Hindawi
Mathematical Problems in Engineering
Volume 2020, Article ID 1243583, 13 pages
https://doi.org/10.1155/2020/1243583

Research Article

Modified Least Trimmed Quantile Regression to Overcome Effects of Leverage Points

Habshah Midi,[1] **Taha Alshaybawee**,[1,2] **and Mohammed Alguraibawi**[3]

[1]*Faculty of Science and Institute for Mathematical Research, Universiti Putra, Malaysia 43400 UPM, Serdang Selangor, Malaysia*
[2]*Department of Statistics, College of Administration and Economics, University of Al-Qadisiyah, Al Diwaniyah, Iraq*
[3]*Technical Institute of Dewaniya, Al-Furat Al-Awsat Technical University, Al-Qadisiyah, Al Diwaniyah, Iraq*

Correspondence should be addressed to Habshah Midi; habshahmidi@gmail.com

Received 25 November 2019; Revised 17 March 2020; Accepted 4 May 2020; Published 12 June 2020

Academic Editor: Emilio Gómez-Déniz

Quantile regression estimates are robust for outliers in y direction but are sensitive to leverage points. The least trimmed quantile regression (LTQReg) method is put forward to overcome the effect of leverage points. The LTQReg method trims higher residuals based on trimming percentage specified by the data. However, leverage points do not always produce high residuals, and hence, the trimming percentage should be specified based on the ratio of contamination, not determined by a researcher. In this paper, we propose a modified least trimmed quantile regression method based on reweighted least trimmed squares. Robust Mahalanobis' distance and GM6 weights based on Gervini and Yohai's (2003) cutoff points are employed to determine the trimming percentage and to detect leverage points. A simulation study and real data are considered to investigate the performance of our proposed methods.

1. Introduction

Quantile regression (QReg) has received much attention since the seminal work of Koenker and Bassett [1]. It can be considered as one of the important statistical breakthroughs in recent decades. The desirable advantages of quantile regression led to its application in wide areas of sciences such as in medicine, financial, economics, agriculture, environment, and others [2, 3]. QReg is an extension of the mean regression model to conditional of the different quantiles of the response variable distribution. Therefore, QReg is able to provide much more detailed stochastic relationship among random variables.

Consider the following regression model:

$$y_i = x_i'\beta + \varepsilon_i, \tag{1}$$

where y_i is an $(n \times 1)$ vector of a response variable, x_i is a $(k \times 1)$ vector of covariates variables, β is a vector of unknown parameters, and ε_i is an $(n \times 1)$ vector of error terms. For any τ-quantiles in the interval (0, 1), the parameter β_τ can be estimated consistently as the solution to the following optimization problem:

$$\min_{\beta} \sum_{i=1}^{n} \rho_\tau (y_i - x_i'\beta_\tau), \tag{2}$$

where $\rho_\tau(.)$ is the check function, defined as

$$\rho_\tau(u) = u(\tau - I(u < 0)), \tag{3}$$

where $I(.)$ denotes the indicator function.

One of the important advantages of quantile regression is the insensitivity for outliers and heavy tailed distribution for error term. This robustness of QReg for outliers arises because of the nature of the check function which is shown in (3) (see [3–5]). Similar to M-estimator regression, QReg is not robust when the predictor variables contain outliers which are called high leverage points (HLPs) [6]. There are some attempts to overcome the effect of HLPs and maximize the breakdown point of QReg. Giloni et al. [7] proposed a weighting method to increase the breakdown point and cope with HLP, based on the blocked adaptive computationally

efficient outlier nominators (BACON) method that is proposed by Billor et al. (2000), in which a clean subset is chosen via their algorithm. The limitation of the weighting method is that it can be used with small numbers of regressers (often, one or two regresser variables). Rousseeuw and Hubert [8] proposed the regression depth as an extended version for regression quantile. They pointed out that the depth quantiles is robust to HLPs. Adrover et al. [9] presented a robust estimation method that is unaffected by leverage points and, at the same time, maximizes the breakdown point. The disadvantages of the weighting method and depth quantiles are computational complexity and nonstandard asymptotic distributions Neykov et al. [10].

Recently, least trimmed QReg is proposed by Neykov et al. [10] to reduce the effects of HLPs. This method is a generalization of the location estimator that was proposed by Tableman [11] and least trimmed absolute deviation proposed by Hawkins and Olive [12]. Neykov et al. [10] proved the consistency of the least trimmed quantile regression method and discussed the breakdown point of the estimators. The limitation of this method is that the trimming percentage is a constant whereby the trimmed data may be lower or higher than the actual contamination percentage of the data. The least trimmed quantile method minimizes the quantile residuals in (2) for the subset (h) out of the sample size (n). However, it is important to mention that the leverage point is not affected by residuals. Therefore, this method does not correctly detect the high leverage points.

In this paper, we proposed a new algorithm to develop the least trimmed quantile regression method and to overcome these disadvantages in the existing methods. The new proposed algorithm integrates the reweighted least trimmed method that proposed by Čížek [13] with QReg to determine the trimming percentage and robust Mahalanobis' distance to identify the HLPs. In addition, we employ the Gervini and Yohai [14] technique to compute the cutoff point and new weights for the QReg. Besides that, RMD is used to detect the leverage points.

2. Least Trimmed Quantile Regression (LTQReg)

Least trimmed squares (LTS) method is a robust estimation technique proposed by Rousseeuw [15] by minimizing the following objective function:

$$\min_{\beta} \sum_{i=1}^{h} \varepsilon_{(i)}^{2}(\beta), \tag{4}$$

where $\varepsilon_{(i)}^{2}(\beta)$ is the i-th order statistic squares residuals, $\varepsilon_{1}^{2}(\beta)$, $\varepsilon_{2}^{2}(\beta)$, ..., $\varepsilon_{n}^{2}(\beta)$, and (h) is a subset out of (n). The trimming constant $h = n(1-\alpha)+1$, where α is a ratio of trimming. When $h = (n+p+1)/2$, the highest breakdown point of the LTS estimator will be achieved (0.50) [16]. Roozbeh and Hamzah [17] developed the LTS method for restricted semiparametric regression models. Based on the least trimmed squares (LTS) method, robust ridge and nonridge type estimators were developed by Roozbeh [18] in semiparametric model regression when the errors are dependent. Roozbeh et al. [19] introduced some alternative robust estimators based on a penalization scheme, whereas a nonlinear integer programming was used.

Neykov et al. [10] proposed the least trimmed quantile regression (LTQReg) as an efficient and robust method to overcome the effect of HLPs on QReg. LTQReg is defined as follows:

$$\widehat{\beta}_{\tau} = \arg\min_{\beta} \sum_{i=1}^{h} \rho_{\tau}(\varepsilon_{i}(\beta)), \tag{5}$$

where $\rho_{\tau}(\varepsilon_{i}(\beta))$ is defined as in (2) and (3). Neykov et al. [10] proved that when the trimming constant $h = (n+N(X)+1)/2$, the breakdown point of LTQReg estimator is asymptotically equal to 0.50, where $N(X)$ is the maximum number of explanatory variables. Müller [20] and Neykov et al. [10] pointed out that $N(X) = p-1$.

The LTQReg method is based on the smallest quantile errors to reduce the influence of leverage points. In this situation, we would like to ask the following question: is the error values of the QReg and LS will be high for all leverage points? We most answer to this question by the following example. Let us consider the simple linear regression model $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, with a sample size of $n = 50$, and the independent variable x_i is uniformly distributed $[-1, 1]$. Let $\beta_0 = \beta_1 = 1$ and $\varepsilon_i \sim N(0, 1)$.

In order to illustrate the effect of leverage points and outliers on the error term, 20% of the observations are contaminated by replacing the first 10 observations with contaminated observations. We consider three cases of contaminations: outliers, HLPs, and both outliers and HLPs simultaneously. The first 10 observations of the explanatory variable and dependent variable are contaminated as follows: $x_i \sim U(-50, 50)$ and $y_i \sim U(50, 100)$. Least squares (LS) and QReg at three quantiles (0.25, 0.50, and 0.75) were then applied to the data. In this example, we want to investigate if the LS and QReg produced high errors in all contamination scenarios which are suitable for LTQReg.

For all the three cases of contaminations, the fitted residuals are plotted as shown in Figures 1–3.

Figures 1–3 clarify influences of outliers, HLPs, and both on LS and QReg in different quantiles. In Figures 1 and 3, we can see clearly that when the data are contaminated by outliers, the first 10 observations have highest residuals for both LS and QReg in different quantiles. On the contrary, in Figure 2, when the data are contaminated by HLPs, the residuals of LS and QReg are not affected by HLPs. From Figures 1–3, we can conclude that the outlier observations have a direct effect on the residuals, whereas the leverage points have no effect on the residuals. Hence, we can say that the LTQReg method is not an effective method to reduce the effect of leverage points because it is based on trimming the highest $(n-h)$ residuals.

