Research Article

Minimum Covariance Determinant-Based Quantile Robust Regression-Type Estimators for Mean Parameter

Usman Shahzad ¹, Nadia H. Al-Noor ², Noreen Afshan,¹ David Anekeya Alilah ³,
Muhammad Hanif,¹ and Malik Muhammad Anas¹

¹Department of Mathematics and Statistics—PMAS-Arid Agriculture University, Rawalpindi, Pakistan

²Department of Mathematics, College of Science, Mustansiriyah University, Baghdad, Iraq

³Department of Mathematics, Masinde Muliro University of Science and Technology, Kakamega, Kenya

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

Received 7 June 2021; Accepted 10 July 2021; Published 21 July 2021

Academic Editor: Adnan Maqsood

Copyright © 2021 Usman Shahzad 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.

Robust regression tools are commonly used to develop regression-type ratio estimators with traditional measures of location whenever data are contaminated with outliers. Recently, the researchers extended this idea and developed regression-type ratio estimators through robust minimum covariance determinant (MCD) estimation. In this study, the quantile regression with MCD-based measures of location is utilized and a class of quantile regression-type mean estimators is proposed. The mean squared errors (MSEs) of the proposed estimators are also obtained. The proposed estimators are compared with the reviewed class of estimators through a simulation study. We also incorporated two real-life applications. To assess the presence of outliers in these real-life applications, the Dixon chi-squared test is used. It is found that the quantile regression estimators are performing better as compared to some existing estimators.

1. Introduction

The use of auxiliary information in survey sampling is as ancient as survey sampling itself (Bulut and Zaman [1]). Neyman's work [2] mentioned the early works in which auxiliary information was used. The problem of improving the efficiency of parameter estimation by the use of auxiliary information has received a lot of attention in sampling theory and practice. The common examples of such methods are ratio, product, and regression estimators. Under the simple random sampling scheme, ratio and product estimation techniques are widely used. For more details about these two estimation techniques, studies by Cochran [3], Murthy [4], Singh [5], and Shalabh and Tsai [6] can be referred. Furthermore, there is a wealth of literature on ratio estimators for population mean. Studies by Koyuncu [7], Abid et al. [8, 9], Irfan et al. [10], Shahzad et al. [11], Ali et al. [12], and Yadav and Zaman [13] are examples of such works.

Both ratio and product techniques have benefits and drawbacks. For instance, a ratio estimator is appropriate when the study and the auxiliary variables have a positive linear relationship/correlation, whereas a product estimator is appropriate when they have a negative linear relationship. This problem is solved by using the regression estimation technique, which yields significantly improved results for both positive and negative correlations. It should be noted that the typical regression estimator is based on the linear least square regression coefficient. For regression estimation, interested readers may refer to Ijaz et al. [14] and Tanış et al. [15].

The linear least squares or ordinary least squares (OLS) regression is the most conventional statistical method commonly used for parameter estimation due to its easy implementation. This method aims to reduce the sum of the squares of the differences or residuals (s_i) between the observed dependent variable and the predictions made by

the independent variable's linear function. The OLS strategy produces the best estimation results for straight-line regression under the ideal circumstances revealed by OLS. On the other hand, the parameter estimates based on OLS is influenced by outliers or extreme values and, therefore, do not produce significantly productive results. The OLS-fitting threshold point is n^{-1} (n , sample size) or 0%, indicating that a single outlier can have a significant impact (Rousseeuw and Leroy [16]). As a result, in the presence of outliers, the mean estimation based on OLS is also affected (Zaman and Bulut [17]). To overcome this issue in sense of mean estimation, the authors used robust regression tools, robust quantile regression tools, and robust covariance matrices (see, among others, recent works of Zaman and Toksoy [18], Zaman [19], and Zaman and Bulut [17, 20] who developed robust regression techniques to monitor the effect of extreme values). In this study, we have attempted to utilize the quantile regression with minimum covariance determinant estimator-based measures of location and propose a class of quantile regression-type mean estimators.

The main parts of this study are organized as follows. First, a review of robust regression tools and the MCD-based ratio is provided. The proposed class of quantile regression-type mean estimators is then introduced, as are its MSE. In addition, numerical illustrations of the existing and proposed class of estimators are included. Finally, the conclusion is provided.

2. Review of MCD-Based Ratio Estimators

In attempts to find estimators for the location parameter that are efficient across a wide range of datasets, the robust approach as the main approach has been developed (Hogg [21]). The purpose of the robust technique is to discover a single estimator that is efficient across a wide range of datasets, even if it is not exactly ideal for any population. For this article's purposes, the following lines give a precise description of some of the robust regression tools and MCD estimator utilized by authors to overcome the issue of extreme values in mean estimation.

