Research Article

Robust Regression-Ratio-Type Estimators of the Mean Utilizing Two Auxiliary Variables: A Simulation Study

**Tolga Zaman,¹ Emre Dünder,² Ahmed Audu,³ David Anekeya Alilah ⁴,
Usman Shahzad ,⁵ and Muhammad Hanif⁵**

¹Çankırı Karatekin University, Faculty of Science, Department of Statistics, Çankırı, Turkey

²Ondokuz Mayıs University, Faculty of Science, Department of Statistics, Samsun, Turkey

³Department of Mathematics, Usmanu Danfodiyo University, Sokoto, Nigeria

⁴Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

⁵Department of Mathematics and Statistics, PMAS Arid Agriculture University, Rawalpindi, Pakistan

Correspondence should be addressed to David Anekeya Alilah; aliladavid2010@gmail.com

Received 22 May 2021; Accepted 28 August 2021; Published 6 September 2021

Academic Editor: Bekir Sahin

Copyright © 2021 Tolga Zaman 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.

Many authors defined the modified version of the mean estimator by using two auxiliary variables. These proposed estimators highly depend on the calculated regression coefficients. In the presence of outliers, these estimators do not give satisfactory results. In this study, we improve the suggested estimators using several robust regression techniques while obtaining the regression coefficients. We compared the efficiencies between the suggested estimators and the estimators presented in the literature. We used two numerical examples and a simulation study to support these theoretical results. Empirical results show that the modified ratio estimator performs well in the presence of outliers when adopting robust regression techniques.

1. Introduction

The classical ratio estimator is the most usual estimator of $\bar{Y}$ when the correlation between study variable y_i and auxiliary variables x_1 and x_2 is positively high. However, when there are outliers in dataset, classical estimators perform poorly by decreasing their efficiency. As a result, the following studies in ratio estimators are available in the literature to lessen the detrimental impact of outlier data. In ratio estimators, Kadilar et al. [1] introduced using Huber-M estimate instead of least squares estimation (LSE). Noor-ul-Amin et al. [2] proposed to use Huber estimate, instead of LSE under double sampling. In ratio estimators, Zaman and Bulut [3] improved the estimators by applying robust regression coefficients. Zaman [4] provided estimators combining ratio estimators for $\bar{Y}$ utilizing the ratio estimators given in Zaman and Bulut [3]. Subzar et al. [5] provided ratio estimators to predict $\bar{Y}$ by using various robust regression techniques. Shahzad et al. [6] presented some ratio estimators to predict $\bar{Y}$ by using of the Zaman and Bulut [3]

ratio estimators for several cases of missing observations. Zaman and Bulut [7] presented ratio estimators by considering robust regression techniques and robust covariance estimations for stratified random sampling. Subzar et al. [8] presented various ratio estimators by considering robust regression tools. Ali et al. [9] provided robust regression-type estimators of the mean for simple random sampling. Grover and Kaur [10] improved the various ratio-type estimators by considering robust regression tools. In this paper, we are suggested to use of some robust regression estimates, instead of LSE, for the improved estimator of $\bar{Y}$ by considering two auxiliary variables for simple random sampling.

In the next section, we will go over ratio-type estimators for $\bar{Y}$ for simple random sampling, as well as their MSEs. In Section 4, we deduce the properties of the suggested estimators. Section 5 compares the efficiency of the estimators offered by Kadilar and Cingi [11] and the suggested estimators based on the MSEs. An empirical study utilizing two datasets and a simulation study are conducted, and we

obtain satisfactory results, both theoretically and numerically. The numerical results are presented in Sections 6 and 7, respectively. We provide conclusion in the last section. In addition, the current study relies on robust estimates for two auxiliary variable studies for simple random sampling.

2. Traditional Ratio Estimator

Abu-Dayyeh et al. [12] provided the following estimator of $\bar{Y}$ utilizing two auxiliary variables for simple random sampling and assuming that $\bar{X}_1$ and $\bar{X}_2$ are known:

$$\bar{y}_{AD} = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right)^{\alpha_1} \left(\frac{\bar{X}_2}{\bar{x}_2} \right)^{\alpha_2}, \quad (1)$$

where α_1 and α_2 are real numbers.

