



OPEN Enhancing performance in the presence of outliers with redescending M-estimators

Aamir Raza¹, Mashal Talib¹, Muhammad Noor-ul-Amin², Nevine Gunaime³, Imed Boukhris⁴ & Muhammad Nabi⁵✉

In real-life situations, we have to analyze the data that contains the atypical observations, and the presence of outliers has adverse effects on the performance of ordinary least square estimates. In this situation, redescending M-estimators, proposed by Huber (1964), are used to tackle the effects of outliers to increase the efficiency of least square estimates. In this study, we introduce a redescending M-estimator designed to generate robust estimates by mitigating the influence of outlier observations, even when the tuning constant is set to low values. This innovative estimator exhibits enhanced linearity at its core and maintains continuity throughout its range. Our proposed estimator stands out for its novelty, simplicity, differentiability, and practical applicability across real-world scenarios. The results of the proposed redescending M-estimators are compared with existing robust estimators using an extensive simulation study. Two examples based on real-life data are also added to validate the performance of the suggested function. The formulated redescending M-estimator produced efficient results as compared to all the considered redescending M-estimators.

Keywords Outliers, Redescending M-estimator, Robust, Ordinary Least Square

The ordinary least square (OLS) technique is usually applied to estimate the parameters of the regression line. The performance of the OLS depends on the assumptions of the error terms. The OLS procedure is widely applied in real-life situations for estimation, and it is regarded as an ideal technique for estimation. Nevertheless, it is recognized that OLS may perform less well when errors are not normally distributed, especially when heavy-tailed distributions or outlying data are included. The OLS's sensitivity to outliers can produce deceptive results. The robust regression technique has been created as an improved alternative in response to these difficulties. Robust regression functions can be used as a supplement to the least squares estimate, address the impact of outliers and these methodologies can also identify valid observations and can provide stability in the presence of substantial data points. Robust regression becomes an essential technique when analyzing datasets impacted by outliers, as it allows the detection of such variations and produces consistent results when data have outliers.

Huber¹ created M-estimation, the most widely used general robust regression approach. The robust estimators that are most commonly used are the following: Huber's M-Estimators Huber¹, MM-estimators (Yohai²), GM-Estimators, Siegel's Repeated Median Estimators (Rousseeuw³), Least Median of Squares (LMS) estimators, Least Trimmed Squares (LTS) estimators (Rousseeuw⁴), S-Estimators (Rousseeuw⁴), Minimum Volume Ellipsoid (MVE) estimators (Rousseeuw³), and Minimum Covariance Determinant (MCD) estimators (Rousseeuw&Driessen⁵), robust estimators for exponential families of distribution (Baraud & Chen⁶) and robust estimators for high directional linear regression models (Ghosh et al.⁷).

M-Estimators

The creation of the M-estimator embodies Huber's fundamental contribution to statistical methods. Huber's approach minimizes a function of residuals that develops more slowly, in contrast to traditional approaches that minimize the sum of the squared residuals. This method was created especially to lessen the impact of outliers in the regression analysis. The maximum likelihood formulations are used in M-estimators, which perform exceptionally well in non-normally distributed settings by figuring out the most optimal weightings for the dataset.

¹Govt. College Women University Sialkot, Sialkot, Pakistan. ²COMSATS University Islamabad-Lahore Campus, Lahore, Pakistan. ³Department of Basic Sciences, College of Science and Theoretical Studies, Saudi Electronic University, 11673 Riyadh, Saudi Arabia. ⁴Department of Physics, Faculty of Science, King Khalid University, P.O. Box 960, Abha, Saudi Arabia. ⁵Khost Mechanics Institute, Khost, Afghanistan. ✉email: m.nabi.com1990@gmail.com

To strengthen the estimator against the impact of outliers and improve the general resilience and reliability of regression analyses, this robust technique entails replacing the squared residuals used in Ordinary Least Squares (OLS) estimation with an alternative function of residuals, resulting in

$$\underset{\hat{\beta}}{\text{Minimize}} \sum_{k=1}^n \rho(p_k) \quad (1)$$

where ρ is a symmetric function with a unique minimum value at zero. The $\rho(p_k)$ is typically chosen to represent a weighting mechanism for the k th residual, deliberately giving outlier observations less weight. By purposefully reducing estimates, this intentional reduction makes estimates less vulnerable to noise, which strengthens the modeling method. Interestingly, giving a weight of zero to an observation correlates with classifying it as an outlier. Certain characteristics of a well-designed ρ function are necessary to ensure its reasonability within the strong regression framework.