As robust regression tools, we considered the following:

LAD: the least absolute deviation estimator is based on minimizing the sum of absolute squared errors (SE).

LMS: the least median of squares estimator is based on minimizing the SE median.

LTS: the least trimmed squares estimator is based on applying OLS to specified initial observations of sorted SE that is; hence, their computations are not affected by extreme values.

M-estimation: the M-estimation, M stands for maximum likelihood type, is based upon the minimization of the objective function $\varnothing(\cdot)$. Some of the designed formulae for the objective function of residuals (s_i) are as follows:

Huber M-estimator: Huber [22] investigated the objective function $\varnothing(s_i)$ with $a = 4.685$, or 6 as

$$\varnothing(s_i) = \begin{cases} \frac{1}{2}s_i^2, & |s_i| \leq a, \\ a\left(|s_i| - \frac{1}{2}a\right), & |s_i| > a. \end{cases} \quad (1)$$

Hampel M-estimator: Hampel [23] investigated the objective function $\varnothing(s_i)$ with $k = 1.7$, $g = 3.4$, and $a = 8.5$ as

$$\varnothing(s_i) = \begin{cases} \frac{1}{2}s_i^2, & 0 < |s_i| \leq k, \\ k|s_i| - \frac{1}{2}s_i^2, & k < |s_i| \leq g, \\ \frac{-k}{2(a-g)}(a-s_i)^2 + \frac{k}{2}(g+a-k), & g < |s_i| \leq a, \\ \frac{k}{2}(g+a-k), & a < |s_i|. \end{cases} \quad (2)$$

Tukey M-estimator: Tukey [24] investigated the objective function $\varnothing(s_i)$ with $a = 4.685$, or 6 as

$$\varnothing(s_i) = \begin{cases} \frac{a^2}{6} \left[1 - \left(1 - \left(\frac{s_i}{a} \right)^2 \right)^3 \right], & |s_i| \leq a, \\ \frac{a^2}{6}, & |s_i| > a. \end{cases} \quad (3)$$

MM-estimator: the MM-estimator, MM stands for modified M-estimator, is another type of a robust regression tool that is also developed in the presence of outliers. It is proposed based on combining the high resistance to outliers of S-estimators with the high efficiency of M-estimators. The MM-estimator is a regression M-estimator with a redescending function, with the initial values of the regression coefficients and the scale estimate coming from the S-estimator, which is based on a robust scale M-estimator minimization. For details, Yohai [25] may be viewed.

MCD: the minimum covariance determinant estimator is one of the earliest and most robust estimators of multivariate location and spread. Although MCD was first introduced in 1984 (Rousseeuw [26, 27]), the development of the computationally efficient fast MCD algorithm in 1999 (Rousseeuw and Van Driessen [28]) marked the beginning of its principal application. The MCD location estimate is the mean of the r^{th} observations, $(n/2) \leq r \leq n$ and $r = n * (1 - \text{trimming ratio})$, for which the sample covariance matrix determinant is as small as possible. For more details, refer studies by Al-Noor and Mohammad [29], Hubert et al. [30], Bulut and Zaman [1], and Zaman and Bulut [17, 20].

Zaman and Bulut [17] incorporated traditional measures of location such as coefficient of variation ψ_x , coefficient of kurtosis $\psi_2(x)$, and arithmetic mean $\bar{X}$. They calculated these characteristics through MCD estimation which is highly sensitive in the absence of normality and in presence of outliers. They defined the following family of MCD-based ratio-type estimators

$$\bar{y}_{bzi} = \frac{\bar{y}_M + \hat{\beta}_{(rob)}(\bar{X}_M - \bar{x}_M)}{(A_1\bar{X}_M + A_2)} (A_1\bar{X}_M + A_2), \quad i = 1, 2, \dots, 35, \quad (4)$$

where $(\bar{X}_M, \bar{Y}_M)$ are the population means and $(\bar{x}_M, \bar{y}_M)$ are the sample means when a simple random sample of size n is gathered from the population. $\hat{\beta}_{(rob)}$ is based on robust regression tools as discussed in previous lines of the current section. Furthermore, A_1 and A_2 are either $(0, 1)$ or some known population measures. The family members of $\bar{y}_{bzi}$ are provided in Table 1. The MSE of $\bar{y}_{bzi}$ is given by

$$MSE(\bar{y}_{bzi}) = \theta [S_y^2 + u_i^2 S_x^2 + 2\beta_{(rob)} u_i S_x^2 + \beta_{(rob)}^2 S_x^2 - 2u_i S_{xy} - 2\beta_{(rob)} S_{xy}], \quad i = 1, 2, \dots, 35, \quad (5)$$

where $u_i = (A_1\bar{Y}_M / A_1\bar{X}_M + A_2)$ and $\theta = ((1-f)/n)$ for $i = 1, 2, \dots, 9$. Furthermore, S_y^2 and S_x^2 are the unbiased variances, and S_{xy} be the covariance of Y and X . All these quantities are calculated through MCD estimation as Bulut and Zaman [1].