3. Modified Least Trimmed Quantile Regression (MLTQReg)

In this section, we will discuss the modified LTQReg method to determine the rate of contamination data and the best

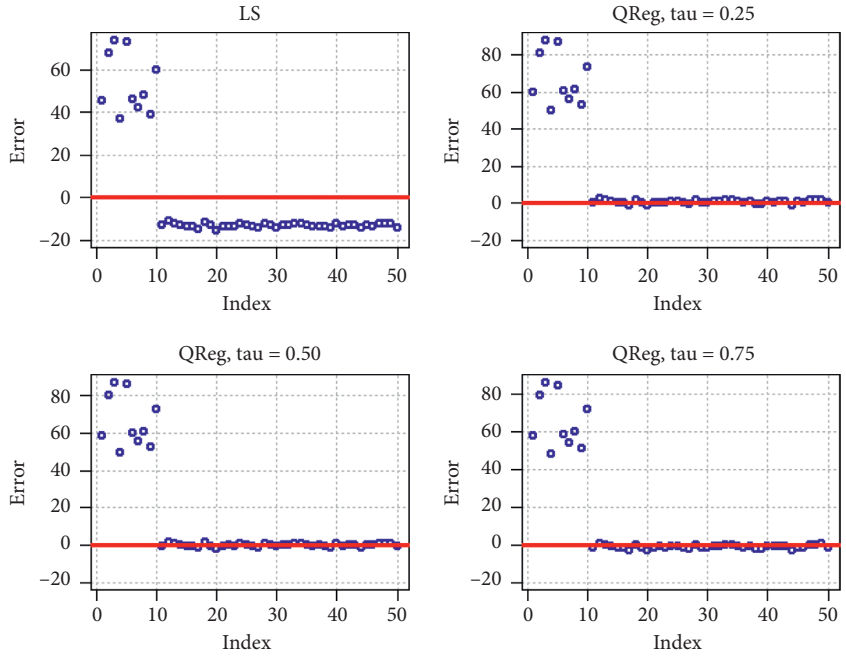

Figure 1: The fitted residuals for LS and QReg at three quantiles (0.25, 0.50, and 0.75) for contaminated data with outliers.

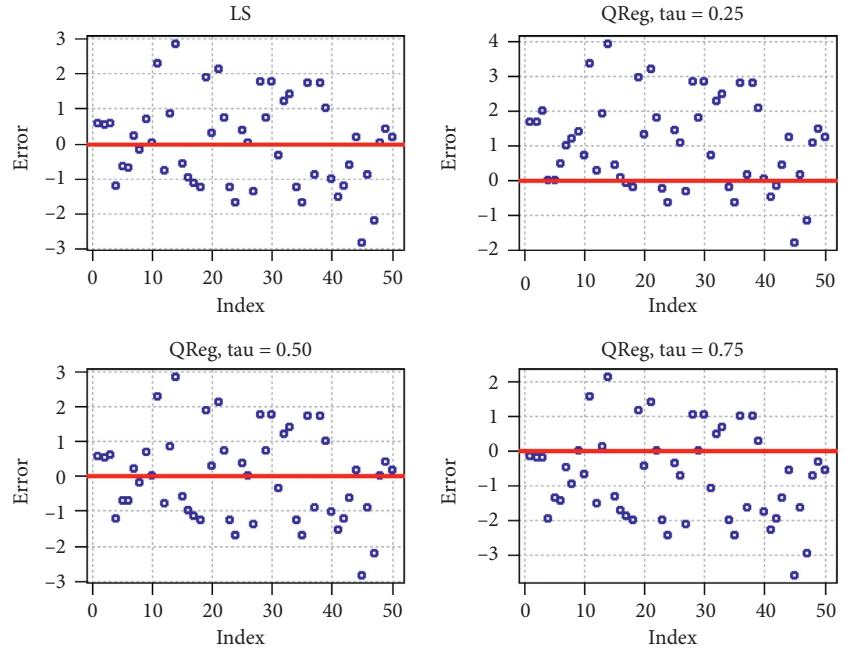

Figure 2: The fitted residuals for LS and QReg at three quantiles (0.25, 0.50, and 0.75) for contaminated data with HLPs.

trimming percentage. Three modified methods will be discussed in this section based on the reweighted least trimmed squares (RWLTS) method which was proposed by Čížk [13] depending on hard rejection weights [16] and combined to the LTQReg method to robustify the weighted least squares method. The hard rejection weights in the RWLTS method are defined as

$$w_i = \begin{cases} 1, & \text{if } |u_i| < t_n, \\ 0, & \text{if } |u_i| \geq t_n, \end{cases} \tag{6}$$

where $u_i's$ are the standardized of regression residuals and $t_n > 0$ is the cutoff point that was adapted by Gervini and Yohai [14]. The cutoff point value is computed by comparing

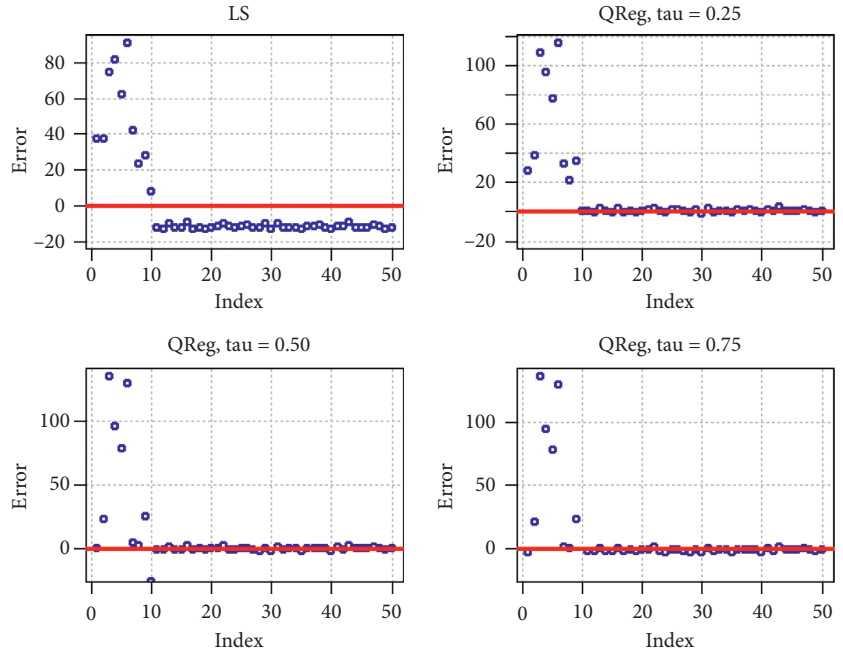

Figure 3: The fitted residuals for LS and QReg at three quantiles (0.25, 0.50, and 0.75) for contaminated data with both outliers and HLPs.

the empirical distribution function G_n^+ of standardized absolute residuals with the distribution function G_0^+ of absolute residuals under the assumed model. The friction of unusual observation in the sample (d_n) can be measured as

$$d_n = \underbrace{\sup}_{t \geq k} \{G_n^+(t) - G_0^+(t)\}^+, \tag{7}$$

where $k = 2.5$ [16]. Therefore, the t_n value is set to the $(1 - d_n)$th quantile of $G_n^+(t)$ as follows:

$$t_n = \min\{t\colon G_n^+(t) \geq 1 - d_n\}. \tag{8}$$

The procedure of the reweighted least trimmed square method [13] can be described in two steps. The first step is determining the trimming constant h based on the weights that are given in (6), defined as

$$h = \sum_{i=1}^{n} w_i. \tag{9}$$

The second step is applying the LTS method depending on the trimming constant h that is computed in the first step.

To increase breakdown points of the proposed method, a high breakdown estimator LTS, LMS, or S are used as initial (see [13, 14]) and the robust weights are used to improve the efficiency.

Next, we will describe three algorithms based on the RWLTS to improve the LTQReg [10].

3.1. Reweighted Least Trimmed Quantile Regression (RWLTQReg). In this method, we combine the RWLTS with LTQReg to determine the trimming constant, and the algorithm for this method can be describe as follows:

Step 1. Consider the LTQReg estimator as an initial estimate with high breakdown point and compute the standardized residual u_i for $i = 1, \ldots, n$.

Step 2. Calculate hard rejection weights for the standardized residuals as

$$w_i^{\text{Res}} = I\{u_i < t_n\}, \tag{10}$$

where t_n is the cutoff point of Gervini and Yohai [14] that is shown in (8).