Considering the ratio estimator presented in (1), Kadilar and Cingi [11] provided the following estimator:

$$\bar{y}_{KC} = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right)^{\alpha_1} \left(\frac{\bar{X}_2}{\bar{x}_2} \right)^{\alpha_2} + b_1 (\bar{X}_1 - \bar{x}_1) + b_2 (\bar{X}_2 - \bar{x}_2). \quad (2)$$

The MSE equation of the estimator given in (2) was obtained as follows:

$$\begin{aligned} \text{MSE}(\bar{y}_{KC}) \cong & \frac{1-f}{n} \left[S_y^2 + (\alpha_1^* R_1 + B_1)^2 S_{x_1}^2 + (\alpha_2^* R_2 + B_2)^2 S_{x_2}^2 - 2(\alpha_1^* R_1 + B_1) S_{yx_1} - 2(\alpha_2^* R_2 + B_2) S_{yx_2} \right. \\ & \left. + 2(\alpha_1^* R_1 + B_1)(\alpha_2^* R_2 + B_2) S_{x_1 x_2} \right], \end{aligned} \quad (3)$$

where $R_1 = (\bar{Y}/\bar{X}_1)$, $R_2 = (\bar{Y}/\bar{X}_2)$, $\alpha_1^* = (S_y \rho_{x_1 x_2} (\rho_{y x_1} \rho_{x_1 x_2} - \rho_{y x_2}) / R_1 S_{x_1} (1 - \rho_{x_1 x_2}^2))$, and $\alpha_2^* = (S_y \rho_{x_1 x_2} (\rho_{y x_2} \rho_{x_1 x_2} - \rho_{y x_1}) / R_2 S_{x_2} (1 - \rho_{x_1 x_2}^2))$.

$B_1 = (S_{y x_1} / S_{x_1}^2)$ and $B_2 = (S_{y x_2} / S_{x_2}^2)$ are obtained by the LS estimate, $f = (n/N)$, $S_{x_1}^2$ and $S_{x_2}^2$ are the variances of population of x_{1i} and x_{2i} , respectively, and $S_{y x_1}$ and $S_{y x_2}$ are the covariance of population between y_i and x_{1i} and between y_i and x_{2i} , respectively [11].

3. Robust Regression Methods

Here, we describe some important and famous robust regression methods.

3.1. Huber-M Estimation. Huber [13] proposed an estimation by using different ρ functions known as M-estimators. This estimate is based on minimizing another function of outliers instead of error squares (e_i).

The objective function of M-estimation is presented as

$$\min \sum_{i=1}^n \rho(e_i), \quad (4)$$

and is a symmetric function of outliers.

Huber's function ρ is designed as

$$\rho(e) = \begin{cases} \frac{e^2}{2}, & |e| \leq m, \\ m|e| - \frac{m^2}{2}, & |e| > m. \end{cases} \quad (5)$$

A derivative of the function presented in (5) is equal in the following equation:

$$\varphi(y) = \begin{cases} e, & |e| \leq m, \\ m \text{sgn}(e), & |e| > m, \end{cases} \quad (6)$$

where $\text{sgn}(\cdot)$ is sign function and specified as

$$\text{sgn}(x) = \begin{cases} -1 & e < m \\ 0 & e = m \\ 1 & e > m \end{cases}.$$

For Huber's estimate, $m = 1.345$ and a constant value [14].

3.2. S Estimation. Rousseeuw and Yohai [15] proposed the regression estimates associated with M-scales to S estimation. S estimate is based upon the residual scale of M estimate. The estimate uses the residual standard deviation to tackle the weaknesses of the median. Function to minimize is given as follows:

$$\min \sum_{i=1}^n \rho \left(\frac{y_i - \sum_{j=1}^n x_{ij} \beta}{\hat{\sigma}_s} \right), \quad (7)$$

where $\sqrt{(1/nL) \sum_{i=1}^n w_i e_i^2}$ in which $L = 0.199$ and $w_i = (\rho(u_i)/u_i^2)$ [16].

3.3. Least Trimmed Squares Estimation (LTS). In LTS estimate, initially squared error terms are sorted. Then, the sum of the first φ of sorted error terms is taken, and function to minimize is given as follows:

$$\min \sum_{i=1}^{\varphi} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2, \quad (8)$$

where $\varphi = \tau/2 + 1$ and τ is the number of observation.