3. Proposed Class of Quantile Regression-Ratio-Type Estimators

Quantile regression is a variant of standard linear regression that calculates the conditional median of the result variable and can be used when the assumptions of linear regression are not fulfilled. Quantiles are points in a distribution that corresponds to the rank order of the distribution's values. The median is the value in the sorted sample that falls in the middle (middle quantile, $q^{50th} = 0.50$). Interested readers may refer to studies by Koenker and Bassett [31], Koenker and Hallock [32], Koenker [33], Hao et al. [34], and Korkmaz and Chesneau [35].

So, with the presence of outliers and based on equation (4), A_1 and A_2 are $(0, 1)$, and we propose a class of quantile regression-ratio-type estimators as

$$\bar{y}_{pi} = \bar{y}_M + \hat{\beta}_{(q)}(\bar{X}_M - \bar{x}_M), \quad i = 1, 2, \dots, 5, \quad (6)$$

with

$$\hat{\beta}_{(q)} = \arg \min_{\beta \in RP} \rho_q(v) \sum_{i=1}^n (y_i - < x_i, \beta >), \quad (7)$$

where $\bar{y}_M, \bar{X}_M, \bar{x}_M$ are based on MCD, $\hat{\beta}_{(q)}$ is the quantile regression, and $\rho_q(v)$ is a continuous piecewise linear function (or asymmetric absolute loss function), for quantile $q \in (0, 1)$, but nondifferentiable at $v = 0$. Note that $\hat{\beta}_{(q)}$ is the quantile regression coefficient for $p = 2$ variables. The MSE of the proposed family of estimators is

$$MSE(\bar{y}_{pi}) = \left(\frac{1-f}{n} \right) \left[\bar{Y}_M^2 \psi_y^2 - 2B_i \bar{X}_M \bar{Y}_M \rho_{xy} \psi_x \psi_y + B_i^2 \bar{X}_M^2 \psi_x^2 \right], \quad i = 1, 2, \dots, 5. \quad (8)$$

For the purposes of the current study, it is worth mentioning that we are using $q^{14th} = 0.14$, $q^{24th} = 0.24$, $q^{34th} = 0.34$, $q^{44th} = 0.44$, and $q^{54th} = 0.54$ quantiles. We see from the consequences of the numerical study conducted in

next section that utilizing the quantile regression coefficients, based on these referenced quantiles, incredibly enhance the efficiencies of proposed estimators. Note that investigation of these five referenced quantiles leads to

propose a class containing five members. For sack of readability, let us provide five members of the proposed class with their MSE in a compact form, as follows:

$$\begin{aligned} \bar{y}_{pi} &= \begin{cases} \bar{y}_{p1} = \bar{y}_M + \hat{\beta}_{(0.14)}(\bar{X}_M - \bar{x}_M); (q) = (0.14), \\ \bar{y}_{p2} = \bar{y}_M + \hat{\beta}_{(0.24)}(\bar{X}_M - \bar{x}_M); (q) = (0.24), \\ \bar{y}_{p3} = \bar{y}_M + \hat{\beta}_{(0.34)}(\bar{X}_M - \bar{x}_M); (q) = (0.34), \\ \bar{y}_{p4} = \bar{y}_M + \hat{\beta}_{(0.44)}(\bar{X}_M - \bar{x}_M); (q) = (0.44), \\ \bar{y}_{p5} = \bar{y}_M + \hat{\beta}_{(0.54)}(\bar{X}_M - \bar{x}_M); (q) = (0.54), \end{cases} \\ \text{MSE}(\bar{y}_{pi}) &= \begin{cases} \left(\frac{1-f}{n} \right) \left[\bar{Y}_M^2 \psi_y^2 - 2B_{(0.14)} \bar{X}_M \bar{Y}_M \rho_{xy} \psi_x \psi_y + B_{(0.14)}^2 \bar{X}_M^2 \psi_x^2 \right]; (q) = (0.14), \\ \left(\frac{1-f}{n} \right) \left[\bar{Y}_M^2 \psi_y^2 - 2B_{(0.24)} \bar{X}_M \bar{Y}_M \rho_{xy} \psi_x \psi_y + B_{(0.24)}^2 \bar{X}_M^2 \psi_x^2 \right]; (q) = (0.24), \\ \left(\frac{1-f}{n} \right) \left[\bar{Y}_M^2 \psi_y^2 - 2B_{(0.34)} \bar{X}_M \bar{Y}_M \rho_{xy} \psi_x \psi_y + B_{(0.34)}^2 \bar{X}_M^2 \psi_x^2 \right]; (q) = (0.34), \\ \left(\frac{1-f}{n} \right) \left[\bar{Y}_M^2 \psi_y^2 - 2B_{(0.44)} \bar{X}_M \bar{Y}_M \rho_{xy} \psi_x \psi_y + B_{(0.44)}^2 \bar{X}_M^2 \psi_x^2 \right]; (q) = (0.44), \\ \left(\frac{1-f}{n} \right) \left[\bar{Y}_M^2 \psi_y^2 - 2B_{(0.54)} \bar{X}_M \bar{Y}_M \rho_{xy} \psi_x \psi_y + B_{(0.54)}^2 \bar{X}_M^2 \psi_x^2 \right]; (q) = (0.54). \end{cases} \end{aligned} \quad (9)$$

