditure may be used as two auxiliary variables. A significant amount of research work has been done to develop new and improved estimators of the population parameters, which include the population mean, total, CDF, median, etc. (for more details, see [2–9] and the references cited therein). To the best of our knowledge, [9–13] have done some recent work on the estimation of finite population mean using auxiliary information. However, because classical estimators are sensitive to extreme values, the outlier problem, which is the presence of extreme values in data, reduces efficiency (see [1] and the reference cited therein).

When there are extreme values in the data, the efficiency of a classical ratio/product-type estimator declines in terms of relative efficiency (RE). Similarly, the regression-type estimators do not perform well in the presence of outliers as it is a well-known phenomenon that ordinary least square (OLS) estimators are sensitive in the presence of outliers. However, extreme values, if known, can be retained in the data and used as auxiliary information to increase the precision of the estimate (for more details, see [6, 9, 14–17] and the references cited therein, to name a few).

In this study, we used extreme values of the auxiliary variable as auxiliary information and retained it in the data and suggested improved ratio/product-type estimators. On the lines of [17, 18] and 20 in simple random sampling (SRS), we introduced an improved class of estimators for predicting the finite population mean based on extreme values, using the lowest and maximum values of the auxiliary variable as auxiliary information.

The rest of the study is as follows: in Section 2, the methodology and notation of the study are described. In Section 3, existing estimators are discussed. In Section 4, we briefly discussed our proposed estimators. Section 5 and Section 6 contain the mathematical and numerical comparison. Finally, in Section 7, we discussed the main findings and concluded the study.

2. Methodology and Notation

Let us consider a finite population $\delta = (\delta_1, \delta_2, \ldots, \delta_N)$ of size N. The values of the study variable Y and the auxiliary variable X, respectively, are yi and xi. Let ri be the values of the auxiliary variable R's corresponding rankings for the i th $(i = 1, 2, 3, \ldots, N)$ units. We use SRS without replacement to choose a sample of size n units from the population δ. Let $\overline{Y} = \sum_{i=1}^{N} y_i/N$, $\overline{X} = \sum_{i=1}^{N} x_i/N$, and $\overline{R} = \sum_{i=1}^{N} R_i/N$ be the population mean of study variable, auxiliary variable and the ranks of the auxiliary variables, respectively. It is further assumed that $S_y^2 = \sum_{i=1}^{N} (y_i - \overline{Y})^2/N - 1$, $S_x^2 = \sum_{i=1}^{N} (x_i - \overline{X})^2/N - 1$, and $S_r^2 = \sum_{i=1}^{N} (r_i - \overline{R})^2/N - 1$ be the corresponding population variances of Y, X and the ranks of the auxiliary variable R, respectively.

Let $C_y = S_y/\overline{Y}$, $C_x = S_x/\overline{X}$, and $C_r = S_r/\overline{R}$ be the population coefficients of variation of the study variable, auxiliary variable, and the ranks of the auxiliary variable, respectively. ρ_{yx}, ρ_{yr}, and ρ_{xr} are the population correlation coefficients between the subscripts.

Let $\overline{y} = \sum_{i=1}^{n} y_i/n$, $\dot{x} = \sum_{i=1}^{n} x_i/n$, and $\overline{r} = \sum_{i=1}^{n} r_i/n$ be the sample means and $\widehat{S}_y^2 = \sum_{i=1}^{n} (y_i - \overline{y})^2/n - 1$, $\widehat{S}_x^2 = \sum_{i=1}^{n} (x_i - \overline{x})^2/n - 1$, and $\widehat{S}_r^2 = \sum_{i=1}^{n} (r_i - \overline{r})^2/n - 1$ be the sample variance of the study variable, auxiliary variable, and the ranks of the auxiliary variable, respectively.

We may use the following relative error terms to determine the biases and MSEs of the existing and proposed class of $\overline{Y}$ estimators. Let

$$\xi_0 = \frac{\overline{y} - \overline{Y}}{\overline{Y}},\quad \xi_1 = \frac{\overline{x} - \overline{X}}{\overline{X}},\quad \xi_2 = \frac{\overline{r} - \overline{R}}{\overline{R}}, \tag{1}$$

such that $E(\xi_0) = E(\xi_1) = E(\xi_2) = 0$.

3. Existing Estimator

In this section, we define the existing estimators of finite population means, which are to be compared with our proposed estimator.

3.1. Usual Unbiased Estimator. The unbiased estimator of a finite population mean with variance that is most commonly used is

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{n} y_i, \tag{2}$$

$$\mathrm{Var}(\overline{y}) = \theta\overline{Y}^2 C_y^2, \tag{3}$$

respectively.

3.2. Cochran's Ratio Estimator. Cochran [2] recommended a ratio type estimator by first estimating the finite population mean in SRS, which is obtained by employing auxiliary information.

$$\overline{y}_R = \overline{y}\left(\frac{\overline{X}}{\overline{x}}\right). \tag{4}$$

Mathematical expression upto the first-order of approximation for the bias and MSE of $\overline{y}_R$ is given by

$$\mathrm{Bias}(\overline{y}_R) \approx \theta\overline{Y}\left(C_x^2 - \rho_{yx}C_yC_x\right), \tag{5}$$

$$\mathrm{MSE}(\overline{y}_R) \approx \theta\overline{Y}^2\left(C_y^2 + C_x^2 - 2\rho_{yx}C_yC_x\right), \tag{6}$$

respectively.