3.4. Least Median of Squares Estimation (LMS). LMS estimate is improved by Rousseeuw and Leroy [17]. In LMS, the median of error squares is minimized. The following equation is minimized:

$$\text{Min median}(\varepsilon_i^2). \quad (9)$$

The estimate is robust against unusual observations in the direction of both x and y , and its breakdown point is 0.5 [17]. For more detailed information about robust regression techniques, Zaman and Bulut [3] work can be investigated.

4. Suggested Estimators

In the presence of unusual observations, we propose utilizing the following 4 robust ratio estimators based on LTS, S, LMS, and Huber-M estimations instead of the ratio estimators stated in (2):

$$\bar{y}_{pr1} = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right)^{\alpha_1} \left(\frac{\bar{X}_2}{\bar{x}_2} \right)^{\alpha_2} + b_{1(\text{LTS})} (\bar{X}_1 - \bar{x}_1) + b_{2(\text{LTS})} (\bar{X}_2 - \bar{x}_2), \quad (10)$$

$$\bar{y}_{pr2} = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right)^{\alpha_1} \left(\frac{\bar{X}_2}{\bar{x}_2} \right)^{\alpha_2} + b_{1(\text{S})} (\bar{X}_1 - \bar{x}_1) + b_{2(\text{S})} (\bar{X}_2 - \bar{x}_2), \quad (11)$$

$$\bar{y}_{pr3} = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right)^{\alpha_1} \left(\frac{\bar{X}_2}{\bar{x}_2} \right)^{\alpha_2} + b_{1(\text{LMS})} (\bar{X}_1 - \bar{x}_1) + b_{2(\text{LMS})} (\bar{X}_2 - \bar{x}_2), \quad (12)$$

$$\bar{y}_{pr4} = \bar{y} \left(\frac{\bar{X}_1}{\bar{x}_1} \right)^{\alpha_1} \left(\frac{\bar{X}_2}{\bar{x}_2} \right)^{\alpha_2} + b_{1(\text{Hub M})} (\bar{X}_1 - \bar{x}_1) + b_{2(\text{Hub M})} (\bar{X}_2 - \bar{x}_2). \quad (13)$$

The MSE equations of the suggested estimators $\bar{y}_{pri}$ where $i = 1, 2, 3, 4$ are assumed to be the same as the expression for MSE in (3), but it is clear that B_1 and B_2 in (3) should be substituted by $B_{1\text{rob}(k)}$ and $B_{2\text{rob}(k)}$, whose values

as computed by LTS, S, LMS, and Huber-M estimates, respectively. The MSE equations for the suggested robust regression-ratio-type estimators belonging to LTS, S, LMS, and Huber-M estimates are obtained as follows:

$$\begin{aligned} \text{MSE}(\bar{y}_{pri}) \cong & \frac{1-f}{n} \left[S_y^2 + (\alpha_1^* R_1 + B_{1\text{rob}(k)})^2 S_{x_1}^2 + (\alpha_2^* R_2 + B_{2\text{rob}(k)})^2 S_{x_2}^2 - 2(\alpha_1^* R_1 + B_{1\text{rob}(k)}) S_{yx_1} - 2(\alpha_2^* R_2 + B_{2\text{rob}(k)}) S_{yx_2} \right. \\ & \left. + 2(\alpha_1^* R_1 + B_{1\text{rob}(k)})(\alpha_2^* R_2 + B_{2\text{rob}(k)}) S_{x_1 x_2} \right], \end{aligned} \quad (14)$$

where $B_{1\text{rob}(k)}$ and $B_{2\text{rob}(k)}$ are obtained from LTS, S, LMS, and Huber-M estimates, respectively ($k = \text{LTS, S, LMS, and Huber-M}$).

5. Efficiency Comparisons

We compare the MSE equation for the traditional estimator, presented in (3), with the MSE equations for the suggested

robust regression-ratio-type estimators, given in (14), to derive the conditions for which the suggested estimators will perform better than traditional estimator for two auxiliary variables for simple random sampling.