4. Numerical Illustration

In this section, the performance of proposed and existing estimators through two real-life applications and simulation study is presented.

4.1. Real-Life Applications

Population 1. In this dataset, “amount of nonreal estate farm loans during 1977” is taken as auxiliary variable (X), while “amount of real estate farm loans during 1977” is taken as study variable (Y). Furthermore, $n = 20$ is selected for $N = 50$. For remaining characteristics of the population interested readers may refer to Singh [36].

Population 2. We use “UScereals” dataset, which describes 65 widely available breakfast cereals in the USA, depending on the information available on the mandatory food label on the packet. The measurements are normalized here to a serving size of one American cup. The data come from ASA Statistical Graphics Exposition and are used by Venables and Ripley [37]. The dataset contains a number of variables. So, “grams of sodium in one portion” is taken as auxiliary

variable (X), while “Number of calories” is taken as study variable (Y). Furthermore, $n = 20$ is selected for $N = 65$. For remaining characteristics of the population, interested readers may refer to Venables and Ripley [37].

For diagnostic checking, we should apply robust regression techniques on the referenced dataset; let us check the presence of outliers in (X, Y) individually through the Dixon chi-squared test for outliers presented by the Dixon [38, 39] test. An R package “outliers” is used for this purpose. The results are provided in Table 2.

Table 2 provides significant results in terms of p values, hence providing clear indication of the presence of outliers in our considered datasets. In light of DT, we can say that traditional OLS is not suitable for our dataset. So, there is a need to incorporate some sort of techniques, which can provide us better results in the presence of outliers. Therefore, we apply robust and quantile regression with MCD estimation. The results associated with Populations 1 and 2 are available, respectively, in Tables 3 and 4.

5. Simulation Study

In the current subsection, an assessment of proposed and some existing estimators through Monte Carlo simulation is

considered. The simulation design is organized as follows. A random variable $X_i \sim G(2.7, 3.9)$ and random variable Y_i is defined as $Y_i = h + RX_i + \varepsilon X_i^p$. Here, it is assumed that $p = 1.7, h = 6, R = 2.2$, and ε has standard normal distribution with population of size $N = 1000$. The simple random sampling (SRS) is considered for $n = 200$. The SRS sampling has been replicated 1000 times. We examine empirical MSEs of $\bar{y}_{bzi}, \bar{y}_{reg}$, and $\bar{y}_{Ni}$ as $MSE = (\sum_{i=1}^{k'} (z_i - \bar{Z})^2 / k')$. The results are available in Table 5.

6. Discussion

As we see, the MSE results real-life examples and simulation study for each existing estimator based on $\beta_{(rob)}$, while their corresponding proposed estimators are based on $\beta_{(q)}$ in

Tables 3–5. Each row of these tables structured as given follows:

$$\begin{aligned}\hat{\theta}_1 &= \bar{y}_{bz1}, \bar{y}_{bz6}, \bar{y}_{bz11}, \bar{y}_{bz16}, \bar{y}_{bz21}, \bar{y}_{bz26}, \bar{y}_{bz31}, \bar{y}_{p1}, \\ \hat{\theta}_2 &= \bar{y}_{bz2}, \bar{y}_{bz7}, \bar{y}_{bz12}, \bar{y}_{bz17}, \bar{y}_{bz22}, \bar{y}_{bz27}, \bar{y}_{bz32}, \bar{y}_{p2}, \\ \hat{\theta}_3 &= \bar{y}_{bz3}, \bar{y}_{bz8}, \bar{y}_{bz13}, \bar{y}_{bz18}, \bar{y}_{bz23}, \bar{y}_{bz28}, \bar{y}_{bz33}, \bar{y}_{p3}, \\ \hat{\theta}_4 &= \bar{y}_{bz4}, \bar{y}_{bz9}, \bar{y}_{bz14}, \bar{y}_{bz19}, \bar{y}_{bz24}, \bar{y}_{bz29}, \bar{y}_{bz34}, \bar{y}_{p4}, \\ \hat{\theta}_5 &= \bar{y}_{bz5}, \bar{y}_{bz10}, \bar{y}_{bz15}, \bar{y}_{bz20}, \bar{y}_{bz25}, \bar{y}_{bz30}, \bar{y}_{bz35}, \bar{y}_{p5}.\end{aligned}\quad (10)$$

