

# Leverage effects on Robust Regression Estimators

David Adedia <sup>\*1</sup> Atinuke Adebajji<sup>2</sup> Simon Kojo Appiah<sup>2</sup>

1. Department of Basic Sciences, School of Basic and Biomedical Sciences, University of Health and Allied Sciences, P.M.B 31, Ho, Ghana

2. Department of Mathematics, College of Science, Kwame Nkrumah University of Science and Technology, P.M.B KNUST, Kumasi, Ghana

## Abstract

In this study, we assess the performance of some robust regression methods. These are the least- trimmed squares estimator (LTSE), Huber maximum likelihood estimator (HME), S-Estimator (SE) and modified maximum likelihood estimator (MME) which are compared with the ordinary least squares Estimator (OLSE) at different levels of leverages in the predictor variables. Anthropometric data from Komfo Anokye Teaching Hospital (KATH) was used and the comparison is done using root mean square error (RMSE), relative efficiencies (RE), coefficients of determination (R-squared) and power of the test. The results show that robust methods are as efficient as the OLSE if the assumptions of OLSE are met. OLSE is affected by low and high percentage of leverages, HME broke-down with leverages in data. MME and SE are robust to all percentage of aberrations, while LTSE is slightly affected by high percentage leverages perturbation. Thus, MME and SE are the most robust methods, while OLSE and HME are the least robust and the performance of the LTSE is affected by higher percentages of leverage in this study.

Keywords: Leverages, estimators, power of the test, coefficient of determination, root mean square error

## 1. Introduction

The development of ordinary least squares estimator is credited to Gauss (1808) as cited by Stigler (1981), for estimating regression parameters. This method performs well when the assumptions impose on the dataset by the method are satisfied. On the contrary, failure of these assumptions renders the OLSE incapable of providing stable results. The OLSE is susceptible to the effects of leverages, which are outliers in the space of the predictors. Good leverages are outlying in the dimension of predictors which are close to the regression line. Bad leverages affect intercept and the slopes of the OLSE, and are far from the fitted line. These aberrations in a dataset do affect numerical measures like coefficient of determination, root mean square error and others that are computed from the dataset. Leverage points in a dataset are identified using hat values, that is, hat values that exceed twice the average hat-value are considered leverage points.

## 2. The Linear Regression Model

The multiple linear regression model is given as

$$y = X\beta + \epsilon \quad (1)$$

where  $y$  is an  $n \times 1$  vector of observed response values,  $X$  is the  $n \times p$  matrix of the predictor variables,  $\beta$  is the  $p \times 1$  vector containing the unknown parameters that must be estimated, and  $\epsilon$  is the  $n \times 1$  vector of random errors. To fit this model to the data, a regression estimator is used to estimate the unknown parameters in  $\beta$ , to obtain  $\hat{\beta}$

where  $\hat{\beta}^T = (\hat{\beta}_1 \dots \hat{\beta}_p)$ . The expected value of  $y_i$ , (i.e the fitted value),  $E(y_i)$  is given by

$$\hat{y} = X^T \beta \quad (2)$$

As a result the residuals are computed using

$$\epsilon_i = y_i - \hat{y}_i \quad (3)$$

where  $i = 1, \dots, n$  and  $n$  is the sample size. Stuart (2011) provides that if  $\epsilon_i$  is normally distributed with mean 0 and variance  $\sigma^2$ , the least squares regression estimator is the maximum likelihood estimator for  $\beta$ .

### 3. The Least Squares Estimator

The least squares estimator aims to minimize the sum of the square residuals as:

$\text{Min} \sum \epsilon^2 = (Y - X\beta)^T (Y - X\beta)$ . Therefore using the OLSE to estimate the regression parameters in the model  $Y = X\beta + \epsilon_i$  we have:

$$\hat{\beta} = (X^T X)^{-1} X^T Y \quad (4)$$

The least squares estimates can be computed directly from any dataset when  $X^T X$  is nonsingular. However, if the assumptions of the OLSE are not met due to leverages, the OLSE cannot be used to estimate the regression parameters. Hence robust methods were developed to resist the effects of aberrations in a dataset.

