

A Modified M -estimator for the Detection of Outliers

Asad Ali

Department of Statistics, University of Peshawar

NWFP, Pakistan

Email: asad_yousafzay@yahoo.com

Muhammad F. Qadir

Department of Statistics, University of Peshawar

NWFP, Pakistan

Abstract

A new ρ -function is proposed in the family of smoothly redescending M -estimators. The ψ -function associated with this new ρ -function attains much more linearity in its central section before it redescends, compared to other ψ -functions such as those of Andrews sine, Tukey's biweight and Qadir's beta function resulting in its enhanced efficiency. The iteratively reweighted least squares (IRLS) method based on the proposed ρ -function clearly detects outliers and ignoring those outliers refines the subsequent analysis. Three examples selected from the relevant literature, are used for illustrative purposes. A comparative simulation study has been conducted to evaluate its general applications. The proposed weighted least squares (WLS) method indeed achieves the goals for which it is constructed, for it gives quite improved results in all situations and is able to withstand substantial amount of outliers.

Introduction

It is well known that the *ordinary least squares* (OLS) method is very sensitive to outlying observations. There are several approaches to deal with this problem. Several statisticians have developed many robust methods based on different approaches. The ever first step came from Edgeworth in 1887 that proposed to use L_1 -criterion instead of OLS (Rousseeuw and Yohai 1984). Afterwards the most commonly used robust estimators are Huber's M -Estimators (Hampel *et al* 1986), MM -estimators (Yohai 1987), GM -Estimators, Siegel's *Repeated Median* Estimators (Rousseeuw and Leroy 1987), *Least Median of squares* (LMS) estimators, *Least Trimmed Squares* (LTS) estimators (Rousseeuw 1984), S -Estimators (Rousseeuw and Yohai 1984), *Minimum Volume Ellipsoid* (MVE) estimators (Rousseeuw and Leroy 1987), and *Minimum Covariance Determinant* (MCD) estimators (Rousseeuw and Van Driessen 1998).

Huber introduced the notion of M -estimators in 1964 (Hampel *et al* 1986), which opened new gates in the theory of classical statistics. Afterwards several M -estimators were proposed from time to time and the theory of M -estimators got enriched by every day passed. A brief discussion over the concept of M -estimators is given the following subsections.

M-Estimators

The name *M*-estimators comes from “generalized maximum likelihood.” It is the simplest approach both computationally and theoretically (Hampel *et al* 1986).

The *M*-estimate $T_n(x_1, \dots, x_n)$ for the function ρ and the sample x_1, \dots, x_n is the value of t that maximizes the objective function $\sum_{i=1}^n \rho(x_i; t)$. A reasonable ρ should have the following properties:

1. $\rho(0) = 0$
2. $\rho(t) \geq 0$
3. $\rho(t) = \rho(-t)$ (Symmetry)
4. for $0 < t_1 < t_2 \Rightarrow \rho(t_1) \leq \rho(t_2)$
5. ρ is continuous (ρ is differentiable)

When ρ can be differentiated with respect to t , a function (which except for a multiplicative constant) we denote by ψ , (i.e. $\psi(x_i; t) = \frac{\partial \rho(x_i; t)}{\partial t}$), we may find it more convenient to calculate T_n by finding the value of t that satisfies

$$\sum_{i=1}^n \psi(x_i; t) = 0 \tag{1}$$

The corresponding *w*-function (weight function) for any ρ is then defined as

$$w(x_i; t) = \frac{\psi(x_i; t)}{t} . \tag{2}$$

Employing this *w*-function in OLS, we get weighted least squares (WLS) method and the resulting estimates are then called the weighted estimates (Hoaglin *et al* 1983).

the weighted estimates are computed by solving the equations

$$\hat{\beta} = (X'WX)^{-1} X'Wy \tag{3}$$

where W is a $n \times n$ diagonal square matrix having the diagonal elements as weights.

M-estimators minimize objective function more general than the familiar sum of squared residuals associated with the sample mean. Instead of squaring the deviations of each observation x_i from the estimate t , we apply the function $\rho(x_i; t)$ and form the objective function by summing over the sample:

$\sum_{i=1}^n \rho(x_i; t)$. The nature of $\rho(x_i; t)$ determines the properties of the M-estimator (Hoaglin *et al* 1983). We briefly elaborate some of the most popular M-estimators in the following lines.