The determination of best estimators is related to minimum values of MSE associated with each estimator.

Regarding Table 3, the MSE of proposed and existing estimators for Pop-1 associated to $\hat{\theta}_1 - \hat{\theta}_5$ can be ordered as follows:

$$\begin{aligned}MSE_{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3} &: \text{Quantile} < \text{LAD} < \text{LMS} < \text{MM} < \text{LTS} < \text{Huber} < \text{Tukey} < \text{Hampel}, \\ MSE_{\hat{\theta}_4} &: \text{LMS} < \text{MM} < \text{LTS} < \text{Quantile} < \text{Huber} < \text{Tukey} < \text{Hampel} < \text{LAD}, \\ MSE_{\hat{\theta}_5} &: \text{LAD} < \text{Quantile} < \text{LMS} < \text{MM} < \text{LTS} < \text{Huber} < \text{Tukey} < \text{Hampel}.\end{aligned}\quad (11)$$

On the other hand, within five values of θ , quantile estimators perform best with $\hat{\theta}_1$, LAD performs best with $\hat{\theta}_3$, and all other estimators perform best with $\hat{\theta}_4$.

Regarding Table 4, the proposed estimators record the best performance among all competitors' estimators for Pop-2, where the MSE of proposed and existing estimators associated to $\hat{\theta}_1 - \hat{\theta}_5$ can be ordered as follows:

$$\begin{aligned}MSE_{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\theta}_5} &: \text{Quantile} < \text{LMS} < \text{LAD} < \text{MM} < \text{Tukey} < \text{LTS} < \text{Hampel} < \text{Huber}, \\ MSE_{\hat{\theta}_4} &: \text{Quantile} < \text{Huber} < \text{LTS} < \text{Hampel} < \text{Tukey} < \text{MM} < \text{LAD} < \text{LMS}.\end{aligned}\quad (12)$$

Also, within five values of θ , quantile estimators perform best with $\hat{\theta}_1$, LAD and LMS estimators perform best with $\hat{\theta}_5$, and all other estimators perform best with $\hat{\theta}_4$.

Based on Table 5, again the proposed estimators record the best performance among all compared estimators in this simulation study, where the MSE of proposed and existing estimators associated to $\hat{\theta}_1 - \hat{\theta}_5$ can be ordered as follows:

$$\begin{aligned}MSE_{\hat{\theta}_1, \hat{\theta}_2} &: \text{Quantile} < \text{LMS} < \text{LTS} < \text{LAD} < \text{Huber} < \text{Tukey} < \text{MM} < \text{Hampel}, \\ MSE_{\hat{\theta}_3} &: \text{Quantile} < \text{MM} < \text{Tukey} < \text{Hampel} < \text{LTS} < \text{LAD} < \text{LMS} < \text{Huber}, \\ MSE_{\hat{\theta}_4} &: \text{Quantile} < \text{Huber} < \text{Hampel} < \text{LMS} < \text{MM} < \text{LAD} < \text{LTS} < \text{Tukey}, \\ MSE_{\hat{\theta}_5} &: \text{Quantile} < \text{Hampel} < \text{Tukey} < \text{LTS} < \text{LAD} < \text{MM} < \text{Huber} < \text{LMS}.\end{aligned}\quad (13)$$

Furthermore, within five values of θ , all estimators appear the best performance with $\hat{\theta}_3$.

Overall, with real data and simulation, the proposed estimators record the best or near-best performance

TABLE 1: Reviewed robust ratio type estimators.

