Research Article

Flexible Robust Regression-Ratio Type Estimators and Its Applications

Muhammad Ijaz, 1 Syed Muhammad Asim, 2 Atta ullah, 2 and Ibrahim Mahariq 3

1 Department of Mathematics and Statistics, University of Haripur, Haripur, Pakistan
2 Department of Statistics, University of Peshawar, Peshawar, Pakistan
3 College of Engineering and Technology, American University of the Middle East, Kuwait

Correspondence should be addressed to Muhammad Ijaz; m.ijaz@uoh.edu.pk

Received 18 August 2022; Accepted 8 September 2022; Published 28 September 2022

Academic Editor: Shabir Ahmad

In real-world situations, the data set under examination may contain uncommon noisy measurements that unreasonably affect the data’s outcome and produce incorrect model estimates. Practitioners employed robust-type estimators to reduce the weight of the noisy measurements in a data set in such a scenario. Using auxiliary information that will produce reliable estimates, we have looked at a few flexible robust-type estimators in this study. In order to estimate the population mean, this study presents unique flexible robust regression type ratio estimators that take into account the data from the midrange and interdecile range of the auxiliary variables. Up to the first order of approximate computation, the bias and mean square were calculated. In order to compare the flexibility of the proposed estimator to those of the existing estimators, theoretical conditions were also obtained. We took into account data sets containing outliers for empirical computation, and it was found that the suggested estimators produce results with higher precision than the existing estimators.

1. Introduction

In practice, collecting all of the information on an object under investigation is challenging; therefore, predictions and decision-making studies are based on samples. Using probability theory, sampling is an art form for determining the dependability of available data. Simple random sampling (SRS) is the most common and simplest approach for selecting samples with equal probability at each selection while avoiding the concentration of auxiliary information. We collect some additional information (X) that is positively or negatively connected to the variable of interest (Y) in real-life situations with the variable of interest (Y). If we incorporate new information into classical estimators, we will get flexible results. Many researchers are presently striving to increase the flexibility of existing estimators by incorporating additional data. For example, Kadilar and Cingi [1] worked on the regression type estimators, Yan and Tian [2], Ijaz et al. [3–5].

The usual estimator of the population mean is defined by

\[ t_0 = \bar{Y} \]  

The bias and mean square error of \( t_j \) up to the first-order approximation are given as

\[ \text{Bias}(t_0) = 0, \]

\[ \text{Var}(t_0) = \frac{1}{n} \bar{Y}^2 C_Y^2. \]  

Kadilar and Cingi [2006] introduced a classical ratio and regression estimator.
where Yan and Tian [2010] suggested the efficient ratio-type estimators

\[ t_6 = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\bar{x} + \beta_1)} (\bar{X} + \beta_1), \]

\[ t_7 = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\bar{x} + \beta_2)} (\bar{X} + \beta_2). \]

The mean square error of \( t_k \)

\[ \text{MSE}(t_k) = \frac{1 - f}{n} \sum \{ C^2_y (1 - \rho^2) + R^2_k C^2_x \}, \quad k = 6, 7, \]

where \( R_k = \bar{X}/\bar{X} + \beta_1, R_2 = \bar{X}/\bar{X} + \beta_2. \)

Ijaz et al. [3] proposed ratio and regression type estimators

\[ t_8 = \frac{\bar{Y}}{(\bar{x})^2} \bar{X}^2, \]

\[ t_9 = \left[ \frac{\bar{X} + (\beta_1 - \beta_2)}{X + (\beta_1 - \beta_2)} \right]. \]

The mean square error is, respectively, given by

\[ \text{MSE}(t_k)_{rpe} = \frac{1 - f}{n} \bar{Y}^2 C^2_y (1 - \rho^2), \]

\[ \text{MSE}(t_k) = \frac{1 - f}{n} \bar{Y}^2 \left[ C^2_y + \delta^2 C^2_x - 2\theta \rho C_x C_y \right]. \]

Other estimators of Ijaz et al. [4, 5] are defined by

\[ t_{10} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{X} C_x)}, \]

\[ t_{11} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{X} C_x)}, \]

\[ t_{12} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{X})}, \]

\[ t_{13} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{X} C_x)}, \]

\[ t_{14} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{M}_d)}, \]

\[ t_{15} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{X} Q D)}, \]

\[ t_{16} = \frac{\bar{Y} + b(\bar{X} - \bar{x})}{(\delta \bar{X} + \bar{M}_d C_x)}. \]

