Insha's Redescending M-estimator for Robust Regression: A Comparative Study

Insha Ullah
Department of Statistics, University of Peshawar
NWFP, Pakistan
Email: inshahullah03@hotmail.com

Muhammad F. Qadir
Department of Statistics, University of Peshawar
NWFP, Pakistan
Email: qadir_mf@yahoo.com

Asad Ali
Department of Statistics, University of Peshawar
NWFP, Pakistan
Email: asad_yousafzay@yahoo.com

Abstract

In this paper we present a new redescending M-estimator "Insha's estimator" for robust regression and outliers detection that overcomes some drawbacks of other M-estimators for robust regression and outliers detection, such as destruction of the good observations and lack of simplicity in applications. The Ψ-function associated with the proposed estimator attains more linearity in the central section before it redescends, resulting in enhanced efficiency. Moreover the estimator is continuous everywhere and can be written in closed form without the use of an indictor function. The estimator is also applied to a real world example taken from the literature. For the purpose of comparison with other well-known redescending M-estimators extensive simulation study has been carried out. The example and simulation study show that using this estimator all the outliers can be successfully detected and is not affected by outliers.

1. Introduction

One of the most important statistical tools is a linear regression analysis for many fields. Nearly all regression analysis relies on the method of least squares for estimation of the parameters in the model. But the least squares method has been constructed under specific assumptions, such as normality of the error distribution. It is also assumed that the underlying model holds for every observation. However, when applying the method in practice these assumptions are rarely met in full, as they are at best only approximations to reality. The most well-known problem in the application of regression is when some observations, called outliers, deviate from the postulated pattern. Outliers can be generated by from a simple operational mistake to including small sample from a different population, and they make serious effects of statistical inference. Even one outlying observation can destroy least squares estimation, resulting in parameter estimates that do not provide useful information for the majority of the data.

Robust regression analyses have been developed as an improvement to least squares estimation in the presence of outliers and to provide us information about what a valid observation is and whether this should be thrown. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. Robust regression is an important tool for analyzing data that are contaminated with outliers. It can be used to detect outliers and to provide resistant (stable) results in the presence of outliers.

Many methods have been developed for these problems. However, in statistical applications of outlier detection and robust regression, the methods most commonly used today are Huber's M-estimation (Hampel et al., 1986), LMS-estimation (Rousseeuw, 1984), LTS-estimation (Rousseeuw, 1984, 1985), S-estimation (Rousseeuw and Yohai, 1984) and MM-estimation (Yohai, 1987) etc.

2. M-estimator

M-estimators use maximum likelihood formulations by deriving optimal weighting for the data set in non-normal conditions. It was introduced by Huber (1973) as a generalization of the familiar least squares criterion, replacing the quadratic loss function with a symmetric function *ρ(.)*, yielding

$$\underset{\hat{\boldsymbol{\beta}}}{\text{Minimize}} \sum_{i=1}^{n} \rho(r_i) \qquad (1)$$

where the function ρ(.) gives the contribution of each residual to the objective function. It is the simplest approach both computationally and theoretically. Although it is not robust with respect to leverage points, it is still used extensively in analyzing data for which it can be assumed that the contamination is mainly in the response direction. The function *ρ(.)* is usually chosen such that it represents some weighting of the ith residual. This weighting means that outlying observations have their weights reduced and thus the estimates are affected less by such noise. A weighting of zero is equivalent to classification as an outlier.

Differentiating (1) with respect to the regression coefficients $\hat{\beta}_j$ yields

$$\sum_{i=1}^{n} \psi(r_i)\mathbf{X}_i = \mathbf{0} \qquad (2)$$

where *ψ(.)* is the derivative of *ρ(.)* and the corresponding M-estimator is the maximum likelihood estimator.

Define the weight function $w(r_i) = \psi(r_i)/r_i$, then the estimating equations may be written as

$$\sum_{i=1}^{n} w(r_i)\mathbf{X}_i = \mathbf{0} \qquad (3)$$

Solving the estimating equations is a weighted least-squares problem. The weights, however, depend upon the residuals, the residuals depend upon the estimated coefficients, and the estimated coefficients depend upon the weights. An iterative solution (called iteratively reweighted least-squares (IRLS)) is therefore required.

