Pak. J. Statist.
2016 Vol. 32(6), 425-441

A COMPARATIVE STUDY OF THREE IMPROVED ROBUST REGRESSION PROCEDURES

Dost Muhammad Khan[1§], Sajjad Ahmad Khan[1], Alamgir[2]
Umair Khalil[1] and **Amjad Ali[3]**
[1] Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan
[2] Department of Statistics, University of Peshawar, Peshawar, Pakistan
[3] Department of Statistics, Islamia College Peshawar, Peshawar, Pakistan
[§] Corresponding author Email: dost_uop@yahoo.com

ABSTRACT

In this study we evaluated three robust regression techniques namely least trimmed square (LTS), least absolute deviation (LTA) and a redescending M-estimator in terms of efficiency and robustness in simple and multiple linear regressions using extensive simulations. The impacts of outlier concentration, geometry and outlying distance in both leverage and residual have been assessed. The simulation scenarios focus on outlier configurations likely to be encountered in practice. The results for each scenario provide insight and restrictions to performance to each procedure. From the simulation results we found that the underlying estimators are robust whenever an acceptable percentage of outliers is present in the data and when it crosses that limit, the robust estimators' breakdown occurs. It is also worth-revealing that no single uniform robust method under study may completely be satisfying all the concerns for regression analysis in the presence of outliers. That is one method may be best for one contamination scenario but might not be good for other. For example, for the full coverage $h = n$ (sample size), LTS perform better than LTA for normal errors and LTA perform better than LTS for Laplace errors. Lastly, we summarize each procedure's performance and make recommendations.

KEYWORDS

Outlier; Robust Regression; Breakdown; Leverage Points; Elemental Set.

1. INTRODUCTION

Outliers have interesting and concerned data analysis for centuries and some history of the problem may be found in Barnett and Lewis (1994) or Hawkins (1980). Sometimes the interest is in locating outliers – mining exploration, for example, aims to identify areas where the concentration of a mineral is outlying relative to the background. In other settings, the outliers are nuisances, caused, for example, by data capture errors or the inadvertent contamination of a sample with some observations from a different population. Here, the objective is to mitigate the effect of the outliers on the inferences drawn from the sample. This is the realm of robust inference.

Early approaches to robust inference conflated these two objectives with the procedure of first deciding whether the sample contained any apparent outliers. If so, these outliers were removed, and the remaining observations treated as a clean sample

from the population of interest. Huber (1964) explicitly introduced the idea of robust estimation, removing the connection between outlier identification and the subsequent parameter estimation. Rousseeuw and Leroy (1987) popularized the idea of high breakdown estimation through the use of the PROGRESS codes.

An operational distinction can perhaps be made between robust estimation and high breakdown estimation. Robust estimation aims for the situation in which there is a modest departure from idealized model assumptions. For example, residuals may follow a Laplace distribution rather than a normal, and so be more prone to some extreme values. Or some small numbers of observations, perhaps selected randomly, were contaminated – for example by a misplaced decimal point. High Breakdown Estimation (HBE) assumes a game against a malicious opponent who is allowed to replace a large fraction of the data by values designed to confound the analysis, and HBE is intended to give decent estimates of the underlying parameters despite this distortion.

For a long time it was supposed that high breakdown methods would also be a solution to the tamer robust estimation problem. For example, an initial high breakdown regression fit by PROGRESS using the least median of squares (LMS) criterion produces an initial high breakdown estimate of the regression coefficients and the corresponding residuals, from which a scale estimate follows. Dividing the residuals by the scale estimate allows one to identify apparent outliers. Deleting these and refitting the regression by least squares ("reweighting for efficiency") was thought to provide the best of both worlds – a fit that could accommodate even maliciously designed bad values, and that would nevertheless reduce to conventional regression if the full weight of HBE was not needed. This belief though has proved to be misplaced, leaving much uncertainty about the actual potential of HB methods in less extreme situations.

2. CRITERIA

Let the regression model

$$Y = \mathrm{X}b + \varepsilon \tag{1}$$

where the dependent variable Y and the vector of true residuals ε are $n \times 1$ and the design matrix X is $n \times p$. Let $\hat{\beta}$ be an estimate of β and $e = Y - \mathrm{X}\hat{\beta}$ be the corresponding fitted residuals.

M estimators aim to minimize $\sum_{i=1}^{n} \rho(e_i)$, where the function ρ is even. Ordinary least squares (OLS) uses $\rho(e) = e^2$, which is notoriously sensitive to outliers, particularly those occurring on high leverage cases. The most familiar robust M estimator is L1 regression, in which $\rho(e) = |e|$. It is well known that the L1 criterion is quite robust to outlying Y values provided these occur on cases of only moderate leverage, but that it can be derailed by outlying Y values occurring on high leverage cases. If the Gaussian model holds, L1 regression is only some 70% efficient relative to OLS, but for quite modestly heavier-tailed error distributions is preferable to OLS.