3.3. Classical Regression Estimator. The classical regression estimator for $\overline{Y}$ under SRS is given by

$$\overline{y}_{lr} = \overline{y} - b_{yx}(\overline{x} - \overline{X}), \tag{7}$$

where $b_{yx} = S_{yx}/S_x^2$ is the regression coefficient between y_i and x_i for $i = 1, 2, \ldots, n$. The MSE of $\overline{y}_{lr}$ upto the first order of approximation is given as under

$$\mathrm{MSE}(\overline{y}_{lr}) \approx \theta S_y^2\left(1 - \rho_{yx}^2\right). \tag{8}$$

3.4. Mohanty and Sahoo Estimator. On the similar lines of [17, 18], as auxiliary information, we gave two finite population mean estimators based on the minimum and maximum values of the auxiliary variable as follows:

$$\overline{y}_{t1} = \overline{y}\left(\frac{\overline{V}}{\overline{v}}\right),$$
$$\overline{y}_{t2} = \overline{y}\left(\frac{\overline{W}}{\overline{w}}\right), \tag{9}$$

respectively. Where $v = (x_i + x_M/x_M + x_m)$ and $w = (x_i + x_M/x_M + x_m)$. Here x_m and x_M are the minimum and maximum values, respectively.

The expression for bias and MSE of $\overline{y}_{t1}$ and $\overline{y}_{t2}$ are given by

$$\text{Bias}(\overline{y}_{t1}) \approx \theta\overline{Y}\left(t_1^2 C_x^2 - t_1 C_{yx}\right), \tag{10}$$

$$\text{MSE}(\overline{y}_{t1}) \approx \theta\overline{Y}^2\left(C_y^2 - 2t_1\rho_{yx}C_yC_x + t_1^2C_x^2\right), \tag{11}$$

$$\text{Bias}(\overline{y}_{t2}) \approx \theta\overline{Y}\left(t_2^2 C_x^2 - t_2 C_{yx}\right), \tag{12}$$

$$\text{MSE}(\overline{y}_{t2}) \approx \theta\overline{Y}^2\left(C_y^2 - 2t_2\rho_{yx}C_yC_x + t_2^2C_x^2\right), \tag{13}$$

respectively. Where $t_1 = \overline{X}/(\overline{X} + X_M)$ and $t_2 = \overline{X}/(\overline{X} + X_M)$

3.5. Walia et al. Estimator. Walia et al. [9] presented some estimators based on known knowledge about the auxiliary variable's minimum and maximum values are provided. The following is the transformation:

$$Z = x + \left(\frac{X_M}{X_m}\right). \tag{14}$$

The estimators of the finite population mean are listed as follows $\overline{Y}$:

$$\overline{y}_{hm1} = \overline{y}\left(\frac{\overline{Z}}{\overline{z}}\right),$$
$$\overline{y}_{hm2} = \overline{y}\left(\frac{\overline{Z} + C_z}{\overline{z} + C_z}\right), \tag{15}$$

respectively. Where $C_z = S_z/\overline{Z} = S_x/(\overline{X} + (X_M + X_m))$ and $S_x^2 = S_z^2$.

The bias and MSE of the above-modified estimators $\overline{y}_{hm1}$ and $\overline{y}_{hm2}$ are calculated as follows:

$$\text{Bias}(\overline{y}_{hm1}) \approx \theta\overline{Y}\left(C_1^2C_x^2 - C_1\rho_{yx}C_yC_x\right), \tag{16}$$

$$\text{MSE}(\overline{y}_{hm1}) \approx \theta\overline{Y}^2\left(C_y^2 - 2C_1\rho_{yx}C_yC_x + C_1^2C_x^2\right), \tag{17}$$

$$\text{Bias}(\overline{y}_{hm2}) \approx \theta\overline{Y}\left(C_2^2C_x^2 - C_2\rho_{yx}C_yC_x\right), \tag{18}$$

$$\text{MSE}(\overline{y}_{hm2}) \approx \theta\overline{Y}^2\left(C_y^2 - 2C_2\rho_{yx}C_yC_x + C_2^2C_x^2\right), \tag{19}$$

respectively. Where $C_1 = \overline{X}/(\overline{X} + (X_M)/X_m)$ and $C_2 = \overline{X}(\overline{X} + (X_M/X_m)/((\overline{X} + (X_M/X_m))^2 + S_x))$.

4. Proposed Estimator

In this section, we develop two auxiliary information-based (AIB) classes of estimators, say ratio and exponential ratio, under the SRS technique, for calculating the mean of a finite population $\overline{Y}$.

4.1. First Proposed Class of Estimator. We present a better class of estimators for estimating $\overline{Y}$ under SRS utilizing known information about the auxiliary variable X's lowest and maximum values, as motivated by [15]. The following is the improved class of estimator:

$$\overline{y}_{Di} = \left[k_1\overline{y}\left(\frac{\overline{X}}{\overline{x}}\right)^{\alpha_1} + k_2\overline{y}\left(\frac{\overline{X}}{\overline{x}}\right)^{\alpha_2}\right]\exp\left[\frac{(a\overline{U} + b_1) - (a\overline{u} + b_1)}{(a\overline{U} + b_1) + (a\overline{u} + b_1)}\right], \tag{20}$$

where k_1 and k_2 are unknown constants whose values must be determined in order to calculate the bias and MSE of the $\overline{y}_{Di}$ minimum and $\overline{u} = \overline{x} + (X_M - X_m/X_M + X_m)$. Further, $a = 1, b_1 = X_M - X_m, b_2 = R_M - R_m$, and α_1, α_2 be the scalar quantities that may assume $(0, -1, 1)$ values. In addition, the sub-cases of the $\overline{y}_{Di}$ are summarized in appendix (given in Table 1).