$$\text{MSE}(\bar{y}_{pri}) < \text{MSE}(\bar{y}_{\text{KC}}); i = 1, 2, 3, 4,$$

$$\begin{aligned} & (\alpha_1^* R_1 + B_{1\text{rob}(k)})^2 S_{x_1}^2 + (\alpha_2^* R_2 + B_{2\text{rob}(k)})^2 S_{x_2}^2 - 2(\alpha_1^* R_1 + B_{1\text{rob}(k)}) S_{yx_1} \\ & - 2(\alpha_2^* R_2 + B_{2\text{rob}(k)}) S_{yx_2} + 2(\alpha_1^* R_1 + B_{1\text{rob}(k)})(\alpha_2^* R_2 + B_{2\text{rob}(k)}) S_{x_1 x_2} \\ & < (\alpha_1^* R_1 + B_1)^2 S_{x_1}^2 + (\alpha_2^* R_2 + B_2)^2 S_{x_2}^2 - 2(\alpha_1^* R_1 + B_1) S_{yx_1} - 2(\alpha_2^* R_2 + B_2) S_{yx_2} + 2(\alpha_1^* R_1 + B_1)(\alpha_2^* R_2 + B_2) S_{x_1 x_2}, \\ & (B_{1\text{rob}(k)} - B_1) \left[S_{x_1}^2 (B_{1\text{rob}(k)} + B_1) + 2S_{x_1}^2 \alpha_1^* R_1 - 2S_{yx_1} + 2S_{x_1 x_2} \alpha_2^* R_2 \right] \end{aligned}$$

$$\begin{aligned}
& + (B_{2\text{rob}(k)} - B_2) \left[S_{x_2}^2 (B_{2\text{rob}(k)} + B_2) + 2S_{x_2}^2 \alpha_2^* R_2 - 2S_{yx_2} + 2S_{x_1 x_2} \alpha_1^* R_1 \right] \\
& + 2S_{x_1 x_2} \left[(B_{1\text{rob}(k)}) (B_{2\text{rob}(k)}) - B_1 B_2 \right] < 0.
\end{aligned} \tag{15}$$

The suggested robust regression-ratio-type estimators in (10)–(13) perform better than the ratio estimator in (2) when condition (15) is satisfied.

6. Numerical Illustrations

In this part, we performed numerical examples on two real datasets. The first dataset (Education) was collected to model the expenditure for public education [18]. The second dataset (Crime) was used to predict the violent crimes in states [19]. These datasets were previously used for investigating the robust techniques in the literature. Education dataset exists in robustbase package in R software [20].

We employed four robust estimators: LTS, LMS, S, and Huber-M, respectively. We compared the efficiency of robust techniques with LSE. While assessing the performance of the estimators, we used the efficiencies relative to LSE. We used R programming language in the implementation phase [21]. We utilized MASS and robustreg packages for robust regression analysis [20, 22]. We obtained the MSE values for each estimator for performance evaluation.

The definitions of the variables existing in datasets are shown in Table 1. The datasets contain two auxiliary and one study variables. Table 2 indicates the descriptive statistics for the real datasets.

In Table 3–6, the covariance and correlation matrices of the variables are given for both datasets. The high positive correlation values satisfy the condition of the applicability for the ratio estimators.

We calculated the Mahalanobis distances for checking the existence of potential outliers. For each observation “ i ,” Mahalanobis distances are calculated as follows:

$$\text{MAHDIST}_i = \sqrt{(x_i - L)C^{-1}(x_i - L)^T}, \tag{16}$$

where “ L ” shows the location matrix and “ C ” indicates the covariance matrix. The cut-off value is $\chi_{p, (1-\alpha)/2}^2$ where p is the number of variables. The α term represents the error level. The observation is considered as the potential outlier when exceeds the cut-off value.

We used minimum covariance determinant (MCD) estimators for avoiding the masking effect [23]. Figures 1 and 2 represent the Mahalanobis distance plots for the real datasets. The cut-off value is given with a straight line in plots. Clearly, we can see that Education and Crime datasets include some potential outliers. Also, these two datasets were used to evaluate the robustness issues in the literature.

Table 7 indicates the regression coefficients which were obtained using the robust estimates and LSE. The coefficients are rather different from LSE for robust estimates. Although there is a positive correlation between x_1 and y in Education dataset, B_1 is negative. This case shows the corruptive effect of outliers in dataset. Robust estimates adroitly overcome

this problem, and all regression coefficients are positive for each robust estimate.