Let $e_{(1)}, e_{(2)}, \ldots, e_{(n)}$ for these n residuals written in ascending order by absolute value. Let h, the "coverage", be an integer between $n/2$ and n. Then conventional high breakdown regression criteria include

- Least median of squares (LMS), (Rousseeuw and Leroy 1987) which minimizes $e_{(h)}^2$.
- Least trimmed squares (LTS), (Rousseeuw and Leroy 1987) which minimizes $\sum_{i=1}^{h} e_{(i)}^2$. This is a modification of OLS to respond only to the more central cases.
- Least trimmed absolute values (LTA) (Hawkins and Olive 1999a) which minimizes $\sum_{i=1}^{h} |e_{(i)}|$. This criterion is a modification of the L1 criterion aimed at solving its sensitivity to high leverage outliers.

As a regression method, LMS is now largely out of favor as it has the problem of being "less efficient" than least trimmed squares regression (Andersen, 2008; Ortiz et al., 2006). This means that it requires a bigger sample size to arrive at the same conclusion in probabilistic terms when the distribution of errors is normal due to its low statistical efficiency relative to LTS.

All three criteria require a choice of the coverage constant h. The choice that maximizes the protection against badly-placed outliers is $h = \lfloor (n + p + 1)/2 \rfloor$ and this is commonly the default used in applying any of the criteria. If, however, there is reason to expect that the number of outliers may be more modest, then increasing h leads to more efficient estimators, but will still protect against the outliers provided they number no more than $n - h$.

It has been suggested (Hawkins and Olive, 1999b) that as the L1 criterion is robust to low-leverage outliers, it may be possible to use larger values of h when using LTA than when using LTS, since there is no particular harm in covering some outliers provided they are of low leverage. This suggestion does not, however, appear to have been investigated in any depth.

Let ψ be the first derivative of ρ. M estimators can be divided into two classes – in the redescending class, $\psi(e) \to 0$ as $|e| \to \infty$; in the non-redescending group this is not the case. L1 is a non-redescender; $\psi(e) = \text{sign}(e)$.

Redescending estimators are potentially high breakdown. Like the trimming estimators LMS, LTS and LTA, they are able to ignore observations that appear to deviate from the consensus model. Unlike the trimming estimators, the amount of trimming is driven by the data; only cases with extreme residuals will be trimmed.

The practical performance of redescending estimators, however, is less clear. The criterion function is not convex, and so convergence requires starting values within the radius of convergence of the global optimum – in other words, it requires a good high breakdown estimator at its start. To deal with this need, redescending estimators are best implemented as follow-on from an initial trimming estimator such as LMS, LTS or LTA.

3. COMPUTATIONAL CONSIDERATIONS

The trimming estimators have combinatorial computational complexity; fitting them exactly requires an exhaustive search over all subsets of a particular size. For LMS, this size is $p+1$; for LTA it is p, and for LTS it is h. With $n > 2p$, this means that an exact solution is least difficult for LTA, followed by LMS, then LTS. This means that an LTA exact fit is thinkable if, say, n is upto 100 and p upto 5, and LMS for the same n and p upto 4, but exact LTS is impossible for sample size of more than a few dozen.

The standard approach for the trimming criteria is to use an approximate method. "Elemental sets" are subsets of cases of size p. Write E for an elemental set, and Y_E and X_E for the elemental set's dependent vector and predictor matrix. Provided X_E is non-singular, the elemental set then gives an exact fit

$$\hat{\beta}_E = X_E^{-1} Y_E .$$

From this elemental set, residuals on all cases can be calculated

$$e(E) = Y - X\hat{\beta}_E$$

Finding the h cases with the smallest absolute residuals provides an initial estimate of the correct subset of cases to cover with the high breakdown estimator.

Applying OLS or L_1 regression to these cases gives a starting candidate for the LTS or LTA fit to the data. Next we apply "concentration" steps. In this step, the residuals from the current candidate LTS or LTA fit are computed for all cases. Sorting these residuals by absolute value gives a refined estimate of the correct subset of cases for the HBE to cover. If this subset differs from the current candidate subset, then it replaced the current candidate subset and OLS or L_1 is applied to it. This process continues until the subset, and consequently the OLS or L_1 fit, stabilizes.

In view of the non-convexity of the criterion, this does not necessarily yield the global optimum, and so the entire procedure needs to be repeated using different starting elemental sets.

An M-estimator for the regression is defined by $\sum_{i=1}^{n} \rho(r_i)$, where r_i represent residuals. If the derivative function $\psi = \rho'$ of ρ is redescending i.e. if it satisfies $\lim_{r_i \to \pm\infty} \psi(r_i) = 0$ then the M-estimator is called redescending.

Redescending estimators are able to reject extreme outliers. They were first introduced by Hampel (Huber, 2004), who used a three part re-descending estimator, with function ρ being bounded and ψ – function being zero for large $|r|$. This estimator has shown to perform well in the Princeton robustness study. But since the Hampel's $\psi(.)$ function is not ideally differentiable and so, a smooth $\psi(.)$ function was required. As a result several smoothly re-descending M-estimators were developed. The most

common of these are Andrew's sine function (Andrews *et al*., 1972), Tukey's biweight function (Beaton and Tukey, 1974). For an overview, see Jajo (2005) and Wu (1985).