Huber's M-estimator uses the following ψ -function

$$\psi(t) = \begin{cases} -a, & t < -a \\ t, & -a \leq t \leq a \\ a, & t > a \end{cases} \quad (5)$$

Both the least squares and Huber objective functions increase without bound as the residual departs from 0, but the least-squares objective function increases more rapidly. Least squares assigns equal weight to each observation; the weights for the Huber estimator decline when $|t| > a$. The Huber's ψ -function takes into account the neighborhood of a normal model in a linear way. It has a *constant-linear-constant* behavior, i.e. it is constant beyond the specified bound ($-a$ to a) and is linear like mean within these bound see Figure 1. Like the OLS it assigns equal weights to all observations within its bound, which surely will result in its high efficiency but distant outliers still have a maximum influence (in the form of constant a), which lead to the efficiency losses of about 10-20 percent in typical cases with outliers (Hampel *et al* 1986). To cope with this problem redescending M-estimators were introduced.

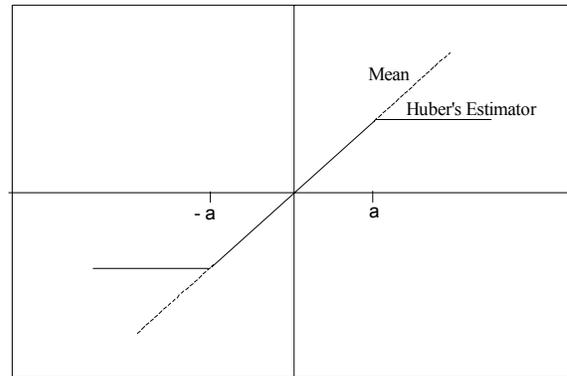


Figure 1 Shape of the ψ -functions of mean and Huber-estimators.

Redescending M-estimators

The redescending M-estimators were introduced by Hampel (Hampel *et al* 1986), who used a *three part-redescending* estimator with ρ -function bounded and ψ -function becoming 0 for large $|t|$ see Figure 2. They reject distant outliers completely, but not suddenly, allowing a transitional zone of increasing doubt, and are therefore much more efficient than “hard” rejection rules; they are usually about as good to clearly better than Huber-estimators (Hampel *et al* 1986). The logic of these estimators is that the very central observations (in the neighborhoods of 0) of the normal neighborhood receive maximum weight and as they depart from center their weight declines, and as they reach the specified bounds their ψ -function becomes 0.

The Hampel’s *three part redescending* ψ -function is defined as

$$\psi(t) = \begin{cases} t & \text{if } |t| < a \\ a \operatorname{sgn}(t) & \text{if } a \leq |t| < b \\ \{(c - |t|)/(c - b)\} a \operatorname{sgn}(t) & \text{if } b \leq |t| \leq c \\ 0 & \text{otherwise ,} \end{cases} \quad (6)$$

Looking at Figure 2 one can easily conclude that the Hampel’s *three part redescending* estimator is still not a good one, as the abrupt changes in its slope are unappealing because of the abrupt changes in the way the data are used. So the need of a ψ -function with a smoothly redescending nature was seriously felt. Several smoothly redescending M-estimators have been proposed from time to time.

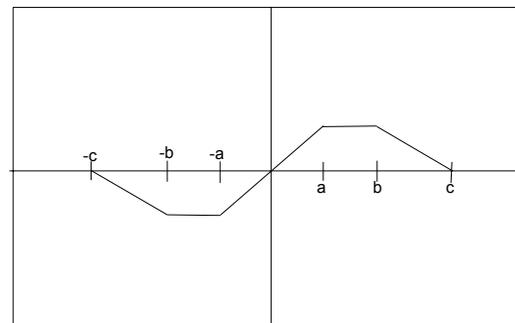


Figure 2 Hampel's Three Part ψ -Function

A real improvement came from Andrews (Andrews 1974) and Tukey (Mosteller and Tukey 1977; Hoaglin *et al* 1983) who used *wave* estimators (also called *sine* estimators) and *biweight* estimators respectively. Both Andrews’ wave and Tukey’s biweight estimators have smoothly redescending ψ -functions, shown in Figure 3 and 4 (Hoaglin *et al* 1983). Afterwards Qadir proposed another smoothly redescending ψ -function shown in Figure 5, whose weight function is a beta function with $\alpha = \beta$ (Qadir 1996). The weights for all these decline as soon as residuals depart from 0, and are 0 for $|t| > a$.

