ISSN 1684 – 8403
Journal of Statistics
Vol: 13, No.1 (2006)

Regression Outliers: New *M*-Class ψ-Functions Based on Winsor's Principle With Improved Asymptotic Efficiency

Asad Ali[1], Muhammad. F. Qadir[2] and Salahuddin[2]

Abstract

The well-known Winsor's principle states that all the distributions are normal in the middle. Two new smoothly redescending ψ-functions based on Winsor's principle are proposed in the family of *M*-estimators. The central sections of both of these new ψ-functions resembles with that of the mean, which is linear and this linearity is the actual reason of highest efficiency of the mean under the assumption of normality. The efficiency of an estimator is inversely related to the severity of its robustness. We show that in the class of redescending *M*-Estimators, this new approach produces asymptotically very efficient ψ-functions than that of any other earlier one, while still robust against outliers. The Iteratively Re-weighted Least Squares (IRLS) method based on the proposed ψ-functions clearly detect outliers and ignoring those outliers by giving then zero weights. Two examples selected from the relevant literature, are used for illustrative purposes. The Weighted Least Squares (WLS) method based on the proposed new ψ-functions indeed achieve the goals for which it is constructed. It gives quite improved and satisfactory results in all situations.

[1] GPG College Swabi NWFP Pakistan
[2] University of Peshawar, Peshawar, NWFP, PAKISTAN
Received Dec. 2006, Revised March 2007.

Key Words OLS Regression, Outliers, *M*-Estimators, Winsor's Principle

1. Introduction

Ordinary Least Squares (OLS) is the optimal regression estimator under a set of assumptions on the distribution of the error term and predictor variables. It is well known that OLS method behaves badly when the error distribution is not normal, particularly when the errors are heavy tailed that is if there exists outlying observations. This sensitivity of OLS to outliers results in very misleading results. To cope with this problem the technique of robust regression was developed. The most common general method of robust regression is *M*-estimation, introduced by Huber (1964). 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).

The aim of the current study is to introduce a new family of asymptotically more efficient, smoothly redescending *M*-estimators. This new approach is based on the well-known Winsor’s principle (Tukey, 1960), which states that all the distributions are normal in the origin. Huber introduced the notion of *M*-estimators in 1964 (Huber, 1964), 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 introduction of *M*-estimators is given below.

2. M-Estimators

M-estimators are based on the idea of replacing the squared residuals used in OLS estimation by another function of the residuals, yielding

$$\underset{\hat{\theta}}{\text{minimize}} \sum_{i=1}^{n} \rho(r_i) = 0 \qquad (2.1)$$

where ρ is a symmetric function with a unique minimum at zero. A reasonable ρ-function 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)

Differentiating Equation (2.1) with respect to the regression coefficients yields

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

where ψ is the derivative of ρ and $\mathbf{x}_i$ is the row vector of explanatory variables of the *i*th observation. The *M*-estimate is obtained by solving this system of 'p' nonlinear equations. The solution is not equivariant with respect to scale. Thus, the residuals should be standardized by means of some estimate of the standard deviation σ so that

$$\sum_{i=1}^{n} \psi\left(r_i / \hat{\sigma}\right)\mathbf{x}_i = 0$$

where must be estimated simultaneously. One possibility is to use the median absolute deviation (MAD) scale estimator:

$$\hat{\sigma} = 1.483 \text{ med}\left(\left|r_i - \text{med}(r_i)\right|\right)$$

The multiplication by 1.483 is made so that for normally distributed data $\hat{\sigma}$ is an estimate of the standard deviation. The corresponding *w*-function (weight function) for any ρ is then defined as

$$w(t_i) = \frac{\psi(t_i)}{t_i}$$

where t_i is the *i*th standardized residual. 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 estimating equations may be written as

$$\sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2 .$$

The weighted estimates are computed by solving the equations

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

where W is a n x n diagonal square matrix having the diagonal elements as weights.

2.1 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 $|t|$. 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. 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 departs 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

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

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. The need of a ψ-function with a smoothly redescending nature. Several smoothly redescending M-estimators have been proposed from time to time.

