Alternative Weighted Least Squares Estimators

Renumas Gulasirima
Ph.D. Candidate, School of Applied Statistics,
National Institute of Development Administration,
Bangkapi, Bangkok 10240, Thailand.
Lecturer, Faculty of Science and Technology, Suan Dusit Rajabhat University, Bangkok 10300, Thailand.
renumas99@yahoo.com

Pachitjanut Siripanich,
Associate Professor, School of Applied Statistics,
National Institute of Development Administration,
118 Serithai Road, Klongchan, Bangkapi,
Bangkok 10240, Thailand.
Tel.662-727-3035, Fax: 662-374-4061
edu@as.nida.ac.th

Abstract

In this paper, alternative regression estimators in a class of weighted least squares estimators were proposed. Follow the idea of Windham (1995) in robustifying model fitting, two types of weight function are considered: One applied to every value of residuals, the other applied partially. Under the assumptions that the error terms are i.i.d. normal with zero mean and constant variance, it is found that these alternative estimators are more resistant or robust than the ordinary least squares estimator in the situation of outliers. Many numerical examples of various situations concerning outliers exhibit that the two alternative estimators are more preferable than the least squares estimator by means of R^2 and MSE.

Keywords: robustifying model fitting, weighted least squares

1. Introduction

One of the main problems in analyzing a data is outliers. Many robust statistical methods have been developed to cope various types of effect and problems caused by outliers. One of those was proposed by Windham [8], such method was called 'Robustifying Model Fitting'. Unknown parameters are estimated from transformed data that is weighted by

$$w(x;\theta)=\frac{g^c(x;\theta)}{E_F[g^c(X;\theta)]}, \tag{1.1}$$

where $\{g(x;\theta)\}$ is assumed to be a parametric unimodal family of densities with unknown parameter θ and c is a positive constant. Windham has studied for continuous univariate families, especially in exponential family and found that a weighted distribution is measured invariance with decreasing variance.

For a sample of size n, each observation x_i is weighted by

$$w_i = w(x_i;\hat{\theta}) = \frac{Kg^c(x_i)}{\sum g^c(x_i)},\quad i=1,2,...,n, \tag{1.2}$$

where K is such that $\sum w_i = 1$, c is a pre-assigned value and g is a density of a chosen distribution.

Robustifying is a data transformation method, so that the effect of outliers is negligible after some numbers of iteration. We have found that if X is a random variable from $N(\mu\ ,\ \sigma^2)$, then the weighted random variable w(x)X, say X_w is distributed as $N\left(\mu\ ,\ \frac{\sigma^2}{c+1}\right)$.

Followed Windham papers, Basu et al. [1], [2] and Choi et al. [3] proposed the estimators which procedures connected to approach suggested by Windham (1995). These estimators are presented in forms of functional.

In our study, we considered a linear regression model: $y = X\beta+\varepsilon$ with some influential observations and $\varepsilon \sim NID(0\ ,\ \sigma^2 V)$ where V is a diagonal matrix. Windham's robustifying method is applied on the residual $r_i = y_i - \hat{y}_i$, $i=1,2,...,n$ where $\hat{y}_i$'s are obtained by least squares (LS) method.

2. Robust Regression Weighting

In regression analysis, one basic way to overcome the problem of outliers is the robust regression in which the effect of outliers is reduced. The first robust estimator might be the least absolute deviation regression estimator proposed by Edgeworth (1887). This estimator is obtained by minimizing sum of absolute residuals and hence it can detect the outlier only in y-direction.

Our interest is focused on weight-type estimators. Huber [5] extended the idea of M-estimator to the M-regression which is optimal if the error is assumed to be normally distribution contaminated by small fraction of other distribution. His criterion is minimization of the maximum possible variance for infinitely large samples. Huber suggested to minimize the objective function of residuals compromised between r^2 and $|r|$. In other words, Huber introduced the weight function

$$w(r)=\begin{cases}1 & \text{for } |r|\le k,\\ k/|r| & \text{for } |r|>k,\end{cases} \tag{2.1}$$

where k is a tuning positive constant.