Andrews wave function
$$\psi(t) = a \sin\left(\frac{t}{a}\right) \quad |t| \leq a \quad (7)$$

$$= 0 \quad \text{otherwise}$$

Tukey's biweight function
$$\psi(t) = t \left[1 - \left(\frac{t}{a}\right)^2 \right]^2 \quad |t| \leq a \quad (8)$$

$$= 0 \quad \text{otherwise}$$

Qadir's beta function
$$\psi(t) = \frac{t}{16} \left[1 - \left(\frac{t}{a}\right)^2 \right]^2 \quad |t| \leq a \quad (9)$$

$$= 0 \quad \text{otherwise}$$

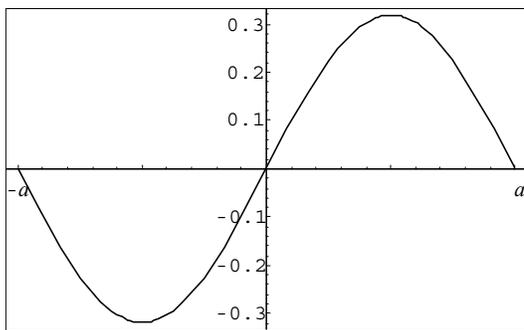


Figure 3 Andrews wave ψ -function

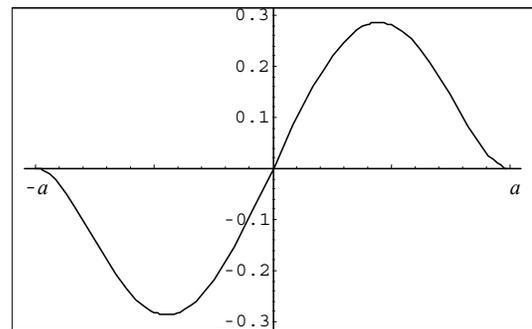


Figure 4 Tukey's Biweight ψ -function

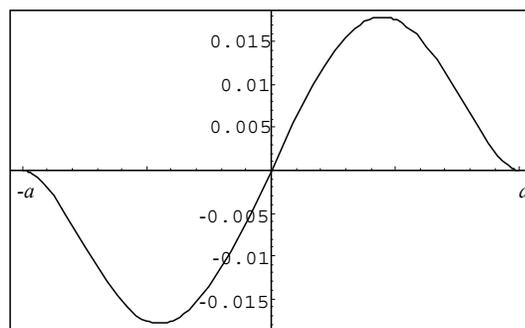


Figure 5 Qadir's beta ψ -function

The Proposed ρ -function

We propose a new ρ -function with its corresponding ψ and w -functions, thus giving development to a new weighted least square method. We discuss the shape and properties of its ψ -function comparative to that of others such as Huber's, Andrews, and Tukey's biweight function.

Description of the Algorithm

Consider the following objective function

$$\rho(t) = \begin{cases} \frac{t^2}{45 a^8} (3t^8 - 10a^4 t^4 + 15a^8) & \text{if } |t| \leq a \\ \frac{8a^2}{45} & \text{if } |t| > a \end{cases} \quad (10)$$

where a is the so-called tuning constant and for i th observation the variable t are the residuals scaled over MAD.

The above ρ -function satisfies the standard properties stated in section 2, generally associated with a reasonable objective function.

The derivative of Eq. (10) with respect to t is

$$\psi(t) = \begin{cases} \frac{2t}{3} \left(1 - \left(\frac{t}{a} \right)^4 \right)^2 & \text{if } |t| \leq a \\ 0 & \text{if } |t| > a \end{cases} \quad (11)$$

The corresponding standard weight function is thus defined as follows

$$w(t) = \begin{cases} \frac{2}{3} \left(1 - \left(\frac{t}{a} \right)^4 \right)^2 & \text{if } |t| \leq a \\ 0 & \text{if } |t| > a \end{cases} \quad (12)$$

The Shape of ψ -function

Before proceeding to use a robust estimator one would naturally wish to know the answers of the two typical questions

1. How robust the estimator is?
2. What is the efficiency of the estimator?

A natural answer to both of these questions is “a compromise”, that is one have to choose an estimator, which has maximum resistance with minimum efficiency losses. One would certainly avoid using a robust estimator on the cost of large efficiency loss neither would use a completely non-robust estimator with high efficiency but would make a compromise between these two options.

It is a well-known statistical truth that the arithmetic mean has the highest possible efficiency among all other estimators but unfortunately it is extremely sensitive to outliers and even a single outlier can have disastrous effects on it. Among M -estimators the most efficient estimator would be Huber’s estimator with carefully chosen bounds. The ψ -function of Huber’s estimator has a constant-linear-constant behavior (Hoaglin *et al* 1983). Figure 1 shows that the central section of Huber’s ψ -function is linear and within specified bounds it coincides with the ψ -function of the mean and in respect of efficiency it is its plus point. Within the specified bounds it is as efficient as mean can be.