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. Afterwards Qadir (1996) proposed a ψ-function, with weight function is a beta function with $\alpha = \beta$. Recently, Asad (2004) proposed another ψ-function that attains more linearity in its central section. The weights for all these decline as soon as residuals departs from 0, and are 0 for $|t| > a$.

Andrews wave function

$$\psi(t) = \begin{cases} a \sin\left(\dfrac{t}{a}\right) & |t| \le a \\ 0 & \text{otherwise} \end{cases}$$

Turkey's biweight function

$$\psi(t) = \begin{cases} t\left[1 - \left(\dfrac{t}{a}\right)^2\right]^2 & |t| \le a \\ 0 & \text{otherwise} \end{cases}$$

Qadir's beta function

$$\psi(t) = \begin{cases} \dfrac{t}{16}\left[1 - \left(\dfrac{t}{a}\right)^2\right]^2 & |t| \le a \\ 0 & \text{otherwise} \end{cases}$$

Asad's ψ-function

$$\psi(t) = \begin{cases} \dfrac{2t}{3}\left[1 - \left(\dfrac{t}{a}\right)^4\right]^2 & |t| \le a \\ 0 & \text{otherwise} \end{cases}$$

2.2 Asymptotic Variance and Efficiency of *M*-Estimators:

M-estimators are statistically more efficient (at a model with Gaussian errors) than L_1 regression, while at the same time they are still robust with respect to outlying y_i. The breakdown point of *M*-estimators is 0% due to the vulnerability to leverage points (Rousseeuw and Leroy 1987). In the univariate case, the asymptotic variance of the *M*-estimators at symmetric distribution *F* is given by

$$V(\psi, F) = \frac{\int_{-\infty}^{\infty} \psi^2 dF}{\left(\int_{-\infty}^{\infty} \psi' \, dF \right)^2}$$

If Gaussian distribution is assumed so that the r_i are *i.i.d.* and $N(0, \sigma^2)$, the multivariate *M*-estimators have the asymptotic covariance matrix given as

$$V(\psi, \Phi) = \frac{E[(X'X)^{-1}]\sigma^2}{e}$$

where '*e*' is the asymptotic efficiency defined as:

$$e = \frac{1}{V(\psi, \Phi)} = \frac{\left(\int_{-\infty}^{\infty} \psi' \, d\Phi \right)^2}{\int_{-\infty}^{\infty} \psi^2 d\Phi} = \frac{[E(\psi')]^2}{[E(\psi^2)]}$$

In real practice one has to estimate $[E(\psi^2)]$ by $\frac{1}{n}\sum_{i=1}^{n} \psi^2$ and $[E(\psi')]^2$ by $\left(\frac{1}{n}\sum_{i=1}^{n} \psi' \right)^2$ combined as

$$\hat{e} = \frac{\left(\frac{1}{n}\sum_{i=1}^{n} \psi' \right)^2}{\frac{1}{n}\sum_{i=1}^{n} \psi^2}.$$

3. Winsor's Principle

Winsor's principle states that all distributions are normal in the middle. Hence, the ψ-function of *M*-estimators should resemble the one that is optimal for Gaussian data in the middle. Since the Maximum Likelihood estimate for Gaussian data is the mean which has a linear ψ-function, it is desired that $\psi(t) \approx k\,t$ for small $|\,t|$, where k is a nonzero constant. In general, a ψ-function is linear in the middle results in better efficiency at the Gaussian distribution (Tukey 1960).

4. The New ψ-Functions

We propose some new ψ-functions and discuss their properties as compared with other ψ-functions: Andrews function and Turkey's biweight function.

The proposed new ψ-functions, which are derived by *trial-and-error* method, are given below.

$$\psi_1(t) = \begin{cases} \dfrac{t}{2}\left(1-\left(\dfrac{t}{a}\right)^6\right)^2 & if \quad |t| \le a \\ 0 & if \quad |t| > a \end{cases} \qquad (4.1)$$

and

$$\psi_2(t) = \begin{cases} \dfrac{t}{2}\left(1-\left(\dfrac{t}{a}\right)^8\right)^2 & if \quad |t| \le a \\ 0 & if \quad |t| > a \end{cases} \qquad (4.2)$$

where a is the so-called tuning constant and for ith observation the variable 't' are the residuals scaled over MAD.