In order to derive approximate mathematical expressions for the bias and MSE of $\overline{y}_{Di}$, we can write $\overline{y} = \overline{Y}(1 + \xi_0)$ and $\overline{x} = \overline{X}(1 + \xi_1)$. Let us express the right-hand side (RHS) of (20) in terms of ξ's to get

$$\overline{y}_{Di} = \left[k_1\overline{Y}(1 + \xi_0)(1 + \xi_1)^{-\alpha_1} + k_2(1 + \xi_1)^{-\alpha_2}\right] \cdot \exp\left[\frac{-g_1\xi_1}{2}\left(1 + \frac{g_1\xi_1}{2}\right)^{-1}\right], \tag{21}$$

where $g_1 = a\overline{X}/(a\overline{X} + (X_M - X_m/X_M + X_m) + b_1)$. Let us expand the RHS of equation (21) and retain terms up to 2nd power of ξ's, we have

$$\overline{y}_{Di} - \overline{Y} \approx -\overline{Y} + k_1\overline{Y}\left[1 + \xi_0 - \xi_1\left(\alpha_1 + \frac{g_1}{2}\right) + \xi_1^2\left(\frac{\alpha_1 g_1}{2} + \frac{3g_1^2}{8} + \frac{\alpha_1(\alpha_1 + 1)}{2}\right) - \xi_0\xi_1\left(\alpha_1 + \frac{g_1}{2}\right)\right] + k_2\left[1 - \xi_1\left(\alpha_2 + \frac{g_1}{2}\right) + \xi_1^2\left(\frac{\alpha_2 g_1}{2} + \frac{3g_1^2}{8} + \frac{\alpha_2(\alpha_2 + 1)}{2}\right)\right]. \tag{22}$$

Table 1: Subcases of the proposed estimator I.

Estimator	α_1	α_2
$\overline{y}_{D1} = [k_1\overline{y} + k_2(\overline{x}/\overline{X})]L$	0	−1
$\overline{y}_{D2} = [k_1\overline{y}(\overline{x}/\overline{X}) + k_2]L$	−1	0
$\overline{y}_{D3} = [k_1\overline{y} + k_2(\overline{X}/\overline{x})]L$	0	1
$\overline{y}_{D4} = [k_1\overline{y}(\overline{X}/\overline{x}) + k_2]L$	1	0
$\overline{y}_{D5} = [k_1\overline{y}(\overline{X}/\overline{x}) + k_2(\overline{x}/\overline{X})]L$	1	−1
$\overline{y}_{D6} = [k_1\overline{y}(\overline{x}/\overline{X}) + k_2(\overline{X}/\overline{x})]L$	−1	1
$\overline{y}_{D7} = [k_1\overline{y}(\overline{X}/\overline{x}) + k_2(\overline{X}/\overline{x})]L$	1	1
$\overline{y}_{D8} = [k_1\overline{y}(\overline{X}/\overline{x}) + k_2(\overline{x}/\overline{X})]L$	−1	−1

Let us take expectation on both sides of equation (22), which is provided by, to get the bias of $\overline{y}_{Di}$ up to the first order of approximation

$$\mathrm{Bias}\left(\overline{y}_{Di}\right) \approx -\overline{Y} + k_1\overline{Y}D + k_2G, \tag{23}$$

where

$$D = \left[1 + \theta C_x^2\left(\frac{\alpha_1 g_1}{2} + \frac{3g_1^2}{8} + \frac{\alpha_1(\alpha_1+1)}{2}\right) - \theta C_{yx}\left(\alpha_1 + \frac{g_1}{2}\right)\right],$$
$$G = \left[1 + \theta C_x^2\left(\frac{\alpha_2 g_1}{2} + \frac{3g_1^2}{8} + \frac{\alpha_2(\alpha_2+1)}{2}\right)\right]. \tag{24}$$

Taking square on both sides of equation (22) and then taking its expectation to get the MSE of $\overline{y}_{Di}$ under first order of approximation, which is given by

$$\mathrm{MSE}\left(\overline{y}_{Di}\right) \approx \overline{Y}^2 + \overline{Y}^2k_1^2A + k_2^2B - 2\overline{Y}^2k_1D - 2\overline{Y}^2k_2G + 2\overline{Y}k_1k_2F, \tag{25}$$