The mean square error of proposed ratio type estimators is

\[ \text{MSE}(t_l) = \frac{1 - f}{n} \bar{Y}^2 [C^2_y (1 - \rho^2) + \theta^2 C^2_x], \]

where \( \theta_{10} = \delta \bar{X}/\delta \bar{X} + \bar{Q} D \times \bar{C}_x, \theta_{11} = \delta \bar{X}/\delta \bar{X} + \bar{X} C_x, \theta_{12} = \delta \bar{X}/\delta \bar{X} + \bar{X}, \theta_{13} = \delta \bar{X}/\delta \bar{X} + \bar{Q} D \times \bar{M}_d, \)

\( \theta_{14} = \delta \bar{X}/\delta \bar{X} + \bar{M}_d, \theta_{15} = \delta \bar{X}/\delta \bar{X} + \bar{Q} D, \theta_{16} = \delta \bar{X}/\delta \bar{X} + \bar{M}_d C_x. \)
\[ t_{17} = \frac{1}{\beta_1 - \beta_2} + C_x \]

\[ t_{18} = \frac{1}{\beta_1 - \beta_2} + (M_d - QD) \]

\[ t_{19} = \frac{1}{\beta_1 - \beta_2} + (\delta - M_d) \]

\[ t_{20} = \frac{1}{\beta_1 - \beta_2} + (\delta - QD) \]

\[ t_{21} = \frac{1}{\beta_1 - \beta_2} + (\delta - QD) \]

\[ t_{22} = \frac{1}{\beta_1 - \beta_2} + (\delta - QD) \]

\[ MSE(t_m) = \frac{1}{n} t^2 [C_x^2 + \theta_1 C_x + 2\theta_2 C_x] \]

\[ m = 17, 18, 19, 20, 21, 22 \]

where \( \beta_1 = X(1 - \beta_2)/[(\beta_1 - \beta_2) + C_x], \theta_1 = X \times QD/X \times QD + (M_d - QD), \theta_2 = X \times QD/X \times QD + (\delta - M_d), \theta_3 = X \times QD/X \times QD + (\delta - QD), \theta_4 = X \times QD/X \times QD + (\delta - QD), \frac{\theta_5}{\beta_1 - \beta_2} = X \times QD/X \times QD, \delta = QD/S_x \]

The Jeelani et al. [2013] recommended ratio estimator is as follows:

\[ t_{23} = \frac{\bar{Y} + b(X - \bar{X})}{(X \beta_1 + QD)} \]

The mean square error of the above estimator is defined by

\[ MSE(t_{23}) = \frac{1}{n} t^2 [C_x^2(1 - \rho^2) + \theta_2 C_x] \]

where \( \theta_2 = X \beta_1 / X \beta_1 + QD \).  

2. Research Problem

In actual, some data sets have a broad range of values known as outliers. The classical estimators will result in an incorrect conclusion and overfitting of the model in such a case. The primary goal of the current work is to create an estimate that will not be significantly impacted by an outlier. This paper used the midrange and interdecile range to investigate novel robust type ratio type estimators.

3. Methodology of the Proposed Estimators

The study is motivated by Kadilar and Cingi [1] where the authors proposed some regression type estimators. The study of Kadilar and Cingi [1] was not taken into account the data sets with an outlier. The current study focused to cover this gap and developed some robust type estimators that are not much effective against outliers. This paper presents new estimators for estimating the population means using the auxiliary information in the forms of midrange (MR) and interdecile range (IDR). The proposed estimators are defined by

\[ p_i = \frac{\bar{Y} + \beta_2(x)(X - \bar{X})}{(X + MR)} \]

where \( W_i = 1, MR, \beta_2(x) \]

\[ p_1 = \frac{\bar{Y} + \beta_2(x)(X - \bar{X})}{(X + IDR)} \]

\[ p_2 = \frac{\bar{Y} + \beta_2(x)(X - \bar{X})}{(X + IDR)} \]

\[ p_3 = \frac{\bar{Y} + \beta_2(x)(X - \bar{X})}{(X + IDR)} \]

\[ p_4 = \frac{\bar{Y} + \beta_2(x)(X - \bar{X})}{(X + IDR)} \]

