

# Robust Regression by Least Absolute Deviations Method

Firas H. Thanoon

Department of Mathematics, Faculty of Science & Education-Akre, University of Duhok, Iraq

**Abstract** The research deals with the choice of approach or the appropriate method of estimation, As a topic to be important, especially if you do not check some assumptions of linear form, such as the hypothesis of normal distribution for random errors, and the state of data pollution with abnormal values, since such cases have a negative impact on the estimation of Linear regression model parameters and thus the consequences would be misleading. Hence came the importance of using techniques or robust methods in matters of estimating parameters of linear regression model methods. Robust methods are known as resistant of abnormal values and other violations of model assumptions and appropriate for a broad category of distributions. A comparison in the study has been made between least absolute deviations method (Approach Iteratively Reweighted Least Squares) and least squares method, it has been made by using the empirical part which was the generation of experimental data depending on comparison criteria (standard mean square error of the model, standard mean square error for the parameters estimated and standard average deviation of absolute error), it was discovered that the least absolute deviations method is more efficient in estimating the parameters in all cases the distribution of errors for the model and especially in the case of following normal distribution and in the case of contamination of the data with abnormal values.

**Keywords**  $L_1$ -norm,  $L_2$ -norm, Least absolute deviations, Ordinary least square, Outliers, Iterative weighted least squares

## 1. Introduction

Least squares regression is sensitive to outlier points. It has dominated the statistical literature for a long time. This dominance and popularity of the least squares regression can be ascribed, at least partially to the fact that the theory is simple, well developed and documented. The computer packages are also easily available.

The least squares regression is optimal and in the maximum likelihood estimators of the unknown parameters of the model if the errors are independent and follow a normal distribution with mean zero and a common (though unknown) variance  $\sigma^2$ .

The least squares regression is very far from the optimal in many non-Gaussian situations, especially when the errors follow distributions with longer tails. For the regression problems Huber (1973) [7] stated "just a single grossly outlying observation may spoil the least squares estimate, and, moreover, outliers are much harder to spot in the regression than in the simple location case". The outliers occurring with extreme values of the regressor variables can be especially disruptive. Andrews (1974) [3] noted that even when the errors follow a normal distribution, alternatives to least squares may be required; especially if the form of the model is not exactly known. Further, least squares are not

very satisfactory if the quadratic loss function is not a satisfactory measure of the loss. Loss denotes the seriousness of the nonzero prediction error to the investigator, where prediction error is the difference between the predicted and the observed value of the response variable. Meyer & Glauber (1964) [9] stated that for at least certain economic problems absolute error may be a more satisfactory measure of loss than the squared error. The least absolute deviation errors regression (or for brevity, absolute errors regression) overcomes the aforementioned drawbacks of the least squares regression and provides an attractive alternative. It is less sensitive than least squares regression to the extreme errors and assumes absolute error loss function. Because of its resistance to outliers, it provides a better starting point than the least squares regression for certain robust regression procedures. Unlike, other robust regression procedures, it does not require (a rejection parameter). It may be noted that the absolute errors estimates are maximum likelihood and hence asymptotically efficient when the errors follow the Laplace distribution.

Many quantitative models are utilized to diagnose and evaluate the response to studies of their regression analysis is an important statistical tool that is routinely applied in most applied sciences. To ease the model formulation and computation, some desired assumptions such as normality of the response variable are made on the regression structure. Out of many possible regression techniques for fitting the model, the ordinary least squares (OLS) method has been traditionally adopted due to the ease of computation. However, there is presently a widespread awareness of the

\* Corresponding author:

drfiras\_hamid@yahoo.com (Firas H. Thanoon)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2015 Scientific & Academic Publishing. All Rights Reserved

dangers posed by the occurrence of outliers in the OLS estimates (Rousseuw and Leroy, 2003) [11]. The robustness method is considered as an alternative to a least squares method, especially if the regression model does not meet the fundamental assumptions or there are violations of it in the least square method. The estimation of significance becomes impaired; furthermore the prediction and estimation of the model may become biased.

## 2. Least Square or $L_2$ -norm Method (OLS)

Utilizing the Ordinary Least Squares (OLS) method, the estimator ( $\beta$ ) is found by minimizing the sum of squared residuals:

$$\min_{\beta} \sum_{i=1}^n (U_i)^2 \quad \text{where} \quad U_i = y_i - \hat{y}_i$$

This gives the OLS estimator for ( $\beta$ ) as:

$$\hat{\beta}_{OLS} = (X'X)^{-1} X'Y \quad (1)$$

The OLS estimate is optimal when the error distribution is assumed to be normal (Hampel 1974 [6] and Mosteller & Tukey, 1977 [10]) in the presence of influential observations, robust regression is a suitable alternative to the OLS, Robust procedures have been the focus of many studies recently, all of which triggered by the ideas of Hampel 1974[6].