One of the well-known robust regressions is the least median of squares estimator (LMS) which is obtained by minimizing the medians of squared residuals, $\min_{\hat{\beta}} \operatorname{med}_i r_i^2$ proposed by Rousseeuw [7] based on the idea of Hample [6]. It turns out that this estimator is very robust with respect to outliers in y as well as in x. Unfortunately, the LMS performs poorly in the sense of asymptotic efficiency. Later, Rousseeuw suggested the WLSE in which the objective function is $\min_{\hat{\beta}} \sum wr^2$, where $w = w(r)$ is defined as

$$w(r) = \begin{cases} 1 & \text{if } \left|r/\hat{\sigma}^*\right| \le 2.5, \\ 0 & \text{if } \left|r/\hat{\sigma}^*\right| > 2.5. \end{cases} \quad (2.2)$$

where $\hat{\sigma}^*$ is Rousseeuw's estimator of standard deviation of residual.

Recently, based on the idea of Rousseeuw (1987), Daniel and Yohai [4] introduced the robust and fully efficient regression estimator (REWLS). The objective function of the REWLS is $\min_{\hat{\beta}} \sum wr^2$ where $w = w(u)$ is a non-increasing function and is defined as

$$w(u) = \begin{cases} 1 & \text{if } u = 0, \\ g(u) & \text{if } 0 < |u| \le 1, \\ 0 & \text{if } |u| > 1, \end{cases} \quad (2.3)$$

where $g(u) > 0$, and u is defined to be proportional to $|r|$.

3. Alternative Weighted Least Squares Estimators

In this paper we concentrate on a linear regression model $y = X\beta + \varepsilon$, where outliers occur one way or another. A method called "Robustifying" proposed by Windham (1995) is applied on an estimated residual vector $r = y - \hat{y}$ where $\hat{y} = X\hat{\beta}$, and $\hat{\beta}$ is an estimator of β. In particular $\hat{\beta}$ is the LS estimator. Since ε_i's are assumed to be i.i.d. normal, hence the "robustifying weight" according to Windham, is

$$w_R(r_j) = \frac{nf^c(r_j)}{\sum_{j=1}^{n} f^c(r_j)} \quad (3.1)$$

where $r_j = y_j - \hat{y}_j$ and $f(r_j)$ is a density function of normal distribution. Note that $\sum_{j=1}^{n} w_R(r_j) = n$. Therefore the robustifying weighted least squares estimator or RWLS1 in short, can be computed from the following formula.

$$\hat{\beta}_R = (X'W_RX)^{-1}X'W_RY \quad (3.2)$$

where $W_R = \mathrm{diag}\left(w_R(r_1), w_R(r_2), ..., w_R(r_n)\right)$.

According to [7] and [4], the partial weighted least squares estimator is introduced. For instant, let k be a positive constant depending on fraction α of influential outliers, i.e., $P[|R_j - \mu| > \sigma k] = \alpha$ where $R_j = Y_j - \hat{Y}_j$ is an estimator of residual, $\mu = E(R_j)$ and $\sigma^2 = \mathrm{Var}(R_j)$ for $j = 1, 2, ..., n$. Thus, the partial robustifying weight is

$$w_P(r_j) = \begin{cases} 1 & ; \quad |r_j - \hat{\mu}| \le k\hat{\sigma}, \\ \min\left(1, \dfrac{nf^c(r_j)}{\sum_{j=1}^{n} f^c(r_j)}\right) & ; \quad |r_j - \hat{\mu}| > k\hat{\sigma}. \end{cases} \quad (3.3)$$

Note that $\hat{\mu} = \frac{\sum w_P r}{\sum w_P} = \frac{\sum w_P (y - \hat{y})}{\sum w_P}$, and $\hat{\sigma}^2 = \frac{\sum w_P (r - \hat{\mu})^2}{\sum w_P}$ where $w_P = w_P(r)$. Hence the partial robustifying weighted least square estimator or RWLS2 can be obtained as follow:

$$\hat{\beta}_P = (X'W_PX)^{-1}X'W_PY \quad (3.4)$$

where $W_P = \mathrm{diag}\left(w_P(r_1), w_P(r_2), ..., w_P(r_n)\right)$.

4. Computation of the Alternative Estimators

Alternative estimators are obtained by weighted least squares method as can be seen in equation (3.2) and (3.4). The weights are computed as follow:

Consider a fraction of data size $(1-\alpha)n$ that has no outlier (α is a proportion of outliers and n is a size of sample). That means outliers must be identified and deleted from the original data. (There are many methods of identifying outliers such as using Cook's distance, etc.) Suppose $\hat{\beta}_0$ is a vector of the LS estimator of regression coefficients applied on a sub-sample without outliers. Calculate all n residuals: $r_j = y_j - \hat{y}_j = y_j - x_j'\hat{\beta}_0$, $j = 1, 2, ..., n$.