We use the MSE values of the conventional and suggested robust regression-ratio-type estimators, as specified in Sections 2 and 4, to calculate the relative efficiency of each suggested robust regression-ratio-type estimator in (10)–(13) in comparison to the classical estimators in (2), utilizing the formulae:

$$\text{RE} = \frac{\text{MSE}(\bar{y}_{pri})}{\text{MSE}(\bar{y}_{KC})}, \quad i = 1, 2, 3, 4. \tag{17}$$

Table 8 denotes the performance results of the estimators. According to efficiencies, the suggested robust regression-ratio-type estimator is better than the estimator presented by Kadilar and Cingi [11]. Especially the suggested robust regression-ratio-type estimators based on S and LTS estimates produced the lowest MSE values for Education and Crime dataset, respectively. The classical estimator has the highest MSE value when comparing with four robust regression-ratio-type estimators in both two datasets. These results are not surprising because condition (15) is satisfied.

7. Simulation

A simulation study is performed to evaluate the efficiencies of the suggested robust regression-ratio-type estimators. Epilepsy dataset is used for the simulation [24]. The dataset include two auxiliary and one study variable. There are $n = 59$ observations in epilepsy dataset. The purpose of this dataset is to predict the epilepsy attacks. Epilepsy dataset exists in “robustbase” package of R programming language.

The description of the variables for epilepsy dataset is shown in Table 9. Sum of Y is the study, and the other variables are auxiliary variables.

Figure 3 demonstrates the Mahalanobis distance plot for epilepsy dataset. The distances are obtained similar to the previous application section. Obviously, we can see that this dataset contains some possible outlier observations.

We randomly selected sample from the datasets for 10000 times randomly and estimated the population means using the traditional and proposed estimators. We computed the MSE equation as follows:

$$\text{MSE}(\hat{\bar{Y}}) = \frac{1}{10000} \sum_{i=1}^{10000} \left(\hat{\bar{Y}}_i - \bar{Y} \right)^2, \tag{18}$$

where $\hat{\bar{Y}}_i$ indicates the estimation of mean for $i = 1, 2, \dots, 9999, 10000$ and $\bar{Y}$ represents the priorly known population mean of the study variable. In Table 10, we reported the MSE ratios of the suggested robust regression-ratio-type estimators with respect to the traditional estimator for each dataset. These values are obtained using (18).

TABLE 1: Description of the auxiliary and study variables for datasets.

Dataset	Variable	Description
Education	x_1	Number of residents per thousand residing
	x_2	Per capita personal income
	y	Per capita expenditure on public education
Crime	x_1^*	Number of murders per 1,000,000 people
	x_2^*	Percentage of the people who are single parents
	y^*	Violent crimes per 100,000 people

TABLE 2: Descriptive statistics of the variables for datasets.

Dataset	Variable	Mean	STD	Min	Max	a_1^*	a_2^*
Education	x_1	657.800	145.016	322	909	-0.405	0.089
	x_2	4675.120	644.506	3448	5889		
	y	284.600	61.340	208	546		
Crime	x_1^*	8.727	10.718	1.600	78.500	-0.149	-2.085
	x_2^*	11.325	2.121	8.400	22.100		
	y^*	612.843	441.100	82.000	2922.000		

TABLE 3: Covariance matrix for Education dataset.

Dataset	Variable	x_1	x_2	y
Education	x_1	21029.755	58128.680	2865.327
	x_2	58128.678	415388.310	24048.702
	y	2865.327	24048.700	3762.612

TABLE 4: Covariance matrix for Crime dataset.

Dataset	Variable	x_1^*	x_2^*	y^*
Crime	x_1^*	114.866	19.529	4189.516
	x_2^*	19.529	4.501	785.012
	y^*	4189.516	785.012	194569.495

TABLE 5: Correlation matrix for Education dataset.

Dataset	Variable	x_1	x_2	y
Education	x_1	1	0.622	0.322
	x_2	0.622	1	0.608
	y	0.322	0.608	1

TABLE 6: Correlation matrix for Crime dataset.

	Variable	x_1^*	x_2^*	y^*
Crime	x_1^*	1	0.859	0.886
	x_2^*	0.859	1	0.839
	y^*	0.886	0.839	1

Numerical results were conducted for the sizes of sample $n = 10, 20, 30, 40$. Computations were run in R software.