Step 3. Calculate the trimming constant (h_n) based on the weights in equation (10), from the formula $h_n = \sum_{i=1}^{n} w_i^{\text{Res}}$.

Step 4. Applying the LTQReg based on the algorithm that proposed by Neykov et al. [10] for the subset of the size h_n, this procedure can be described as follows:

(i) Set $r = 0$, select a subset with the size h_n from the sample.

(ii) For the subset h_n, use the QReg to estimate the coefficients $(\widehat{\beta}_\tau^r)$.

(iii) For all observations in the sample, compute the residuals and then order the residuals as $u_{(1)}(\widehat{\beta}_\tau^r) \leq u_{(2)}(\widehat{\beta}_\tau^r), \ldots, \leq u_{(n)}(\widehat{\beta}_\tau^r)$, $i = 1, \ldots, n$.

(iv) Then, set $r = r + 1$ and the new subset is considered the first h_n.

(v) For the new subset, Steps (ii), (iii), and (iv) were repeated. This procedure is repeated until convergence.

3.2. Modified Least Trimmed Quantile Regression Based on RMD (RMD-LTQReg). In this algorithm, we used the modified least trimmed quantile regression (MLTQReg) and

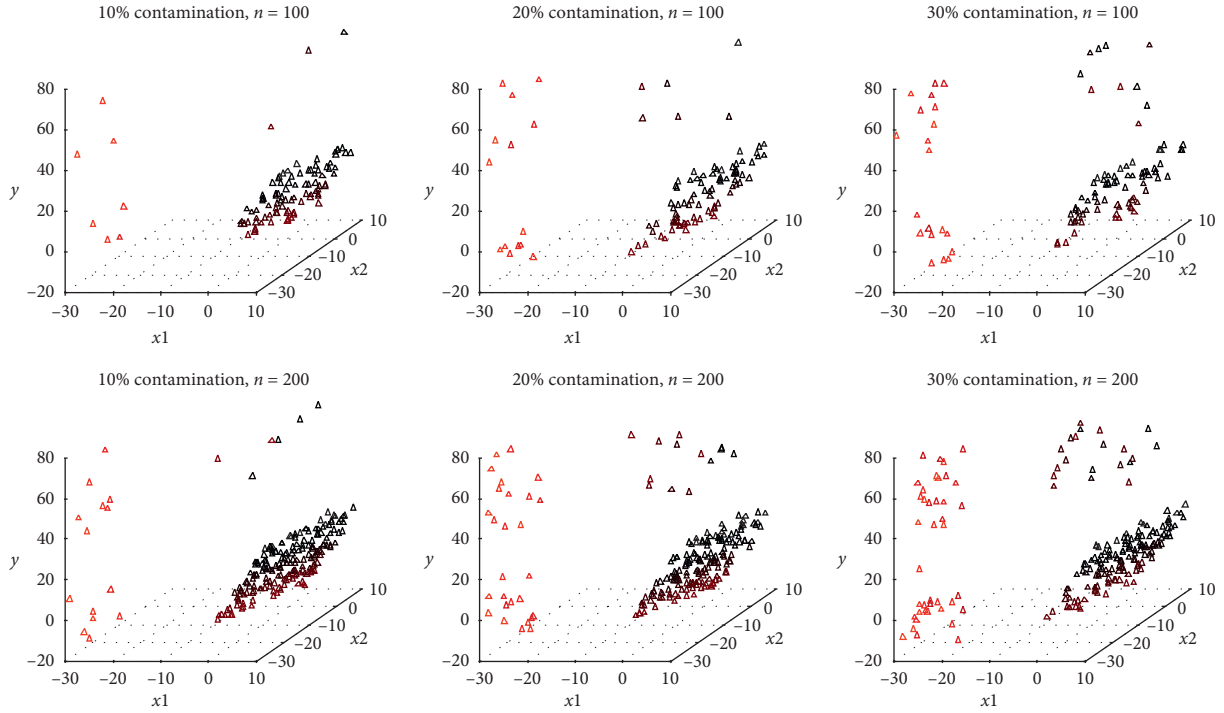

FIGURE 4: The spread shape of the generated data in the first experiment at the three levels of contamination (0.10, 0.20, and 0.30) when the sample size $n = 100$ and 200.

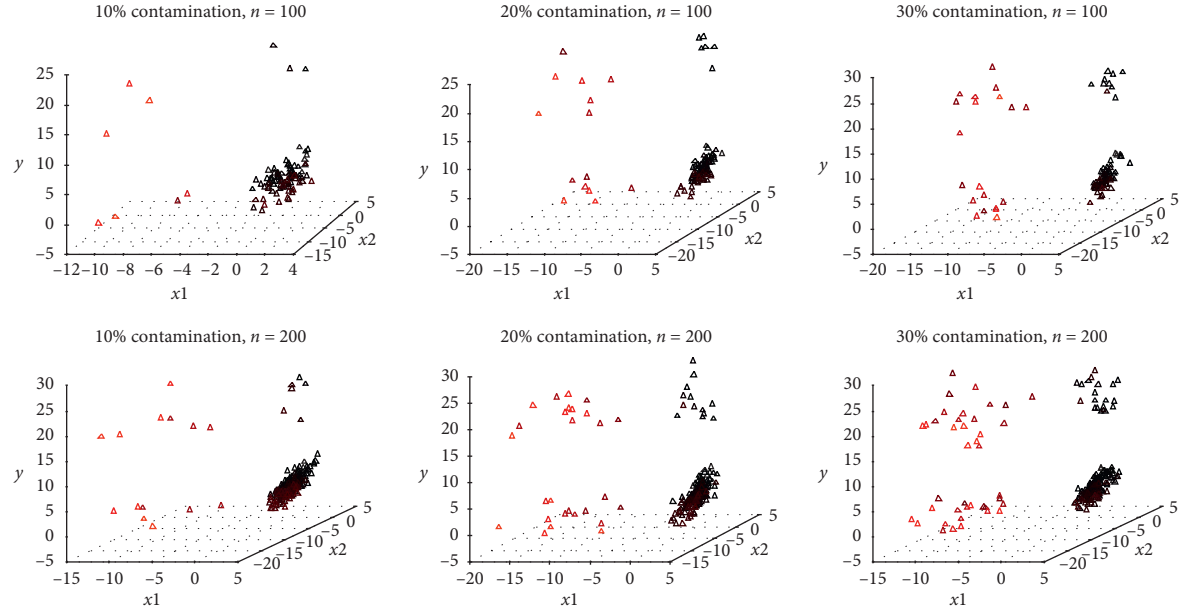

FIGURE 5: The spread shape for generated data in the second experiment at the three levels of contamination (0.10, 0.20, and 0.30) when the sample size $n = 100$ and 200.

RMD to detect the leverage points. The algorithm is presented as follows:

Step 1. Compute the RMD as follows:

$$\mathrm{RMD}(x_i) = \sqrt{(x_i - T(X))' C(X)^{-1} (x_i - T(X))}, \quad (11)$$

where $T(X)$ and $C(X)$ are the location and the shape estimates of MVE.

Step 2. Compute $w_i^{\mathrm{RMD}} = I\{\mathrm{RMD}(x_i) < K\}$, where K is the cutoff point computed as follows:

$$K = \mathrm{median}(\mathrm{RMD}(x_i)) + c * \mathrm{mad}(\mathrm{RMD}(x_i)), \quad (12)$$

Table 1: RMSE and MAE values for QReg, LTQReg (20%), LTQReg (30%), RWLTQReg, RMD-LTQReg, and GM6-LTQReg at the three quantiles 0.25, 0.50, and 0.75, in the different contamination levels 10%, 20%, and 30% for the first experiment.