2.1 Redescending M-estimator

On the other hand redescending M-estimators are those M-estimators that are able to reject extreme outliers completely. That is, an M-estimator is called a Redescending M-estimator, If the derivative $\psi = \rho'$ of ρ is redescending, i.e. satisfies $\lim_{r_i \to \pm\infty} \rho'(r_i) = 0$.

It was first introduced by Hampel (Andrews et al., 1972), who used a three part-redescending estimator, with ρ-function bounded and *ψ*-function becoming zero for large $|r|$. The *ψ*-function of Hampel's estimator is given by

$$\psi(r) = \begin{cases} r & for \quad |r| \le a \\ a\, sign(r) & for \quad a < |r| \le b \\ a\dfrac{c-|r|}{c-b} sign(r) & for \quad b < |r| \le c \\ 0 & for \quad |r| > c \end{cases} \tag{4}$$

where a, b, and c are positive constants and $0 < a \le b < c < \infty$. This estimator demonstrated good performance in the Princeton Robustness Study. The lack of differentiability of $\psi(.)$ is not ideal, however, and a smooth ψ-function would be preferred. This led to the development of smoothly redescending M-estimators. Several smoothly redescending M-estimators have been proposed.

The most commonly used smoothly redescending M-estimators are Andrew's sine function introduced in the Princeton Robustness Study (Andrews et al., 1972), the *Ψ*-function of which is given by

$$\psi(r) = \begin{cases} c\sin(r/c) & for \quad |r| \le c\pi \\ 0 & for \quad |r| > c\pi \end{cases} \tag{5}$$

Tukey's biweight function by Beaten and Tukey (1974) with *Ψ*-function given by

$$\psi(r) = \begin{cases} r[1-(r/c)^2]^2 & for \quad |r| \le c \\ 0 & for \quad |r| > c \end{cases} \tag{6}$$

and Qadir Beta function (Qadir 1996) the *Ψ*-function of which is

$$\psi(r) = \begin{cases} \dfrac{r}{16c^4}(c+r)^2(c-r)^2 & for \quad |r| \le c \\ 0 & for \quad |r| > c \end{cases} \tag{7}$$

Ali proposed a modified form of Tukey's biweight function (Ali and Qadir 2005) the *Ψ*-function of which is

$$\psi(r) = \begin{cases} \dfrac{2r}{3}[1-(r/c)^4]^2 & for \quad |r| \le c \\ 0 & for \quad |r| > c \end{cases} \tag{8}$$

For regression analysis, some of the redescending M-estimators can attain the maximum breakdown point. Moreover, some of them are the solutions of the problem of maximizing the efficiency under bounded influence function when the regression coefficient and the scale parameter are estimated simultaneously. Hence redescending M-estimators satisfy several outlier robustness properties.

3. Insha's redescending M-estimator

We propose a new redescending M-estimator. This estimator covers some drawbacks of existing redescending M-estimators and should be regarded as new tool for robust regression and outlier detection. We discuss the shape and properties of its *ρ*-function and the corresponding *Ψ*-function and weight function.

Consider the following objective function

$$\rho(r) = \frac{c^2}{4}\left[Arc\tan\left(r/c \right)^2 + \frac{c^2 r^2}{c^4 + r^4} \right] \qquad for \qquad |r| \geq 0 \qquad (9)$$

where c is the tuning constant and for ith observation the variable r_i are the residuals scaled over MAD.

The above *ρ*-function satisfies the standard properties generally associated with a reasonable objective function of redescending M-estimator, namely,

- $\rho(r_i) \geq 0$
- $\rho(0) = 0$
- $\rho(r_i) = \rho(-r_i)$
- $\rho(r_i) \geq \rho(r_j)$ for $|r_i| \geq |r_j|$
- *ρ* is continuous (*ρ* is differentiable)

Taking derivative of (8) with respect to *r*, we get the corresponding *Ψ*-function that is

$$\psi(r) = r\left[1 + \left(\frac{r}{c}\right)^4\right]^{-2} \qquad for \qquad |r| \geq 0 \qquad (10)$$

Dividing the above *Ψ*-function by *r* we get the corresponding weight function as given below

$$w(r) = \left[1 + \left(\frac{r}{c}\right)^4\right]^{-2} \qquad for \qquad |r| \geq 0 \qquad (11)$$

We graph the *ρ*-function, *Ψ*-function and the weight function in Fig. 1.