Table 10 denotes the simulation results. According to results, the suggested robust regression-type estimators apparently outperform the estimator presented in Kadilar

and Cingi [11] in all sample sizes. In most cases, the suggested robust regression-type estimator based on LTS estimate has the lowest MSE value. As the sample size grows, all estimators produce lower MSE values. Generally, the suggested regression-type estimators based on robust

FIGURE 1: Mahalanobis distance plot for Education dataset.

FIGURE 2: Mahalanobis distance plot for Crime dataset.

TABLE 7: Regression coefficients for each estimator.

Dataset	Estimators	B_1	B_2
Education	LSE	-0.039	0.063
	LTS	0.116	0.029
	S	0.100	0.034
	LMS	0.118	0.029
	Huber-M	0.042	0.045
Crime	LSE	25.999	61.607
	LTS	43.655	71.956
	S	46.210	57.910
	LMS	50.293	58.780
	Huber-M	22.235	83.478

techniques perform better than traditional estimator and overcome the outliers problem. These simulation findings support the theoretical results in Table 8.

8. Discussion

Tables 8 and 10 clearly show that proposed robust regression-ratio-type estimators for estimating $\bar{Y}$ by

utilizing outliers data for simple random sampling that are more efficient. The estimators of (10)–(13) provide lower MSEs than the MSE of traditional ratio estimator of (3). This means that the estimators of (10)–(13) show high performance than the estimator presented by Kadilar and Cingi [11]. These results have been demonstrated theoretically and supported by both empirical and simulation results.

TABLE 8: Theoretical results for the efficiency values of each suggested robust regression-ratio-type estimator with respect to the traditional estimator.

Dataset	Estimator	RE
Education	$\bar{y}_{KC}$	1
	$\bar{y}_{pr1}$	0.601
	$\bar{y}_{pr2}$	0.525
	$\bar{y}_{pr3}$	0.618
	$\bar{y}_{pr4}$	0.749
Crime	$\bar{y}_{KC}$	1
	$\bar{y}_{pr1}$	0.621
	$\bar{y}_{pr2}$	0.767
	$\bar{y}_{pr3}$	0.764
	$\bar{y}_{pr4}$	0.879

TABLE 9: Description of variables.

Variable	Description
Age	Age values of the patients
Base	Number of epileptic attacks recorded which is prior to randomization
Sum of Y	Total number of epilepsy attacks

FIGURE 3: Mahalanobis distance plot for epilepsy dataset.

TABLE 10: Simulation results for the efficiency values of each suggested robust regression-ratio-type estimator with respect to the traditional estimator.

Sample size	Estimators	RE
$n = 10$	$\bar{y}_{KC}$	1
	$\bar{y}_{pr1}$	0.44
	$\bar{y}_{pr2}$	0.467
	$\bar{y}_{pr3}$	0.438
	$\bar{y}_{pr4}$	0.682
$n = 20$	$\bar{y}_{KC}$	1
	$\bar{y}_{pr1}$	0.503
	$\bar{y}_{pr2}$	0.53
	$\bar{y}_{pr3}$	0.512
	$\bar{y}_{pr4}$	0.719

TABLE 10: Continued.

Sample size	Estimators	RE
$n = 30$	$\bar{y}_{KC}$	1
	$\bar{y}_{pr1}$	0.583
	$\bar{y}_{pr2}$	0.577
	$\bar{y}_{pr3}$	0.59
	$\bar{y}_{pr4}$	0.763
$n = 40$	$\bar{y}_{KC}$	1
	$\bar{y}_{pr1}$	0.682
	$\bar{y}_{pr2}$	0.699
	$\bar{y}_{pr3}$	0.691
	$\bar{y}_{pr4}$	0.816

9. Conclusion

Traditional ratio estimators suffer from the outliers because of the distorting effect. In this study, we proposed robust regression-ratio-type estimators using several robust estimates to handle the robustness task for the estimator proposed by Kadilar and Cingi [11]. We aimed to improve the performance of the suggest estimators by adopting robust coefficients. We used two real dataset examples containing possible outliers for comparing the suggested robust regression-ratio-type estimators with traditional estimator. According to our findings, the suggested robust regression-ratio-type estimators based on all robust techniques have lower MSE values when comparing with the traditional estimator. Numerical results demonstrate that the suggested robust regression-ratio-type estimators provide more efficient results than the traditional estimator. We hope that in the future we will expand the estimators presented here to other sampling designs.