### 4. The Robust Linear Regression Estimators

A regression method is as robust if it is efficient as OLSE when the errors are normal, it is not impaired by small deviation from normality, and it is not incapacitated by large deviations (Tiku & Akkaya 2004). According to Adedia *et. al.* (2016a) and Adedia *et. al.* (2016b), robust methods are able to resist the effect of aberrations: outliers, contaminations, non-normality, and both outliers and leverages based on their level of resistance to these aberrations. Huber (1973), Rousseeuw (1984), Rousseeuw & Yohai (1984), and Yohai (1987) developed some robust methods, of which some are presented in the following:

#### 4.1 Least Trimmed Squares Estimator

Rousseeuw (1984), developed the least trimmed squares estimator (LTSE) that is given by,

$$\hat{\beta} = \min \sum_{i=1}^h (\epsilon_i^2) \quad (5)$$

LTSE is computed by minimizing the  $h$  ordered squared residuals, where  $h = \left(\left[\frac{n}{2}\right] + 1\right)$ , for  $n$  and  $h$  being the sample size and the trimming constant, respectively AL-Noor & Mohammad (2013). The estimates have a high breakdown point of 50% but a low efficiency of 7.13%, when we use the  $h$  trimmed data (Matias & Yohai, 2006). A trimming constant of  $h = [n[1 - \alpha] + 1]$  where  $\alpha$  is the trimmed percentage was suggested by (Rousseeuw & Leroy, 1987). The LTSE can be fairly efficient if the trimmed percentage is chosen carefully to discard only the outliers in the data. However, in situations where there are more outliers and only some are trimmed this method can perform as poorly as the OLSE. LTSE has a break-down point of 50%, making it a high break-down method of estimation. This implies that half of the data has to be unusual points before estimates of

the LTSE can be affected when the method of the ordinary least squares is applied. The breakdown value is  $\frac{n-h}{n}$  for the LTSE estimate.

#### 4.2 Huber Maximum Likelihood Estimator

The class of M-estimator models contains all models that are derived to be maximum likelihood models. The most common method of robust regression is M-estimation, developed by Huber (1973) that is almost as efficient as OLSE is the Huber maximum likelihood estimator. This M-estimate minimizes a function  $\rho$  of the errors rather than minimizes the sum of squared errors as the objective. The M-estimate objective function is,

$$\sum_{i=1}^n \rho\left(\frac{\epsilon_i}{s}\right) = \sum_{i=1}^n \rho\left(\frac{y_i - x_i^T \beta}{s}\right) \quad (6)$$

where  $s = \frac{\text{median}|r_i - \text{median}(r_i)|}{.6745}$  is an estimate of scale from a linear combination of the residuals AL-Noor & Mohammad (2013). The function  $\rho$  gives the contribution of each residual to the objective function. A reasonable  $\rho$  according to Alma (2011), and AL-Noor & Mohammad (2013) should satisfy the following properties:  $\rho(\epsilon) \geq 0$ ,  $\rho(0) = 0$ ,  $\rho(\epsilon) = \rho(-\epsilon)$ ,  $\rho(\epsilon_i) \geq \rho(\epsilon'_i)$  for  $|\epsilon_i| \geq |\epsilon'_i|$ , and  $\rho$  is continuous.

The objective function of the least squares estimation is given by  $\rho(\epsilon) = \epsilon^2$ . The system of normal equations to solve this minimization problem is found by taking partial derivatives of (6) with respect to  $\beta$  and equating to 0. Equation (6) is minimized with respect to each of the  $p$  parameters, to obtain

$$\sum_{i=1}^n x_{ij} \psi\left(\frac{\epsilon_i}{s}\right) = \sum_{i=1}^n x_{ij} \psi\left(\frac{y_i - x_i^T \beta}{s}\right) = 0 \quad (7)$$

where  $j = 1, 2, \dots, p$  and  $i = 1, 2, \dots, n$ , and  $\psi(u) = \frac{\partial \rho}{\partial u}$  is the score function. The weight function is  $w(u) = \frac{\psi(u)}{u}$ , where  $w(u) = \frac{y_i - x_i^T \beta}{s}$ , which results in  $w_i = \frac{w(\epsilon_i)}{s}$ ,

for  $i = 1, 2, \dots, n$  and  $w_i = 1$  if  $\epsilon_i = 0$ . Substituting this in equation (7) yields

$$\sum_{i=1}^n x_{ij} w_i \left(\frac{y_i - x_i^T \beta}{s}\right) = 0 \quad (8)$$

Since  $s \neq 0$ , we define the weight matrix  $W = \text{diag}(w_i : i = 1, \dots, n)$ , with *diag* implying diagonal. Solving (8) above and making  $\beta$  the subject of formula gives

$$\hat{\beta} = (X^T W X)^{-1} X^T W Y \quad (9)$$

The weight function in equation (9) for HME, down-weights the effects of vertical outliers, but cannot minimize the effect of leverages in the predictor variables (Adedia *et. al.* 2016).