1. $\rho(0) = 0$
2. $\rho(p) \geq 0$
3. $\rho(p) = (-p)(\text{symmetry})$
4. For $0 < p_1 < p_2 \Rightarrow \rho(p_1) \leq \rho(p_2)$
5. ρ is continuous (is differentiable)

By differentiating Eq. (1) with respect to the residuals p , resulting the psi function given as

$$\sum_{i=1}^n \Psi(p_i) X_i = 0 \quad (2)$$

where $\Psi(\cdot)$ is the derivative of $\rho(\cdot)$ and the maximum likelihood estimator serves as the corresponding M-estimator. After that, the weight function is obtained by dividing the psi function by the residuals that correspond to it, which is, $w_i = \frac{\Psi(p)}{p_i}$. Then the above estimated weight function can be written as

$$\sum_{i=1}^n w(p_i) X_i = 0 \quad (3)$$

Iterative methods are necessary to solve the above-described nonlinear equation system to compute M-estimators. When it comes to optimization tactics, the Iterative Reweighted Least Squares (IRLS) approach is widely acknowledged as a popular and widely used method. Because IRLS is iterative, it is a reliable and effective technique for handling the complexity involved in calculating M-estimators. The valuable contribution in the field of M-estimators is done by many researched and renowned names are Raza et al.⁸, Mukhtar et al.⁹, Luo et al.¹⁰, Anekwe & Onyeagu¹¹, Noor-ul-Amin et al.¹², Khalil et al.¹³, Alamgir et al.¹⁴, Ullah et al.¹⁵, Ali & Qadir¹⁶, Qadir¹⁷, Hampel¹⁸, Andrews¹⁹ and Beaton & Tukey²⁰.

Redescending M-estimators

One of the unique capabilities of redescending M-estimators is their complete ability to eliminate the impact of severe outliers. These estimators exhibit robustness when handling data with prominent outliers, as evidenced by their notable non-decreasing behaviour around the origin. The following list describes a few popular and highly known redescending M-estimators.

A first attempt at M-estimator specifically for regression tasks was presented by Andrews in 1974 by giving Andrews-Sine function that was a unique redescending M-estimator. This estimator is simply redescending and stands out for its increased robustness against outliers. The Ψ —function of Andrews- sine estimator is presented as

$$\Psi(p) = \begin{cases} h \operatorname{sign}\left(\frac{p}{h}\right) & |p| \leq h\pi \\ 0 & |p| > h\pi \end{cases} \quad (4)$$

where p is the OLS error term and h is the tuning constant.

Beaton & Tukey²⁰ introduced yet another redescending M-estimator known as Tukey's bi-weight function, which has gained widespread utility. The Tukey's bi-weight estimator's Ψ —function can be obtained by

$$\Psi(p) = \begin{cases} p \left[1 - \left(\frac{p}{k} \right)^2 \right]^2 & |p| \leq k \\ 0 & |p| > k \end{cases} \quad (5)$$

where k is tuning constant. Certain drawbacks of the Andrews Ψ function are efficiently addressed by Tukey's bi-weight and. It's crucial to remember that these substitutes provide some logical compliance a lesser weight.

Hampel¹⁸ is credited with introducing Hampel's three-piece-wise redescending M-estimator, commonly referred to as Hampel's three-part redescending estimators. These estimators are characterized by a psi-function that becomes 0 in the presence of significant residuals. Notably, the objective function of these M-estimators is a stepwise function, illustrating their efficacy in addressing outliers and efficiently handling regression problems. The Ψ - function of Hampel's M estimator is,

$$\Psi(p) = \begin{cases} \frac{p}{k \operatorname{sign}(p)} & |p| \leq k \\ k \frac{n-|p|}{n-h} \operatorname{sign}(r) & h < |p| \leq n \\ 0 & |p| > n \end{cases} \quad (6)$$

where h , k and n are tuning constants, $0 < k \leq h < n < \infty$. The Princeton Robustness Study revealed that his estimator performed admirably. The function is not optimal and it is not perfectly differentiable. One would prefer a smoother Ψ -function despite its success.