Sample size	Contamination rate		10%		20%		30%	
	Tau	Methods	RMSE	MAE	RMSE	MAE	RMSE	MAE
100	0.25	QReg	23.7652	7.9886	34.5250	15.6395	29.6037	18.6166
		LTQReg (20%)	1.0111	0.7997	1.3567	0.9870	17.5208	7.0508
		LTQReg (30%)	0.8256	0.6668	0.9691	0.7708	1.2864	0.9324
		RWLTQReg	0.5645	0.4622	0.6352	0.5171	4.5239	3.7996
		RMD-LTQReg	0.6183	0.5049	0.6562	0.5310	0.6784	0.5492
		GM6-LTQReg	0.9434	0.7543	1.0365	0.8177	1.2492	0.9029
	0.50	QReg	23.2466	7.7603	22.8781	13.7647	26.8218	17.0167
		LTQReg (20%)	0.7788	0.6114	1.1549	0.8252	8.3452	6.7991
		LTQReg (30%)	0.5973	0.4803	0.7407	0.5839	6.7834	5.5281
		RWLTQReg	0.3919	0.3157	1.3174	1.0960	4.0146	3.2788
		RMD-LTQReg	0.4389	0.3544	0.4641	0.3739	0.4857	0.3917
		GM6-LTQReg	0.6926	0.5512	0.9085	0.6744	7.6986	6.1182
	0.75	QReg	17.1983	11.0353	23.7787	19.3859	28.4886	24.1319
		LTQReg (20%)	4.0678	3.3538	14.9824	12.4491	22.1116	18.6020
		LTQReg (30%)	3.6118	2.9751	13.1301	10.9304	19.6022	16.4446
		RWLTQReg	4.1062	3.3109	10.1311	8.1356	15.6859	12.7729
		RMD-LTQReg	0.6267	0.5088	0.6854	0.5560	0.7710	0.6227
		GM6-LTQReg	2.9753	1.2881	16.1177	13.1634	24.4130	20.5151
200	0.25	QReg	24.1181	8.1252	32.1859	14.4208	31.0170	18.4829
		LTQReg (20%)	1.0604	0.8539	1.4644	1.0847	15.3258	6.3742
		LTQReg (30%)	0.8840	0.7276	1.0497	0.8465	1.4660	1.0884
		RWLTQReg	0.6256	0.5240	0.7207	0.5994	0.8308	0.6854
		RMD-LTQReg	0.6479	0.5396	0.7301	0.6030	0.8101	0.6651
		GM6-LTQReg	1.0388	0.8399	1.1369	0.9069	1.2498	0.9808
	0.50	QReg	23.5486	7.9332	23.5468	12.8543	26.4878	16.6496
		LTQReg (20%)	0.8420	0.6735	1.2309	0.8951	7.9088	5.8414
		LTQReg (30%)	0.6562	0.5365	0.8053	0.6448	1.2468	0.8983
		RWLTQReg	0.4006	0.3322	0.7993	0.6658	1.1482	0.9557
		RMD-LTQReg	0.4748	0.3919	0.5244	0.4302	0.5696	0.4662
		GM6-LTQReg	0.7638	0.6176	0.8888	0.7017	1.2270	0.8142
	0.75	QReg	17.2519	11.0336	24.1359	19.3725	28.2692	23.0904
		LTQReg (20%)	1.0653	0.8564	12.7057	10.6259	19.9887	16.5240
		LTQReg (30%)	0.8839	0.7258	10.9569	9.1502	17.4184	14.3180
		RWLTQReg	2.7474	2.2629	8.4783	6.8885	13.6166	11.0773
		RMD-LTQReg	0.6651	0.5525	0.7408	0.6122	0.8233	0.6760
		GM6-LTQReg	1.3506	1.0279	13.5634	11.3055	21.4860	17.7558

Table 2: RMSE and MAE values for QReg, LTQReg (20%), LTQReg (30%), RWLTQReg, RMD-LTQReg, and GM6-LTQReg at the three quantiles 0.25, 0.50, and 0.75, in the different contamination levels 20%, 30%, and 40% for the second experiment.

Sample size	Contamination rate		10%		20%		30%	
	Tau	Methods	RMSE	MAE	RMSE	MAE	RMSE	MAE
100	0.25	QReg	7.60072	2.71137	12.63955	5.40649	15.29570	7.76377
		LTQReg (20%)	0.96439	0.79551	1.12911	0.90803	2.57547	1.52490
		LTQReg (30%)	0.83776	0.69587	0.94068	0.77194	1.10528	0.89317
		RWLTQReg	0.61278	0.50790	0.70393	0.58018	0.76134	0.62801
		RMD-LTQReg	0.61412	0.50472	0.63576	0.51919	0.60821	0.49545
		GM6-LTQReg	1.00462	0.82615	1.04673	0.85208	1.07611	0.87303
	0.50	QReg	7.03403	2.49861	9.50562	4.90036	11.69555	6.27158
		LTQReg (20%)	0.75641	0.61555	1.34668	1.08131	2.43778	1.87786
		LTQReg (30%)	0.62499	0.51160	1.04447	0.85431	1.67420	1.36147
		RWLTQReg	0.39408	0.32070	0.68870	0.56272	1.01328	0.82816
		RMD-LTQReg	0.44134	0.35757	0.45001	0.36310	0.45131	0.36386
		GM6-LTQReg	0.75162	0.61254	1.10578	0.90201	1.49676	1.22331
	0.75	QReg	6.48270	2.98638	10.87028	5.68746	13.19635	6.99549
		LTQReg (20%)	1.27777	1.04650	2.76438	2.25273	3.10530	2.52145
		LTQReg (30%)	1.19103	0.98439	2.32918	1.91100	2.54114	2.09051
		RWLTQReg	0.93936	0.77860	1.59655	1.31319	1.73952	1.43453
		RMD-LTQReg	0.62520	0.51422	0.64571	0.52944	0.64247	0.52357
		GM6-LTQReg	1.33548	1.08809	2.51296	2.05925	2.42449	1.99687
200	0.25	QReg	8.09625	2.93802	12.50005	5.39671	15.54302	7.89287
		LTQReg (20%)	0.97660	0.81400	1.12126	0.91353	2.58969	1.53480
		LTQReg (30%)	0.85519	0.71915	0.93652	0.78048	1.10876	0.90673
		RWLTQReg	0.65875	0.55684	0.69650	0.58754	0.76961	0.64687
		RMD-LTQReg	0.62150	0.52297	0.61801	0.51920	0.62247	0.52132
		GM6-LTQReg	1.03788	0.85881	1.05223	0.86622	1.08009	0.88672
	0.50	QReg	7.46737	2.69586	9.44630	4.89712	11.81101	6.27112
		LTQReg (20%)	0.76073	0.62729	1.14512	0.92500	2.45982	1.97166
		LTQReg (30%)	0.62627	0.52152	0.92752	0.76556	1.81734	1.48822
		RWLTQReg	0.41374	0.34492	0.68605	0.57105	1.07019	0.89003
		RMD-LTQReg	0.41887	0.34717	0.43697	0.36076	0.41705	0.34530
		GM6-LTQReg	0.77188	0.63616	1.01386	0.82994	1.73916	1.42738
	0.75	QReg	7.45019	3.97614	10.92026	5.62669	13.21633	7.00179
		LTQReg (20%)	1.61894	1.34617	2.74805	2.27645	3.17620	2.61515
		LTQReg (30%)	1.41970	1.19561	2.36234	1.98466	2.65588	2.22840
		RWLTQReg	1.09951	0.93351	1.74217	1.48032	1.94745	1.65112
		RMD-LTQReg	0.61526	0.51805	0.63496	0.53380	0.64102	0.53548
		GM6-LTQReg	1.78427	1.46644	2.57393	2.14842	2.57785	2.16814