As discussed before the main drawback of Huber’s ψ -function is that distant outliers still have maximum (though bounded) influence and would lead certainly to certain percent efficiency losses. To avoid these losses, one can use smoothly re-descending M -estimators such as Andrew’s sine function; see Figure 3, or Tukey’s biweight function, see Figure 4, with ρ being bounded and ψ -function continuously becoming zero for large absolute residual. They reject distant outliers completely, but not suddenly, allowing a transitional zone of increasing doubt, and are therefore much more efficient than “hard” rejection rules; they are usually about as good to clearly better than Huber-estimators.

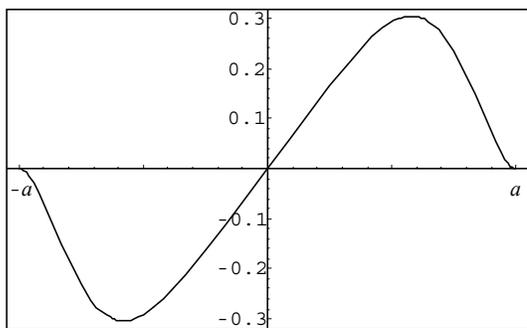


Figure 6 Shape of Asad’s ψ -function.

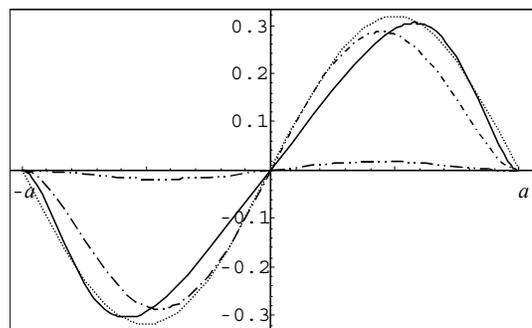


Figure 7 Combined plots of Andrews (Dots), Tukey’s (Dash-Dot) Qadir’s (Dash-Dot-Dot) and Asad’s (Solid) ψ -functions.

Now we discuss the shape of the proposed ψ -function. The ψ -function given in Figure 6 (also in Figure 7) has a different behaviour as compared to that of other redescending estimators such as Andrews's wave and Tukey's biweight estimators. Recalling that the ψ -function of the arithmetic mean is just a linear straight-line rendering it theoretically the most efficient estimator. The proposed ψ -function capture the property of longer linear central section from the ψ -function of mean and behaves linearly for large number of the central values as compared to other smoothly redescending ψ -functions. This increased linearity certainly responses in enhanced efficiency. The ψ -function then redescends gradually for increasing values of residuals and becomes zero for values lying outside the specified band. The difference can be clearly seen in Figure 7.

The Method

The method has a very similar procedure as used for a typical M -estimator. First an ordinary least squares model is fitted to the data and the residuals obtained from the fit are standardized over the initial scale MAD while subsequent scaling is made over Huber's proposal 2 described in Hampel *et al* (1986) and Street *et al* (1988). The scaled residuals are transformed using the proposed w -function and the initial estimates of regression parameters are calculated. Then using the Huber's proposal 2 by IRLS method the final estimates are obtained. Experiences show that the new WLS method of estimation is quite insensitive to the presence of outliers and can be applied to detect outliers.

Applications

We applied the proposed WLS method to different examples to verify its effectiveness in detecting outliers. The results of LMS and LTS were obtained by using the program PROGRESS proposed by Rousseeuw and Leroy (1987).

Example 1

The first example is that of Sparrows Wing Length data, taken from Zar (1999) in which the wing length is the response variable and age is the explanatory variable. The actual data is quite a good data with no outliers at all as the OLS model $\hat{y} = 0.830 + 0.265x$ with a residuals sum of squares equal (RSS) to 0.720 seems reasonable.

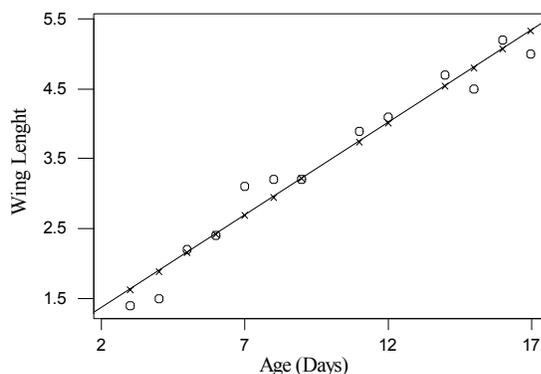


Figure 8 OLS line fitted to the actual data.

Figure 7 shows that the fitted line represent the data well. However let us suppose we mistakenly record the third value of the response variable as 3.5 instead of 2.2 thus producing an outlier in y-direction. The summary

Table 1 Sparrow’s Wing Length data fitted by OLS and different robust methods including the proposed method.