There has been a notable trend towards the development of techniques characterized by smoother mathematical properties and increased robustness. This trend has led to the creation of smoothly redescending M-estimators, which have been influenced by the preference for such characteristics. Qadir¹⁷ acquainted another redescending M-estimator comprehended as Qadir Beta function. The Ψ function of Qadir Beta estimator is given by

$$\Psi(p) = \begin{cases} \frac{p}{16k^4} (k+p)^2 (k-p)^2 & |p| \leq k \\ 0 & |p| > k \end{cases} \quad (7)$$

According to research done by Ali & Qadir¹⁶, the author suggested a modified version of Tukey's bi-weight function. This updated version's Ψ -function is shown as.

$$\Psi(p) = \begin{cases} \frac{2p}{3} \left[1 - \left(\frac{p}{k} \right)^4 \right]^2 & |p| \leq k \\ 0 & |p| > k \end{cases} \quad (8)$$

A redescending M-estimator was developed by Ullah et al.¹⁵ to identify outliers. The Ψ -function of Ullah's redescending M-estimator is presented by

$$\Psi(p) = p \left[1 + \left(\frac{p}{k} \right)^4 \right]^{-2} \text{ for } |p| \geq 0 \quad (9)$$

Alamgir et al.¹⁴ indicated another redescending M-estimator for robust regression. The Ψ -function of Alamgir's redescending M-estimator is presented by

$$\Psi(p) = \begin{cases} \frac{16pe^{-2\left(\frac{p}{k}\right)^2}}{\left(1+e^{-\left(\frac{p}{k}\right)^2}\right)^2} & |p| \leq k \\ 0 & |p| > k \end{cases} \quad (10)$$

Khalil et al.¹³ also introduces a redescending M-estimator. The Ψ -function of his estimator is presented by

$$\Psi(p) = \begin{cases} p \left(\frac{3}{2} \right) \left\{ 1 - \left(\frac{p}{k} \right)^4 \right\}^2 \sin \left[\left(\frac{2}{3} \right) \left\{ 1 - \left(\frac{p}{k} \right)^4 \right\}^2 \right] & |p| \leq k \\ 0 & |p| > k \end{cases} \quad (11)$$

Noor-ul-Amin et al.¹² proposed a new redescending M-estimator for robust regression whose Ψ -function is given as

$$\Psi(p) = \frac{c^2}{4} \left[\frac{\tan^{-1} \left(\frac{2p}{c} \right)^2}{4} + \frac{p^2 c^2}{c^4 + 16p^4} \right] \text{ for } |p| \geq 0 \quad (12)$$

Another new redescending M-estimator was introduced by Raza et al.⁸, he claimed that his psi function provided more efficient results. The Ψ -function of Raza given below

$$\Psi(p) = \frac{k^2}{2a} \left[1 - \left\{ 1 + \left(\frac{p}{k} \right)^2 \right\}^{-a} \right] \text{ for } |p| \geq 0 \quad (13)$$

For regression analysis, attaining the highest breakdown point is a desirable attribute of several redescending M-estimators. The M-estimator proposed by Ali et al. and Alamgir et al. are noteworthy due to its total rejection of observations with greater residuals. Although Ullah et al. attempt to offer an estimator that overcomes this constraint, it is not very generic. As a result, we have created an M-estimator with redescending characteristics that successfully fixes the issues with the estimators that were previously provided.

Proposed redescending M-estimator

We present a novel redescending M-estimator with the characteristics of redescending estimators to improve the robust regression outlier identification. Specifically created to overcome the drawbacks of traditional M-estimators, this new estimator is a cutting-edge tool for robust regression. The following explanation delves further into the characteristics and forms of the relevant Ψ -function, weight function, and ρ -function, offering a clear understanding of the unique aspects of our suggested methodology.

The proposed M-estimator's function is specified as

$$\rho(p) = \begin{cases} \frac{p^2}{810a^8} [p^8 - 30a^2p^4 + 405a^2] & \text{if } |p| \leq a \\ \frac{188a^2}{405} & \text{if } |p| > a \end{cases} \quad (14)$$

where p are residuals obtained from the OLS a works the tuning constant. We have discussed in details the shape of the objective function in Fig. 1. The suggested function is recursive and meets the required criteria. The typical characteristics are:

- $\rho(0) = 0$
- $\rho(p) \geq 0$
- $\rho(p_i) = \rho(-p_i)$ (symmetry)
- For $0 < p_i < p \Rightarrow \rho(p_i) \leq \rho(p_j)$
- ρ is continuous (is differentiable)

A series of residuals are produced using the R-program to show how well the suggested M-estimator performs. The objective function's resulting graph is shown in Fig. 1, and it clearly shows a declining trend. This graphic sheds light on the properties and performance of the suggested objective function.