Apply Windham's robustifying model fitting on $r_1, r_2, ..., r_n$ by means of normal distribution with initial parameters estimated by the maximum likelihood estimators (MLE), that are

$$\hat{\mu}_0 = \frac{1}{n}\sum_{j=1}^{n} r_j = \frac{1}{n}\sum_{j=1}^{n}\left(y_j - x_j'\hat{\beta}_0\right) \text{ and } \hat{\sigma}_0^2 = \frac{1}{n}\sum_{j=1}^{n}\left(r_j - \hat{\mu}_0\right)^2.$$

Thus the initial weight is $w_0(r) = \frac{nf_0^c(r)}{\sum_{\text{all } r} f_0^c(r)}$,

$$\text{where } f_0(r) = \frac{1}{\hat{\sigma}_0\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{r - \hat{\mu}_0}{\hat{\sigma}_0}\right)^2}. \quad (4.1)$$

For normal distribution Windham suggested that $c = 0.5$. After weighting the residuals new estimators of parameter μ and σ^2 are obtained. At the k^{th} iteration we may have

$$\hat{\mu}_k = \frac{\sum_{j=1}^{n} w_{k-1}\left(r_j\right) r_j}{\sum_{j=1}^{n} w_{k-1}\left(r_j\right)}, \text{ and}$$

$$\hat{\sigma}_k^2 = \left(c+1\right)\frac{\sum_{j=1}^{n} w_{k-1}\left(r_j\right)\left(r_j - \hat{\mu}_k\right)^2}{\sum_{j=1}^{n} w_{k-1}\left(r_j\right)}, \; k = 1, 2, \ldots \tag{4.2}$$

and the process is terminated at the k^{th} iteration if $\left|\hat{\mu}_k - \hat{\mu}_{k-1}\right|$ less than some pre-assigned value. At the last iteration, k^{th} iteration the weighted estimators of the residuals are $\hat{\mu}_w = \hat{\mu}_k$ and $\sigma_w^2 = \hat{\sigma}_k^2$, and the fitted density is

$$f(r_j) = \left(2\pi\hat{\sigma}_w^2\right)^{-1/2} e^{-\frac{1}{2}\left(\frac{r_j - \hat{\mu}_w}{\hat{\sigma}_w}\right)^2}. \tag{4.3}$$

Substitute (4.3) in equation (3.1) and (3.3) , the proposed weight functions, we then obtain two estimators RWLS1 and RWLS2 as appear in equation (3.2) and (3.4) respectively.

5. Numerical Examples

Eight examples are selected from 'Robust Regression and Outlier Detection' by Rousseeuw, P.J. and Leroy, A.M. (1987). These data sets consist of outliers in various situations, some have no influence data points, some have outliers in y-direction and/or x-direction. In each data set, the LS and the LMS estimators are already given. The rest is to compute alternative estimators and then compare with the LS and the LMS estimators.

In the first example, the data set was restored from Daniel and Wood (1971). The response is the titration determined by the acid content, and the explanatory is the extraction and weighting determined by the organic acid content. We can see in Figure 1 that there is no outlier in this data set.

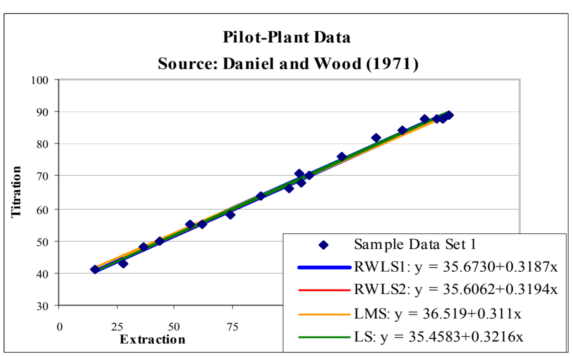

Figure 1 Observations and regression lines for data set 1: Pilot-Plant Data [7, p.22]

The second data set is in the field of astronomy, it is the Hertzsprung-Russell diagram of the star cluster CYG OB1 which contains 47 stars in the direction of Cygnus. The response is the logarithm of light intensity (L/L_0) and the explanatory is the logarithm of the effective temperature at the surface of the star (T_c). The data were given to Rousseeuw and Leroy by Doom who extracted the raw data from Humphreys (1978) and performed the calibration according to Vansina and De Greve (1982). The result of this study is shown as follows.