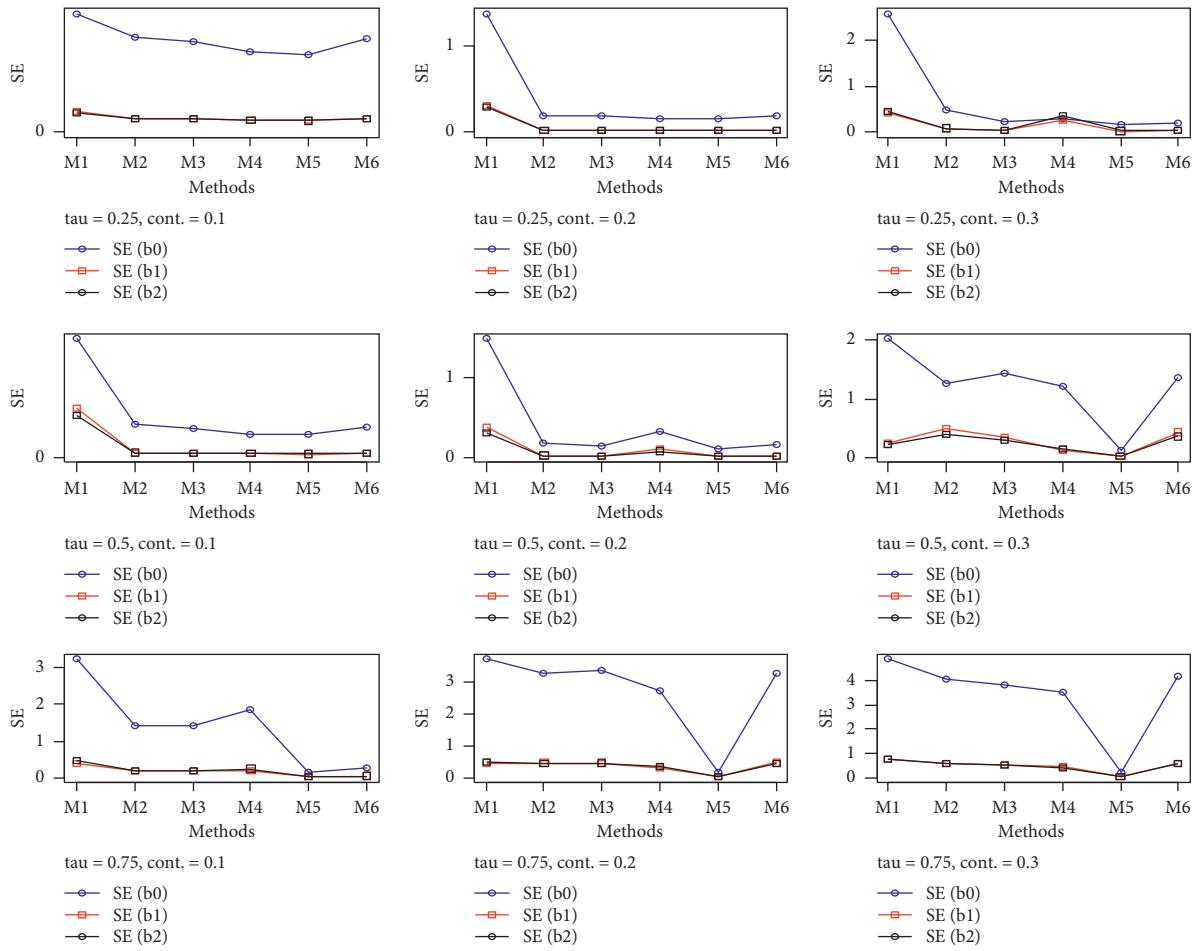

Figure 6: SE values for the estimated parameters at each quantile and contamination level for all methods (M1 = QReg, M2 = LTQReg (20%), M3 = LTQReg (30%), M4 = RWLTQReg, M5 = RMD-LTQReg, and M6 = GM6-LTQReg) in the first experiment when $n = 100$.

where $\mathrm{mad}(y_i) = \mathrm{med}\{|y_i - \mathrm{med}\, y_j|\}$ and c is a constant, 2 or 3.

Step 3. As in Step 1 and 2 of the RWLTQReg algorithm, we compute w_i^{Res} weights.

Step 4. Find the final weights by combining w_i^{RMD} with w_i^{Res} as follows:

$$w_i^{\mathrm{Fin}} = \begin{cases} w_i^{\mathrm{RMD}}, & \text{if } w_i^{\mathrm{RMD}} = w_i^{\mathrm{Res}}, \\ 0, & \text{if } w_i^{\mathrm{RMD}} \neq w_i^{\mathrm{Res}}. \end{cases} \tag{13}$$

Step 5. Hence, the trimming constant ...

Step 6. We will apply Step 4 in the RWLTQReg algorithm for the subset of the size h_n from the sample and set probability zero for the leverage points to ensure that we will not start with the bad subset (contains leverage points), Rousseeuw and Van Driessen [21], which means the condition of $w_i^{\mathrm{RMD}} \neq 0$ is satisfied (clean of leverage points).

3.3. Modified Least Trimmed Quantile Regression Based on GM6 Method (GM6-LTQReg). The GM-estimator is proposed by Schweppe (see [22]) to reduce the influence of leverage points. Adrover et al. [9] showed that the breakdown of the GM-estimator was never higher than $1/(p+1)$. Coakley and Hettmansperger [23] proposed the GM6-estimator to increase the breakdown point of the GM-estimator by using the least trimmed squares (LTS) as initial and RMD based on MVE to downweight leverage points. In this paper, we suggest using the GM6 weights to modify the LTQReg, and the procedure of this modification can be determined by following algorithm:

Step 1. For $i = 1, \ldots, n$, compute an initial estimate for the coefficients and the corresponding residuals (ε_i), the high breakdown estimators (LTQReg), Neykov et al. [10].

Step 2. Compute the scale of the residuals (se), as follows:

$$\mathrm{se} = 1.4826\big(\text{median of largest } (n-p) \text{ of the } |\varepsilon_i|\big). \tag{14}$$

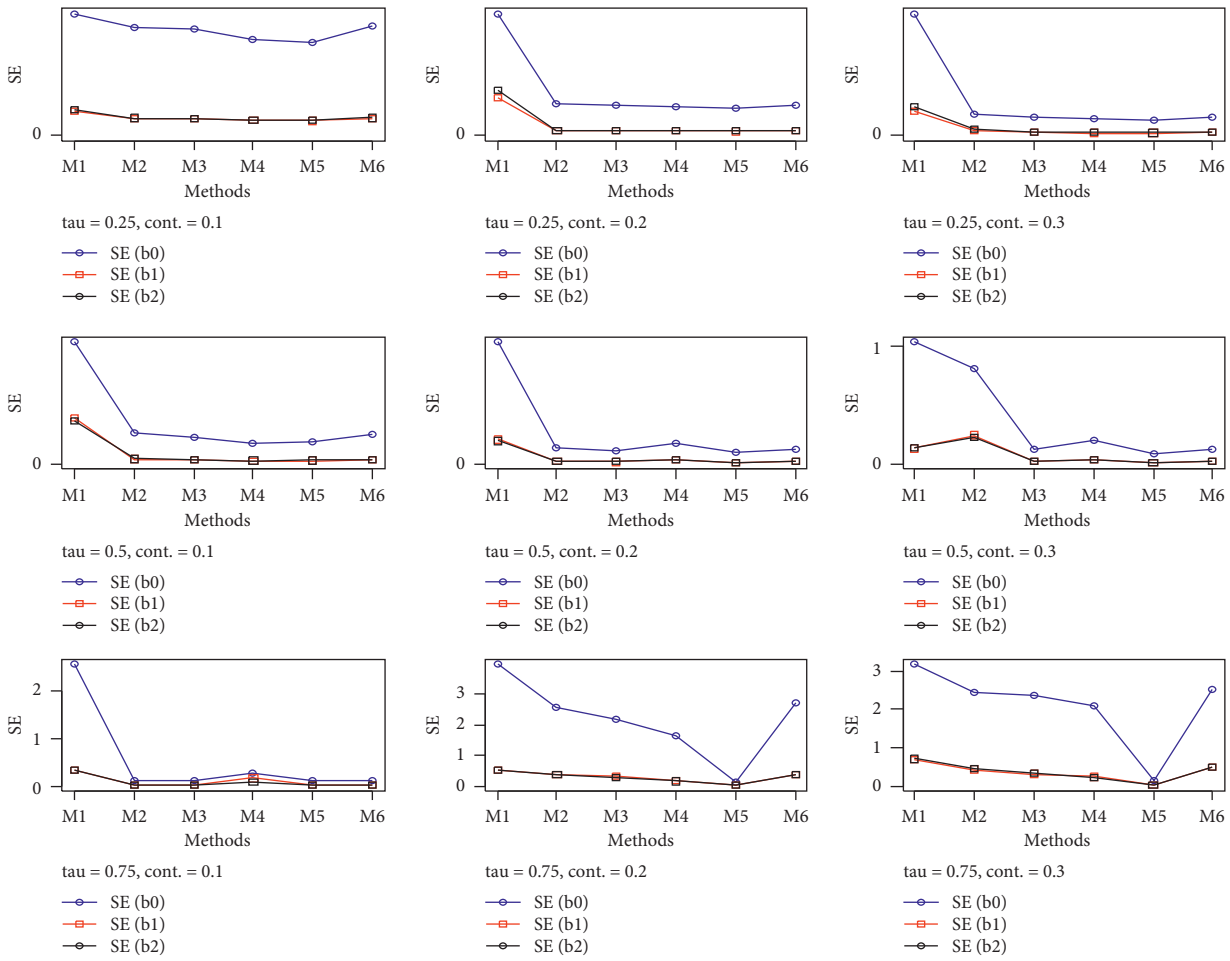

Figure 7: SE values for the estimate parameters at each quantile and contamination level for all methods (M1 = QReg, M2 = LTQReg (20%), M3 = LTQReg (30%), M4 = RWLTQReg, M5 = RMD-LTQReg, and M6 = GM6-LTQReg) in the first experiment when $n = 200$.