Insha Ullah et al. (2006) proposed a re-descending M-estimator which covers some of the drawbacks of the fore mentioned redescenders. The ρ function defined by Insha is

$$\rho(r) = \frac{c^2}{4}\left[Arc\tan\left(\frac{r}{c}\right)^2 + \frac{c^2 r^2}{c^2 + r^2}\right], \quad \text{for } |r| \geq 0 \tag{2}$$

with psi function $\psi(.)$ and weights $w(.)$ being

$$\psi(r) = r\left[1+\left(\frac{r}{c}\right)^4\right]^{-2} \quad \text{for } |r| \geq 0 \tag{3}$$

and

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

where c is the tuning constant and determines the properties of the associated estimators (such as efficiency, influence function, and gross-error sensitivity). Smaller values of 'c' make more resistance to outlier, but at the expense of lower efficiency when the errors are normally distributed. Here ρ function satisfies the standard properties of the re-descending estimators. "The main attraction in the Insha's ψ function is that it performs linearly for large number of central values as compared to other smoothly redesceding ψ functions. This increased linearity in the center certainly responses in enhanced efficiency" (Ali and Qadir, 2005; Ali et al., 2006; Ullah et al., 2006).

As there is no closed form solution for any choice of ρ or ψ function, an Iteratively Reweighted Least Square (IRWLS) fitting algorithm is commonly used to solve for the regression coefficient $\hat{\beta}$ (see, e.g., Holland and Welsch 1977; Fox 2002) and is outlined below.

- Start with initial parameter estimates $\hat{\beta}_{LTS}$ of β and $\hat{\sigma}$ of σ. The Fast Least Trimmed Squares regression is used as a starting estimator.
- Use the initial estimates to form the scaled residuals $r_i = \dfrac{y_i - x_i^t\hat{\beta}}{\hat{\sigma}}$
- Define weights $w_i = \psi(r_i)/r_i \quad = \left[1+\left(\dfrac{r_i}{c}\right)^4\right]^{-2}$
  $= 0$ Otherwise
- Update the estimate $\hat{\beta}$ with the weights w_i by using a weighted least squares estimation.
- Iterate the procedure until convergence.

- In general the tuning constant is chosen to have reasonably maximum efficiency for normal errors. In particular, $c = 5$, for the Insha weight function that give 95% efficiency in case of normally distributed errors and still offers protection against outliers. For detailed discussion about tuning constant and robust estimation see Hogg (1979), Wilcox (2005).

For a given parameter estimate $\hat{\beta}$, there are two common estimates of σ. The first one is the median absolute deviation (MAD). The alternative estimate of σ is Huber's proposal (Street et al., 1988), which is the solution of the form

$$(n-p)^{-1}\sum_{i=1}^{n}\psi_i^2\left(\frac{y_i - x_i^t\hat{\beta}}{\hat{\sigma}}\right) = E_Z\psi^2 = \delta \qquad (5)$$

where $E_z\psi^2(\varepsilon)$ is the expected value of $\psi^2(\varepsilon)$ when ε has a standard normal distribution.

The right side of Eq. (5) is again chosen so that for normally distributed data, $\hat{\sigma}$ estimates the standard deviation. Solving (5) requires iteration. If $\hat{\sigma}_0$ is the present estimate of σ, then the next step in the iteration is defined by

$$\hat{\sigma}^2 = \frac{1}{\delta(n-p)}\sum_{i=1}^{n} w_i^2 r_i^2\left(\hat{\beta}\right) \qquad (6)$$

where $w_i = \psi^2\left(r_i^0\right)\Big/ r_i^0$ with $r_i^0 = \left(y_i - x_i^t\hat{\beta}\right)\Big/\hat{\sigma}_0$.

We compute $\hat{\sigma}$ and $\hat{\beta}$ and iterate the computation steps until convergence. The method thus combines the resistance of the robust Fast-LTS method in the presence of outliers with the efficiency of **redescender o**nce the outliers have been identified.

4. COMPUTATIONAL ALGORITHM

The computational algorithm consists of two steps of optimization to get the final estimates:

I. Elemental Generation Step

- Select *' p '* cases randomly out of n (elemental set) & find its fit θ_E
- Predict all n cases in the data and get the corresponding ordered residuals $r_{(i)}, i = 1, 2, \ldots n$
- Select $h = \left[\frac{n+p+1}{2}\right]$ cases with smallest $r_{(i)}$
- Find $Q_i = \sum_{i=1}^{h}\left|r_{(i)}\right|$
- Keep track the smallest such Q, its covered cases (h) and elemental fit

II. Concentration Step

- Apply L_1 / L_2 to the h covered cases, calculate $r_{(i)}$ for all n cases, smallest h residuals spotted and calculate $Q_j = \sum_{i=1}^{h} \left| r_{(i)} \right|$
- If $Q_j < Q_i$, take these h cases with the smallest residuals and fit L_1 / L_2 and repeat this process until Q is stable. Report final step as a candidate for LTS/LTA solution.

