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

Copyright © 2022 Muhammad Ijaz 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.

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

\[ t_0 = \bar{Y}. \]  

(1)

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

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

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

(2)

Kadilar and Cingi [2006] introduced a classical ratio and regression estimator.
The mean square error of \( t_k \)

\[
MSE(t_k) = \frac{1}{n} \sum_{i=1}^{n} \left[ C_y^2 \left( 1 - \rho^2 \right) + R_i^2 C_x^2 \right], \quad k = 6, 7, \tag{8}
\]

where \( R_k = \tilde{X}/\hat{X} + \beta_1, R_\ell = \tilde{X}/\hat{X} \beta_1 + \beta_2 \).

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

\[
t_8 = \frac{Y}{(\bar{X})}, \tag{9}
\]

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

The mean square error is, respectively, given by

\[
MSE(t_{9,0}) = \frac{1}{n} \sum_{i=1}^{n} \left[ C_y^2 \left( 1 - \rho^2 \right) \right], \tag{10}
\]

\[
MSE(t_{9,1}) = \frac{1}{n} \sum_{i=1}^{n} \left[ C_y^2 + \theta^2 C_x^2 - 2\theta \rho C_y C_x \right]. \tag{11}
\]

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

\[
t_{10} = \frac{Y + b(\bar{X} - \bar{x})}{(\bar{X} + \rho)}, \tag{12}
\]

\[
t_{11} = \frac{Y + b(\bar{X} - \bar{x})}{(\bar{X} + \rho)}, \tag{13}
\]

\[
t_{12} = \frac{Y + b(\bar{X} - \bar{x})}{\delta X + \bar{x}}, \tag{14}
\]

\[
t_{13} = \frac{Y + b(\bar{X} - \bar{x})}{\delta X + \bar{X}}, \tag{15}
\]

\[
t_{14} = \frac{Y + b(\bar{X} - \bar{x})}{\delta X + \bar{x}}, \tag{16}
\]

\[
t_{15} = \frac{Y + b(\bar{X} - \bar{x})}{\delta X + \bar{X}}, \tag{17}
\]

\[
t_{16} = \frac{Y + b(\bar{X} - \bar{x})}{\delta X + \bar{X}}, \tag{18}
\]

The mean square error of proposed ratio type estimators is

\[
MSE(t_l) = \frac{1}{n} \sum_{i=1}^{n} \left[ C_y^2 \left( 1 - \rho^2 \right) + \theta_l C_x^2 \right], \tag{19}
\]

where \( \theta_{10} = \delta X / \delta X + QD \times C_x, \theta_{11} = \delta X / \delta X + X \bar{C}_x, \theta_{12} = \delta X / \delta X + \bar{X}, \theta_{13} = \delta X / \delta X + QD \times M_d, \theta_{14} = \delta X / \delta X + M_d, \theta_{15} = \delta X / \delta X + QD, \theta_{16} = \delta X / \delta X + M_d C_x. \)

Yan and Tian [2010] suggested the efficient ratio-type estimators

\[
t_6 = \frac{Y + b(\bar{X} - \bar{x})}{(\bar{X} + \beta_1)}, \tag{7}
\]

\[
t_7 = \frac{Y + b(\bar{X} - \bar{x})}{(\bar{X} + \beta_1)}, \tag{7}
\]

\[
t_8 = \frac{Y}{(\bar{X})}, \tag{9}
\]