where

$$A = \left[1 + \theta\left\{C_y^2 + C_x^2\left\{\left(\alpha_1 + \frac{g_1}{2}\right)^2 + \left(\alpha_1 g_1 + \frac{3g_1^2}{8} + \alpha_1(\alpha_1+1)\right)\right\} - 4C_{yx}\left(\alpha_1 + \frac{g_1}{2}\right)\right\}\right],$$
$$B = \left[1 + \theta C_x^2\left\{\left(\alpha_2 + \frac{g_1}{2}\right)^2 + \left(\alpha_2 g_1 + \frac{3g_1^2}{8} + \alpha_2(\alpha_2+1)\right)\right\}\right],$$
$$F = \left[\begin{array}{c} 1 + \theta\left\{C_x^2\left\{\left(\alpha_1 + \frac{g_1}{2}\right)\left(\alpha_2 + \frac{g_1}{2}\right) + \left(\alpha_1 g_1 + \frac{3g_1^2}{8} + \alpha_1(\alpha_1+1)\right) + \left(\alpha_2 g_1 + \frac{3g_1^2}{8} + \alpha_2(\alpha_2+1)\right)\right\}\right\} 1 + \theta\left\{-C_{yx}(\alpha_1 + \alpha_2 + g_1)\right\} \\ +\theta\left\{C_x^2\left\{\left(\alpha_1 + \frac{g_1}{2}\right)\left(\alpha_2 + \frac{g_1}{2}\right) + \left(\alpha_1 g_1 + \frac{3g_1^2}{8} + \alpha_1(\alpha_1+1)\right) + \left(\alpha_2 g_1 + \frac{3g_1^2}{8} + \alpha_2(\alpha_2+1)\right)\right\}\right\} \end{array}\right]. \tag{26}$$

By reducing equation (25) with regard to k_1 and k_2, the optimum values of k_1 and k_2 are determined

$$k_{1(opt)} = \frac{\mathrm{BD} - \mathrm{FG}}{\mathrm{AB} - F^2},$$
$$k_{2(opt)} = \frac{\overline{Y}(\mathrm{AG} - \mathrm{DF})}{\mathrm{AB} - F^2}. \tag{27}$$

Substituting the optimum values of k_1 and k_2 in equations (23) and (25), we get the minimum bias and MSE of $\overline{y}_{Di}$, respectively

$$\mathrm{Bias}\left(\overline{y}_{Di}\right)_{\min} \approx -\overline{Y}\left[1 - \frac{\left(\mathrm{AB}^2 + \mathrm{BD}^2 - 2\mathrm{DFG}\right)}{\mathrm{AB} - F^2}\right], \tag{28}$$

$$\mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \approx \overline{Y}^2\left[1 - \frac{\left(\mathrm{AB}^2 + \mathrm{BD}^2 - 2\mathrm{DFG}\right)}{\mathrm{AB} - F^2}\right]. \tag{29}$$

4.2. Second Proposed Estimator. On the similar lines of 6, we propose another improved class of exponential-type estimator for estimating $\overline{Y}$ using supplementary information in terms of minimum and the maximum values of X under SRS scheme. The improved estimators are given by

$$\overline{y}_{De} = \overline{y}\exp\left[k_3\left\{\frac{(a\overline{U}+b_1)-(a\overline{u}+b_1)}{(a\overline{U}+b_1)+(a\overline{u}+b_1)}\right\}\right]\exp\left[k_4\left\{\frac{(a\overline{R}+b_2)-(a\overline{r}+b_2)}{(a\overline{R}+b_2)+(a\overline{r}+b_2)}\right\}\right], \tag{30}$$

where k_3 and k_4 are unknown constants whose values must be chosen so that the biases and MSE of $\overline{y}_{Di}$ are as small as possible and where k_3 and k_4 are unknown constants whose values must be set so that the biases and MSE of $\overline{y}_{Di}$ are as little as feasible, and $\overline{u} = \overline{x} + (X_M - X_m/X_M + X_m)$ and $\overline{U} = \overline{X} + (X_M - X_m/X_M + X_m)$. Further, $a = 1$, $b_1 = X_M - X_m$, and $b_2 = R_M - R_m$ be the known values.

Let us express the RHS of equation (30) in terms of ξ's to acquire the following approximate mathematical equations for the bias and MSE of $\overline{y}_{De}$ as

$$\overline{y}_{De} = \overline{Y}(1+\xi_0)\exp\left[k_3\left\{\frac{-g_1\xi_1}{2}\left(1+\frac{g_1\xi_1}{2}\right)^{-1}\right\}\right]\exp\left[k_4\left\{\frac{-g_2\xi_2}{2}\left(1+\frac{g_2\xi_2}{2}\right)^{-1}\right\}\right], \tag{31}$$

where $g_2 = a\overline{R}/(a\overline{R} + (R_M - R_m/R_M + R_m) + b_2)$. Using the Taylor series' first-order approximation, we have

$$\overline{y}_{De} - \overline{Y} \approx \left[\xi_0 - \frac{k_3g_1\xi_1}{2} - \frac{k_4g_2\xi_2}{2} + \left(\frac{k_3g_1^2}{4} + \frac{k_3^2g_1^2}{8}\right)\xi_1^2 + \left(\frac{k_4g_2^2}{4} + \frac{k_4^2g_2^2}{8}\right)\xi_2^2\left(\frac{k_4g_2^2}{4} + \frac{k_4^2g_2^2}{8}\right)\xi_2^2 - \frac{k_3g_1\xi_0\xi_1}{2} - \frac{k_4g_2\xi_0\xi_2}{2} + \frac{k_3k_4g_1g_2\xi_1\xi_2}{4}\right]. \tag{32}$$

Simplifying and applying expectation on equation (32), we have the final expression of bias of $\overline{y}_{De}$, given by

$$\text{Bias}(\overline{y}_{De}) \approx \theta\overline{Y}\left[\left(\frac{k_3g_1^2}{4} + \frac{\alpha_1^2g_1^2}{8}\right)C_x^2 + \left(\frac{k_4g_2^2}{4} + \frac{\alpha_2^2g_2^2}{8}\right)C_r^2 - \frac{k_3g_1}{2}C_{yx} - \frac{k_4g_2}{2}C_{yr} + \frac{k_3k_4g_1g_2}{2}C_{xr}\right]. \tag{33}$$