To derive the estimator bias, and mean square error, we consider

\[ e_0 = Y - \bar{Y} / \sqrt{\bar{Y}}, \ e_1 = X - \bar{X} / \sqrt{\bar{X}}, \text{ and } E(e_0) = E(e_1) = 0 \]

Then,  

\[ p_i = \frac{\bar{Y} + e_0 - \beta_2(x)}{(X + MR)} \]

\[ p_i = \frac{\bar{Y} + e_0 - \beta_2(x)}{(X + IDR)} \]

\[ p_i = \frac{\bar{Y} + e_0 - \beta_2(x)}{(X + IDR)} \]

\[ p_i = \frac{\bar{Y} + e_0 - \beta_2(x)}{(X + IDR)} \]

applying expectations on both sides, we get the bias of \( p_i \) which is given by

\[ Bias(p_i) = \frac{1}{n} [\bar{Y} (U_i C_x^2 - U_i \rho C_x C_y) + T \bar{X} U_i C_x^2] \]

\[ (i = 1, 2, ..., 5) \]

Squaring and applying expectations on both sides of equation (26), we get

\[ E(p_i - \bar{Y})^2 = \bar{Y}^2 E(e_0^2 + U_i e_1^2 - 2U_i \rho e_1, e_1) \]

\[ + \beta_2(x)^2 \bar{X}^2 E(e_1^2) - 2\beta_2(x) \bar{X} E(e_0 e_1) - U_i e_1^2) \]

\[ (28) \]

The mean square error of \( p_i \) up to the first order approximation is given as
\[ MSE(p_i) = \frac{1}{n} \left[ \frac{\bar{Y}^2(C_y^2 + U_i^2C_x^2 - 2U_i\rho C_y C_x) + T_2\bar{X}^2C_x^2}{-2T\bar{Y}\bar{X}(\rho C_y C_x - U_iC_x)} \right] \]

where \( T = Q.D/MR, U_1 = \bar{X}/\bar{X} + MR, U_2 = \bar{X}/\bar{X} + IDR, U_3 = \bar{X}MR/\bar{X} + IDR, U_4 = \bar{X}\bar{\beta}_2(\xi)/\bar{\beta}_2(\xi) + MR, U_5 = \bar{X}\bar{\beta}_2(\xi)/\bar{\beta}_2(\xi) + IDR, IDR = D_2 - D_1, MR = X_1 + X_2/2, X_1 \) is the minimum value, and \( X_2 \) is the maximum value of a data set.

\[ (i = 1, 2, ..., 5), \] (29)

3.1. Theoretical Conditions. In this section, theoretical conditions are derived so that to assess the performance of the proposed estimators as compared to the existing estimators. The MSE of the proposed estimators is given in equation (29) with the usual mean estimator given in equation (2) can be compared in the following way.

\[ MSE(p_i) < Var(t_0), \] if

\[ MSE(p_i) < MSE(t_j), \] if

\[ \frac{1}{n} \left[ \frac{\bar{Y}^2(C_y^2 + U_i^2C_x^2 - 2U_i\rho C_y C_x) + T_2\bar{X}^2C_x^2}{-2T\bar{Y}\bar{X}(\rho C_y C_x - U_iC_x)} \right] < \frac{1}{n} \left[ \frac{\bar{Y}^2C_y^2}{T_2\bar{X}^2C_x^2} \right], \] 

(30)

Similarly, the Mse of the proposed estimator given in equation (29) can be compared with that of the Mse given in equation (29), we have the following.

\[ \frac{1}{n} \left[ \frac{\bar{Y}^2(C_y^2 + U_i^2C_x^2 - 2U_i\rho C_y C_x) + T_2\bar{X}^2C_x^2}{-2T\bar{Y}\bar{X}(\rho C_y C_x - U_iC_x)} \right] < \frac{1}{n} \left[ \frac{\bar{Y}^2(C_y^2(1 - \rho^2_{yx}) + R_2^2C_x^2)}{-2T\bar{Y}\bar{X}(\rho C_y C_x - U_iC_x)} \right] < 0. \] 

(31)

The proposed estimator leads to a better performance as compared to others if the above conditions are satisfied. Table 1 defines the result of theoretical conditions using population data sets 1 and 2.