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

The mean square error is, respectively, given by

\[
MSE(t_{9,0}) = \frac{1}{n} \sum_{i=1}^{n} \left[ C_y^2 \left( 1 - \rho^2 \right) \right], \tag{10}
\]

\[
MSE(t_{9,1}) = \frac{1}{n} \sum_{i=1}^{n} \left[ C_y^2 + \theta^2 C_x^2 - 2\theta \rho C_y C_x \right]. \tag{11}
\]
where

$$t_{17} = \frac{Y}{X} \left[ \frac{X(\beta_1 - \beta_2) + C_x}{X(\beta_1 - \beta_2) + C_x} \right].$$

$$t_{18} = \frac{X \times QD + (M_d - QD)}{X \times QD + (M_d - QD)}.$$

$$t_{19} = \frac{X \times QD + (\delta - M_d)}{X \times QD + (\delta - M_d)}.$$

$$t_{20} = \frac{X \times QD + (\delta - QD)}{X \times QD + (\delta - QD)}.$$

$$t_{21} = \frac{X(\beta_1 - \beta_2) + QD}{X(\beta_1 - \beta_2) + QD}.$$

$$MSE(t_m) = \frac{1}{n} \sum \left[ C_x^2 + \beta_m C_x + 2\beta_m p C_y C_x \right],$$

where

$$\theta_{07} = \frac{X(\beta_1 - \beta_2) + C_x}{X(\beta_1 - \beta_2) + C_x}, \theta_{08} = \frac{X \times QD/X \times QD + (M_d - QD)}, \theta_{09} = \frac{X \times QD/X \times QD + (\delta - M_d)}, \theta_{10} = \frac{X \times QD/X \times QD + (\delta - QD)}, \theta_{21} = \frac{X \times C_x / X \times C_x + (\beta_1 - \beta_2)}, \theta_{22} = \frac{X(\beta_1 - \beta_2) / X(\beta_1 - \beta_2) + QD}.$$}

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

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

The mean square error of the above estimator is defined by

$$MSE(t_{23}) = \frac{1}{n} \sum \left[ \frac{C_x^2 (1 - \rho^2) + \theta_{23} C_x^2} \right],$$

where

$$\theta_{23} = X \beta_1 / X \beta_1 + QD.$$}

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 estimator 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

$$p_i = \frac{Y + \beta_{2(i)} (X - \bar{X})}{(XW_i + V_i)} (XW_i + V_i).$$

where

$$W_i = 1, MR, \beta_{2(i)} V_i = MR, IDR,$$

$$p_1 = \frac{Y + \beta_{2(i)} (X - \bar{X})}{(X + MR)},$$

$$p_2 = \frac{Y + \beta_{2(i)} (X - \bar{X})}{(X + IDR)},$$

$$p_3 = \frac{Y + \beta_{2(i)} (X - \bar{X})}{(XMR + IDR)},$$

$$p_4 = \frac{Y + \beta_{2(i)} (X - \bar{X})}{(XMR + IDR)(XMR + IDR)},$$

$$p_5 = \frac{Y + \beta_{2(i)} (X - \bar{X})}{(XMR + IDR)(XMR + IDR)} (XMR + IDR).$$

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

$$e_0 = Y - \bar{Y}, \ e_1 = X - \bar{X}, \ \ and \ E(e_0) = E(e_1) = 0.$$}

Then,

$$p_1 = \frac{Y(1 + e_0) - \beta_{2(i)} (X)}{(XW_i + V_i)(1 + XW_i / (XW_i + V_i))} (XW_i + V_i),$$

$$p_1 = \frac{Y(1 + e_0) - \beta_{2(i)} (X)}{1 + U_i e_1},$$

$$p_1 = \frac{Y(1 + e_0) - \beta_{2(i)} (X)}{1 + U_i e_1} (1 - U_i e_1 + U_i^2 e_1^2).$$

$$(p_1 - Y) = \frac{Y(e_0 - U_i e_1 + U_i^2 e_1^2 - U_i e_0 e_1)}{-\beta_{2(i)}(X)(e_1 - U_i e_1^2)}.$$

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

$$Bias(p_i) = \frac{1}{n} \sum \left[ \frac{Y(U_i^2 C_x - U_i p C_y C_x) + T X U_i C_x^2} \right],$$

$$(i = 1, 2, \ldots, 5).$$

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

$$E(p_i - Y)^2 = \frac{Y^2}{n} E(e_0^2 + U_i^2 e_1^2 - 2U_i e_0 e_1)$$

$$+ \beta_{2(i)}^2 (X^2 E(e_1^2) - 2\beta_{2(i)} Y X E(e_0 e_1 - U_i^2 e_1^2)).$$

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

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

where

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

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.