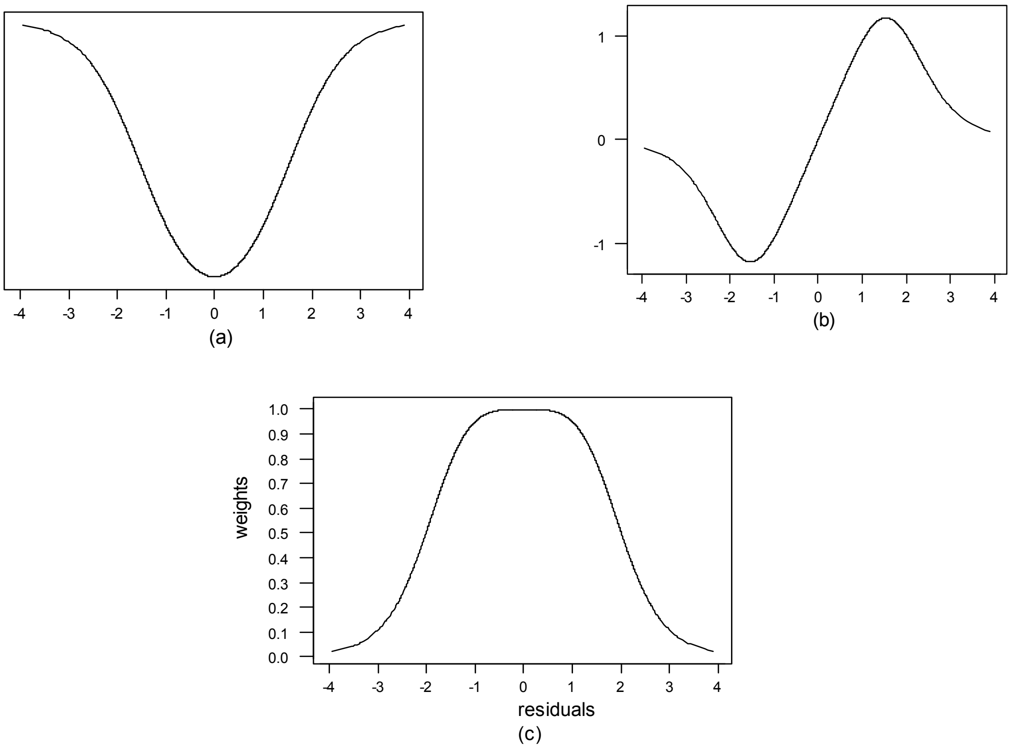

Figure 1: Graph of Insha's function (a) Objective function (b) *Ψ*-function (c) Weight function.

4. Comparison

All of the above mentioned redescending M-estimators works well in detecting outliers and eliminating their influence on the estimates but with some drawbacks. Hampel's three part function requires the user to choose three tuning parameters in the *Ψ*-function which is undesirable. Second the lack of differentiability of its *Ψ*-function is not ideal. Andrew's sine function and Tukey's biweight function covers the drawbacks of Hampel's three part function up to some extent and give robust estimates but on the sacrifice of many good observations. They can give full weight i.e. one only to the observations that have zero residuals and starts to downweight as they slightly departure from zero. According to Hampel, et al. (1986) we can divide the observations into three categories: clear outliers, doubtful outliers and good observations. A good redescending M-estimator will be that which specify some boundary for the good observations and gives full weight within that specified bound and starts to downweight outside the boundary of good observations (doubtful observations) and finally gives zero weight to outliers. There exists such an estimator called *hyperbolic tangent estimator*. This estimator divides the data into three categories mentioned above but the problem is that one has to choose two tuning constants while using this estimator also its *Ψ*-function lacks differentiability and is not very easy to apply.

Also the well-known Winsor's principle states that all the distributions are Normal in the middle. Hence the *Ψ*-function of M-estimators should resemble the one that is optimal for Normal data in the middle. Since the maximum likelihood estimate for Normal data is the mean which has a linear *Ψ*-function, it is desired that $\psi(r) \approx kr$ for small $|r|$, where k is a nonzero constant. In general, *Ψ*-function that is linear in the middle results in better efficiency at the normal distribution (Tukey, 1960). So we need such an estimator that treats the central observations linearly like OLS and then redescends.