Step 3. Determine the standardized residuals $u_i = [\varepsilon_i / (w_i^0 \times \text{se})]$, where w_i^0 is an initial weight computed as follows:

$$w_i^0 = \min\left[1, \left(\frac{\chi^2_{(0.95,p+1)}}{\text{RMD}^2}\right)\right]. \tag{15}$$

Step 4. Hard rejection weights for the standardized residuals can be computed as follows:

$$w_i = I\{u_i < t_n\}, \tag{16}$$

where t_n is a Gervini and Yohai [14] cutoff point that is shown in (8).

Step 5. Hence, the trimming parameter will be computed as $h_n = \sum_{i=1}^{n} w_i$.

Step 6. Step 4 in the RWLTQReg algorithm for the subset of the size h_n.

4. Simulation Study

In this section, the Monte Carlo simulation study is presented to compare the performances of some existing methods such as LTQReg [10] and QR [1] with our proposed methods RWLTQReg, RMD-LTQReg, and GM6-LTQReg.

Following Neykov et al. [10], two explanatory variables (x_{i1} and x_{i2}) are generated with large sample size ($n = 100$ and 200) from the following classical heteroscedastic multiple linear regression model:

$$\begin{aligned} y_i &= b_0 + b_1 x_{i1} + b_2 x_{i2} + \sigma_i \varepsilon_i, \\ \sigma_i &= \left[\exp\left(0.11\left(x_{i1} + x_{i2}\right)\right)\right]^{1/2}, \end{aligned} \tag{17}$$

where we assume that the coefficients $b_0 = b_1 = b_2 = 1$, and the error term (ε_i) is distributed as $N(0, 1)$. Also, two experiments are considered with different distribution for explanatory and response variables with three levels of contamination ($\delta = 10\%, 20\%$, and 30%). The trimming percentage for the LTQReg will be considered as (0.20, 0.30).

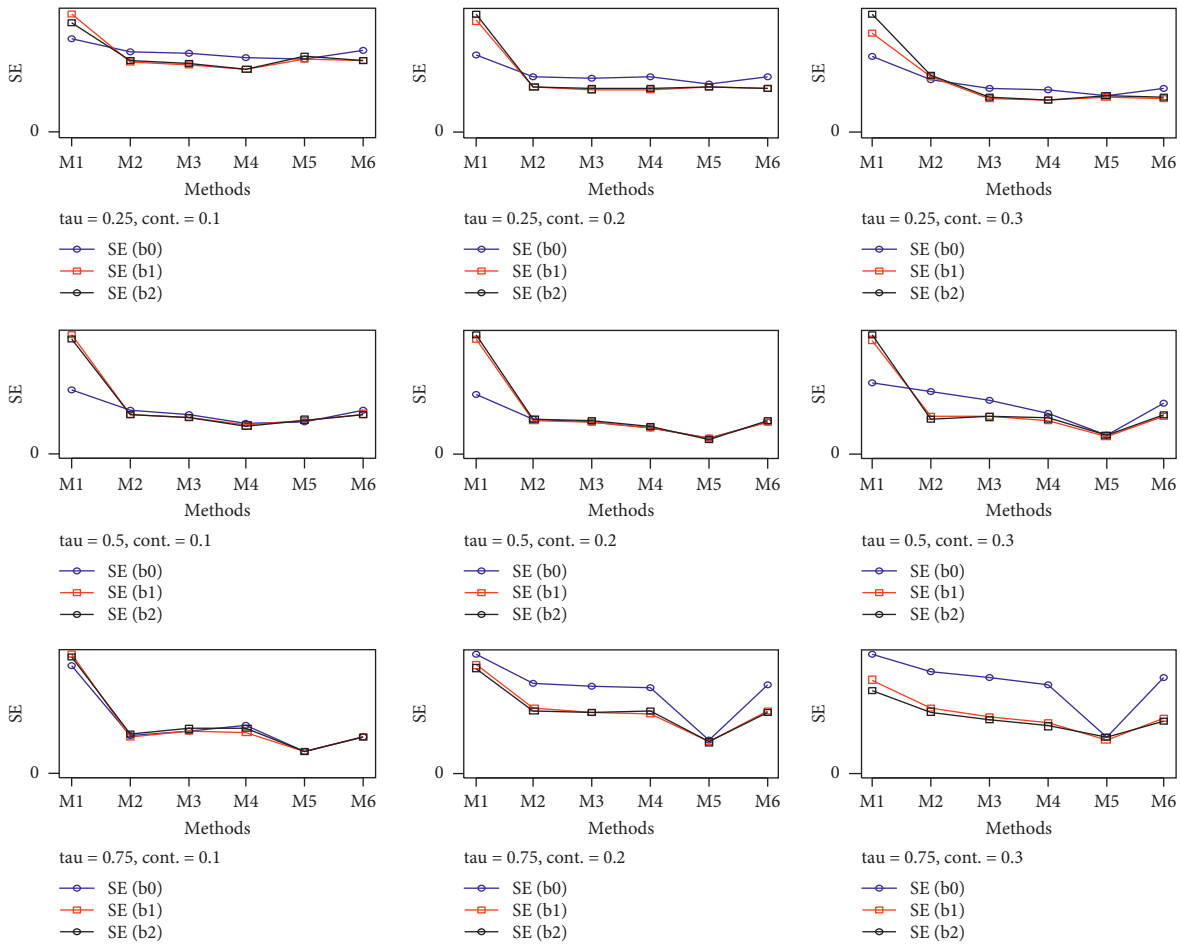

Figure 8: The SE values for the estimated parameters at each quantile and contamination level for all methods (M1 = QReg, M2 = LTQReg (20%), M3 = LTQReg (30%), M4 = RWLTQReg, M5 = RMD-LTQReg, and M6 = GM6-LTQReg) in the second experiment when $n = 100$.

4.1. The First Experiment. In this experiment, a distribution of the variables in the model is the uniform distribution (Unif) with parameters $[-10, 10]$. The variables are contaminated with different percentages, where the explanatory variables are contaminated as $x_{ij} \sim \text{Unif}(-30, -20)$, $j = 1, 2$, and the response variable is contaminated as $y_i \sim \text{Unif}(40, 80)$.

4.2. The Second Experiment. In this experiment, a distribution of explanatory variables is set as normal distribution $N(0, 1)$. The variables are contaminated with different percentages, where the explanatory variables are contaminated as $x_{ij} \sim N(-10, 3)$, $j = 1, 2$, and the response variable is contaminated as $y_i \sim N(20, 3)$.

The contamination is done by replacement of clean data by outlying data in both explanatory and response variables. Let $m = \{\delta \times n\}$, and the explanatory variables are contaminated by replacing $i = \text{intger}(m/3), \ldots, m$ clean observations with outlying observations, whereas the response variable y_i is contaminated by replacing $i = 1, \ldots, \text{intger}(2m/3)$ clean observations with outlying observations.

Figures 4 and 5 show the spread shape of generating data. It can be seen clearly that these data contain leverage points (outlying in the x direction) and outliers (outlying in the y direction) and influence observations (outlying in both x and y directions in the same time).

At the three quantiles 0.25, 0.50, and 0.75, the generated data in different contamination percentages (10%, 20%, and 30%) are fitting via the proposed methods (RWLTQReg, RMD-LTQReg, and GM6-LTQReg) and the existing methods (QReg and LTQReg). Root of mean squares errors $(\text{RMSE} = \sqrt{(1/n)\sum_{i=1}^{n}(y_i - \widehat{y}_i)^2})$ and mean absolute errors $(\text{MAE} = (1/n)\sum_{i=1}^{n}|y_i - \widehat{y}_i|)$ for the model and standard error for the parameters $(\text{SE} = \sqrt{(1/100)\sum_{r=1}^{100}(\widehat{\beta}_j^{(r)} - \beta_j^{\text{true}})^2})$ are computed to evaluate our proposed methods.

In Tables 1 and 2, we reported the RMSE and MAE values for the first and second experiments. In these tables, we can see that RMSE and MAE values for all the methods at three quantiles are shown in the rows and three levels of contamination are shown in the columns. LTQReg (20%) and LTQReg (30%) show the least trimmed quantile method with 20% and 30% trim, respectively. The results in these tables are the average of 100 replications for the two experiments of the simulation study.