#### 4.3 The S-Estimator

The S-estimator is a high breakdown method introduced by Rousseeuw & Yohai (1984) that minimizes the standard deviation of the residuals. The SE was introduced to take care of the low breakdown point of the M-estimators due to outliers especially leverages. The high breakdown SE possesses a desirable property, that is, it is affine, scale and regression equivariant (Matias & Yohai 2006). As the least squares estimator minimizes the

variance of the residuals, SE is the  $\hat{\beta}$  that makes the dispersion of the residuals minimal (Adedia *et. al.* 2016).

$$\frac{1}{n} \sum_{i=1}^n \rho \left( \frac{\epsilon_i}{s} \right) = k \tag{10}$$

Differentiating (10) we obtain the estimating equations for S-estimator:

$$\frac{1}{n} \sum_{i=1}^n x_i \psi \left( \frac{\epsilon_i}{s} \right) = 0 \tag{11}$$

where  $\psi$  is replaced with an appropriate weight function. Biweight or Huber function is usually used as with most M-estimation procedures. Although SE have a breakdown point (BDP) of 0.5, it comes at the cost of a very low relative efficiency (Verardi & Croux 2009).

The choice of the tuning constant is  $a=1.548$ , with  $k=0.1995$  for 50% breakdown and about 29% asymptotic efficiency. To increase the efficiency of the S-estimator, if  $a = 5.182$ , the Gaussian efficiency rises to 96.6% and unfortunately the breakdown point drops to 10%. Tradeoffs between breakdown and efficiency are based on the selection of tuning constant  $a$ , and a constant  $k$ . The final scale estimate,  $s$ , is the standard deviation of the residuals from the fit that has the minimum dispersion of the residuals, where  $k$  is a constant and the objective function  $\rho$  satisfies the following conditions:

1.  $\rho$  is symmetric, continuously differentiable and  $\rho(0) = 0$ .
2. There exists a  $a > 0$  such that  $\rho$  is strictly increasing on  $[0, a]$  and constant on  $[a, \infty)$ .
3.  $\frac{k}{\rho(a)} = \frac{1}{2}$

Therefore an S-estimator is the estimator  $\hat{\beta}$  that is the solution of (4.6), with  $s$  being the smallest. The second condition on the objective function means that the associated score function will be redescending. To obtain a breakdown point of 50%, the third condition is required even though it is not strictly necessary (Stuart 2011). The choice of  $k$  is done so that the resulting  $s$  is an estimate for  $\sigma$  when the errors are normally distributed. To do this, we set  $k$  such that  $k = E\rho(u)$ , which is the expected value of the objective function if it is assumed that  $u$  has a standard normal distribution (Rousseeuw & Leroy 1987). To use the Tukey bisquare objective function, Rousseeuw & Yohai (1984) stated that if the tuning constant is  $a = 1.547$ , the third condition is satisfied, and hence makes the S-estimator has a 50% BDP.

#### 4.4 The Modified Maximum Likelihood Estimator

The MME is also type of M-estimation developed by (Yohai 1987). It is a combination of high breakdown value estimation and efficient estimation. It was the first estimator with a high breakdown point and high efficiency when errors are normally distributed. Yohai in (Yohai 1987) described three stage procedures for MME as:

1. First and foremost, a high breakdown estimator is used to find an initial estimate, which we denote  $\tilde{\beta}$  which should be efficient. We compute the residuals,  $r_i(\tilde{\beta}) = y_i - x_i^T \tilde{\beta}$  using this estimate.
2. Making use of these residuals from the robust fit and (4.6), we standard error with 50% BDP, denoting  $s(r_1(\tilde{\beta}) \dots r_n(\tilde{\beta}))$  by  $s_n$ . We labeled the objective function used in this stage as  $\rho_0$ .
3. The MME is then defined as an M-estimator of  $\beta$  using a redescending score function,  $\psi_1(u) = \frac{\partial \rho_1(u)}{\partial u}$

and the  $s_n$  obtained from Stage 2. Therefore an MME  $\hat{\beta}$  is defined as a solution to

$$\sum_{i=1}^n x_{ij} \psi_1 \left( \frac{y_i - x_i^T \beta}{s_n} \right) = 0 \quad (12)$$

where  $j = 1, \dots, p$ . The objective function  $\rho_1$  that resulted in this score function may not be the same as  $\rho_0$  but it must satisfy these conditions:

- $\rho$  is symmetric and continuously differentiable, and  $\rho(0) = 0$ .
- There exists  $a > 0$  such that  $\rho$  is strictly increasing on  $[0, a]$  and constant on  $[a, \infty)$ .
- $\rho'(u) \leq \rho'(0)$

In addition, the final condition that must be satisfied by the solution to (12) is that

$$\sum_{i=1}^n x_{ij} \psi_1 \left( \frac{y_i - x_i^T \hat{\beta}}{s_n} \right) \leq \sum_{i=1}^n x_{ij} \psi_1 \left( \frac{y_i - x_i^T \tilde{\beta}}{s_n} \right) \quad (13)$$

#### 4.4.1 Properties of MM-estimators

The first two stages of the MM-estimation process are responsible for the estimator having high breakdown point, whilst the third stage aims for high asymptotic relative efficiency. This is why  $\rho_0$  and  $\rho_1$  need not be the same, and why the estimator chosen in stage 2 can be inefficient. Yohai (1987) showed that when estimating MM-estimator, using an estimator with 50% BDP at the first stage will result in the final MME having 50% BDP. The MME is very resistant to multiple leverage points and vertical outliers. The MME is also equivariant and hence it transforms “properly” in some sense (Rousseeuw & Leroy 1987).

## 5. Data set

Datasets (anthropometric measurements) from Komfo Anokye Teaching Hospital (KATH) on total body fat as the dependent variable and height, Body Mass Index (BMI), Triceps skin-fold (TS), and Arm Fat as percent composition of the body (parmfat), as the predictors. To evaluate the effects of leverages on the estimators, we considered seven sample cases (dataset with normal errors, 5% leverages in BMI and parmfat, 5% leverages in TS and height, 5% leverages in BMI, parmfat, TS and height; also, 10% leverages in BMI and parmfat, 10% leverages in TS and height, 10% leverages in BMI, parmfat, TS and height). By using the R statistical software, the perturbations are carried out by introducing percentage of leverage points which replaces the same percentage of the original data points. The datasets are analyzed using R statistical package.

## 6. Results and Discussion

To assess the sensitivity of the methods of estimation above, we used a dataset from KATH. Moreover, the results below are presented according to (Adedia 2014). In order to assess the effects of leverages on the sensitivity of the estimators discussed, we will use root mean square error AL-Noor & Mohammad (2013), relative efficiency, coefficient of determination Alma (2011) and power of the test. The power of the test is the probability of correctly rejecting the Null hypothesis or the probability of avoiding the occurrence of type II error (Erceg-Hurn & Vikki, 2008). The post hoc power analysis helps to assess in retrospective how the effects are found assuming they exist.

### 6.1 A dataset with normal errors

Tables 1 below presents the criteria for comparing the robust estimators with the OLSE.

Table 1: Root mean square error, relative efficiency, coefficient of determination and the power of the test for original dataset with normal errors

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	1.0650	1.0000	0.9696	1.0000
LTSE	0.9641	1.2203	0.9641	1.0000
HME	1.2050	0.7811	0.9574	1.0000
SE	1.0720	0.9870	0.9586	1.0000
MME	1.0670	0.9963	0.9577	1.0000

From Table 1 it is observed that all the estimators perform well, since the errors are normally distributed. This confirms the previous studies [Yohai (1987), Schumacker *et. al.* (2002), and Adedia *et. al.* (2016a)], that all estimators perform well under normal errors. The root mean square errors, relative efficiencies and the coefficients of determination showed that when the errors are normal, all the estimators do well.

### 6.2 5% leverages in BMI and parmfat

Perturbing the Body Mass Index (BMI) and Arm Fat as percent composition of the body (parmfat) reported the following values in Table 2.