5. BEHAVIOR OF THE PROCEDURES

We investigate three procedures – LTS, LTA and a redescender. The two coverage-based methods LTS and LTA are assessed for a range of coverages h – the maximum breakdown choice $\left\lfloor (n + p + 1)/2 \right\rfloor$, along with $h = 0.6n,\ 0.7n,\ 0.8n,\ 0.9n$ and n, thus illustrating the interplay between the protection and the efficiency.

The primary simulations use $n = 30,\ 300$ with $p = 2$ and $n = 100, p = 5$ illustrating a setting where finding "the right" answers should be relatively straightforward. We will assess all fits by the mean squared Euclidean norm of the error in the vector of slopes.

5.1 Efficiency and Robustness

The first set of simulations are of errors that are $N(0,1)$ and Laplace respectively. We know that for the full coverage $h = n$, LTS is better than LTA for normal errors and LTA is better than LTS for Laplace errors. What's interesting is to see the impact of the different coverages given below.

5.1.1 A Little-known Property of High Breakdown Estimators (HBE)

That high breakdown estimator's produce parameter estimates consistent with the majority of the data does not mean quite what many assume. Consider, for example, a data set of size $n = 300$ in which in 180 of the cases (60%) form a "planet" in which the predictors are $N(0,\ I)$, and the dependent values Y are independent $N(0,\ I)$. Clearly, the true regression of this majority of the cases has slope vector 0, with sampling standard deviation approximately $1/\sqrt{180} = 0.07$. Now imagine that the remaining 40% of cases are replaced by a dense "moon" of cases in which the predictors follow a compact distribution centered at Δ, and the dependent variables are $N(\alpha\Delta, \tau^2)$ where τ is some value close to zero. Many users assume that high breakdown estimators will return estimated slopes of close to 0, but this is not true. Rather, if α is not too large, the contaminated cases will be included in the covered set, and an equal number of "good" cases excluded, yielding a slope estimate of approximately α rather than 0. This is not surprising, but what may be surprising is how large α must be before the contaminating cases are trimmed rather than accommodated. Some straightforward theoretical calculations show that with a single predictor, the cutoff occurs at $|\alpha| \approx 5$. While a slope

of no more than 5 is bounded, as proved in the theory of high breakdown estimation, it is nevertheless some 70 standard errors from what many users might have expected. We illustrate this with an example.

5.1.2 The Effectiveness of HBE in Some Different Contamination Scenarios

Next, we explore the operation of LTS and LTA, applied with different coverages h, and of a redescender in analyzing data sets that include contamination. All our data sets have ($n = 300$ and $p = 2$), ($n = 30$ and $p = 2$), ($n = 100$ and $p = 5$), and include a "planet" of 0.6n points with predictors distributed as $N(0, I)$ and independent dependents that are $N(0,1)$. The true line through these good points therefore has slope vector **0**, and the individual coefficients have sampling standard deviations of approximately 0.07, 0.23 and 0.1 for the data sets having $n = 300$, 30, 100 respectively.

Several different configurations are used for the contamination. The following are the descriptions of the four scenarios:

- First Scenario: Clean data without outliers.
  All the input variables are normally distributed i.e. $X_{ij} \sim N(0,1)$ and $Y_i \sim N(0,1)$.
- Second Scenario: Outliers in the Y-direction.
  We considered a contaminated distribution, where all X_{ij} and 60% of the cases of the response variable Y_i are generated as in the first design and 40% of the Y_i cases are generated from normal N (1000,1).
- Third Scenario: Outliers in both X and *Y*-direction.
  As in the first Scenario, but 20% of the X_{ij} are replaced by the values generated from N (20, 1) and 40% of Y_i are replaced by N (1000, 1) – generated values.
- Fourth Scenario: Outliers in both X and *Y*-direction.
  As in the first design, but 20% of the X_{ij} are replaced by the values generated from normal N (0, 10) and 40% of Y_i are replaced by N (1000, 1) - generated values.

The simulation scenarios place a bunch of outliers at a particular position shifted in *X*-direction (outlying in the explanatory variables only), shifted in *Y*-direction (outlying in the response variable only) and also shifted in both *XY*-directions.

We assess the quality of the estimates by their mean squared error – that is, the squared Euclidean norm of the fitted slope vector. This is estimated by 2000 simulations of each of our configurations. Code development and simulations were done using R for windows version 2.8.1 (Statistical Package).