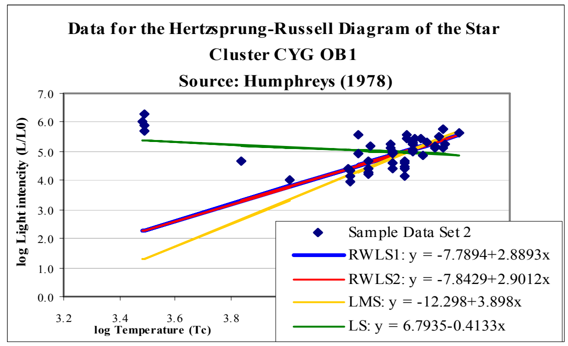

Figure 2 Observations and regression lines for data set 2: Data for the Hertzsprung - Russell Diagram of the Star Cluster CYG OB1 [7, p.27]

The third data set is the total number (in tens of millions) of international phone calls from Belgium in the years 1950-1973, provided by the Belgian Statistical Survey. Unusual data points occurred in the year 1964-1969 (see Figure 3).

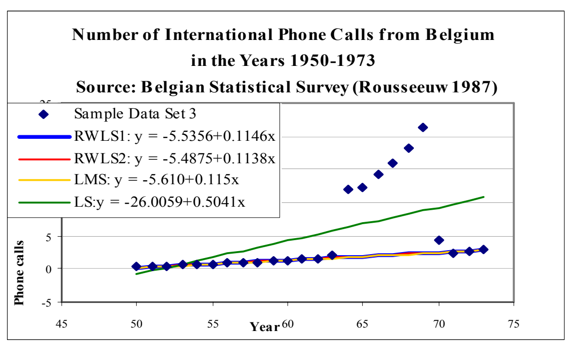

Figure 3 Observations and regression lines for data set 3: Number of International Calls from Belgium (in tens of millions) [7, p.26]

The fourth data is the set of annual rates of growth of the average prices in the main cities of Free China from 1940 to1948 (Simkin 1978). The data contains one outlier in y-direction (1948) that caused by hyperinflation (a result of government spending large amount, the budget deficit, and the war) as shown in Figure 4.

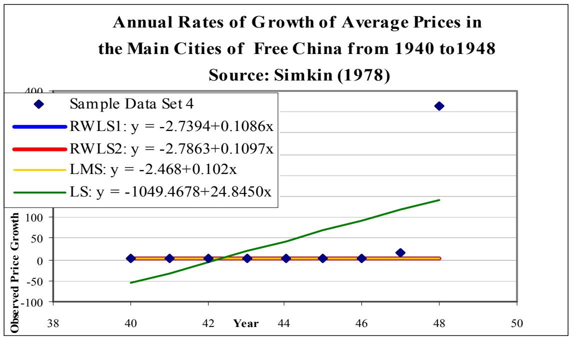

Figure 4 Observations and regression lines for data set 4: Annual Rates of Growth of Average Prices in the main Cities of Free China, 1940-1948[7, p.51]

The fifth data set consists of the brain weight (in grams) and the body weight (in kilograms) of 28 animals taken from larger data sets of Weisberg 1980 and Jerison 1973. It was investigated that transforming of data in logarithm of base 10 was more appropriate (see figure 5). Hence the transformed data will be considered in this case.

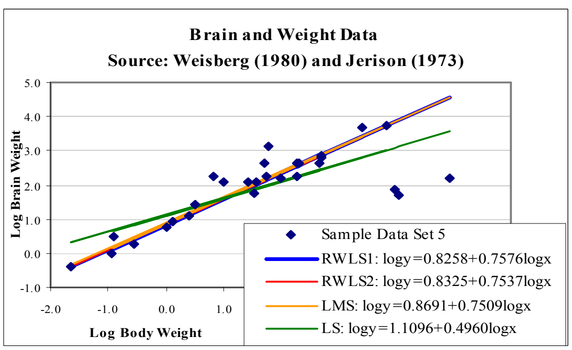

Figure 5 Observations and regression lines for data set 5: Body and Brain Weight of 28 Animals [7, p.57]

Data set 6 is the same data points as in Data set 1, except that the sixth observation has been registered as 370 instead of 37. This was done by Rousseeuw and Leroy so that this data set consists of single outlier (see Figure 6).