Method	Coefficients		Cases Used	SS of Residuals
	Constant	x		
OLS	1.154	0.242	13	2.260
RWLMS	0.816	0.266	12	0.712
RWLTS	0.531	0.299	9	0.155
Qadir ($a = 1.5$)	1.124	0.244	12	0.925
Tukey ($a = 4.0$)	0.802	0.267	12	0.718
Asad ($a = 3.0$)	0.811	0.266	12	0.717

results from different fits for this contaminated data set are given in Table 1. The new RSS is now taken to 2.260, which is significantly different from the old RSS. Except the LTS the robust fits from all the other methods detect one outlier and give an RSS almost equal to that of the old RSS.

The LTS declares four observations as outliers and deleting these gives an RSS of 0.155, which is not a desirable case as we can see that fits from the other robust methods depicts the same trend as that of the actual data. The fit from Qadir beta function gives the intercept term very close to that of OLS (contaminated). The OLS and different robust fits to the contaminated data are sketched in Figure 9. The figure shows that Qadir’s fit and the OLS fits just coincide although it detects one outlier. The reason for this is that it uses the sample standard deviation as the scaling tool for standardizing the OLS residuals, which is itself a non-robust estimator for scale. The fits from LMS, Biweight and Asad’s methods also coincide with each other, as there is a very little difference in their estimated coefficients.

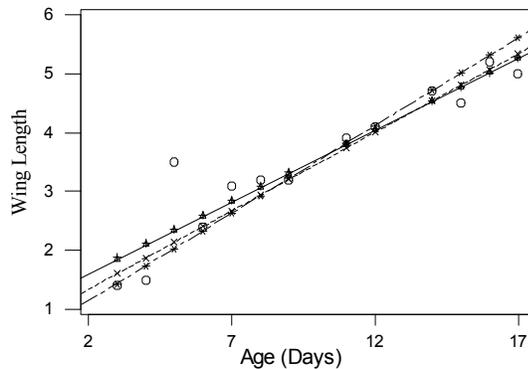


Figure 9 OLS and Qadir (Solid), LTS (Dash-Dot), LMS, Biweight and Asad (Dot) fits to the contaminated data.

Example 2 Telephone Calls Data

The second example is that of *Phone Calls* data. This is a real data set with a few outliers present in the data. The data set is taken from Belgian Statistical Survey. The dependent variable is the number telephone calls made from Belgium and the independent variable is the year (Rousseeuw and Leory 1987). The scatter plot of the data along with different fits is shown in Figure 10. From the plot it is clear that the observation from 1964 to 1969 are outliers. Rousseeuw and Leory (1987) state that actually from the year 1964 to 1969, another recording system was used, giving the total minutes of calls (The years 1963 and 1970 are also partially affected because the transactions did not happen exactly on New Year's Day).

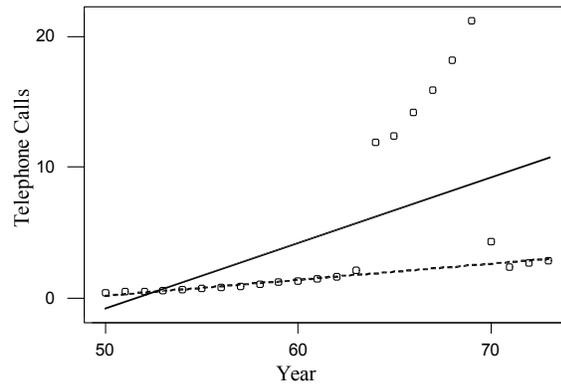


Figure 10 Phone Calls data plot with OLS fit (solid line): Robust fits by Asad's WLS and Tukey's biweight (dashed lines).

The fits from OLS and other robust methods along with the proposed method are given in Table 2. The OLS fit is highly influenced by outliers as it has a very large residual sum of squares (RSS), thus the fit represent neither good nor bad data points well. This is what one could obtain by not looking critically at those data and by applying the OLS method in routine. Except the Qadir's WLS all of the other robust fits ignore 8 outlying observations with a negligible difference among their RSS. Here it is to be noted that throughout our study we use the unweighted RSS so that a real comparison can be made among different robust methods.