Differentiating $\rho(\cdot)$ with respect to residuals, we obtained psi-function that is represented in Eq. (15)

$$\Psi(p) = \begin{cases} p \left[1 - \left(\frac{p}{\sqrt{3}a} \right)^4 \right]^2 & \text{if } |p| \leq a \\ 0 & \text{if } |p| > a \end{cases} \quad (15)$$

Using data in Fig. 1, the graphical display of $\Psi(p)$ is presented in Fig. 2.

The recommended psi-function given in Fig. 2 is a more linear and differentiable profile at the center of the curve than the considered psi-functions used for robust regression to tackle the outliers. Crucially, the proposed psi-function fulfills all the necessary conditions to build an iterative M-estimator. Additionally, this function highlights how well it captures the desired qualities for robust regression analysis by giving greater weight to values closer to the center and decreasing weight for values away from the center.

We found the corresponding weight function, which is provided, by dividing the Ψ -function by residual "r".

$$w(p) = \begin{cases} \left[1 - \left(\frac{p}{\sqrt{3}a} \right)^4 \right]^2 & \text{if } |p| \leq a \\ 0 & \text{if } |p| > a \end{cases} \quad (16)$$

The graphical representation of the weight function is displayed in Fig. 3.

Figure 3 demonstrates the proposed weight function's special robustness attributes.

Graphical comparison of proposed redescending M-estimator

A comparative study is carried out between the suggested redescending M-estimator and other well-known redescending Ψ -functions using graphical representations of Ψ -functions. Especially, the suggested M-function showed linearity greater than the considered regressive functions, guaranteeing continuous differentiability everywhere. This property increases the suggested M-estimator's overall efficacy. The graphical comparison between the proposed and existing Ψ -functions is clearly illustrated in Fig. 4, which further clarifies the beneficial aspects of the recommended methodology.

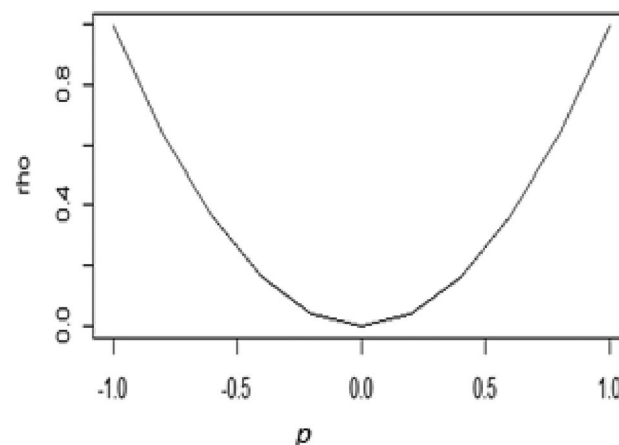


Figure 1. Graph of the proposed objective function.

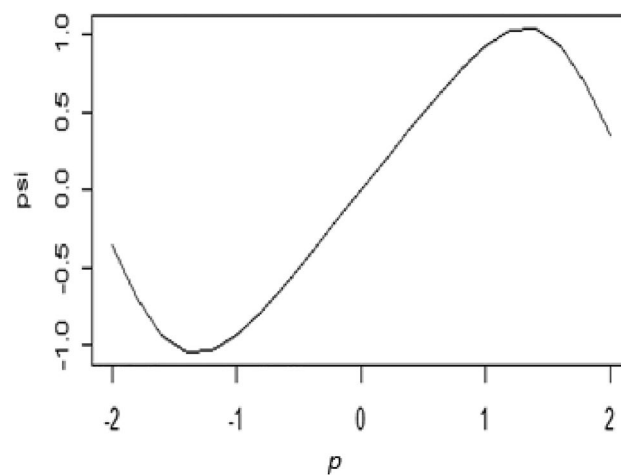


Figure 2. Graph of Ψ -function for the suggested M-Estimator.