By squaring and applying expectation on both sides of equation (32), we obtain the MSE up to first-order of approximation as

$$\text{MSE}(\overline{y}_{De}) \approx \theta\overline{Y}^2\left[C_y^2 + \frac{k_3g_1^2}{4}C_x^2 + \frac{k_4g_2^2}{4}C_r^2 - k_3g_1C_{yx} - k_4g_2C_{yr} + \frac{k_3k_4g_1g_2}{2}C_{xr}\right]. \tag{34}$$

The k_3 and k_4 optimum values are derived by minimizing the equation (34), respectively, given by

$$k_{3(opt)} = \frac{2C_y}{g_1C_x}\left[\frac{\rho_{yx} - \rho_{yr}\rho_{xr}}{1-\rho_{xr}^2}\right],$$
$$k_{4(\text{opt})} = \frac{2C_y}{g_2C_r}\left[\frac{\rho_{yr} - \rho_{yx}\rho_{xr}}{1-\rho_{xr}^2}\right]. \tag{35}$$

We get the minimal bias and MSE of $\overline{y}_{De}$ substituting the best values for k_3 and k_3 in equations (33) and (34), respectively.

$$\text{Bias}(\overline{y}_{De})_{\min} \approx \theta\overline{Y}C_y\left[\frac{g_1C_x(\rho_{yx} - \rho_{yr}\rho_{xr}) + g_2C_r(\rho_{yr} - \rho_{yx}\rho_{xr})}{1-\rho_{xr}^2} - R_{y,xr}^2\right], \tag{36}$$

where $R_{y,xr}^2 = (\rho_{yx}^2 + \rho_{xr}^2 - 2\rho_{yx}\rho_{xr}\rho_{yr})/(1-\rho_{yx}^2)$ is the coefficient of multiple determination of Y on X and R.

$$\mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \approx \theta \overline{Y}^2 C_y^2 \left(1 - R_{y,xr}^2\right). \tag{37}$$

5. Mathematical Comparison

We compared the proposed estimators mathematically to the existing estimator in Section 3 in this section.

5.1. First Proposed Estimator

Condition 1. From equations (2) and (28)

$$\begin{aligned} &\mathrm{Var}\left(\overline{y}\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if} \\ &\text{only } \theta C_y^2 + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{38}$$

Condition 2. From equations (5) and (28)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_R\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if,} \\ &\text{only } \theta\left(C_y^2 + C_x^2 - 2C_{xy}\right) + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{39}$$

Condition 3. From equations (7) and (28)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{lr}\right)_{\min} > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if,} \\ &\text{only } \theta C_y^2\left(1 - \rho_{yx}^2\right) + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{40}$$

Condition 4. From equations (11) and (28)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{t1}\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if} \\ &\text{only } \theta\left(C_y^2 + t_1^2 C_x^2 - 2t_1 C_{yx}\right) + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{41}$$

Condition 5. From equations (13) and (28)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{t2}\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if} \\ &\text{only } \theta\left(C_y^2 + t_2^2 C_x^2 - 2t_2 C_{yx}\right) + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{42}$$

Condition 6. From equations (18) and (28)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{hm1}\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if,} \\ &\text{only } \theta\left(C_y^2 + C_1^2 C_x^2 - 2C_1 C_{yx}\right) + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{43}$$

Condition 7. From equations (20) and (28)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{hm2}\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if,} \\ &\text{only } \theta\left(C_y^2 + C_2^2 C_x^2 - 2C_2 C_{yx}\right) + \left(\frac{\mathrm{BD}^2 + \mathrm{AG}^2 - 2\mathrm{DFG}}{\mathrm{AB} - F^2}\right) > 1. \end{aligned} \tag{44}$$

5.2. Second Proposed Estimator

Condition 8. From equations (2) and (36)

$$\begin{aligned} &\mathrm{Var}\left(\overline{y}\right) > \mathrm{MSE}\left(\overline{y}_{Di}\right)_{\min} \quad \text{if,} \\ &\text{only } R_{y,xr}^2 > 0. \end{aligned} \tag{45}$$

Condition 9. From equations (5) and (36)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_R\right) > \mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \quad \text{if,} \\ &\text{only } R_{y,xr}^2 + C_x^2 - 2C_{yx} > 1. \end{aligned} \tag{46}$$

Condition 10. From equations (7) and (36)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{lr}\right)_{\min} > \mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \quad \text{if,} \\ &\text{only } R_{y,xr}^2 - \rho_{yx}^2 > 0. \end{aligned} \tag{47}$$

Condition 11. From equations (11) and (36)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{t1}\right) > \mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \quad \text{if,} \\ &\text{only } C_y^2 R_{y,xr}^2 + t_1^2 C_x^2 - 2t_1 C_{yx} > 0. \end{aligned} \tag{48}$$

Condition 12. From equations (13) and (36)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{t2}\right) > \mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \quad \text{if,} \\ &\text{only } C_y^2 R_{y,xr}^2 + t_2^2 C_x^2 - 2t_2 C_{yx} > 0. \end{aligned} \tag{49}$$

Condition 13. From equations (18) and (36)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{hm1}\right) > \mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \quad \text{if,} \\ &\text{only } C_y^2 R_{y,xr}^2 + C_1^2 C_x^2 - 2\rho_{yx} C_1 C_y C_x > 0. \end{aligned} \tag{50}$$

Condition 14. From equations (20) and (36)

$$\begin{aligned} &\mathrm{MSE}\left(\overline{y}_{hm2}\right) > \mathrm{MSE}\left(\overline{y}_{De}\right)_{\min} \quad \text{if,} \\ &\text{only } C_y^2 R_{y,xr}^2 + C_2^2 C_x^2 - 2C_2 C_{yx} > 0. \end{aligned} \tag{51}$$

6. Numerical Comparison

In this section, simulated and real datasets are considered, and the percentage relative efficiencies (PREs) of the proposed estimator are computed.