6. SIMULATION RESULTS

Table 1
MSE's for Robust Estimators based on Clean Data from Normal Errors:
$\left(X_{ij} \sim N(0,1), Y_i \sim N(0,1)\right)$

Coverage								
Criterion	p	n	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
LTS	2	30	0.626	0.596	0.511	0.430	0.352	0.277
		300	0.248	0.209	0.169	0.135	0.106	0.082
	5	100	0.587	0.540	0.462	0.385	0.312	0.230
LTA	2	30	0.644	0.623	0.556	0.483	0.419	0.345
		300	0.270	0.231	0.193	0.159	0.128	0.102
	5	100	0.611	0.568	0.506	0.440	0.365	0.286
Redescender	2	30	0.278	-	-	-	-	-
		300	0.082	-	-	-	-	-
	5	100	0.229	-	-	-	-	-

Table 2
MSE's for Robust Estimators based on Clean Data Generated from Laplace Errors
$\left(X_{ij} \sim N(0,1); Y_i \sim Laplace(0,1)\right)$

Coverage								
Criterion	p	n	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
LTS	2	30	0.574	0.548	0.478	0.415	0.386	0.388
		300	0.147	0.136	0.126	0.118	0.111	0.114
	5	100	0.478	0.432	0.378	0.342	0.320	0.327
LTA	2	30	0.618	0.586	0.508	0.447	0.389	0.350
		300	0.152	0.133	0.119	0.107	0.099	0.092
	5	100	0.511	0.462	0.401	0.352	0.313	0.281
Redescender	2	30	0.368	-	-	-	-	-
		300	0.107	-	-	-	-	-
	5	100	0.308	-	-	-	-	-

Table 3
MSE's for the Robust Estimators based on Contaminated Data from Normal Errors
$\left[40\%\ of\ Y \sim N(1000,\ 1) + 60\%\ of\ Y \sim N\ (0,1),\ X \sim N\ (0,1)\right]$

Coverage								
Criterion	*p*	*n*	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
LTS	2	30	0.444	0.378	332	333	264	136
		300	0.162	0.106	162	188	146	40
	5	100	0.419	0.305	247	305	244	113
LTA	2	30	0.526	0.470	41	162	209	163
		300	0.194	0.139	0.204	45	53	43
	5	100	0.494	0.384	0.626	38	107	76
Redescender	2	30	0.278	-	-	-	-	-
		300	0.109	-	-	-	-	-
	5	100	0.310	-	-	-	-	-

Table 4
MSE's for Robust Estimators based on Contaminated Data from Normal Errors
$\left[80\%\ of\ X \sim N(0,1) + 20\%\ of\ X \sim N(20,1); 60\%\ of\ Y \sim N(0,\ 1) + 40\%\ of\ Y \sim N(1000,1)\right]$

Coverage								
Criterion	*p*	*n*	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
LTS	2	30	0.421	0.369	36	36	119	84
		300	0.164	0.109	35	35	88	35
	5	100	0.423	0.311	23	22	136	83
LTA	2	30	0.521	0.470	37	36	37	36
		300	0.194	0.137	35	35	35	34
	5	100	0.505	0.384	23	23	23	23
Redescender	2	30	0.310	-	-	-	-	-
		300	0.111	-	-	-	-	-
	5	100	0.313	-	-	-	-	-

Table 5
MSE's for the Robust Estimators based on Contaminated Data from Normal Errors
$\left[80\%\ of\ X \sim N(0,1) + 20\%\ of\ X \sim N(0,10);\ 60\%\ of\ Y \sim N\ (0,1) + 40\%\ of\ Y \sim N(1000,1)\right]$

Coverage								
Criterion	*p*	*n*	**50%**	**60%**	**70%**	**80%**	**90%**	**100%**
LTS	2	30	0.421	0.369	95	130	113	48
		300	0.164	0.109	85	86	66	10
	5	100	0.424	0.311	88	120	105	36
LTA	2	30	0.521	0.470	81	96	99	79
		300	0.194	0.137	63	68	67	30
	5	100	0.505	0.384	75	88	90	68
Redescender	2	30	0.310	-	-	-	-	-
		300	0.111	-	-	-	-	-
	5	100	0.313	-	-	-	-	-

7. DISCUSSION ON THE SIMULATION RESULTS

The simulation plan considers the effect of dimensions of the data, sample size and proportion of outliers in the data. Here, the number of predictors, p, are 2 and 5. Three sample sizes are considered: $n = 30$, 100 and 300.

The simulation results in the above tables speak for themselves. We have the following findings from the simulation results:

For a fixed number of predictors, the larger the sample size, the smaller the value of the mean square error is. The dimension also plays a significant role in the behaviors of the robust estimators in our study. The relation of the sample size and dimension becomes a significant factor when the percentage of outliers is relatively high. These results are very analogous to common conclusions in the literature of robustness in which the dimension, sample size, and proportion of outliers in the data are the focal factors on the influence of performance of robust estimators.