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

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

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

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.

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

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

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.

| Estimators | Population I | Population II |
|------------|--------------|---------------|
| P₁ vs t₀   | −10264.81    | −18618.73     |
| P₂ vs t₀   | −11445.01    | −19321.21     |
| P₃ vs t₀   | −12782.44    | −22537.39     |
| P₄ vs t₀   | −12723.93    | −23096.04     |
| P₅ vs t₀   | −12786.60    | −23034.91     |
| P₁ vs t₁   | −6363.179    | −18188.94     |
| P₂ vs t₁   | −7543.376    | −18891.41     |
| P₃ vs t₁   | −8880.813    | −22107.6      |
| P₄ vs t₄   | −8822.300    | −22666.22     |
| P₅ vs t₅   | −8885.067    | −22605.4      |

The results of Table 1 clearly demonstrate that the aforementioned theoretical requirements are met for both data sets; hence, it is anticipated that the suggested estimators will perform better for these two data sets than for others.

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

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

Tables 4 and 5 define the Bias, Mse, and PRE of the proposed and other existing estimators. The results
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 (PRE) 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**

[1] C. Kadilar and H. Cingi, “An improvement in estimating the population mean by using the correlation coefficient,” *Hacettepe Journal of Mathematics and Statistics*, vol. 35, 2006.

[2] Z. Yan and B. Tian, “International Conference on Information Computing and Applications,” in *Proceedings of the 2010 ICICA 2010: Proceedings of the First International Conference on Information Computing and Applications*, Tangshan China, October 2010.

[3] M. Ijaz, A. U. Khan, S. M. Asim et al., “A new ratio type estimator to estimate the population mean using auxiliary information,” *Advances and Applications in Statistics*, vol. 63, no. 1, pp. 97–108, 2020.

[4] M. Ijaz, T. Zaman, H. Bulut, A. Ullah, and S. M. Asim, “An improved class of regression estimators using the auxiliary information,” *Journal of Science Iranica*, vol. 20, no. 4, pp. 789–800, 2020.

[5] M. Ijaz, A. Ullah, and T. Zaman, “On the development of the ratio type estimators using auxiliary information,” *Journal of Modern Mathematical Sciences*, vol. 6, p. 174, 2013.

[6] M. I. Jeelani, S. Maqbool, and S. A. Mir, *International Journal of Modern Mathematical Sciences*, vol. 35, 2006.

[7] IBEP, “Improved Ratio Estimators of Variance Based on Robust,” 2004, https://www.os.servatorio.nazionaleri-uti.it/enleri/etipodoc.asp?Atipodoc=6.

[8] M. Abid, S. Ahmad, M. Tahir, H. Z. Nazir, and M. Riaz, “Improved ratio estimator of variance based on robust measures,” *Journal of Science Iranica*, vol. 26, no. 4, pp. 2484–2494, 2019.
[9] W. G. Cochran, Wiley Eastern Limited, Hoboken, NJ, USA, Third Edition, 1940.

[10] C. Kadilar and H. Cingi, “Ratio estimators in simple random sampling,” *Applied Mathematics and Computation*, vol. 151, no. 3, pp. 893–902, 2004.

[11] H. Koç, “Ratio-type estimators for improving mean estimation using Poisson regression method,” *Communications in Statistics - Theory and Methods*, vol. 50, no. 20, pp. 4685–4691, 2021.

[12] C. Kadilar, M. Candan, and H. Cingi, “Ratio estimators using robust regression. Hacettepe,” *Journal of Mathematics and Statistics*, vol. 36, no. 2, pp. 181–188, 2007.

[13] E. Oral and C. Kadilar, “Robust ratio-type estimators in simple random sampling,” *Journal of the Korean Surgical Society*, vol. 40, no. 4, pp. 457–467, 2011.

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

[15] 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.

[16] 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.