The ρ-functions corresponding to the above ψ-function satisfies the standard properties, generally associated with a reasonable objective function.

Robustness and Efficiency

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

- How robust the estimator is?
- What is the efficiency of the estimator?

One should be aware of the fact that the efficiency and robustness of an estimator are inversely related. Then 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.

Now the proposed new ψ-functions given by Equations (4.1) and (4.2) have a different behaviour as compared to that of other redescending estimators such as Tukey’s biweight estimators. The proposal is based on the so-called Winsor’s principle stated in section (3). Recalling that the ψ-function of the arithmetic mean is just a linear straight-line rendering it theoretically the most efficient estimator. The proposed new 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 the enhanced efficiency. The ψ-function then redescends gradually for increasing values of residuals and becomes zero for values lying outside the specified band.

In empirical situations a routine data set typically contains (1-10)% or up to 25% outliers in it. Certainly it is more realistic in applied regression to consider data sets with roughly 10% of outliers than with 50% outliers (Hampel *et al* 1986). In any case the proportion of outliers cannot exceed 50%, as then it will be very difficult to distinguish good data points from bad data points. In other words, we can say that at least 50% of the observations in a data set can still be considered as good observations (no matter how heavier tailed the error distribution is). This new approach takes this fact into account and the increased linearity of the central section of the proposed ψ-functions means that the central observations (at least 50%) still receives almost equal weights like OLS.

Asymptotic Efficiency of the Proposed *M*-Estimators

It should be noted that a smoothly redescending *M*-estimator behaves very badly if the errors are really normally distributed. We know that the application of *M*-estimators.

produces bouncing residuals thus making the detection of outliers easier. Now Winsor's principle states that all the distributions are normal in the middle; this means that the observations in the middle can still be treated by OLS method, which gives equal weights to all observations, which is the secret of its highest efficiency. But on the contrary almost all of the currently used smoothly redescending M-estimators produces bouncing (although a very little bounce) residuals for these middle observations too, by giving some weights to these observations which in turn increases residual's sum of squares. The proposal of the current M-estimators is aimed to treat the observations in the middle of the distribution linearly like OLS, which results in high efficiency gain. The proposed new ψ-functions given by Equations (4.1) and (4.2) have the highest ever-attained linearity in their central sections. We evaluate the asymptotic efficiency at the standard Gaussian distribution for the selected estimators, as shown in Figure 4.1.

From the results presented in Table 4.1 we observe that as the tuning constant increase, linearity of the centre of all ψ-functions increases, which results in a higher efficiency of the estimator resembling with that of the mean. But the increment in the values of tuning constants will result in a wider band and the larger no of the observations to be included in the analysis, thus the killing the purpose of robustification, as the outliers may then be escaped. Of course one has to consider 2.5 as the cut off value in the case of standard Gaussian errors.