6.1. Study of Simulation. We undertake simulation research using the notion from [19] to compare the performance of our recommended estimators to the comparable current estimators. We used the following distributions to construct six datasets of size $N\,(=10,000)$ for the auxiliary variable X:

$$\begin{aligned}&\text{Population 1: } X \sim \text{Exponential}(\lambda = 2),\\&\text{Population 2: } X \sim \text{Exponential}(\lambda = 5),\\&\text{Population 3: } X \sim \text{Uniform}(b_2 = 0, b_3 = 1),\\&\text{Population 4: } X \sim \text{Uniform}(b_2 = 0, b_3 = 2),\\&\text{Population 5: } X \sim \text{Gamma}(\alpha_3 = 1, \alpha_4 = 2),\\&\text{Population 6: } X \sim \text{Gamma}(\alpha_3 = 2, \alpha_4 = 4),\end{aligned} \tag{52}$$

and the study variable Y by

$$Y = r_{yx}^2 \times X + \varepsilon, \tag{53}$$

where r_{yx}^2 is the sample correlation coefficient between study and auxiliary variables, and ε is the random error term, which has $\mu = 0$ and $\sigma^2 = 1$ and follows a conventional normal distribution.

We considered the following steps in *R*Studio to get the results of mentioned estimators in this study:

Step 1: the first step was to create different populations (as an auxiliary variable) of 1000 units using the different distributions, and then Y is computed using the model given in equation (53).

Step 2: the unknown constants' optimal values for the suggested estimators are obtained using the datasets computed in **Step** 1.

Step 3: we use SRS without replacement to draw a sample of size $n\,(=250)$ and calculate Var/MSE for all the estimators covered in this research.

Step 4: under the same environment, the variances and MSE of the mean estimators are computed by drawing 50 thousand samples from each population under SRS given in **Step** 3. The variances/MSE of the proposed and existing estimators based on SRS are approximated by using the following formulae:

$$\begin{aligned}\operatorname{Var}(\overline{y}) &\approx \frac{1}{(\psi-1)}\sum_{t=1}^{\psi}\left(\overline{y}_t - \overline{Y}\right)^2,\\ \operatorname{MSE}(\overline{y}_{Di}) &\approx \frac{1}{(\psi-1)}\sum_{t=1}^{\psi}\left(\overline{y}_{Di,t} - \overline{Y}\right)^2,\\ \operatorname{MSE}(\overline{y}_{De}) &\approx \frac{1}{(\psi-1)}\sum_{t=1}^{\psi}\left(\overline{y}_{De,t} - \overline{Y}\right)^2,\end{aligned} \tag{54}$$

where $\psi = 50,000$ and $i = 1, 2, \ldots, 8$. On similar lines, the MSE of other estimators $\overline{y}_R$, $\overline{y}_{lr}$, $\overline{y}_{t1}$, etc., given in Section 3, are obtained. The PRE of $\operatorname{MSE}(\overline{y}_{Di})$ with respect to $\operatorname{Var}(\overline{y})$ is given by

$$\operatorname{PRE}\left(\operatorname{MSE}(\overline{y}_{Di}), \operatorname{Var}(\overline{y})\right) = \frac{\operatorname{Var}(\overline{y})}{\operatorname{MSE}(\overline{y}_{Di})} \times 100. \tag{55}$$

On similar lines, the PREs of the other estimator based on SRS may be computed. The REs of these proposed and existing estimators are reported in Table 2. It can be seen that the proposed estimators are more efficient than usual unbiased estimators and existing estimators as well in terms of PRE, i.e., all values of the PREs are greater than a hundred. The effect of increasing the number of sample size n is precluded. However, generally, with an increase in the sample size, the PREs tend to increase and vice versa.

6.2. Real-Life Data. We used three real datasets to compare the PREs of all these estimators to see how well they performed compared to the comparable existing estimators. These datasets' descriptions and summary statistics are listed as follows.