Table 1: Numerical values of theoretical conditions using data sets 1 and 2.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Population I</th>
<th>Population II</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_1 vs t_0</td>
<td>-10264.81</td>
<td>-18618.73</td>
</tr>
<tr>
<td>P_2 vs t_0</td>
<td>-11445.01</td>
<td>-19321.21</td>
</tr>
<tr>
<td>P_3 vs t_0</td>
<td>-12782.44</td>
<td>-22537.39</td>
</tr>
<tr>
<td>P_4 vs t_0</td>
<td>-12723.93</td>
<td>-23096.04</td>
</tr>
<tr>
<td>P_5 vs t_0</td>
<td>-12786.60</td>
<td>-23034.91</td>
</tr>
<tr>
<td>P_1 vs t_1</td>
<td>-6363.179</td>
<td>-18188.94</td>
</tr>
<tr>
<td>P_2 vs t_1</td>
<td>-7543.376</td>
<td>-18891.41</td>
</tr>
<tr>
<td>P_3 vs t_1</td>
<td>-8880.813</td>
<td>-22107.6</td>
</tr>
<tr>
<td>P_4 vs t_1</td>
<td>-8822.300</td>
<td>-22666.22</td>
</tr>
<tr>
<td>P_5 vs t_2</td>
<td>-8885.067</td>
<td>-22605.4</td>
</tr>
</tbody>
</table>

The proposed estimator leads to a better performance as compared to others if the above conditions are satisfied. The data sets were obtained from the Italian Bureau of the Environment Protection [7] and recently cited by Abid et al. [8]. The data statistics are given in Tables 2 and 3.

The percentage relative efficiency (PRE) is shown in Tables 4 and 5 computed with the following mathematical formula:

\[ PRE = \frac{MSE(\bar{Y}_{existing})}{MSE(\bar{Y}_{pro(n)})} \times 100. \] 

(32)

3.2. Applications. The paper proposed the robust type estimators, and hence, we considered two data sets with outliers. The results

\[ \begin{aligned} \text{PRE} &= \frac{\text{MSE}(\bar{Y}_{existing})}{\text{MSE}(\bar{Y}_{pro(n)})} \times 100. \end{aligned} \]
To evaluate the model performance, we derived theoretical conditions and used real-world data sets to back up our findings. Furthermore, the proposed as well as alternative estimators’ results for the bias, mean square error, and percentage relative efficiency (PRF) are computed. The proposed estimators are superior to others in terms of PRE, but their MSE is the lowest of all. It is obvious that the suggested estimators outperform other methods in terms of results.

### Data Availability

The data set is used to evaluate the real significance of the proposed estimator and is given in the manuscript.

### Conflicts of Interest

The authors declare that they have no conflicts of interest.

### References


Table 2: Data statistics.

<table>
<thead>
<tr>
<th>$N = 103$</th>
<th>$n = 40$</th>
<th>$C_y = 1.4588$</th>
<th>$C_x = 1.0963$</th>
<th>$P_{xy} = 0.7298$</th>
<th>$IDR = 757.087$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MR = 1906.840$</td>
<td>$Q_1 = 62.621$</td>
<td>$Q_2 = 259.3830$</td>
<td>$S_y = 91.3550$</td>
<td>$S_x = 610.1643$</td>
<td>$\beta_{2(a)} = 17.8738$</td>
</tr>
<tr>
<td>$Md = 373.82$</td>
<td>$Q_4 = 628.0235$</td>
<td>$Q_5 = 368.6405$</td>
<td>$Q_6 = 184.3203$</td>
<td>$Q_{7} = 443.7033$</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Data statistics.