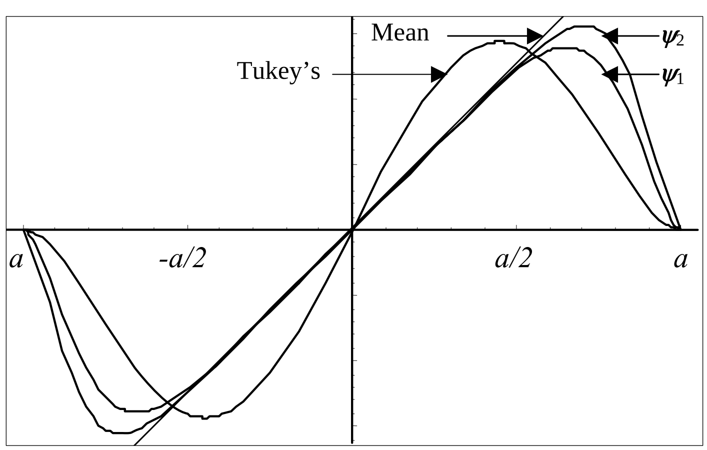


Figure 4.1. **A comparative sketch of different ψ-functions**

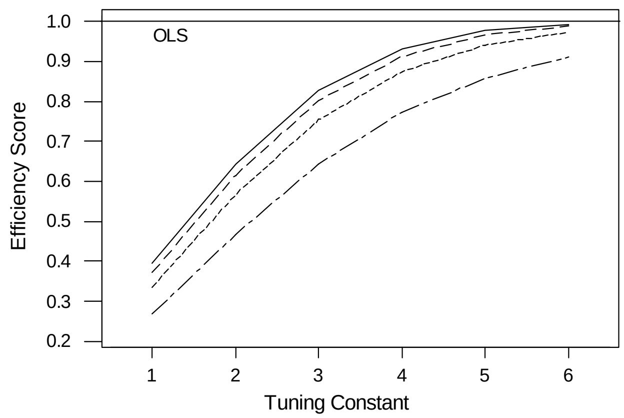
We use strictly theoretical approach, using limited integration with tuning constants as the limits, to pursue the asymptotic efficiency of the selected estimators with the assumption of standard Gaussian normality for noise. We see that asymptotic variance and efficiency of both of the new ψ-functions, that is, ψ_1 and ψ_2 are much improved as compared to other versions. This approach has the benefit of being independent of the sample size.

One has just to evaluate the expectations of the two functions involved in the equation continuously. The above results can be easily verified by generating standard Gaussian random samples of different sizes e.g. 10, 20, 50 and so on, and then using equation 2.16 we can find the numerical values for $\hat{e}$, for different tuning constants.

In Figure (4.2), we see that the efficiency score curves of the proposed ψ_1 (Dashed curve) and ψ_2 (Solid curve) functions are above all others, approaching to that of OLS for a larger tuning constant. Actually Asad's proposal [Asad (2005); Asad and Qadir (2005)] was also an attempt to gain higher efficiency with considerable robustness against outliers. But we see that the new ψ-functions attains the highest efficiency than any other previous

Figure4.2 **Asymptotic Efficiency of Tukey's Biweight (Dash-Dot-Dash), Asad's ψ-function (Dots), ψ_1 (Dashes) and ψ_2 (Solid) compared to OLS (Solid Straight Line).**

ψ-functions, while still having excellent robustness property.



5. 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. Simulation studies show that the new method of estimation is quite insensitive to the presence of outliers and can be applied to detect outliers with a higher efficiency.

6. Real examples:

Telephone Calls Data

Our first example is the familiar telephone calls data set, which is a good example of real regression data with a few outliers in y-direction. 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 Leroy 1987). The data set is executed and analysed by many researchers including Rousseeuw and Leroy (1987), Qadir (1996) and Asad (2004). The scatter plot of the data along with different fits is shown in Figure (6.1). From the plot it is clear that the observations from 1964 to 1969 are outliers. Rousseeuw and Leroy (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).

The fits from OLS and other robust methods along with the proposed method are given in Table (6.1). 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. The fits from the OLS, Tukey's biweight and the proposed method are sketched in Figure (6.1). It is obvious that with the proposed method, the model fits the data 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, where as the fit with the proposed method shows very much robustness and fits a model which represents the majority of the observations and avoids outliers.

Table 6.1 Telephone Calls data fitted by OLS and different robust methods including the proposed method.

Method	Outliers Detected	SS of Residuals
OLS	0	695.44
RWLTS	8	0.1313
RWLMS	8	0.1313
Tukey (a = 3.8)	8	0.1362
Asad (a = 3.0)	8	0.1314
ψ_1 (a = 2.7)	8	0.1313
ψ_2 (a = 2.6)	8	0.1313

The robust fits by the proposed method and Tukey's biweight differ very little therefore in Figure (6.1), the two robust lines cannot be differentiated from one another.