Another outcome of the coverage-based robust estimators (LTS and LTA) is the choice of the coverage parameter h. It can be concluded from the simulation results that the higher the value of h, the higher the efficiency of LTS and LTA will be, and the more stable the identification of outliers is, provided that the value of h is not large enough to include the existing outliers. Moreover, large values of the coverage parameter h may be needed for higher efficient estimates and the stability of identification of outliers. For instance, we compare coverage [0.5n] and [0.6n] for the values of h when the data is clean (Table 1 and 2) and when there is high proportion of outliers in x and y -direction (Table 3, 4 and 5). The MSE's of the estimates are smaller for $h = [0.6n]$ than

that for $h=\lfloor 0.5n \rfloor$ when 40% of the outliers exist in the data with different dimensions and sample sizes. For example, for $n=300$ and $p=2$, the MSE of the LTS regression coefficient is 0.162 as reported in Table 3 for *h* = [0.5n], whereas the result is improved to be 0.106 for $h=\lfloor 0.6n \rfloor$.

The relative performance of the three robust procedures is evident from the simulation results. The first set of simulation (Table 1) where the data are all clean $N(0,1)$. For the same dimension and sample size $(n=30,\ p=2)$, we get the efficiency by dividing the absolute constant 0.277 (the MSE of full-sample OLS) to that of the estimator. So LTS is 44.2% efficient for $h=\lfloor 0.5n \rfloor$, while the efficiency of LTA is 54%. But as the coverage increases LTS outperform LTA and Redescender. Similarly, when the data are all clean $N(0,1)$, $n=30$, $p=2$, LTA with 100% coverage gets the efficiency of 0.277/0.345=80% as compared to LTS and it retains the same efficiency for other dimension and sample sizes like for $n=300,\ p=2$ and $n=100,\ p=5$, the LTA is 80% efficient. Redescender with 50% coverage achieves similar efficiency as the LTS does with 100% coverage.

However, when the data are all clean from Laplace (0, 1), $n=30,\ p=2$, LTA outperforms LTS with 100% coverage as the latter gets the efficiency of 0.350/0.388=90%. Similarly, LTS with 100% coverage gets an efficiency of 80% with $p=2,\ n=300$ and an efficiency of 85% with $p=5$ and $n=100$.

The simulation results in above tables are used to compare how some factors influence the performance of the robust procedures. The smaller value of coverage parameter h provides a higher breakdown to prevent a relatively higher proportion of outliers when keeping other conditions the same.

8. APPLICATIONS OF THE ROBUST PROCEDURES TO REAL DATA SETS

In this section, we apply the underlying high breakdown estimators to some real data sets taken from the previous literature in order to assess the relative performance of these estimators against each other and to deliver accurate underlying estimates in the presence of outliers. In each data, the OLS, LTS, LTA and Redescender estimators are applied and the trimmed sum of square are noted.

8.1 Belgium International Phone Calls Data

We consider a real data set from a Statistical Survey conducted in Belgium from 1950 to 1973 with a few outliers present in the data. The data, we use, consist of the total number of international telephone calls made from Belgium. From 1950 to 1962, it is purely the number of calls, but from somewhere in the end of 1963 to 1969, they recorded the total number of minutes of these calls and again from somewhere in the mid of 1970 till 1973, they recorded total number of calls. The dependent variable is the number of telephone calls made from Belgium and the independent variable is the year (Rousseeuw and Leroy, 1987). It is observed from the results drawn in Table 6 that the estimated values are almost same

for this particular data set under the LTS, LTA and Redescender methods and identify outliers whereas the LS estimates are highly influenced by the presence of outliers. It can also be observed from the Figure 1 that contrary to what happens with the LS, the outliers do not influence LTS, LTA and Redescender fits.

Table 6
Phone Calls Data Fitted by OLS, LTS, LTA and Redescending Estimators

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	TrimRSS
OLS	-26.01	0.504	52.296
LTS	-5.652	0.116	0.0343
LTA	-5.550	0.115	0.0380
Redescender	-5.190	0.109	0.0574

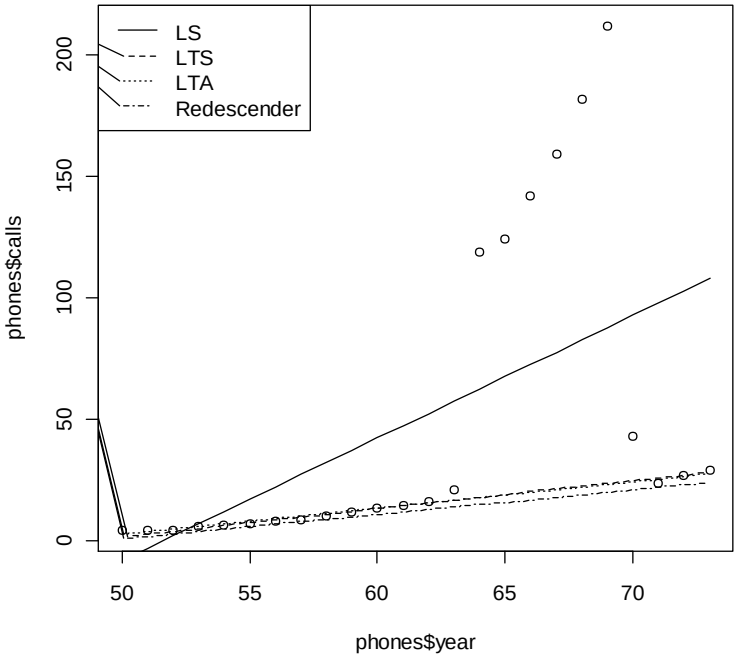

Figure 1: Scatter Plot of the Phone Calls Data with OLS Fit and Robust Fits by LTS, LTA and Redescender