Table 1 and 2 show that the RMSE and MAE values for the QReg method at the different quantiles and different

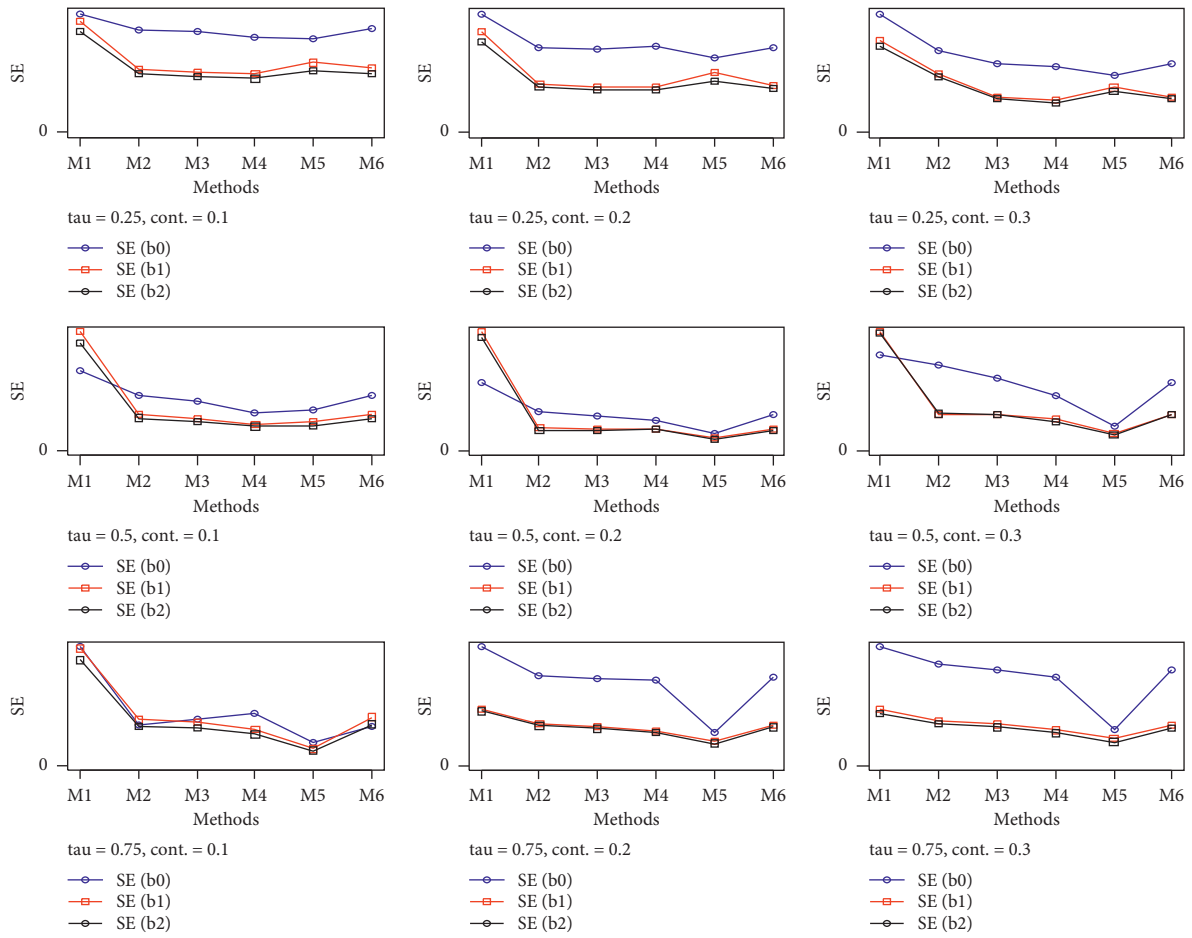

Figure 9: The SE values for the estimated parameters at each quantile and contamination levels for all methods (M1 = QReg, M2 = LTQReg (20%), M3 = LTQReg (30%), M4 = RWLTQReg, M5 = RMD-LTQReg and M6 = GM6-LTQReg) in the second experiment when $n = 200$.

Table 3: RMSE and MAE values for QReg, LTQReg (20%), LTQReg (30%), RWLTQReg, RMD-LTQReg, and GM6-LTQReg at the three quantiles 0.25, 0.50, and 0.75 for the Star Cluster CYG OB1 dataset.

Tau	0.25		0.50		0.75	
Methods	RMSE	MAE	RMSE	MAE	RMSE	MAE
QReg	0.9329833	0.6158796	0.5779843	0.46691973	0.7019406	0.5386939
LTQReg (20%)	0.3657860	0.2983784	0.3587201	0.29214575	0.4894114	0.3793147
LTQReg (30%)	0.3319383	0.2686921	0.2974692	0.24057749	0.3701667	0.2945234
RWLTQReg	0.2230726	0.1754545	0.1675843	0.13943182	0.2453746	0.2111627
RMD-LTQReg	0.1913885	0.1509400	0.1200552	0.09658939	0.1835636	0.1476853
GM6-LTQReg	0.4200258	0.3388333	0.3868742	0.31637312	0.3510218	0.2818068

levels of contamination are the highest. That is, this method is more affected than the other methods by outlying data that fall in both x and y directions. On the contrary, we can see that the proposed method RMD-LTQReg has the lowest values of RMSE and MAE in most cases. This indicates that the performance of RMD-LTQReg is better than the others. On the other hand, the RWLTQReg has better performance than the other methods except the RMD-LTQReg. In addition, we can see when the contamination levels are 20% and 30%, the GM6-LTQReg performance is better than LTQReg (20%) in most cases and LTQReg (30%) in the 30% contamination level.

In Figures 6–9, we can see that the SE values for the parameters that estimated by the RMD-LTQReg method are the smallest in almost all cases, which indicates the performance of the RMD-LTQReg method is the best among all studied methods. Also, these figures show clearly that the QReg method has high SE leads to worse performance. In addition, the rest of the methods used showed close results in most cases and were varying in some other cases. Therefore,

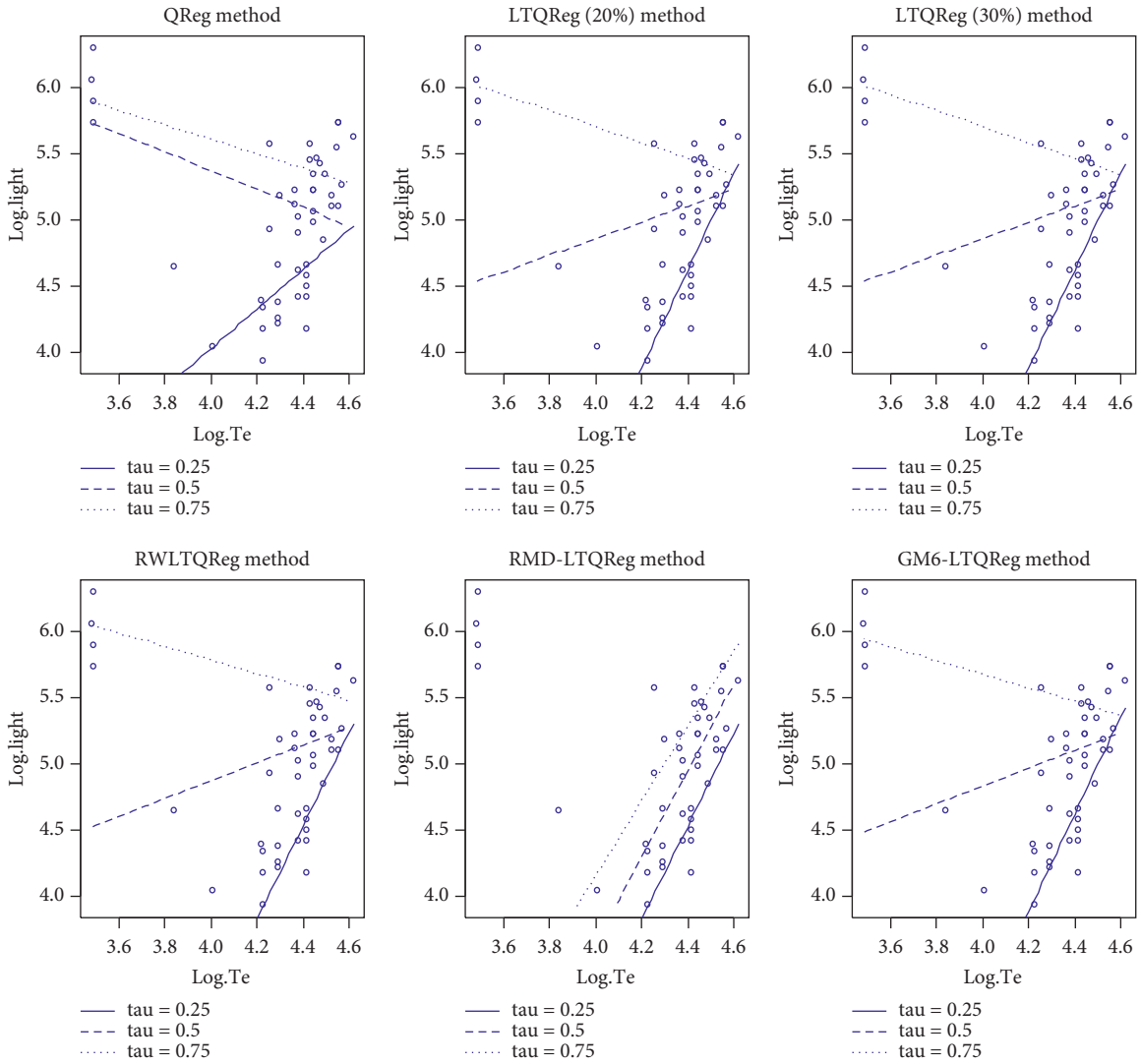

Figure 10: Regression quantile for QReg, LTQReg (20%), LTQReg (30%), RWLTQReg, RMD-LTQReg, and GM6-LTQReg for the Star Cluster CYG OB1 dataset.

it could be argued that it is difficult to determine which one is better than others.