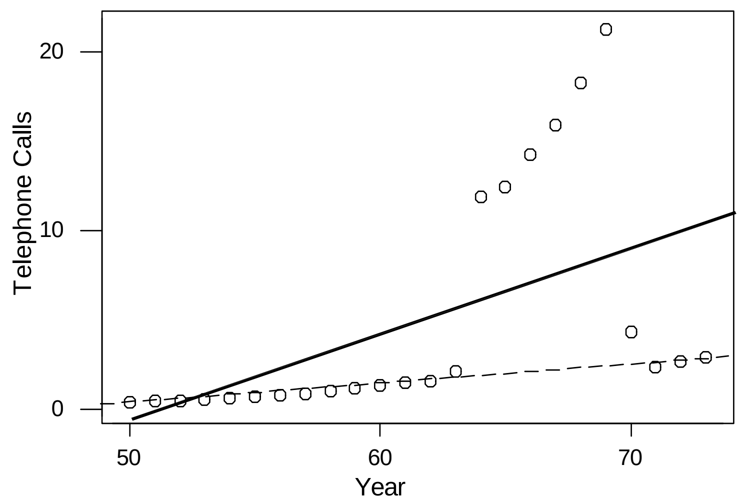


Figure 6.1 **Telephone Calls Data fitted with OLS (solid line) and new WLS based on ψ_1, ψ_2-functions (dashed lines).**

Hadi-Simonoff Data:

Hadi and Simonoff (1993) generated an artificial data set by using the equation given below.

$$y = x_1 + x_2 + e$$

The data contains 25 observations with the errors for first three cases, while for the rest of the 22 observations; the errors were generated from N (0, 1). This data set has three outliers in the response direction. The results summarized in Table (6.2) reveal that the proposed method is quite effective in detecting the artificially introduced outliers. If we exclude the three outlying observations from the actual data set then the OLS fit results in the model,

$$y = -0.535 + 1.012x_1 + 1.055x_2 \qquad (6.2)$$

The Residual Sum of Squares (RSS) obtained from this fitted model is equal to 13.594.

Now the RWLTS and RWLMS both detect several observations as outliers other than the actual outliers. Our new method also detects the same number of outliers for a narrow band of tuning constants. Of course this results in a more robust fit but on the cost

Table 6.2 Hadi-Simonoff data fitted by OLS and different robust methods including the proposed methods.

Method	Outliers Detected	SS of Residuals
OLS	---	35.800
RWLTS	8	3.176
RWLMS	10	1.265
ψ_1 (a = 2.0)	10	1.270
ψ_1 (a = 2.5)	9	2.167
ψ_1 (a = 3.0)	5	9.517
ψ_1 (a = 4.0)	3	13.628
ψ_2 (a = 2.0)	10	1.266
ψ_2 (a = 2.5)	7	4.409
ψ_2 (a = 3.0)	5	9.482
ψ_2 (a = 4.0)	3	13.600

of reduced sample size and hence the reduced efficiency of the estimators. For a = 4.0 both of our functions gives results very similar to that of OLS (Equation 6.2) with three observations removed, declaring it as efficient as the OLS. The rest of the summary is self-explanatory. We know that RWLTS and RWLMS are actually OLS fits to the refined data set after the outliers removed. The residuals from the proposed method are actually the weighted residuals and hence their sum of squares is a weighted RSS, which is still much more similar to that of OLS-RSS with deleted observations. For example see results for a = 2.0 and a = 4.0.

7. Simulation Study:

We report a Monte Carlo study in this section, which is designed to investigate the performance of the newly proposed M-Estimators. 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,