6.2.1. Population I. This dataset is taken from [20] page 226 and was conducted in Pakistan during the year 2012, which comprised 33 divisions. This dataset may be downloaded from the Pakistan Bureau of Statistics web page via the link: https://www.pbs.gov.pk/content/microdata. The study variable Y corresponds to the employment level by divisions in 2012 and the number of registered factories in 2012, respectively, while R corresponds to the rank number of registered factories in 2012. Here, our objective is to estimate the finite population mean under extreme values in SRS. The population constants are $N = 33$; $n = 10$; $\overline{Y} = 27.4909$; $\overline{X} = 72.5455$; $\overline{R} = 17$; $S_y = 10.1308$; $S_x = 10.5770$; $S_r = 9.638$; $X_M = 95$; $X_m = 58$; $R_M = 33$; $R_m = 1.5$; $C_y = 0.3685$; $C_x = 0.1459$; $C_r = 0.5669$; $g_1 = 0.6608$; $g_2 = 0.1151$; $\rho_{yx} = 0.2522$; $\rho_{yr} = 0.1950$; $\rho_{xr} = 0.9810$; and $\theta = 0.0697$

6.2.2. Population II. Another dataset is taken from [20] page 135, conducted in Pakistan during the year 2012, which comprised 33 divisions. This dataset may be downloaded from the Pakistan Bureau of Statistics web page via the link: https://www.pbs.gov.pk/content/microdata. The number of pupils enrolled in each division and the total number of government primary and secondary schools for boys and girls in each division are the research variables Y and X in 2012, respectively, and R correspond to the rank of auxiliary variable X. Here, our objective is to estimate the finite population mean under extreme values in SRS. The population constants are $N = 36$; $n = 10$; $\overline{Y} = 148718.7$; $\overline{X} = 1054.39$; $\overline{R} = 18.5$; $S_y = 182315.1$; $S_x = 402.6098$; $S_r = 10.535$; $X_M = 2370$; $X_m = 388$; $R_M = 36$; $R_m = 1$; $C_y = 1.2259$; $C_x = 0.3818$; $C_r = 0.5695$; $g_1 = 0.3472$; $g_2 = 0.3498$; $\rho_{yx} = 0.1799$; $\rho_{yr} = 0.188$; $\rho_{xr} = 0.9378$; and $\theta = 0.0722$

6.2.3. Population III. This dataset is taken from [2], which comprises 36 units of food cost and weekly income of families. The study variable Y and auxiliary variable X are the food cost of

Table 2: PREs of proposed class of estimators using populations 1–6.

Estimator	Population 1	Population 2	Population 3	Population 4	Population 5	Population 6
$\overline{y}$	100	100	100	100	100	100
$\overline{y}_R$	146.7758	101.3152	103.5414	129.8623	120.5604	107.7454
$\overline{y}_{lr}$	155.5963	104.1005	104.9525	131.1766	122.7052	108.5657
$\overline{y}_{t1}$	154.7504	195.4096	116.0315	134.3762	121.4452	106.4048
$\overline{y}_{t2}$	144.7504	165.8892	114.0315	132.6784	119.0782	109.4256
$\overline{y}_{hm1}$	100.9828	104.1228	108.2917	122.7292	124.5897	112.4089
$\overline{y}_{hm2}$	100.1832	103.7872	107.8894	121.2048	123.1645	110.5284
$\overline{y}_{D1}$	5158.967	3002.202	1619.323	519.6661	625.7129	1095.498
$\overline{y}_{D2}$	251313.9	143590.0	43483.91	24482.38	44110.72	47409.46
$\overline{y}_{D3}$	3932.387	1880.868	1023.353	605.6147	710.3697	910.2687
$\overline{y}_{D4}$	210009.3	133592.9	32736.92	10462.91	29291.28	35035.28
$\overline{y}_{D5}$	4822.239	3126.554	1735.952	673.8917	695.2891	1201.333
$\overline{y}_{D6}$	4815.077	2162.504	1343.679	983.9214	989.9492	1259.309
$\overline{y}_{D7}$	3233.523	1708.126	810.6883	330.1844	439.2591	668.9487
$\overline{y}_{D8}$	6050.281	3023.401	1723.625	660.7379	761.1324	1211.036
$\overline{y}_{De}$	228.7653	170.0789	130.9133	207.0891	133.9556	361.1317

Table 3: Using real datasets, PRE of the proposed class of estimators.

Estimator	Population I	Population II	Population III
$\overline{y}$	100	100	100
$\overline{y}_R$	117.6365	101.5271	104.4765
$\overline{y}_{lr}$	117.651	103.3431	106.7879
$\overline{y}_{t1}$	117.5989	103.1005	106.6716
$\overline{y}_{t2}$	104.1569	102.5963	106.0517
$\overline{y}_{hm1}$	116.9265	101.5758	104.7442
$\overline{y}_{hm2}$	116.9108	101.5783	104.7662
$\overline{y}_{D1}$	812.3034	1519.968	1604.261
$\overline{y}_{D2}$	263419.4	46122.15	6313.68
$\overline{y}_{D3}$	1460.442	1004.19	405.2918
$\overline{y}_{D4}$	139601.2	36976.19	4762.583
$\overline{y}_{D5}$	5208.603	1061.294	603.8153
$\overline{y}_{D6}$	2848.716	1292.207	765.8887
$\overline{y}_{D7}$	1014.97	806.8052	382.3507
$\overline{y}_{D8}$	917.2686	1539.184	1683.153
$\overline{y}_{De}$	119.3596	118.1962	115.8183

families' employment and weekly income of families, respectively, and R correspond to the rank of weekly income of families. For more detail, we can refer to [2] page 24. To estimate the finite population mean under extreme values, population constants are $N = 36; n = 10; \overline{Y} = 148718.7; \overline{X} = 1054.39; \overline{R} = 18.5; S_y = 182315.1; S_x = 402.6098;$ $S_r = 10.535; X_M = 2370; X_m = 388; R_M = 36; R_m = 1; C_y = 1.2259; C_x = 0.3818; C_r = 0.5695; g_1 = 0.3472; g_2 = 0.3498; \rho_{yx} = 0.1799; \rho_{yr} = 0.188;$ $\rho_{xr} = 0.9378;$ and $\theta = 0.0722$.