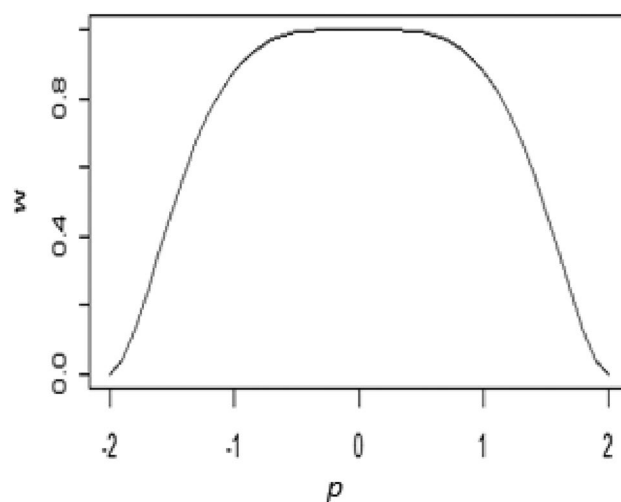


Figure 3. Graph of the weight w -function for Proposed M-estimator.

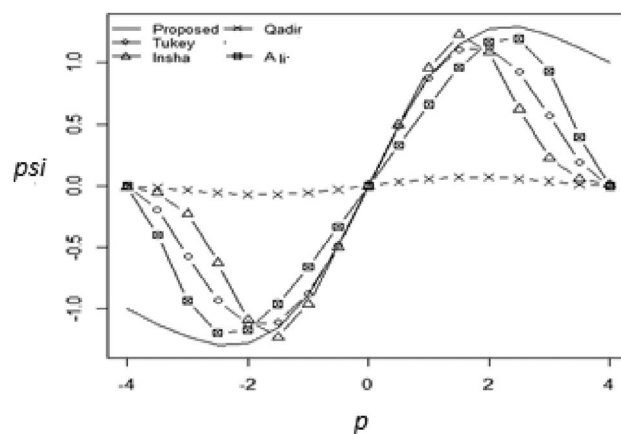


Figure 4. Combined plot of Tukey, Qadir, Ali, Ullah and Proposed Ψ -functions.

The Fig. 4 representations showed that the suggested psi function gave more weight to the central observations as compared to the considered estimators and less weight to the values that have larger outliers which is the primary objective of the proposed redescending M-estimator. The suggested estimator is continuous and differentiable everywhere within the graph.

Practical applications

In this section performance of the suggested M-estimator is demonstrated by adding detailed comparisons among the considered redescending M-estimators by providing a thorough study of real-life data examples along with simulation results. By evaluating real-world data and simulating situations, our goal is to offer a thorough grasp of the effectiveness and relative benefits of the suggested estimator in real-world applications. To attain this objective R-programing is used.

Example-1: yearly average price growth in China 1940–1948

In case I, the data is taken from the Rousseeuw’s³. This data had previously been employed by Ullah et al.¹⁵ and Raza et al.⁸ in their studies on robust regression. The dataset includes nine average annual rates of price growth that were observed between nineteen forty to nineteen forty eight. The variable of interest is the annual growth rate, and years are taken as the predictors. The average growth rates were 1.620%, 1.630%, 1.900%, 2.640%, 2.050%, 2.130%, 1.940%, 15.500%, and 364.0%.

Additionally, the war, the budget deficit, and increased government expenditure during that time all contributed to an exponential rise in prices that resulted in an important increase in rate jump to 364.00% in 1948.. The suggested methodology is compared to other well-known robust methods using the cited data. Table 1 provides the estimates of sum of squared of errors and regression coefficients for every method. A thorough examination of the outcomes shows that OLS has serious performance issues and consistently showed adverse results. This emphasizes how sensitive OLS is towards the outliers, as it shown by the highest sum of squared errors (78532.88).

On the other hand, other strong methods successfully lessen the influence of irregularities. Interestingly, the suggested robust function performs better on outlier-filled data, as evidenced by the lowest sum of square residuals. This emphasizes how reliable and effective the suggested estimator is as compared to other alternative choices.