Table 2: Root mean square error, relative efficiency, coefficient of determination and power of the test for 5% leverages in BMI and parmfat

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	2.8030	1.0000	0.7893	1.0000
LTSE	1.0770	6.7735	0.9669	1.0000
HME	2.1810	1.6517	0.8611	1.0000
SE	1.2860	4.7508	0.9458	1.0000
MME	1.2860	4.7508	0.9466	1.0000

When we perturbed BMI and parmfat with 5% leverages, Table 2 showed that some of the estimators broke-down. Estimators like OLSE and HME assumed values which are quite different from when the errors were normal. Considering all the criteria for the comparison, OLSE and HME were affected with 5% leverages. The residual standard errors of these two estimators were inflated, which led to small relative efficiency of these methods.

### 6.3 5% leverages in height and TS

Table 3 below contains the reported statistics from 5% leverages in height and triceps skin-fold (TS) for comparing the regression estimators.

Table 3: Root mean square error, Relative Efficiency, Coefficient of Determination and Power of the test 5% leverages in height and TS

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	1.6570	1.0000	0.9238	1.0000
LTSE	1.2610	1.7267	0.9533	1.0000
HME	1.6620	0.9940	0.9195	1.0000
SE	1.2210	1.8417	0.9519	1.0000
MME	1.2210	1.8417	0.9520	1.0000

In this section 6.3, we examined the effects of 5% leverages in height and TS on the estimators. Leverages in height and TS made some estimators to breakdown slightly. The OLSE and HME are the estimators that are really affected, since they have low relative efficiencies and coefficients of determination.

#### 6.4 5% leverages in BMI, parmfat, height and TS

The results of perturbing all predictors with 5% leverages are shown in Table 4.

Table 4: Root mean square error, Relative Efficiency Coefficient of Determination and Power of the test for 5% leverages in all predictors

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	4.6600	1.0000	0.4176	1.0000
LTSE	3.9020	1.4263	0.5603	1.0000
HME	4.6860	0.9889	0.3597	1.0000
SE	1.2860	13.1308	0.9458	1.0000
MME	1.2860	13.1308	0.9466	1.0000

With 5% leverages in all the independent variables, we can see that, MME and SE resisted the influence of the leverages. However, OLSE and HME performed badly. This is as a result of the fact that HME and OLSE lack the resistance to leverages. The root mean square errors of OLSE, LTSE and HME are affected. Therefore, these estimators have large root mean square errors which had affected their relative efficiencies and the coefficients of determination.

#### 6.5 10% leverages in BMI and parmfat

The results in Table 5 below are the estimates of the regression parameters computed for 10% leverages in BMI and parmfat.

Table 5: Root mean square error, relative efficiency, coefficient of determination and power of the test

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	2.8610	1.0000	0.7805	1.0000
LTSE	2.0250	1.9961	0.8749	1.0000
HME	2.2940	1.5554	0.8466	1.0000
SE	1.3930	4.2183	0.9435	1.0000
MME	1.3640	4.3995	0.9457	1.0000

Having 10% leverages in BMI and parmfat reduce the efficiency of the estimators, OLSE, HME and LTSE. However, the MME and SE are not affected like other estimators. The leverages have affected the RMSE and RE of OLSE, LTSE and HME. As a result, they have high residual standard errors making the fitted models provided

by these methods unreliable.

### 6.6 10% leverages in height and TS

Table 6 shows the results of comparison of estimators when 10% leverages are introduced into height and triceps skin-fold (TS).

Table 6: Root mean square error, relative efficiency coefficient of determination and power of the test for 10% leverages in height, TS

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	1.7010	1.0000	0.9224	1.0000
LTSE	1.6150	1.1093	0.9284	1.0000
HME	1.7840	0.9091	0.9072	1.0000
SE	1.3590	1.5666	0.9432	1.0000
MME	1.3460	1.5970	0.9422	1.0000

The S-estimator and the modified maximum likelihood estimator have been consistent in being robust to the effects of the leverages in the dataset. Moreover, their results have not differed from when the errors were normally distributed. However, ordinary least squares estimator and Huber maximum likelihood estimator have since been affected by the leverages. Least trimmed squares estimator assumes different values some times because of the trimming.

### 6.7 10% leverages in BMI, parmfat, height and TS

Table 7 displays the results for 10% leverages in all predictors.