Method	Coefficients		Cases Used	SS of Residuals
	Constant	x		
OLS	-26.01	0.504	24	695.44
RWLTS	-5.164	0.108	16	0.1313
RWLMS	-5.164	0.108	16	0.1313
Qadir ($a = 1.0$)	-5.505	0.115	15	3.5347
Tukey ($a = 3.8$)	-5.306	0.111	16	0.1362
Asad ($a = 3.0$)	-5.181	0.109	16	0.1314

The fits from the OLS, Tukey's biweight and the proposed method are sketched in Figure 10. It is obvious that with the proposed method the model fits very well. The OLS line (solid line) is pulled toward the middle of the two groups of observations which is the effect of y values associated with years 1964-69, rendering it a completely unrepresentative fit, whereas the fit with the proposed method shows very much robustness and fits a model which represents the majority of the observations and avoids outliers. The robust fits by the proposed method and Tukey's biweight differ very little therefore in Figure 10 the two robust lines cannot be differentiated from one another.

Example 3 Stackloss Data

The third example is the well-known *Stackloss* data set taken from Rousseeuw and Leory (1987). We choose this example because it is a real data set and is examined by many statisticians such as Denial and Wood (1971), Andrews (1974), Andrews and Pregibon (1978), Cook (1979), Draper and Smith (1981), Dempster and Gasko-Green (1981), Atkinson (1982), Rousseeuw and Leroy (1984), Carroll and Rupert (1985), Qadir (1996) and several others by means of different methods. The data describes operation of a plant for the oxidation of ammonia to nitric acid and consist of 21 four-dimensional observations. The stackloss (y) has to be explained by the rate of operation (x_1), the cooling water inlet temperature (x_2), and the acid concentration (x_3). Summarizing the findings cited in the literature, it could be said that most people concluded that observations 1, 3, 4, and 21 were outliers. According to some people observation 2 is also reported as an outlier (Rousseeuw and Leroy 1987). The standardized residuals plot shown in Figure 11 shows that there are no outliers at all as the horizontal band encloses the standardized residuals between -2.5 and 2.5 . So it can be easily concluded the data set contains no outliers at all because all the standardized residuals fall nicely within the band. However we observe that robust fits give a different idea about the nature of the data. The LTS method shows a severe robustness by detecting 6 observations as outliers and gives the

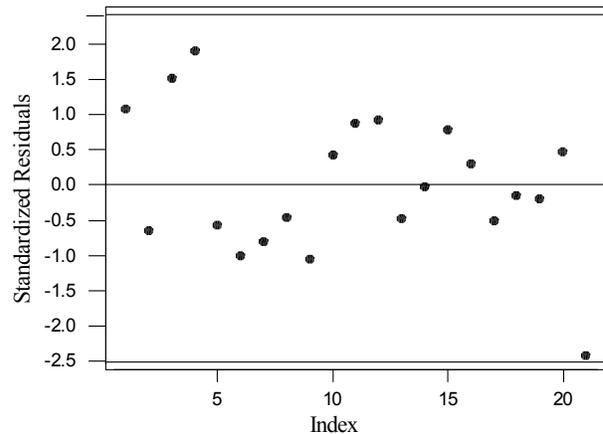


Figure 11 Standardized residuals plotted against the index of the observations for Stackloss data.

smallest RSS among all other robust methods. For a tuning constant equal to 2.0 the proposed method gives an RSS = 12.827 with 5 observations deleted but for a small tuning constant more robust results can be obtained. The OLS and other robust fits are given in Table 3.

Table 3 Stackloss data fitted by OLS and different robust methods including the proposed method.

Method	Coefficients				Cases Used	SS of Residuals
	Constant	x_1	x_2	x_3		
OLS	-39.920	0.716	1.295	-0.152	21	178.830
RWLTS	-34.057	0.757	0.453	-0.052	15	10.273
RWLMS	-37.652	0.798	0.577	-0.067	17	20.401
Qadir ($a = 1.5$)	-37.537	0.726	0.923	-0.103	17	27.882
Tukey ($a = 3.0$)	-36.809	0.850	0.432	-0.075	16	12.760
Asad ($a = 2.0$)	-37.003	0.849	0.425	-0.071	16	12.827

Table 4 shows the residuals from all the five methods. The residuals from our WLS seem very close to that of the other robust methods. The outliers can be detected by looking at the residuals of the five robust fits. The residuals of the observations 1, 3, 4, 21 are higher in all the fits. The LTS tells a different story from others as it further declares observations 2 and 13 as outliers and gives large residuals for these observations. The proposed method gives high residuals values for observations 1, 3, 4, 13, 21 and deletes all of these observations from the subsequent analysis.