Example-2: data of telephone phone calls from Belgium from 1950–1973

The second case study, which comes from Rousseeuw’s³ article, looks at the Belgium Statistical Survey’s year-by-year count of international telephone calls (in 10 millions) made from Belgium between 1950 and 1973 (Table 2). There are few outliers in the response variable in this data. The years are the independent variable (represented by X), and the dependent variable (represented by Y) is the annual telephone call count. Several writers have

Methods	Coefficients estimates			
	A	b	Values used	SSE
OLS	− 1049	24.850	9	78,532.88
LMS	− 2.470	0.1020	7	0.69534
Tukey(2)	− 2.7535	0.10896	7	0.616652
Andrews(0.58)	− 2.7542	0.10898	7	0.616687
Ali (3.0)	− 2.7792	0.10956	7	0.617851
Qadir (1.0)	− 2.7779	0.10950	7	0.617785
Khalil (2.0)	− 2.7851	0.10970	7	0.618139
Alamgir (3.0)	2.7853	0.10970	7	0.61815
Ullah (1.5)	− 2.7851	0.10970	7	0.615962
Raza(8, 2)	− 2.6486	0.10656	7	0.611112
Proposed(1)	− 2.7832	0.10965	7	0.611111

Table1. Comparison Among the existing and Proposed M-Estimators using data of Annual Average growth rate of China.

Years	1950	1951	1952	1953	1954	1955	1956	1957
Calls	0.440	0.470	0.470	0.590	0.660	0.730	0.810	0.880
Years	1958	1959	1960	1961	1962	1963	1964	1965
Calls	1.060	1.200	1.350	1.490	1.610	2.120	11.90	12.40
Years	1966	1967	1968	1969	1970	1971	1972	1973
Calls	14.20	15.90	18.20	21.20	4.300	2.400	2.700	2.900

Table 2. Number of telephone calls (10 million) from the Belgium.

previously used this dataset in their research: Qadir¹⁷, Ali & Qadir¹⁶, Khalil et al.¹³, and Raza et al.⁸. Table 2 presents a study of the performance of the suggested redescending M-estimator, taking into account estimators that are relevant to this dataset.

Table 3 presents a comprehensive comparison between our proposed robust estimator and Ordinary Least Squares (OLS) alongside other robust estimators. The regression coefficients were derived by applying all considered functions, with the corresponding sum of squares displayed in the table. Notably, outlier influence significantly compromises OLS estimates, resulting in the highest total of squared errors (659.44) and potentially misleading insights across the dataset. Conversely, outlier effects are effectively mitigated by all other M-estimators, including the robust estimator we have introduced. It is important to highlight that our proposed estimator exhibits the lowest sum of squared errors, indicating superior performance and yielding a model with minimized SSE. This underscores the efficacy of the recommended robust estimator in enhancing the accuracy and reliability of the regression model.

The Fig. 5, showed a graphical comparison of all the consider robust estimators along with proposed robust estimator. It showed that the proposed robust estimator has least sum of square of errors among all the considered robust estimators hence it produced the most efficient results for the data having contamination of outliers.

Simulation results

The evaluation of performance among the considered estimators is frequently conducted through simulation, a valuable technique, particularly when the true parameter values of the generated data are known. Utilizing a simulation strategy devised by Rousseeuw³, we have assessed the effectiveness of the proposed method. We contrasted it with several prominent redescending M-estimators to gauge its performance against established benchmarks. For the simulation study, following OLS model is incorporated to generate the data

$$y_i = \alpha + \beta x_i + u_i$$

where x follows a normal distribution with $\mu = 20$ & $\sigma^2 = 10$, $\alpha = 2, \beta = 1$ and $u_i \sim N(0, 1)$. With the help of the previously mentioned linear model, we have produced a population of 10,000 values. The R-Program has then been used to randomly select 100 values from this population as a sample. The parameter estimates have been computed using the considered M-estimators in addition to the suggested redescending M-estimator. Table 4 presents the results, providing a summary of the parameter estimates derived from the various estimators that may be compared. In this study, we assess the performance of the specified functions under two distinct

Method	Coefficients			SSEs
	<i>a</i>	<i>b</i>	<i>n</i>	
LMS	– 5.1640	0.1080	16	0.13130
Tukey (3.8)	– 5.2439	0.1102	16	0.13665
Andrews (1.5)	– 4.9072	0.1033	16	0.17308
Ali (3.0)	– 5.2092	0.1094	16	0.13404
Qadir (1.0)	– 5.2092	0.1122	16	0.13404
Ullah (4)	– 5.2430	0.1102	16	0.14060
Khalil (4.0)	– 5.2343	0.1099	16	0.13863
Alamgir (3.0)	– 5.2343	0.1102	16	0.14127
Raza(6, 3.5)	– 5.5109	0.1094	16	0.13256
Proposed(2)	– 5.16568	0.1085	16	0.13120

Table 3. Comparison among the existing and proposed M-estimators on the data of telephone calls.