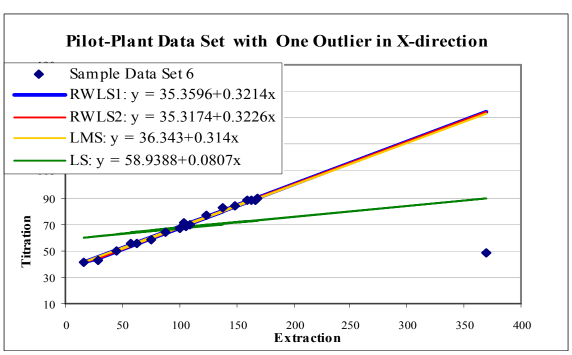

Figure 6 Observations and regression lines for data set 6: Pilot-Plant Data Set with One Outlier [7, p.24]

The seventh data set is of Mickey et al. (1967). The response is the Gesell adaptive score corresponding to the explanatory, age (in month) of 21 children when they uttered their first word. This is a contaminated data sample which outliers appear in both of x and y directions.

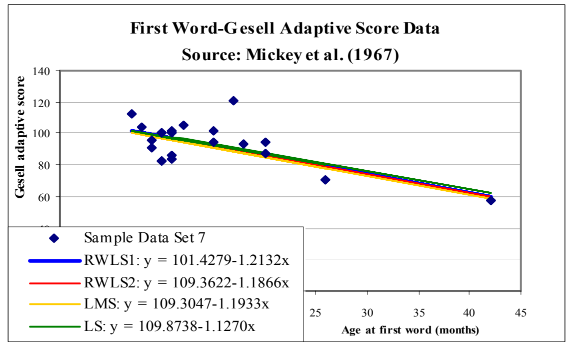

Figure 7 Observations and regression lines for data set 7: First Word-Gesell Adaptive Score Data [7, p.47]

The last example is Siegel's data set (1982), a counterexample for the resistant line estimator devised by Siegel, A. Rousseeuw and Leroy got this data set from Emerson and Hoaglin (1983) who suggested that the line with zero slope would be reasonable summary. In this data set, six out of nine points lie on the line with zero slope and zero intercept (see Figure 8).

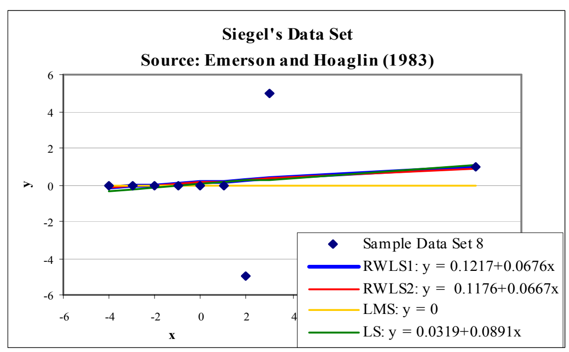

Figure 8 Observations and regression lines for data set 8: Siegel's Data Set [7, p.61]

After fitting the model, the coefficient of determinations (R^2) and the mean squared errors (MSE) for each of the estimation methods are computed. The four methods of estimation that will be compared are least squares method (LS), least median of squares method (LMS), and the two alternative methods: RWLS1 and RWLS2. For the last two estimators, the weighted error is adjusted so that the sum is zero and the corresponding MSE are computed and called the adjusted MSE (Adj.MSE). The R^2's, MSE's, and Adj.MSE's for all eight examples are exhibited in Table 1-3.

When there is no outlier, it is known that the LS estimator is very satisfied for it is BLUE under some classical assumptions. Data set 1 is an example of this case. It can be seen in Figure 1 that the regression lines obtained from four different methods presented here are almost the same. However the R^2 for RWLS1 and RWLS2 are slightly higher than that of the LS and the LMS and the MSE as well as Adj.MSE of the alternative methods are little lower.

Table 1 The R-squares (Coefficient of Determinations) corresponding to various estimators for each of Data Set 1-8.

Data Set	R-square			
	LS	**LMS**	**RWLS1**	**RWLS2**
1. Normal	0.9947	0.9936	0.9960	0.9956
2. Outliers in Y	0.0443	0.0080	0.5803	0.5382
3. Outliers in Y	0.2959	0.7932	0.9938	0.9916
4. Outliers in Y	0.3205	0.1462	0.9521	0.9378
5. Outliers in X	0.6076	0.9637	0.9800	0.9747
6. Outliers in X	0.1410	0.9941	0.9954	0.9942
7. Outliers in X-Y	0.4100	0.5716	0.6086	0.5516
8. Outliers in X-Y	0.0282	0.0000	0.9205	0.9039

Table 2 The Mean Square Errors (MSEs) corresponding to various estimators for each of Data Set 1-8.