Table 4 Residuals from OLS and different robust fits to Stackloss data

Observations	1	2	3	4	5	6
i	OLS	RWLTS	RWLMS	Qadir	Tukey	Asad
1	3.235	<u>7.895</u>	<u>6.218</u>	<u>5.671</u>	<u>5.881</u>	<u>5.918</u>
2	-1.918	<u>2.843</u>	1.151	0.568	0.805	0.846
3	4.555	<u>7.639</u>	<u>6.428</u>	<u>6.235</u>	<u>6.068</u>	<u>6.086</u>
4	5.698	<u>8.776</u>	<u>8.174</u>	<u>7.309</u>	<u>8.318</u>	<u>8.339</u>
5	-1.712	-0.317	-0.671	-0.844	-0.818	-0.810
6	-3.007	-0.770	-1.249	-1.768	-1.250	-1.236
7	-2.390	0.089	-0.424	-1.073	-0.230	-0.233
8	-1.390	1.089	0.576	-0.073	0.770	0.767
9	-3.144	-0.743	-1.058	-1.863	-0.852	-0.838
10	1.267	0.160	0.359	1.033	-0.218	-0.210

11	2.636	0.629	0.963	-1.961	0.460	0.431
12	2.780	0.031	0.473	-1.781	-0.183	-0.214
13	-1.429	<u>-2.736</u>	-2.507	-1.761	<u>-3.067</u>	<u>-3.068</u>
14	-0.051	-1.616	-1.346	-0.551	-1.671	-1.709
15	2.361	0.685	1.344	1.771	1.256	1.226
16	0.905	-0.472	0.143	0.462	0.030	0.012
17	-1.520	-0.655	-0.373	-0.904	-0.457	-0.412
18	-0.455	-0.290	0.097	-0.183	0.071	0.087
19	-0.598	0.309	0.586	-0.003	0.714	0.733
20	1.412	1.871	1.934	1.845	1.767	1.780
21	-7.238	<u>-8.257</u>	<u>-8.630</u>	<u>-7.396</u>	<u>-9.448</u>	<u>-9.469</u>

Discussion

It is evident from the results of all the three examples that the proposed method is quite effective in detecting outliers. In all of the three examples the outliers were detected successfully and the subsequent analysis after their deletion showed clear improvements. We tried such priorities (e.g. selection of tuning constants, selection of scale estimator etc) so that to get similar results as those of the other robust methods such as LTS, LMS, biweight, and Qadir's beta function, to verify the effectiveness of the proposed method in detecting the unusual observations. Of course different results can be obtained with different priorities as it totally depends on user, what kind of results he or she wants.

The actual Sparrows Wing Length data has no outliers at all and the OLS line fits the data well with a reasonable RSS of 0.720. We intentionally contaminated it for experimental purposes by replacing 2.2 (the third value of the response variable) by 3.5. The RSS is now taken to 2.260. The model parameters also differ from those of the first one (non-contaminated). We applied the robust methods and observed that the proposed method detected that outlier successfully with an RSS of about 0.717, quite a good fit.

Both the visual sketch and the large OLS RSS for Telephone Calls data set argue that the data are highly influenced by outliers. It is evident from the graphical sketch of the data as the OLS line is pulled towards the middle of the two groups of the data points rendering it a very unrepresentative line. The fit from our new method ignores outlying observations, which in turn decrease residuals sum of squares to a considerable extent and the robust line fits the remaining observations very well.

Stackloss data set is a very popular data set and several statisticians studied it from time to time. Most of them declared observations 1, 3, 4, and 21 as outliers. The new WLS method also confirms the fact, as the residuals of these observations are higher in all the fits. It discards these observations from the data and gives a residual sum of squares considerably decreased.

Simulation

A simulation strategy described by Rousseeuw and Leroy (1987) has been adopted to verify the performance of the proposed method. The strategy consists of two steps. The first one is the normal situation,

$$y_i = 1 + x_{i,1} + x_{i,2} + \dots + x_{i,p} + e_i$$

in which $e_i \sim N(0, 1)$ and the explanatory variables are generated as $x_{i,j} \sim N$

Table 5. Simulation results for Simple and Multiple Regression

	Values	OLS _N	OLS	RWLMS	RWLTS	Tukey	Asad	
Simple Regression	$n = 20$	θ_0	0.910	3.166	1.072	1.072	0.997	0.992
	Outliers =	θ_1	0.997	0.993	1.010	1.010	1.010	1.008
	4							
		RSS	18.556	357.440	13.518	13.518	13.610	13.640
	Outliers Detected	----	----	4	4	4	4	
Simple Regression	$n = 1000$	θ_0	0.974	1.473	0.987	0.997	0.998	0.990
	Outliers =	θ_1	1.003	0.995	1.001	1.001	1.002	1.002
	50							
		RSS	995.540	5611.726	838.374	828.172	945.428	945.221
	Outliers Detected	----	----	65	67	50	50	
Multiple Regression	$n = 50$	θ_0	1.166	3.047	1.180	1.180	1.231	1.222
	Outliers =	θ_1	0.992	1.022	0.993	0.993	0.989	0.990
	10							
		θ_2	1.001	0.900	0.990	0.990	0.989	0.989
		RSS	41.949	738.753	37.730	37.730	37.891	37.832
	Outliers Detected	----	----	10	10	10	10	
Multiple Regression	$n = 1000$	θ_0	1.125	1.594	1.000	1.062	0.957	0.968
	Outliers =	θ_1	0.932	1.032	0.935	0.935	0.942	0.941
	50							
		θ_2	1.049	0.953	1.058	1.051	1.055	1.055
		RSS	1089.50	5918.65	912.70	907.19	1040.67	1040.54
	Outliers Detected	---	---	65	66	50	50	