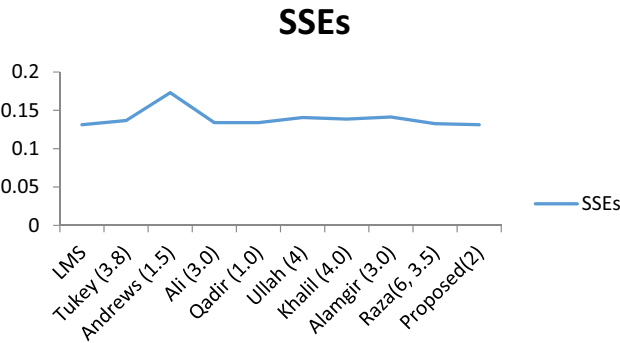


Figure 5. Graphical Comparison among the Existing and Proposed M-Estimators.

Method	Case-I		Case-II	
	Const	X	Const	X
OLS	1.99842	1.00002	7.04260	0.99794
Tukey (4.69)	1.99742	1.00011	1.99586	1.00018
Hample (4)	1.99742	1.00008	1.99586	1.00045
Huber (4)	1.99834	1.00002	1.99760	1.00023
Andrew(3.2)	1.99801	1.00021	2.19465	1.00004
Ullah (4)	2.00924	0.99964	1.99743	0.99764
Ali (4)	2.00303	0.99984	1.99923	1.00011
Khalil (5)	2.02813	0.99846	2.00358	0.999761
Raza (2)	1.99736	1.00009	1.99331	1.00047
Proposed	2.004547	0.99971	2.00341	0.999912

Table 4. Comparison among the existing and proposed M-estimators for n = 100.

scenarios. Firstly, parameter estimation is carried out using clean, outlier-free normal data. Secondly, estimation is performed with the addition of 10% outliers in the Y direction. As discussed by Norazan²¹, M-estimators often yield suboptimal estimates in the presence of outliers in the X direction. The results presented in Table 4 represent the average outcomes of 50,000 simulated data iterations, each comprising 100 samples. All methodologies demonstrate consistent performance in Case I, where outliers are absent, underscoring their reliability in standard settings. In contrast, our proposed M-estimator yields estimates closely resembling the actual parameter values utilized in the simulation for Case II, where outliers are intentionally introduced. With the exception of the OLS approach, which exhibits unreliable outcomes in the presence of outliers, the considered estimators consistently deliver efficient results, corroborating the findings outlined in Section. “Graphical Comparison of Proposed Redescending M-Estimator”. The demonstrated effectiveness of the suggested M-estimator, particularly in scenarios with limited sample sizes, underscores its reliability and robustness for sampling and estimation tasks.

Conclusion

The main purposed of this work is to obtain reliable and efficient estimates when data contain outliers. It has been consistently shown by the results from earlier sections that the proposed redescending M-estimator is more adaptable and effective than the redescending M-estimators considered in this study. The proposed M-estimator exhibits much more continuous behavior before redescending as compared to the previously developed redescending M-estimators. Greater breadth, flexibility, simplicity, and faster convergence characterize the proposed estimator above its predecessors. The suggested redescending M-estimator is more successful, producing the minimum sum of squared errors in the presence of outliers, according to real data applications. The results of simulation tests demonstrate that the coefficients produced by the suggested robust estimator are in good agreement with the genuine parameters, demonstrating the robust estimator’s comparability with other well-known estimators including Huber, Hampel, Andrews, and Beaton & Tukey. Furthermore, the suggested estimator performs no less well than the Ordinary Least Squares (OLS) method when there are no outliers. Together, these findings highlight the robustness and superiority of the proposed redescending M-estimator in a variety of scenarios and applications. The proposed redescending M-estimators can be utilized for combined monitoring the population mean and variance in simple random sampling, in two phase sampling, in ranked set sampling etc. This M-estimator can also be used to produce robust statistical quality control tools for SPC.