5. Real Data Applications

In this section, the "Star Cluster CYG OB1" dataset is considered to verify the performance of our proposed methods.

5.1. Star Cluster CYG OB1. The Star Cluster CYG OB1 dataset was used by many researchers such as Rousseeuw and Leroy [24], Adrover et al. [9], and Neykov et al. [10]. This dataset contains 47 observations with one explanatory variable which is the logarithm of the effective temperature at the surface of stars. The independent variable is the logarithm of its light intensity. Rousseeuw and Leroy [24] presented that the scatterplot of this dataset shows two groups of observations. The first group includes the majority of data that contain 43 stars, whereas the second group includes the remaining four stars (the observations are 11, 20, 30, and 34). The observations 11, 20, 30 and 34 are classified as leverage points [24].

In this example, we consider three quantiles (0.25, 0.50, and 0.75) to examine robustness of our proposed methods.

Table 3 presents the RMSE and MAE values for all proposed and existing estimation methods at each quantiles. It is clear to see that the QReg method has the highest RMSE and MAE values, whereas, the RMD-LTQReg method following by the RWLTQReg method has the better performance due to they have the smallest RMSE and MAE values, whereas the RMD-LTQReg method have detected the HLPs correctly. Also, we can see that the LTQReg (30%) is better than both of GM6-LTQReg and LTQReg (20%).

Figure 10 shows the fitted residuals of regression quantiles for the existing and proposed methods. We can see that the QReg method is dramatically affected by the leverage points. Even though the RWLTQReg has lowest RMSE and MAE values in some cases, it is also affected by leverage points evident by trimming the observations that have high residuals, but it failed to trim leverage points. However, LTQReg (20%), LTQReg (30%), and GM6-LTQReg methods showed convergence in the chart and

illustrated that these methods were also affected by the HLPs but better than QReg. The proposed method RMD-LTQReg shows a good performance due to its ability to trim the leverage points.

6. Conclusions

In this paper, we proposed a new estimation method to overcome the impacts of leverage points in data. The new estimation method is called modified least trimmed quantile regression. In addition, we proposed three methods based on hard rejection weights that are used in reweighted least trimmed squares (Čížek [13]) to determine the trimming constant and to reduce the leverage point influence. In our proposed methods, the cutoff point of Gervini and Yohai [14] is employed for QReg. Moreover, Reweighted least trimmed, GM6 weights and robust Mahalanobis' distance are developed for quantile regression.

To investigate the performances of our proposed methods, a simulation study and real data are considered. The results indicate that the LTQReg has bad performance with data having leverage points due to it trims observations that have high residuals, whereas leverage points do not always have high residuals. Although, the RWLTQReg has good performance, evident by small RMSE, MAE and SE values, but it is not able to get rid of the leverage points. It is the same for the GM6-LTQReg that even though it is able to determine the trimming parameters, it is also affected by leverage points. From the results, it is clear to see that the RMD-LTQReg method is the best estimation method which can avoid the effect of leverage points.

Data Availability

The "Star Cluster CYG OB1" dataset is considered to verify the performance of our proposed method. This dataset is obtained form the basis for the main sequence in a Hertzsprung–Russell diagram of the Star Cluster CYG OB1, and it has been used by many researchers such as Rousseeuw and Leroy [24], Adrover et al. [9], and Neykov et al. [10]. It is available at package "robustbase" in *R*.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] R. Koenker and G. Bassett, "Regression quantiles," *Econometrica*, vol. 46, no. 1, pp. 33–50, 1978.

[2] K. Yu, Z. Lu, and J. Stander, "Quantile regression: applications and current research areas," *Journal of the Royal Statistical Society: Series D (The Statistician)*, vol. 52, no. 3, pp. 331–350, 2003.

[3] R. Koenker, "Quantile regression (no. 38)," *Econometric Society Monographs*, Cambridge University Press, Cambridge, UK, 2005.

[4] L. Hao and D. Q. Naiman, "Quantile regression," *Quantitative Applications in the Social Sciences*, Elsevier, Amsterdam, Netherlands, 2007.

[5] C. Davino, M. Furno, and D. Vistocco, *Quantile Regression: Theory and Applications*, John Wiley & Sons, Hoboken, NJ, USA, 2013.

[6] X. He, J. Jureckova, R. Koenker, and S. Portnoy, "Tail behavior of regression estimators and their breakdown points," *Econometrica*, vol. 58, no. 5, pp. 1195–1214, 1990.

[7] A. Giloni, J. S. Simonoff, and B. Sengupta, "Robust weighted LAD regression," *Computational Statistics & Data Analysis*, vol. 50, no. 11, pp. 3124–3140, 2006.

[8] P. J. Rousseeuw and M. Hubert, "Regression depth," *Journal of the American Statistical Association*, vol. 94, no. 446, pp. 388–402, 1999.

[9] J. Adrover, R. A. Maronna, and V. J. Yohai, "Robust regression quantiles," *Journal of Statistical Planning and Inference*, vol. 122, no. 1-2, pp. 187–202, 2004.

[10] N. M. Neykov, P. Čížek, P. Filzmoser, and P. N. Neytchev, "The least trimmed quantile regression," *Computational Statistics & Data Analysis*, vol. 56, no. 6, pp. 1757–1770, 2012.

[11] M. Tableman, "The asymptotics of the least trimmed absolute deviations (LTAD) estimator," *Statistics & Probability Letters*, vol. 19, no. 5, pp. 387–398, 1994.

[12] D. M. Hawkins and D. Olive, "Applications and algorithms for least trimmed sum of absolute deviations regression," *Computational Statistics & Data Analysis*, vol. 32, no. 2, pp. 119–134, 1999.

[13] P. Čížek, "Reweighted least trimmed squares: an alternative to one-step estimators," *Test*, vol. 22, no. 3, pp. 514–533, 2013.

[14] D. Gervini and V. J. Yohai, "A class of robust and fully efficient regression estimators," *The Annals of Statistics*, vol. 30, no. 2, pp. 583–616, 2002.

[15] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, 1984.

[16] P. Rousseeuw, "Multivariate estimation with high breakdown point," *Mathematical Statistics and Applications*, Springer, vol. 8, pp. 283–297, Berlin, Germany, 1985.

[17] M. Roozbeh and N. A. Hamzah, "Feasible robust estimator in restricted semiparametric regression models based on the LTS approach," *Communications in Statistics—Simulation and Computation*, vol. 46, no. 9, pp. 7332–7350, 2017.

[18] M. Roozbeh, "Robust ridge estimator in restricted semiparametric regression models," *Journal of Multivariate Analysis*, vol. 147, pp. 127–144, 2016.

[19] M. Roozbeh, S. Babaie-Kafaki, and A. Naeimi Sadigh, "A heuristic approach to combat multicollinearity in least trimmed squares regression analysis," *Applied Mathematical Modelling*, vol. 57, pp. 105–120, 2018.

[20] C. H. Müller, "Breakdown points for designed experiments," *Journal of Statistical Planning and Inference*, vol. 45, no. 3, pp. 413–427, 1995.

[21] P. J. Rousseeuw and K. Van Driessen, "Computing LTS regression for large data sets," *Data Mining and Knowledge Discovery*, vol. 12, no. 1, pp. 29–45, 2006.

[22] R. W. Hill, "Robust regression when there are outliers in the carriers: the univariate case," *Communication in Statistics—Theory and Methods*, vol. 11, no. 8, pp. 849–868, 1982.

[23] C. W. Coakley and T. P. Hettmansperger, "A bounded influence, high breakdown, efficient regression estimator," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 872–880, 1993.

[24] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection* vol. 589, John Wiley & Sons, Hoboken, NJ, USA.