On the abovementioned datasets, the PREs of these proposed and current estimators are provided in Table 3. In terms of PRE, it can be seen that the proposed estimators are more efficient than the standard unbiased estimator and existing estimators, i.e., all values are more than one hundred.

7. Conclusion

In this paper, we present some effective estimators for estimating the finite population mean using known information about the minimum and maximum values of auxiliary data. We have identified certain theoretical situations in which the recommended estimators outperform existing estimators. Tables 2 and 3 offer the PREs for all estimators over the mean per unit estimator. According to our findings, the recommended estimators $\overline{y}_{Di}$ outperform the estimators evaluated in this research. They are recommended among the suggested classes of estimators because of their high PREs for all populations.

Appendix

where

$$L = \exp\left[\frac{(a\overline{U} + b_1) - (a\overline{u} + b_1)}{(a\overline{U} + b_1) + a\overline{u} + b_1}\right]. \tag{A.1}$$

Data Availability

All the data used in this study are available within the manuscript.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*, John Wiley & Sons, NJ, USA, 2013.

[2] W. G. Cochran, *Sampling Techniques*, John Wiley & Sons, New York, NY, USA, 3rd edition, 1963.

[3] N. Garg and M. Srivastava, "A general class of estimators of a finite population mean using multi-auxiliary information

under two stage sampling scheme," *Journal of Reliability and Statistical Studies*, vol. 2, no. 1, pp. 103–118, 2009.

[4] M. Khoshnevisan, R. Singh, P. Chauhan, and N. Sawan, "A general family of estimators for estimating population mean using known value of some population parameter (s)," *Far East Journal of Theoretical Statistics*, vol. 22, no. 2, pp. 181–191, 2007.

[5] M. M. Rueda, A. Arcos, M. D. Martínez-Miranda, and Y. Román, "Some improved estimators of finite population quantile using auxiliary information in sample surveys," *Computational Statistics & Data Analysis*, vol. 45, no. 4, pp. 825–848, 2004.

[6] C.-E. Särndal and C.-E. Sarndal, "Sample survey theory vs. General statistical theory: estimation of the population mean," *International Statistical Review/Revue Internationale de Statistique*, vol. 40, no. 1, pp. 1–12, 1972.

[7] B. V. Sukhatme, "Some ratio-type estimators in two-phase sampling," *Journal of the American Statistical Association*, vol. 57, no. 299, pp. 628–632, 1962.

[8] S. Tarima and D. Pavlov, "Using auxiliary information in statistical function estimation," *ESAIM: Probability and Statistics*, vol. 10, pp. 11–23, 2006.

[9] G. Walia, H. Kaur, and M. Sharma, "Ratio type estimator of population mean through efficient linear transformation," *American Journal of Mathematics and Statistics*, vol. 5, no. 3, pp. 144–149, 2015.

[10] S. Hussain, S. Ahmad, M. Saleem, and S. Akhtar, "Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling," *Plos one*, vol. 15, no. 9, Article ID e0239098, 2020.

[11] R. Jabeen, A. Sanaullah, and M. Hanif, "Efficient class of exponential estimators for population mean in two-stage cluster sampling," *Pakistan Journal of Statistics*, vol. 31, no. 6, 2015.

[12] H. P. Singh, S. Singh, and M. Kozak, "A family of estimators of finite-population distribution function using auxiliary information," *Acta Applicandae Mathematica*, vol. 104, no. 2, pp. 115–130, 2008.

[13] R. Singh, P. Chauhan, N. Sawan, and F. Smarandache, "Improvement in estimating the population mean using exponential estimator in simple random sampling," *International Journal of Statistics & Economics*, vol. 3, no. A09, pp. 13–18, 2009.

[14] H. O. Cekim and H. Cingi, "Some estimator types for population mean using linear transformation with the help of the minimum and maximum values of the auxiliary variable," *Hacettepe Journal of Mathematics and Statistics*, vol. 46, no. 4, pp. 685–694, 2016.

[15] U. Daraz, J. Shabbir, and H. Khan, "Estimation of finite population mean by using minimum and maximum values in stratified random sampling," *Journal of Modern Applied Statistical Methods*, vol. 17, no. 1, p. 20, 2018.

[16] M. Khan and J. Shabbir, "Some improved ratio, product, and regression estimators of finite population mean when using minimum and maximum values," *TheScientificWorldJOURNAL*, vol. 2013, Article ID 431868, 7 pages, 2013.

[17] S. Mohanty and J. Sahoo, "A note on improving the ratio method of estimation through linear transformation using certain known population parameters," *Sankhya: The Indian Journal of Statistics*, vol. 57, pp. 93–102, 1995.

[18] B. Sisodia and V. Dwivedi, "Modified ratio estimator using coefficient of variation of auxiliary variable," *Journal of the Indian Society of Agricultural Statistics*, vol. 33, no. 2, pp. 13–18, 1981.

[19] G. K. Agarwal, S. M. Allende, and C. N. Bouza, "Double sampling with ranked set selection in the second phase with nonresponse: analytical results and Monte Carlo experiences," *Journal of Probability and Statistics*, vol. 2012, Article ID 214959, 12 pages, 2012.

[20] Bureau of Statistics P, "Punjab Development Statistics," 2013.