(0, 100) for $j = 1, \dots, p$. In the second step we construct outliers in y -direction. For this purpose, we generate samples where some of the observations (e.g. 90%, 80%, etc) are as in the first situation and the remaining are contaminated by using the error term $e_i \sim N(10, 1)$. The results for different regressions are given in Table 5. From the table it is clear that the proposed method is quite effective in detecting outliers and reduces the RSS to a reasonable extent.

Conclusion

We proposed a new ρ -function. The WLS method based on such ρ -function produced quite good results. The above three examples and the simulation studies show that the proposed WLS method is quite resistive to outliers and the subsequent analysis is refined by ignoring those outliers. We advance this new method to be used for robust regression as an alternative to other M -estimators.

References

1. Andrews, D. F. (1974), "A Robust Method for Multiple Linear Regression," *Technometrics*, 16, 523-531.
2. Andrews, D. F. and Pregibon, D. (1978), "Finding the Outliers that Matter," *Journal of Royal Statistical Society, Series B*, 40, 85-93.
3. Atkinson, A. C. (1982), "Regression Diagnostics, Transformations and Constructed Variables," *Journal of Royal Statistical Society, Series B*, 44, 1-36.
4. Bickel, P. J. (1976), "Another Look at Robustness: a Review of Reviews and some new Developments," *Scandinavian Journal of Statistics*, 3, 145-168.
5. Carroll, R. J., and Rupert, D. (1985), "Transformations in Regression: A Robust Analysis," *Technometrics*, 27, 1-12.
6. Cook, R. D., (1979), "influential observations in regression," *Journal of the American Statistical Association*, 74, 169-174.
7. Dempster, A. P. and Gasko-Green, M. (1981), "New Tools for Residual Analysis," *Annals of Statistics*, 9, 945-959.
8. Deniel, C. and Wood, F. S. (1971), *Fitting Equations to Data*, John Wiley and Sons, New York.
9. Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis*, John Wiley and Sons, New York.
10. Hampel, F. R., Ronchetti, E.M., Rousseeuw, P.J. and Stahel, W.A. (1986), *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, New York.
11. Hoaglin, D. C, Mosteller, F., and Tukey, J. W. (1983), *Understanding Robust and Exploratory Data Analysis*, John Wiley and Sons, New York.
12. Mosteller, F. and Tukey, J. W. (1977), *Data Analysis and Regression*, Addison-Wesley Publishing Company, Inc. Philippines.
13. Qadir, M. F. (1996), "Robust Method for Detection of Single and Multiple Outliers," *Scientific Khyber*, 9, 135-144.

14. Rousseeuw, P.J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, 79, 871-880.
15. Rousseeuw, P.J. and Hubert, M. (1996), "Recent Development in PROGRESS," *Computational Statistics and Data Analysis*, 21, 67-85.
16. Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.
17. Rousseeuw, P.J. and Van Driessen, K. (1998), "Computing LTS Regression for Large Data Sets," Technical Report, submitted to University of Antwerp, Belgium.
18. Rousseeuw, P.J. and Van Driessen, K. (1999), "A Fast Algorithm for the Minimum Covariance Determinant Estimator," *Technometrics*, 41, 212-223.
19. Rousseeuw, P.J. and Yohai, V. (1984), "Robust Regression by Means of S-estimators," in *Robust and Nonlinear Time Series Analysis*, edited by J. Franke, W. Härdle, and R.D. Martin, Lecture Notes in Statistics 26, Springer Verlag, New York, 256-274.
20. Street, J. O., Carroll, R. J., and Rupert, D. (1988), "A Note on Computing Robust Regression Estimates Via Iteratively Reweighted Least Squares," *The American Statistician*, 42, 152-154.
21. Yohai, V.J. (1987), "High Breakdown Point and High Efficiency Robust Estimates for Regression," *Annals of Statistics*, 15, 642-656.
22. Zar, J. H. (1999), *Biostatistical Analysis*, Prentice Hall, Inc. New Jersey.