Estimator	$b_{(i)}$	a	b
$\bar{y}_{bz1}$	$b_{(lad)}$	1	0
$\bar{y}_{bz2}$	$b_{(lad)}$	1	ψ_x
$\bar{y}_{bz3}$	$b_{(lad)}$	1	$\psi_2(x)$
$\bar{y}_{bz4}$	$b_{(lad)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz5}$	$b_{(lad)}$	ψ_x	$\psi_2(x)$
$\bar{y}_{bz6}$	$b_{(lms)}$	1	0
$\bar{y}_{bz7}$	$b_{(lms)}$	1	ψ_x
$\bar{y}_{bz8}$	$b_{(lms)}$	1	$\psi_2(x)$
$\bar{y}_{bz9}$	$b_{(lms)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz10}$	$b_{(lms)}$	ψ_x	$\psi_2(x)$
$\bar{y}_{bz11}$	$b_{(lts)}$	1	0
$\bar{y}_{bz12}$	$b_{(lts)}$	1	ψ_x
$\bar{y}_{bz13}$	$b_{(lts)}$	1	$\psi_2(x)$
$\bar{y}_{bz14}$	$b_{(lts)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz15}$	$b_{(lts)}$	ψ_x	$\psi_2(x)$
$\bar{y}_{bz16}$	$b_{(hbm)}$	1	0
$\bar{y}_{bz17}$	$b_{(hbm)}$	1	ψ_x
$\bar{y}_{bz18}$	$b_{(hbm)}$	1	$\psi_2(x)$
$\bar{y}_{bz19}$	$b_{(hbm)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz20}$	$b_{(hbm)}$	ψ_x	$\psi_2(x)$
$\bar{y}_{bz21}$	$b_{(hpm)}$	1	0
$\bar{y}_{bz22}$	$b_{(hpm)}$	1	ψ_x
$\bar{y}_{bz23}$	$b_{(hpm)}$	1	$\psi_2(x)$
$\bar{y}_{bz24}$	$b_{(hpm)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz25}$	$b_{(hpm)}$	ψ_x	$\psi_2(x)$
$\bar{y}_{bz26}$	$b_{(tkz)}$	1	0
$\bar{y}_{bz27}$	$b_{(tkz)}$	1	ψ_x
$\bar{y}_{bz28}$	$b_{(tkz)}$	1	$\psi_2(x)$
$\bar{y}_{bz29}$	$b_{(tkz)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz30}$	$b_{(tkz)}$	ψ_x	$\psi_2(x)$
$\bar{y}_{bz31}$	$b_{(hmm)}$	1	0
$\bar{y}_{bz32}$	$b_{(hmm)}$	1	ψ_x
$\bar{y}_{bz33}$	$b_{(hmm)}$	1	$\psi_2(x)$
$\bar{y}_{bz34}$	$b_{(hmm)}$	$\psi_2(x)$	ψ_x
$\bar{y}_{bz35}$	$b_{(hmm)}$	ψ_x	$\psi_2(x)$

TABLE 2: Results of the Dixon test (DT).

Pop-1		
For X	DT = 7.9097	P value = 0.004917
For Y	DT = 9.1764	P value = 0.002452
Pop-2		
For X	DT = 17.73	P value = 2.546e-05
For Y	DT = 21.679	P value = 3.224e-06

TABLE 3: MSE of the proposed and existing estimators for Pop-1.

θ	LAD	LMS	LTS	Huber	Hampel	Tukey	MM	Quantile
$\hat{\theta}_1$	2319.381	4321.468	4391.462	4929.630	5058.167	4986.624	4388.699	1698.552
$\hat{\theta}_2$	2312.594	4307.505	4377.314	4914.131	5042.362	4970.989	4374.559	1700.399
$\hat{\theta}_3$	1714.319	2619.321	2660.991	2992.650	3074.471	3028.818	2659.339	1701.123
$\hat{\theta}_4$	2354.633	1698.216	1699.574	1737.322	1752.655	1743.848	1699.504	1703.688
$\hat{\theta}_5$	1724.877	2689.750	2732.966	3075.895	3160.265	3113.199	2731.253	1931.589

TABLE 4: MSE of the proposed and existing estimators for Pop-2.

θ	LAD	LMS	LTS	Huber	Hampel	Tukey	MM	Quantile
$\hat{\theta}_1$	62.87724	62.81852	98.56723	107.89060	106.55715	98.36911	98.24331	30.71491
$\hat{\theta}_2$	62.70337	62.64480	98.31496	107.62158	106.29046	98.11721	97.99164	31.24573
$\hat{\theta}_3$	49.18216	49.13758	77.86217	85.68823	84.56317	77.69694	77.59206	31.70550
$\hat{\theta}_4$	42.15273	42.18798	33.24866	32.61970	33.67139	34.26920	34.28240	31.18453
$\hat{\theta}_5$	38.19166	38.16310	58.92870	65.05818	64.16924	58.80075	58.71956	31.91410

TABLE 5: MSE of proposed and existing estimators for the simulation study.