Data Set	MSE			
	LS	**LMS**	**RWLS1**	**RWLS2**
1. Normal	1.5128	1.8206	1.1686	1.1425
2. Outliers in Y	0.3188	0.2579	0.1025	0.1022
3. Outliers in Y	31.6107	0.2449	0.0038	0.0029
4. Outliers in Y	11218.7639	0.1234	0.0028	0.0020
5. Outliers in X	0.4424	0.0477	0.0274	0.0258
6. Outliers in X	243.3163	1.6817	1.2913	1.4862
7. Outliers in X-Y	121.5045	74.4458	66.4371	75.6834
8. Outliers in X-Y	7.0648	8.3333	0.0134	0.0117

Table 3 The Adjusted MSEs corresponding to various estimators for each of Data Set 1-8.

Data Set	Adjusted MSE	
	RWLS1	**RWLS2**
1. Normal	1.1689	1.2736
2. Outliers in Y	0.1027	0.1024
3. Outliers in Y	0.0068	0.0047
4. Outliers in Y	0.0267	0.0152
5. Outliers in X	0.0384	0.0342
6. Outliers in X	1.2998	1.4943
7. Outliers in X-Y	66.6943	75.9296
8. Outliers in X-Y	0.0135	0.0118

For the cases that outliers exist, the alternative methods (RWLS1 and RWLS2) yield obviously better than the other two methods (LS and LMS), both in R^2 and MSE/Adj.MSE. Data set 2-4 are examples that outliers are appeared in y-direction as indicated by Rousseeuw and Leroy. The R^2 obtained by the LS method are 4.43%, 29.59% and 32.05% respectively. These even worst for the LMS method for data set 2 and 4 (the R^2 are 0.8% and 14.62%, respectively). For Data set 2 the R^2 corresponding to RWLS1 and RWLS2, though not too high, are more than 50%, and, for Data set 3-4, they are greater than 90%.

In comparing the MSE of those four methods of estimation, it is found also that the proposed methods yield very small value of MSE/Adj.MSE while that of the LS method are very high especially for Data set 3-4.

The results of outliers in x-direction case are presented in two examples, Data set 5-6 and of outliers in x-y direction case are presented in Data set 7-8. All cases are in the same manner as above. That is, R^2 obtained by the proposed estimators are higher than 90% in all Data set except in Data set 7 and MSE/Adj.MSE are relatively low compared with that of the LS and the LMS.

Comparing between RWLS1 and RWLS2, the differences are slightly: the R^2 obtained from RWLS1 for all cases are higher and the MSE corresponding to RWLS2 are smaller for almost every cases. In addition, both of alternative estimators perform better than the LMS in R^2 and MSE/Adj.MSE in all cases except for Data set 7 particularly for RWLS2. In such case, the LMS has larger R^2 and smaller MSE than those of RWLS2 but the differences seem not to be significant.

6. Conclusion

Though only eight examples are selected to present some performance of the proposed estimators (RWLS1 and RWLS2) compare to the LS and the LMS, all support in the same manner. That is, whenever outliers appear in data the proposed methods are preferable than the LS and probably, LMS as well. Simulation for more various situations will be done in the future.

References

[1] Basu, A., Harris, I.R., Hjort,N.L. and Jones, M.C. Robust and efficient Estimation by minimising a density power divergence. *Biometrika*, 1998, 85, 549-559.

[2] Basu, A., Harris, I.R., Hjort, N.L. and Jones, M.C. A comparison of related density-based minimum divergence estimators. *Biometrika*, 2001, 88, 865-873.

[3] Choi,E., Hall, P. and Presnell, B. Rendering parametric procedures more robust by empirically tilting the model. *Biometrika*, 2000, 87, 453-465.

[4] Daniel, G. and Yohai, V.J. A class of robust and fully efficient regression estimators. *Annals of Statistics,* 2002, vol. 30, No. 2, 583-616.

[5] Huber, P.J. *Robust Statistics*. New York, John Wiley & Sons, Inc., 1981.

[6] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. *Robust Statistics: The Approach Based on Influence Function.* New York: John Wiley& Sons, Inc., 1985.

[7] Rousseeuw, P.J. and Leroy, A.M. *Robust Regression and Outlier Detection.* New York: John Wiley & Sons, Inc., 1987.

[8] Windham, M.P. Robustifying model fitting. Journal of the Royal Statistical Society, 1995, 57 (Series B), No.3, 599-609.