8.2 Brownlee's Stack Loss Data

The second example is that of Stack loss data, a well-known data set taken from Draper and Smith (1998). These are observations from 21 days operation of a plant for the oxidation of ammonia as a stage in the production of nitric acid. The variables are:

X1: air flow
X2: cooling water inlet temperature
X3: 10(acid concentration -50)
Y: stack loss; ten times the percentage of ingoing ammonia escaping unconverted.

This data set has been extensively analyzed in the statistical literature by many statisticians such as Cook (1979), Rousseeuw and Leroy (1987), Carroll and Rupert (1985), Qadir (1996), Ali *et al*., (2005), Ullah *et al*., (2006) and several others by means of different robust methods. One general conclusion is that it is a paradigm of robust regression that observations, 1, 3, 4 and 21 are outliers. According to some analysts, observation 2 is also reported as an outlier. We analyzed this data set by applying LS, LTS, LTA and redescending regression procedures. The results in Table 7 show the estimates of the regression coefficients under various methods of estimation along with trimmed sum of squares. The coverage based estimators completely ignore the influential outliers whereas the redescender down weights the influence of the outliers as can be seen in Figure 2.

Table 7
Stackloss Data Fitted by OLS, LTS, LTA and Redescending Estimators

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	**TrimRSS**
OLS	-39.92	0.716	1.295	-0.152	29.161
LTS	-37.32	0.741	0.392	0.011	2.9324
LTA	-36.67	0.696	0.425	0.025	3.2542
Redescender	-37.51	0.819	0.546	-0.075	2.9323

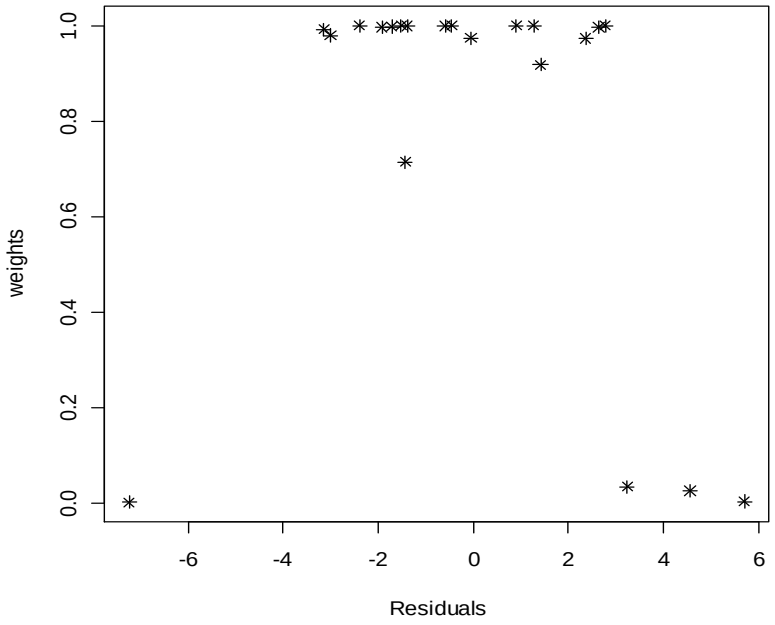

Figure 2: Phone Calls Data: Residuals from OLS Fit vs. Weights given by Redescender's Function

8.3 Hawkins, Bradu and Kass's Artificial Data

This is an Artificial Data set created by Hawkins et al., (1984), which consists of 75 observations with one response and three predictor variables. It provides an excellent example of the masking effect. Here, the outliers are being created in two groups: 1-10 are outliers and 11-14 are swamped inliers making a total of 14 observations as outliers. Only observations 12, 13 and 14 appear as outliers when using conventional methods, but all outliers can be easily unmasked using robust regression procedures. The results in Table 8 show the estimates of the regression coefficients along with trimmed sum of squares under different robust regression procedures. The trimmed residuals sum of square of LTS is smaller than LS, LTA and redescender.

Table 8
HBK Artificial Data Fitted by OLS, LTS, LTA and Redescending Estimators

Method	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	**TrimRSS**
OLS	-0.3875	0.2392	-0.3345	0.3833	9.169046
LTS	-0.61152	0.25487	0.04786	-0.10577	2.947302
LTA	-0.53846	0.30559	0.10883	-0.16100	3.202253
Redescender	-0.18866	0.08493	0.04108	-0.05356	4.258959

9. CONCLUSION

Here, we presented few ways to compare the different methodologies in identifying extreme values or influential observations. In this study, we have covered the possible questions posed by different authors regarding the high breakdown estimate with high efficiency. This work aims at showing a possible way forward in understanding and solving this issue. Here, we have reviewed three robust procedures LTS, LTA and a Redescender for the purpose of comparative effectiveness against the problem of outliers. The performance, its stability and effectiveness of these estimators is evident from both simulations and real data applications. The high breakdown estimate is essentially needed when a relatively high proportion of outlying observations exist in the data. The simulation results also show that the underlying estimators provide a very robust estimation procedure wherever an acceptable proportion of outliers is present in the data and when it crosses that limit, the robust estimators' breakdown occurs.