θ	LAD	LMS	LTS	Huber	Hampel	Tukey	MM	Quantile
$\hat{\theta}_1$	0.4013	0.3930	0.3986	0.4029	0.4061	0.4050	0.4052	0.0080
$\hat{\theta}_2$	0.0993	0.09531	0.0980	0.1001	0.1017	0.1011	0.1013	0.0054
$\hat{\theta}_3$	0.0036	0.0037	0.0035	0.0039	0.0033	0.0032	0.0030	0.0022
$\hat{\theta}_4$	0.0046	0.0044	0.0048	0.0041	0.0043	0.0049	0.0045	0.0027
$\hat{\theta}_5$	0.0055	0.0059	0.0053	0.0058	0.0050	0.0052	0.0057	0.0032

compared to other robust regression estimators. In view of both these results, the proposed estimators can be considered as very robust and reasonable estimators.

7. Conclusion

Bulut and Zaman used robust minimum covariance determinant (MCD) estimation to create a new class of robust regression-type ratio estimators. In this article, drawing inspiration from Bulut and Zaman's work, we propose to use quantile regression with MCD estimator-based location measures under a simple sampling scheme to introduce a class of quantile regression-type mean estimators to the assessment of the population mean with the appearance of outliers. The MSEs of the proposed estimators are also obtained. The performance of the proposed and some existing robust regression estimators are assessed through simulation and two real-life applications, where the Dixon chi-squared test is considered to assess the existence of outliers in the real-life datasets. Based on numerical comparisons, it is obvious that proposed estimators outperform or near outperform existing robust estimators across the considered variety of datasets. As a result, the proposed estimators may be of interest, and they will almost certainly increase the possibility of getting additional accurate estimates of the population mean when outliers exist. Also, these estimators may be developed under the ranked set sampling method, as given by Al-Omari [40], Al-Omari and Almanjahie [41], and Haq et al. [42].