<table>
<thead>
<tr>
<th>$N = 49$</th>
<th>$n = 20$</th>
<th>$C_y = 0.8508$</th>
<th>$C_x = 1.0435$</th>
<th>$P_{xy} = 0.9817$</th>
<th>$IDR = 210.800$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$MR = 254.500$</td>
<td>$Q_1 = 127.7959$</td>
<td>$Q_2 = 103.1329$</td>
<td>$S_y = 123.1212$</td>
<td>$S_x = 104.4051$</td>
<td>$\beta_{2(a)} = 5.9878$</td>
</tr>
<tr>
<td>$Md = 64.000$</td>
<td>$Q_3 = 43.000$</td>
<td>$Q_4 = 120.000$</td>
<td>$Q_5 = 77.000$</td>
<td>$Q_6 = 38.500$</td>
<td>$Q_{7} = 81.500$</td>
</tr>
</tbody>
</table>

Table 4: Bias, MSE, and PRE of the proposed and other estimators using Population I.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Bias</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>0</td>
<td>127.6071</td>
<td>100</td>
</tr>
<tr>
<td>$t_1$</td>
<td>1.147847</td>
<td>131.5218</td>
<td>97.0235</td>
</tr>
<tr>
<td>$t_2$</td>
<td>1.148111</td>
<td>131.5384</td>
<td>97.0113</td>
</tr>
<tr>
<td>$t_3$</td>
<td>1.144672</td>
<td>131.3230</td>
<td>97.1704</td>
</tr>
<tr>
<td>$t_4$</td>
<td>1.150690</td>
<td>131.6999</td>
<td>96.89232</td>
</tr>
<tr>
<td>$t_5$</td>
<td>1.055885</td>
<td>125.7631</td>
<td>101.4662</td>
</tr>
<tr>
<td>$P_1$</td>
<td>0.02962096</td>
<td>60.57783</td>
<td>210.6498</td>
</tr>
<tr>
<td>$P_2$</td>
<td>0.15196360</td>
<td>66.44281</td>
<td>191.4792</td>
</tr>
<tr>
<td>$P_3$</td>
<td>1.02040400</td>
<td>116.3789</td>
<td>109.6479</td>
</tr>
<tr>
<td>$P_4$</td>
<td>0.70222700</td>
<td>97.74705</td>
<td>130.5483</td>
</tr>
<tr>
<td>$P_5$</td>
<td>0.8740394</td>
<td>107.7786</td>
<td>118.3975</td>
</tr>
</tbody>
</table>

Table 5: Bias, MSE, and PRE of the proposed and other estimators using Population II.

<table>
<thead>
<tr>
<th>Estimators</th>
<th>Bias</th>
<th>MSE</th>
<th>PRE</th>
</tr>
</thead>
<tbody>
<tr>
<td>$t_0$</td>
<td>0</td>
<td>349.833</td>
<td>100</td>
</tr>
<tr>
<td>$t_1$</td>
<td>4.040599</td>
<td>529.0587</td>
<td>66.12366</td>
</tr>
<tr>
<td>$t_2$</td>
<td>4.043777</td>
<td>529.4648</td>
<td>66.07294</td>
</tr>
<tr>
<td>$t_3$</td>
<td>4.034300</td>
<td>528.2538</td>
<td>66.22442</td>
</tr>
<tr>
<td>$t_4$</td>
<td>4.113499</td>
<td>538.3750</td>
<td>64.97942</td>
</tr>
<tr>
<td>$t_5$</td>
<td>2.974909</td>
<td>392.8679</td>
<td>89.04595</td>
</tr>
<tr>
<td>$P_1$</td>
<td>-0.463065</td>
<td>92.69699</td>
<td>377.3941</td>
</tr>
<tr>
<td>$P_2$</td>
<td>-0.4732379</td>
<td>76.95468</td>
<td>454.5961</td>
</tr>
<tr>
<td>$P_3$</td>
<td>1.2816110</td>
<td>64.48238</td>
<td>542.5249</td>
</tr>
<tr>
<td>$P_4$</td>
<td>0.7250457</td>
<td>33.81653</td>
<td>1034.503</td>
</tr>
<tr>
<td>$P_5$</td>
<td>0.8101430</td>
<td>37.97263</td>
<td>921.2766</td>
</tr>
</tbody>
</table>

Table 4: Bias, MSE, and PRE of the proposed and other estimators using Population I.

Finally, we concluded that the proposed estimators have a fewer MSE and high PRE as compared to others and thus lead to a preferable fit.

### 4. Discussion and Conclusion

The traditional or common estimators are consistently inadequate for estimating the population parameter in practice. The traditional estimators overestimate the sets of data that contain an outlier(s). The current work examines various innovative estimators that are less susceptible to these outliers as a result. In this paper, novel regression-ratio type estimators are proposed by using the midrange (MR) and interdecile range (IDR). To evaluate the model...