It is also worth-mentioning from the simulation results that there is no single uniform method under study that may fully be fulfilling all the concerns or criteria for regression analysis in the presence of outliers. That is one approach may be best for one contamination scenario but might not be good for other. Several issues, related to the current work, are still to be resolved and need further exploration. More work needs to be done on this, not only to allow better comparison of methodologies, but also to provide tools for assessing the quality of the implementation of outlier methodology in production.

REFERENCES

1. Andersen, R. (2008). *Modern Methods for Robust Regression,* Sara Miller McCune, SAGE publications, The United States of America.
2. Andrews, D.F., Bickel, P.J., Hampel, F.R., Huber, P.J., Rogers, W.H., Tukey, J.W., (1972). *Robust Estimates of Location: Survey and Advances,* Princeton University Press, Princeton, New Jersey.
3. Ali, A. and Qadir, M.F. (2005). A Modified M-estimator for the detection of outliers. *Pakistan Journal of Statistics and Operations Research,* 1, 49-64.
4. Ali, A., Qadir, M.F. and Salahuddin, A.Q. (2005). Regression Outliers: New M-Class Ψ-Functions based on Winsor's Principle with Improved Asymptotic Efficiency. In *Proceedings of the 8th Islamic Countries Conference on Statistical Sciences*, Lahore, Pakistan Vol. 13, 313-319.
5. Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*, 3rd Ed, Wiley Series, New York.
6. Beaton, A.E. and Tukey, J.W. (1974). The fitting of power series, meaning polynomials, illustrated on hand-spectroscopic data. *Technometrics,* 16, 147-192.
7. Carroll, R.J. and Ruppert, D. (1985). Transformations in regression: a robust analysis. *Technometrics,* 27, 1-12.
8. Cook, R.D. (1979). Influential Observations in Linear Regression, *Journal of American Statistical Association,* 74, 169-174.
9. Draper, N.R and Smith, H. (1998). *Applied Regression Analysis*, 3rd Edition, Wiley Series in Probability and Statistics, New York.
10. Fox, J. (2002). *Robust Regression, Appendix to An R and S-PLUS Companion to Applied Regression,* Thousand Oaks, CA, Sage Publications.
11. Hawkins, D.M. (1980). *Identification of Outliers,* Chapman and Hall, London.
12. Hawkins, D.M., Bradu, D. and Kass, G.V. (1984). Location of several outliers in multiple regression data using elemental sets, *Technometrics,* 26, 197-208.
13. Hawkins, D.M. and Olive, D. (1999a). Applications and Algorithms for Least Trimmed Sum of Absolute Deviations Regression. *Computational Statistics and Data Analysis,* 32, 119-134.
14. Hawkins, D.M. and Olive, D.J. (1999b). Improved Feasible Solution Algorithms for High Breakdown Estimation. *Computational Statistics and Data Analysis,* 30, 1-11.
15. Hogg, R.V. (1979). Statistical Robustness: One View of its use in Application Today, *The American Statistician,* 33, 108-115.
16. Holland, P. and Welsch, R. (1977). Robust Regression Using Interactively Reweighted Least-Squares. *Commun. Statist. Theor. Meth.,* 6, 813-827.
17. Huber, P.J. (1964). Robust Estimation of a Location Parameter. *Annals of Mathematical Statistics,* 35, 73-101.
18. Huber, P.J. (2004). *Robust Statistics*, John Wiley & Sons, New York.
19. Jajo, N.K. (2005). A Review of Robust Regression and Diagnostic Procedures in Linear Regression. *Acta Mathematicae Applicatae Sinica, English Series,* 21(2), 209-224.
20. Ortiz. M.C., Sarabia L.A. and Herrero, A. (2006). Robust regression techniques: A useful alternative for the detection of outlier data in chemical analysis, *Talanta,* 70, 499-512.

21. Qadir, M.F. (1996), Robust Method for Detection of Single and Multiple Outliers. *Scientific Khyber*, 9, 135-144.
22. Rousseeuw, P.J. and Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, Wiley-Interscience, New York.
23. 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.
24. Ullah, I., Qadir, M.F. and Ali, A. (2006). Insha's redecending M-estimator for robust regression: A comparative study. *Pakistan Journal of Statistics and Operations Research*, II(2), 135-144.
25. Wilcox, R.R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*, 2nd Edition, Elsevier academic press, USA.
26. Wu, L.L. (1985). Robust M-estimation of Location and Regression. *Sociological Methodology*, 15, 316-388.