Data Availability

The data used to support this study are included within the article.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

References

- [1] C. Kadilar, M. Candan, and H. Cingi, "Ratio estimators using robust regression," *Haceteppe Journal of Mathematics and Statistics*, vol. 36, no. 2, pp. 181–188, 2007.
- [2] M. Noor-ul-Amin, M. Q. Shahbaz, and C. Kadilar, "Ratio estimators for population mean using robust regression in double sampling," *Gazi University Journal of Science*, vol. 29, no. 4, pp. 793–798, 2016.
- [3] 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.
- [4] T. Zaman, "Improvement of modified ratio estimators using robust regression methods," *Applied Mathematics and Computation*, vol. 348, pp. 627–631, 2019.
- [5] M. Subzar, C. N. Bouza, and A. I. Al-Omari, "Utilization of different robust regression techniques for estimation of finite population mean in SRSWOR in case of presence of outliers through ratio method of estimation," *Investigación Operacional*, vol. 40, no. 5, pp. 600–609, 2019.
- [6] U. Shahzad, N. H. Al-Noor, M. Hanif, I. Sajjad, and M. Muhammad Anas, "Imputation based mean estimators in case of missing data utilizing robust regression and variance-covariance matrices," *Communications in Statistics - Simulation and Computation*, vol. 44, pp. 1–20, 2020.
- [7] 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.
- [8] M. Subzar, A. Ibrahim Al-Omari, and A. R. A. Alanzi, "The Robust regression methods for estimating of finite population mean based on SRSWOR in case of outliers," *Computers, Materials & Continua*, vol. 65, no. 1, pp. 125–138, 2020.
- [9] N. Ali, I. Ahmad, M. Hanif, and U. Shahzad, "Robust-regression-type estimators for improving mean estimation of sensitive variables by using auxiliary information," *Communications in Statistics - Theory and Methods*, vol. 50, no. 4, pp. 979–992, 2021.
- [10] L. K. Grover and A. Kaur, "An improved regression type estimator of population mean with two auxiliary variables and its variant using robust regression method," *Journal of Computational and Applied Mathematics*, vol. 382, pp. 1–18, 2021.
- [11] C. Kadilar and H. Cingi, "A new estimator using two auxiliary variables," *Applied Mathematics and Computation*, vol. 162, no. 2, pp. 901–908, 2005.
- [12] W. A. Abu-Dayyeh, M. S. Ahmed, R. A. Ahmed, and H. A. Muttalak, "Some estimators of a finite population mean using auxiliary information," *Applied Mathematics and Computation*, vol. 139, no. 2-3, pp. 287–298, 2003.
- [13] P. J. Huber, "Robust regression: asymptotics, conjectures and monte carlo," *Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.
- [14] J. Fox, *Robust Regression: Appendix to an R and S-PLUS Companion to Applied Regression*, 2002.
- [15] P. Rousseeuw and V. Yohai, "Robust regression by means of S-estimators," *Robust and Nonlinear Time Series Analysis*, vol. 26, pp. 256–272, 1984.
- [16] M. Salibian-Barrera and V. J. Yohai, "A fast algorithm for S-regression estimates," *Journal of Computational & Graphical Statistics*, vol. 15, no. 2, pp. 414–427, 2006.
- [17] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, Hoboken, NJ, USA, 2005.
- [18] S. Chatterjee and A. S. Hadi, *Regression Analysis by Example*, John Wiley & Sons, Hoboken, NJ, USA, 2015.
- [19] A. Agresti and B. Finlay, *Statistical Methods for the Social Sciences*, Pearson Education, London, UK, 4th edition, 2013.

- [20] M. Maechler, P. Rousseeuw, C. Croux et al., *Basic Robust Statistics R Package Version 0.92-7*, 2016.
- [21] R. R Core Team, *A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2017, <https://www.R-project.org/>.
- [22] W. N. Venables and B. D. Ripley, "Tree-based methods," *Modern Applied Statistics with S*, vol. 45, pp. 251–269, 2002.
- [23] P. J. Rousseeuw and B. C. Van Zomeren, "Unmasking multivariate outliers and leverage points," *Journal of the American Statistical Association*, vol. 85, no. 411, pp. 633–639, 1990.
- [24] P. F. Thall and S. C. Vail, "Some covariance models for longitudinal count data with overdispersion," *Biometrics*, vol. 46, no. 3, pp. 657–671, 1990.