The first improvement of the proposed function over others is that it has a continuous derivative everywhere and the second one is that its *Ψ*-function capture the property of longer linear central section from the *Ψ*-function of least squares and behaves linearly for large number of central values as compared to other smoothly redescending *Ψ*-functions. This increased linearity certainly responses in enhanced efficiency (Ali et. al., 2005). The *Ψ*-function then redescends gradually for increasing value of residuals. Moreover, the proposed *Ψ*-function is much more convenient than the previously mentioned *Ψ*-functions because we can write it in closed form without the use of an indictor function. This saves a few steps in programming an iterative method to minimize the objective function. A comparative graph of some of the well known redescending *Ψ*-functions together with the Insha's *Ψ*-function is presented in Fig. 2.

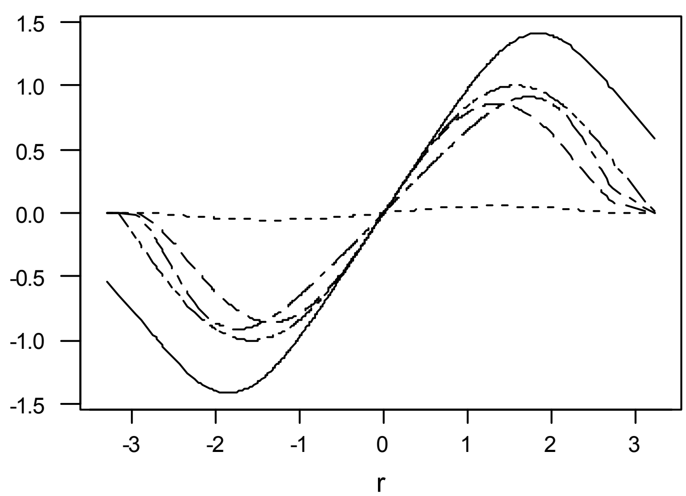

Figure 2: Combined plots of Andrew's (Dash 2-Dot), Tukey's (Dash), Qadir's (Dot), Asad's (Dash 1-Dot) and Insha's (Solid) Ψ-functions.

5. Applications

To illustrate the performance of the proposed estimator as compared to other redescending M-estimators an analysis of a real data set and some simulation results are presented below.

5.1 Example: Annual Rates of Growth of Prices in China

This example is taken from Rousseeuw and Leroy (1987) and is given in Table 1. The response variable is the growth of the average prices in the main cities of Free China while the predictor variable is the year. For instance in 1940 prices went up 1.62% as compared to the previous year. In 1948 a huge jump occurred as a result of enormous government spending, the budget deficit, and the war, leading to what is called *hyperinflation*.

The LS regression equation is given by $\hat{y} = -1049 + 24.85x$, whereas the LMS equation is $\hat{y} = -2.47 + 0.102x$. The equations estimated by other redescending M-estimators such as Tukey's biweight function, Andrew's sine function and Qadir's beta function are $\hat{y} = -2.75 + 0.109x$, $\hat{y} = -2.72 + 0.108x$ and $\hat{y} = -2.75 + 0.109x$ respectively. The estimate by our method is $\hat{y} = -2.64 + 0.106x$, which effectively ignores the outliers and gives closed estimates to other robust methods.

Table 1: Annual rates of growth of average prices in the main cities of Free China from 1940 to 1948.

Year	Growth of prices	Estimated Growth						
(x_i)	(y_i)	LS	LMS	Tukey's	Andrew's	Qadir's	Asad's	Insha's
40	1.62	-55.67	1.61	1.60	1.61	1.60	1.60	1.61
41	1.63	-30.82	1.71	1.71	1.71	1.71	1.71	1.71
42	1.90	-5.98	1.82	1.82	1.82	1.82	1.82	1.82
43	2.64	18.87	1.92	1.93	1.93	1.93	1.93	1.93
44	2.05	43.71	2.02	2.04	2.04	2.04	2.03	2.03
45	2.13	68.56	2.12	2.15	2.15	2.15	2.14	2.14
46	1.94	93.40	2.22	2.26	2.25	2.26	2.25	2.24
47	15.50	118.25	2.33	2.37	2.36	2.37	2.36	2.35
48	364.00	143.09	2.43	2.48	2.47	2.48	2.47	2.46