Table 7: Root mean square error, relative efficiency, coefficient of determination and power of the test for 10% leverages in BMI, parmfat, height and TS

Estimators	RMSE	RE	R-squared	Power of the test
OLSE	4.9930	1.0000	0.3314	0.9900
LTSE	4.9760	1.0068	0.3209	1.0000
HME	5.3110	0.8838	0.1776	0.9900
SE	1.4050	12.6290	0.9432	1.0000
MME	1.3640	13.3997	0.9457	1.0000

The RMSE has been largely affected by leverages in OLSE, HME and LTSE, and the coefficients of determination of these estimators have also changed significantly. HME performed like the OLSE because, HME does not bound the effect of leverages. These estimators report results which are the average of good and bad data points. According to a study by Alma (2011), the SE performed better than the MME because, MME has problems with high leverages in small sample datasets. However, in this study, MME counteracts the effects of leverages.

## 7. Conclusion

From the results above, it is justified that OLSE and HME are susceptible to the effects of leverage points. The criteria for the comparison of these estimators (RMSE, RE and R-squared) assume values which are different from the numerical measures when the errors are normal. On the contrary, robust methods such as SE and MME do very well by counteracting the effects of leverages as in (Adedia *et. al.* 2016b). LTSE is affected in some cases and in other cases, it is slightly affected. This explains the point that LTSE works well when all unusual observations are discarded (Adedia *et. al.* 2016a). Also, by using standard error, relative efficiency, coefficients



of determination and the power of the test, it was evident that MME and SE perform better than the OLSE, LTSE and HME.

## 8. Competing Interests

The authors declare that there is no competing interests regarding the publication of this manuscript.

## References

- Gauss C. F. (1809). *Theoria motus corporum coelestium in sectionibus conicis solem ambientium*.  
URL:<http://www.historyofinformation.com/expanded.php?id=270>
- Stigler, S. M. (1981). Gauss and the invention of least squares. *The Annals of Statistics*. 465-474.
- Stuart, C. (2011) Robust Regression, Department of Mathematical Sciences, *Durham University*. 1-69.
- Tiku, M. L. Akkaya, A. D. (2004). Robust estimation and hypothesis testing. *New Age International (P) Ltd*.
- Adedia, D., Adebajji, A., Labeodan, M., Adeyemi, S. (2016a). Ordinary least squares and robust estimators in linear regression: Impacts of outliers, error and response contaminations. *British Journal of Mathematics & Computer Science*. 13(4), 1-11. DOI: 10.9734/BJMCS/2016/22876.
- Adedia, D., Adebajji, A., Okyere, E., Agyen, J. K. (2016b). Leverages, Outliers and the Performance of Robust Regression Estimators. *British Journal of Mathematics & Computer Science*. 15(3): 1-14. DOI: 10.9734/BJMCS/2016/24281.
- Huber, P. J. (1973). Robust regression: asymptotics, conjectures and monte carlo. *The Annals of Statistics*, 799-821.
- Rousseeuw, P. J. (1984). Least median of squares regression. *Journal of the American Statistical Association*, 79: 871-880.
- Rousseeuw P. J. & Yohai V. J. (1984). Robust regression by means of s-estimators. *Springer*, 256- 272.
- Yohai V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*. 15, 642-656.
- AL-Noor H. N., Mohammad A. A. (2013). Model of robust regression with parametric and nonparametric methods. *Mathematical Theory and Modeling*, 3: 27-39.
- Matias S. B., Yohai V. J. (2006). A fast algorithm for s-regression estimates. *Journal of Computational and Graphical Statistics*, 15(2): 414-427.
- Rousseeuw P. J. & Leroy A. M. (1987). Robust Regression and Outlier Detection. *JOHN WILEY and SONS*.
- Alma O. G. (2011). Comparison of robust regression methods in linear regression. *International Journal for contemp Maths and Science*; 6: 409-421.
- Verardi, V. & Croux, C. (2009). Robust regression in stata. *The Stata Journal*. 3: 439-453.
- D. Adedia. Comparison of Robust Regression Estimators, Unpublished MPhil thesis, *KNUST, Kumasi*.  
<http://ir.knust.edu.gh/bitstream/123456789/6629/1/david%20adedia.pdf>, 2014.
- David M. Erceg-Hurn, Vikki M. M. (2008). Modern Robust Statistical Methods: An Easy Way to Maximize the Accuracy and Power of Your Research. *American Psychologist*, Vol. 63(7), 591-601. DOI: 10.1037/0003-066X.63.7.591
- Schumacker, R. E., Monahan, M. P., Mount, R. E. (2002). A Comparison of OLS and Robust Regression using S-PLUS. *Multiple Linear Regression Viewpoints*; Vol. 28(1), 10-13.