Data Availability

The datasets used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] H. Bulut and T. Zaman, "An improved class of robust ratio estimators by using the minimum covariance determinant estimation," *Communications in Statistics—Simulation and Computation*, pp. 1–7, 2019.
- [2] J. Neyman, "Contribution to the theory of sampling human populations," *Journal of the American Statistical Association*, vol. 33, no. 201, pp. 101–116, 1938.
- [3] W. G. Cochran, "The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce," *The Journal of Agricultural Science*, vol. 30, no. 2, pp. 262–275, 1940.
- [4] M. N. Murthy, "Product method of estimation," *The Indian Journal of Statistics, Series A*, vol. 26, no. 1, pp. 69–74, 1964.
- [5] M. P. Singh, "Ratio cum product method of estimation," *Metrika*, vol. 12, no. 1, pp. 34–42, 1967.
- [6] S. Shalabh and J. R. Tsai, "Ratio and product methods of estimation of population mean in the presence of correlated measurement errors," *Communications in Statistics—Simulation and Computation*, vol. 46, no. 7, pp. 5566–5593, 2017.
- [7] N. Koyuncu, "Efficient estimators of population mean using auxiliary attributes," *Applied Mathematics and Computation*, vol. 218, no. 22, pp. 10900–10905, 2012.
- [8] M. Abid, N. Abbas, and M. Riaz, "Improved modified ratio estimators of population mean based on deciles," *Chiang Mai Journal of Science*, vol. 43, no. 1, pp. 1311–1323, 2016.
- [9] M. Abid, N. Abbas, H. Z. Nazir, and Z. Lin, "Enhancing the mean ratio estimators for estimating population mean using non-conventional location parameters," *Revista Colombiana de Estadística*, vol. 39, no. 1, pp. 63–79, 2016.
- [10] M. Irfan, M. Javed, M. Abid, and Z. Lin, "Improved ratio type estimators of population mean based on median of a study variable and an auxiliary variable," *Haceteppe Journal of Mathematics and Statistics*, vol. 47, no. 3, pp. 659–673, 2018.
- [11] U. Shahzad, P. F. Perri, and M. Hanif, "A new class of ratio-type estimators for improving mean estimation of non-sensitive and sensitive variables by using supplementary information," *Communications in Statistics—Simulation and Computation*, vol. 48, no. 9, pp. 2566–2585, 2019.
- [12] 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, 2019.
- [13] S. K. Yadav and T. Zaman, “Use of some conventional and non-conventional parameters for improving the efficiency of ratio-type estimators,” *Journal of Statistics and Management Systems*, pp. 1–24, 2021.
 - [14] 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 and Arts*, vol. 20, no. 4, pp. 789–800, 2020.
 - [15] C. Tanış, B. Saraçoğlu, C. Kuş, A. Pekgör, and K. Karakaya, “Transmuted lower record type fréchet distribution with lifetime regression analysis based on type I-censored data,” *Journal of Statistical Theory and Applications*, vol. 20, no. 1, pp. 86–96, 2021.
 - [16] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, Wiley Series in Probability and Mathematical Statistics, New York, NY, USA, 1987.
 - [17] 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.
 - [18] T. Zaman and E. Toksoy, “Improvement in estimating the population mean in simple random sampling using information on two auxiliary attributes and numerical application in agricultural engineering,” *Fresenius Environmental Bulletin*, vol. 28, no. 6, pp. 4584–4590, 2019.
 - [19] T. Zaman, “Improvement of modified ratio estimators using robust regression methods,” *Applied Mathematics and Computation*, vol. 348, pp. 627–631, 2019.
 - [20] 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.
 - [21] R. V. Hogg, “Adaptive robust procedures: a partial review and some suggestions for future applications and theory,” *Journal of the American Statistical Association*, vol. 69, no. 348, pp. 909–923, 1974.
 - [22] P. J. Huber, “Robust regression: asymptotics, conjectures and monte carlo,” *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.
 - [23] F. R. Hampel, “A general qualitative definition of robustness,” *The Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887–1896, 1971.
 - [24] J. W. Tukey, *Exploratory Data Analysis*, Addison-Wesley, Boston, MA, USA, 1977.
 - [25] V. J. Yohai, “High breakdown-point and high efficiency robust estimates for regression,” *Annals of Statistics*, vol. 15, no. 2, pp. 642–656, 1987.
 - [26] P. J. Rousseeuw, “Least median of squares regression,” *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, 1984.
 - [27] P. Rousseeuw, “Multivariate estimation with high breakdown point,” in *Mathematical statistics and applications*, W. Grossmann, G. Pflug, I. Vincze, and W. Wertz, Eds., vol. B, pp. 283–297, Reidel Publishing Company, Dordrecht, Netherlands, 1985.
 - [28] P. J. Rousseeuw and K. V. Driessen, “A fast algorithm for the minimum covariance determinant estimator,” *Technometrics*, vol. 41, no. 3, pp. 212–223, 1999.
 - [29] N. H. Al-Noor and A. A. Mohammad, “Model of robust regression with parametric and non-parametric methods,” *Mathematical Theory and Modeling*, vol. 3, no. 5, pp. 27–39, 2013.
 - [30] M. Hubert, M. Debruyne, and P. J. Rousseeuw, “Minimum covariance determinant and extensions,” *WIREs Computational Statistics*, vol. 10, no. 3, pp. 1–11, 2018.
 - [31] R. Koenker and G. Bassett, “Regression quantiles,” *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.
 - [32] R. Koenker and K. F. Hallock, “Quantile regression,” *Journal of Economic Perspectives*, vol. 15, no. 4, pp. 143–156, 2001.
 - [33] R. Koenker, *Quantile Regression*, Cambridge University Press, New York, NY, USA, 2005.
 - [34] L. Hao, D. Q. Naiman, and D. Q. Naiman, *Quantile regression*, p. 149, Sage Publications, Thousand Oaks, CA, USA, 2007.
 - [35] M. Ç. Korkmaz and C. Chesneau, “On the unit burr-XII distribution with the quantile regression modeling and applications,” *Computational and Applied Mathematics*, vol. 40, no. 1, pp. 1–26, 2021.
 - [36] S. Singh, *Advanced Sampling Theory with Applications. How Michael Selected Amy*, Kluwer Academic Publishers, Dordrecht, Netherlands, 2003.
 - [37] W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S-PLUS*, Springer, Berlin, Germany, Third Edition, 1999.
 - [38] W. J. Dixon, “Analysis of extreme values,” *The Annals of Mathematical Statistics*, vol. 21, no. 4, pp. 488–506, 1950.
 - [39] W. J. Dixon, “Ratios involving extreme values,” *The Annals of Mathematical Statistics*, vol. 22, no. 1, pp. 68–78, 1951.
 - [40] A. I. Al-Omari, “Maximum likelihood estimation in location-scale families using varied L ranked set sampling,” *RAIRO — Operations Research*, vol. 55, pp. S2759–S2771, 2021.
 - [41] A. Ibrahim Al-Omari and I. M. Almanjahie, “New improved ranked set sampling designs with an application to real data,” *Computers, Materials and Continua*, vol. 67, no. 2, pp. 1503–1522, 2021.
 - [42] A. Haq, J. Brown, E. Moltchanova, and A. I. Al-Omari, “Paired doubleranked set sampling,” *Communications in Statistics-Theory and Methods*, vol. 45, no. 1, pp. 2873–2889, 2016.