To see the relative performance, Table 1 lists the estimated values by all these methods. All the robust methods together with the Insha's function provide a fair approximation to the majority of the data, except of course for the last two years, where the observed y_i go astray. On the other hand, the LS fit is bad everywhere: The estimated $\hat{y}_i$ is even negative for the first three years, after which it becomes much too large, except for the 1948 value, which it cannot match either. Least squares smear out the effect (of nonlinearity of the original data) over the whole column, whereas all the other robust methods fits the majority of the data (where it is indeed linear) and allows the discrepancy to show up in those two years where actually something went wrong.

5.2 Simulation Results

One of the popular ways for the comparison of different estimators is to carry out simulation, because in such a situation one knows the true parameter values of the generated data. In this section we present some of the simulation results to check the performance of new redescending M-estimator as compared to other well known redescending M-estimators. For this purpose we have resorted to three types of configurations. First one is the normal situation,

$$y_i = 2 + x_{i1} + ... + x_{ip} + e_i ,$$

in which e_i ~ $N(0, 1)$ and the explanatory variables are generated as x_{ij} ~ $N(0, 100)$ for $j = 1,..., p$. The least squares estimates are obtained for the generated model, then the data is contaminated by replacing 20% of the observations by outliers in the y-direction generated according to the above model but using an error term e_i ~ $N(50, 1)$.

Finally, in the third situation we introduce outliers in the x-direction in such a way that 90% of the observations are again as in the first situation. In the remaining 10% the y_i are generated as before, but afterwards the x_{i1} are replaced by values that are now normally distributed with mean 500 and variance 10.

The purpose of our simulation is to measure to what extent the estimates by our new method differ from the true values i.e. $\beta_0 = 2$ and $\beta_1 = \beta_2 = ... = \beta_p = 1$ and from the estimates provided by Andrew's sine function and Tukey's biweight function the two well known redescending M-estimators. In our simulation we performed many replications keeping in view the number of predictor variables and the sample size n. Some results are presented in Table 2, 3, and 4. From these tables it is clear that the results of the new redesceding M-estimator are very similar to that of OLS without outliers, Andrew's sine function and Tukey's biweight function and is not seriously effected by outliers in both x and y directions.

Table 2: Simulation Results of Regression with Intercept, for n=50 and p=2 (including intercept term).

Method Used	Normal		Outliers in y		Outliers in x	
	β_0	β_1	β_0	β_1	β_0	β_1
OLS	2.04	0.999	11.97	1.04	-28.18	0.264
Andrew (1.5)	2.06	0.999	2.07	0.998	2.03	0.999
Tukey (4)	2.07	0.999	2.08	0.998	2.03	0.998
Qadir (4)	2.07	0.999	2.08	0.998	2.03	0.998
Asad (4)	2.06	0.999	2.07	0.998	2.03	0.999
Insha (4)	2.05	0.999	2.07	0.998	2.03	0.999

Table 3: Simulation Results of Regression with Intercept, for n=100 and p=3 (including intercept term).

Method Used	Normal			Outliers in y			Outliers in x		
	β_0	β_1	β_2	β_0	β_1	β_2	β_0	β_1	β_2
OLS	2.07	1.00	1.00	12.14	1.02	0.993	-18.70	0.32	0.99
Andrew(1)	2.05	1.00	1.00	2.04	1.00	0.999	2.04	1.00	1.00
Tukey(3)	2.05	1.00	1.00	2.04	1.00	1.000	2.04	1.00	1.00
Qadir (3)	2.05	1.00	1.00	2.04	1.00	1.000	2.04	1.00	1.00
Asad (3)	2.05	1.00	1.00	2.05	1.00	1.000	2.05	1.00	1.00
Insha(3)	2.05	1.00	1.00	2.05	1.00	0.999	2.05	1.00	1.00

Table 4: Simulation Results of Regression with Intercept, for n=500 and p=4 (including intercept term).