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

Table 7.1 **Simulation results for Simple and Multiple Regression**

		Values	OLS No Outliers	OLS	RWLMS	RWLTS	Tukey	Asad	ψ_1	ψ_2
Simple Regression	n =20	β_o	0.912	3.236	1.032	1.032	0.987	0.991	0.998	0.999
	Outliers = 4	β_l	0.996	0.994	1.012	1.012	1.011	1.008	1.004	1.008
		RSS	18.756	358.330	13.214	13.214	13.415	13.285	13.245	13.225
	Outliers Detected		----	----	4	4	4	4	4	4
	n = 1000	β_o	0.984	1.461	0.993	0.997	0.998	0.990	0.994	0.999
	Outliers = 50	β_l	1.003	0.995	1.002	1.001	1.002	1.002	1.002	1.002
		RSS	995.540	5611.72	839.144	828.172	945.428	945.221	945.324	945.054
	Outliers Detected		----	----	65	67	50	50	50	50
Multiple Regression	n = 50	β_o	1.165	3.154	1.182	1.182	1.221	1.220	1.174	1.180
	Outliers = 10	β_l	0.981	1.032	0.989	0.989	0.968	0.991	0.987	0.991
		β_2	1.021	0.910	0.994	0.994	0.989	0.987	0.995	0.994
		RSS	41.895	739.765	37.745	37.745	37.895	37.865	37.754	37.750
	Outliers Detected		----	----	10	10	10	10	10	10
	n = 1000	β_o	1.121	1.591	1.001	1.060	0.944	0.966	0.997	0.998
	Outliers = 50	β_l	0.923	1.052	0.936	0.931	0.955	0.943	0.931	0.935
		β_2	1.051	0.952	1.051	1.055	1.058	1.066	1.041	1.061
		RSS	1088.50	5945.65	924.70	912.19	1045.67	1041.61	1040.21	1040.15
	Outliers Detected		---	---	65	66	50	50	50	50

in which $e_i \sim N(0, 1)$ and the explanatory variables are generated as $x_{i,j} \sim N(0, 100)$ for j = 1,…, 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 number of simulations used in this study was 5000. The results for both ψ_1 and ψ_2 with different (simple and multiple) regressions are given in Table (7.1). From the Table it is clear that the proposed method is quite effective in detecting outliers and reduces the RSS to a reasonable extent.

8. Conclusions

We proposed two new ψ-functions in the class of smoothly redescending M-*e*stimators. We showed that under the normality assumption of errors, the proposed estimators are asymptotically more efficient than any other earlier version of smoothly redescending M-estimator. The theoretical results are verified by two well-known numerical examples. The simulation results strengthen our results. The estimators are quite successful in detecting outliers with a weighted RSS equivalent to that of OLS with omitted outliers hence gaining efficiency equivalent to that of OLS in the center of errors distribution.

References:

1. Andrews, D. F. (1974), "A Robust Method for Multiple Linear Regression," *Technometrics*, **16**, 523-531.
2. Asad, A. (2004), "A Modified *M*-Class Robust Regression Criterion for Single and Multiple Outliers," Unpublished M.Phil thesis, Department of Statistics, University of Peshawar.
3. Asad, A. and Qadir, M. F. (2005), "A Modified *M*-Estimator for the Detection of Outliers," *Pakistan Journal of Statistics and Operation Research*. **1**,
4. Hadi, A. S. and Simonoff, J. S. (1993), "Procedures for the Identification of Outliers in Linear Models," *Journal of the American Statistical Association*, **88**, 1264-1972.
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. Hoaglin, D. C, Mosteller, F., and Tukey, J. W. (1983), *Understanding Robust and Exploratory Data Analysis*, John Wiley and Sons, New York.
7. Mosteller, F. and Tukey, J. W. (1977), *Data Analysis and Regression*, Addison-Wesley Publishing Company, Inc. Philippines.

8. Qadir, M. F. (1996), "Robust Method for Detection of Single and Multiple Outliers," *Scientific Khyber*, **9**, 135-144.
9. Rousseeuw, P.J. (1984), "Least Median of Squares Regression," *Journal of the American Statistical Association*, **79**, 871-880.
10. Rousseeuw, P.J. and Leroy, A.M. (1987), *Robust Regression and Outlier Detection*, John Wiley & Sons, New York.
11. Rousseeuw, P.J. and Van Driessen, K. (1998), "Computing LTS Regression for Large Data Sets," Technical Report, submitted to University of Antwerp, Belgium.
12. 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.
13. 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.
14. 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, Essays in Honor of Herald Hotelling*, Stanford, CA: Stanford University Press.
15. Yohai, V.J. (1987), "High Breakdown Point and High Efficiency Robust Estimates for Regression," *Annals of Statistic*s, **15**, 642-656.