Data availability

The datasets used and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 15 April 2024; Accepted: 6 June 2024

Published online: 12 June 2024

References

1. Huber, P. J. Robust estimation of a location parameter. *Ann. Math. Stat.* **35**(1), 73–101 (1964).
2. Yohai, V. J. High breakdown-point and high efficiency robust estimates for regression. *Ann. Stat.* <https://doi.org/10.1214/aos/1176350366> (1987).
3. Rousseeuw, P. J. & Leroy, A. M. *Robust regression and outlier detection* (Wiley, 1987).
4. Rousseeuw, P. J. Least median of squares regression. *J. Am. Stat. Assoc.* **79**(388), 871–880 (1984).
5. Rousseeuw, P. J. & Driessen, K. V. A fast algorithm for the minimum covariance determinant estimator. *Technometrics* **41**(3), 212–223 (1999).
6. Baraud, Y. & Chen, J. Robust estimation of a regression function in exponential families. *J. Stat. Plan. Inference* **233**, 106167 (2024).
7. Ghosh, A., Jaenada, M. & Pardo, L. Robust adaptive variable selection in ultra-high dimensional linear regression models. *J. Stat. Comput. Simul.* **94**(3), 571–603 (2024).
8. Raza, A., Noor-ul-Amin, M., Ayari-Akkari, A., Nabi, M. & Aslam, M. U. A redescending M-estimator approach for outlier-resilient modeling. *Sci. Rep.* **14**(1), 7131 (2024).

9. Mukhtar, M., Ali, N. & Shahzad, U. An improved regression type mean estimator using redescending M-estimator. *Univ. Wah J. Sci. Technol. (UWJST)* **7**, 11–18 (2023).
10. Luo, R., Chen, Y. & Song, S. On the M-estimator under third moment condition. *Mathematics* **10**(10), 1713 (2022).
11. Anekwe, S. & Onyeagu, S. The redescending M estimator For detection and deletion of outliers in regression analysis. *Pak. J. Stat. Oper. Res.* <https://doi.org/10.18187/pjsor.v17i4.3546> (2021).
12. Noor-Ul-Amin, M., Asghar, S. U. D., Sanaullah, A. & Shehzad, M. A. Redescending M-estimator for robust regression. *J. Reliab. Stat. Stud.* <https://doi.org/10.13140/RG.2.2.13906.30400> (2018).
13. Khalil, U., Ali, A., Khan, D. M., Khan, S. A. & Qadir, F. Efficient UK's re-descending m-estimator for robust regression. *Pak. J. Stat.* **32**(2), 125–138 (2016).
14. Alamgir, A. A., Khan, S. A., Khan, D. M. & Khalil, U. A new efficient redescending M-estimator: Alamgir redescending M-estimator. *Res. J. Recent Sci.* **2277**, 2502 (2015).
15. Ullah, I., Qadir, M. F. & Ali, A. Insha's Redescending M-estimator for robust regression: A comparative study. *Pak. J. Stat. Oper. Res.* <https://doi.org/10.18187/pjsor.v2i2.97> (2006).
16. Ali, A. & Qadir, M. F. A modified M-estimator for the detection of outliers. *Pak. J. Stat. Oper. Res.* <https://doi.org/10.18187/pjsor.v1i1.117> (2005).
17. Qadir, M. F. Robust method for detection of single and multiple outliers. *Sci. Khyber* **9**(2), 135–144 (1996).
18. Hampel, F. R. (1986). E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel. *Robust Statistics: The Approach Based on Influence Functions*. J. Wiley & Sons.
19. Andrews, D. F. A robust method for multiple linear regression. *Technometrics* **16**(4), 523–531 (1974).
20. Beaton, A. E. & Tukey, J. W. The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics* **16**(2), 147–185 (1974).
21. Norazan, M. R. *Weighted maximum median likelihood estimation for parameters in multiple regression model* (Unpublished PhD, 2008).

Acknowledgements

The authors extend their appreciation to the Deanship of Research and Graduate Studies at King Khalid University, Saudi Arabia for funding this work through Large Groups Project under grant number R.G.P2/177/45.

Author contributions

A.R., M.T., and M.N.-u-A. conceived the study and designed the experiments. A.R. and M.T. implemented the methodology and conducted the simulations. N.G. refined the simulation settings and ensured thorough verification of results. I.B. provided technical corrections and contributed to the application section. M.N. managed communications regarding the manuscript and oversaw the revision process. All authors collectively reviewed and endorsed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to M.N.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International

License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024