Method Used	Normal				Outliers in y				Outliers in x			
	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3	β_0	β_1	β_2	β_3
OLS	1.96	1.00	1.00	1.00	8.15	0.878	0.821	0.826	-9.17	0.280	0.915	0.994
Andrew(1.5)	1.96	1.00	1.00	1.00	1.91	0.999	0.999	1.000	1.93	1.00	0.999	0.999
Tukey(4)	1.96	1.00	1.00	1.00	1.91	0.999	0.999	1.000	1.93	1.00	0.999	0.999
Qadir (4)	1.96	1.00	1.00	1.00	1.91	0.999	0.999	1.000	1.93	1.00	0.999	0.999
Asad (4)	1.95	1.00	1.00	1.00	1.91	0.999	0.999	1.000	1.93	1.00	0.999	0.999
Insha(4)	1.95	1.00	1.00	1.00	1.91	0.999	0.999	1.000	1.93	1.00	0.999	0.999

6. Conclusion

Our purpose is to compare the Insha's redescending M-estimator with that of some other redescending M-estimator for robust regression and outliers detection. The estimator is very easy to apply as compared to other redescending M-estimators as its Ψ-function is continuous everywhere and one has to choose only one tuning parameter while using this estimator. It is also clear from Fig. 1 (c) that the estimator gives more opportunity to the observations with small residuals (good observations) to take part fully (with weight equal to 1) in model fitting. We have also shown how this approach can provide an alternative to other robust regression methods. The above example and simulation study show that our estimator is not affected by outliers. The method proposed here requires only modest computational resources and can be executed with standard statistical software because iterative procedure is involved like in other M-estimators. This method requires the calculation of an ordinary least square fit, finding weights from the residuals and applying iteratively reweighted least squares technique to find the estimates of the parameters.

References

1. Ali, A., Qadir, M. F. (2005), A Modified M-estimator for the detection of outliers. *Pakistan journal of statistics and operation research*, 1, 49-64.
2. Ali, A., Qadir, M. F., Ullah, I. (2005), Regression Outliers: New M-Class Ψ-Functions based on Winsor's Principle with Improved Asymptotic Efficiency. Proceedings of the 8th Islamic countries conference on statistical sciences in Lahore, Pakistan.
3. Andrews, D. F., Bickel, P. J., Hampel, F. R., Huber, P. J., Rogers, W. H., and Tukey, J. W. (1972), *Robust Estimates of Location: Survey and Advances*. Princeton University Press, Princeton, New Jersey.
4. Beaton, A.E., and Tukey, J.W. (1974), The firing of power series, meaning polynomials, illustrated on bandspectroscopic data. *Technometrics*, 16, 147-185.
5. 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.

6. Huber P.J. (1973), Robust regression: Asymptotics, conjectures and Monte Carlo, *The Annals of Statistics*, 1, 799–821.
7. Qadir, M.F. (1996), Robust Method for Detection of Single and Multiple Outliers, *Scientific Khyber*, 9 (2): 135-144.
8. Rousseeuw P.J. (1985), "Multivariate estimation with high breakdown point," in *Mathematical Statistics and Applications, Vol. B* (W. Grossmann, G. P.ug, I. Vincze, and W. Wertz, eds.), Reidel Publishing Company, Dordrecht, pp. 283–297.
9. Rousseeuw P.J. and Yohai V.J. (1984), Robust regression by means of S-estimators, in *Robust and Nonlinear Time Series Analysis*, (J. Franke, W. Hardle, and R. Martin, eds.), Lecture Notes in Statistics No. 26, Springer-Verlag, New York, pp. 256–272.
10. Rousseeuw, P.J. (1984), Least Median of Squares Regression, *Journal of the American Statistical Association*, 79, 871–880.
11. Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*. Wiley-Interscience, New York.
12. Tukey, J. W. (1960), A survey of sampling from contaminated distributions, In I.Olkin, S. G. Goraye, W. Hoeffding, W. G. Madow, and H. B. Mann (Eds), Contribution to Probability and statistics, Essaye in Honor of Harald Hotelling, Stanford, CA: Stanford University Press.
13. Yohai, V.J. (1987), High Breakdown Point and High Efficiency Robust Estimates for Regression, *The Annals of Statistics*, 15, 642-656.