### 2.1. Mean Square Error for Model

$$\begin{aligned} (MSE)_{OLS} &= (\sigma^2)_{OLS} \\ &= SST - SSR = Y'Y - (\hat{\beta})' X'Y \\ &= \frac{\sum_{i=1}^n U_i^2}{d.f(error)} \end{aligned} \quad (2)$$

### 2.2. Mean Square Error for Estimator

$$MSE \left( \hat{\beta} \right)_{OLS} = (\sigma^2)_{OLS} \text{tr} (X'X)^{-1} \quad (3)$$

### 2.3. Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^n |U_i|}{n} \quad (4)$$

## 3. Least Absolute Deviation (or $L_1$ -norm) Method (LAD)

This estimator obtains a higher efficiency than OLS through minimizing the sum of the absolute errors:

$$\min_{\beta} \sum_{i=1}^n |U_i|$$

Once LAD estimation is justified and its edge over the OLS estimation (in an appropriate condition) is established, an efficient algorithm to obtain LAD estimates has a practical significance. A progress in this direction was made by Abdelmalek (1971, 1974) [1, 2], Fair (1974) [5], Schlossmacher (1973) [13] and Spyropoulos, Kiountouzis & Young (1973) [14]. They also proposed an improved algorithm for L1 estimation that is very similar to iterative weighted least squares.

Even though calculus cannot be used to obtain an explicit formula for the solution to the L1-regression problem, Robert (2001) [12], it can be used to obtain an iterative procedure when properly initialized, converges to the solution of the L1-regression problem. The resulting iterative process is called iteratively Reweighted least squares. In this section, we briefly discuss this method. We start by considering the objective function for L1-regression, Robert (2001) [12]:

$$f(\beta) = \|Y - \beta X\|_1 \quad (5)$$

$$f(\beta) = \sum_{i=1}^n \left| Y_i - \sum_{j=1}^m \beta_j X_{ij} \right| \quad (6)$$

Differentiating this objective function is a problem, since it involves absolute values. However, the absolute value function:

$$g(z) = |z|$$

is differentiable everywhere except at one point:  $z = 0$ . Furthermore, we can use the following simple formula for the derivative, where it exists

$$g'(z) = \frac{z}{|z|}$$

Using this formula to differentiate  $f$  with respect to each variable, and setting the derivatives to zero, we get the following equations for critical points

$$\frac{\partial f}{\partial \beta_r} = \sum_{i=1}^n \frac{Y_i - \sum_{j=1}^m \beta_j X_{ij}}{\left| Y_i - \sum_{j=1}^m \beta_j X_{ij} \right|} (-X_{ir}) = 0 \quad (7)$$

Where  $r=1, 2, \dots, m$   
Can rewrite (7) as:

$$\sum_{i=1}^n \frac{\beta_j X_{ij}}{U_i} = \sum_{i=1}^n \sum_{j=1}^m \frac{\beta_j X_{ir} X_{ij}}{U_i} \quad (8)$$

Let  $W$  denote the diagonal matrix, DasGupta & Mishra (2004) [4], where:

$$w_{ij} = \frac{1}{|U_i|} \quad \text{for } i = j$$

$$w_{ij} = 0 \quad \text{for } i \neq j$$

We can write these equations in matrix notation as follows:

$$(X' W Y) = X' W X \beta$$

This equation can't be solved for  $x$  as we were able to do in L2-regression because of the dependence of the diagonal matrix on  $(\beta)$ . But let us rearrange this system of equations by multiplying both sides by of:

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

Table (1). OLS& IRWLS estimators

Distribution of errors	Estimator	OLS				IRWLS			
		n=20	n=50	n=100	n=200	n=20	n=50	n=100	n=200
Normal Distribution	$\hat{\beta}_0$	1.2023	1.0936	1.2879	1.0758	0.7389	0.7240	1.1928	0.9544
	$\hat{\beta}_1$	0.9567	1.04437	0.9643	0.9985	1.1602	1.175	0.9740	1.0116
	MSE	1.029	0.8600	1.06	1.0000	0.5909	0.8261	0.8632	0.7664
	$MSE(\hat{\beta})$	0.7947	0.1245	0.0566	0.0225	0.0157	0.00017	0.0016	0.00015
	MAE	0.78634	0.7643	0.8340	0.7808	0.7680	0.7574	0.8323	0.7779
Contaminated Distribution	$\hat{\beta}_0$	10.681	20.621	10.00	-0.802	2.8103	0.7576	1.7359	1.1890
	$\hat{\beta}_1$	-3.131	-3.419	-0.2534	1.1120	0.1918	1.1589	0.9265	0.9895
	MSE	17.87	643.7	631.3	925.9	2.4956	7.0667	6.8354	8.4235
	$MSE(\hat{\beta})$	13.7952	92.633	33.8891	21.4186	0.3329	0.0075	0.0146	0.00012
	MAE	2.8484	10.3330	8.5242	8.4985	2.2460	6.7910	6.8761	8.2779
Laplace Distribution	$\hat{\beta}_0$	-0.843	-0.5235	0.4891	0.5647	0.7892	0.4808	0.6186	0.7657
	$\hat{\beta}_1$	1.5608	1.3463	1.03926	1.03832	0.9258	1.0492	1.0236	1.0120
	MSE	6.477	6.12	4.18	3.8	1.8253	1.7463	1.3321	1.0287
	$MSE(\hat{\beta})$	5.0020	0.8812	0.2244	0.0889	0.00025	0.0003	0.00097	0.0051
	MAE	1.7173	1.5193	1.2943	1.2172	1.6797	1.4578	1.2931	1.2011
Cauchy Distribution	$\hat{\beta}_0$	-7.458	-22.67	3.213	5.694	-1.0363	-0.6567	1.1605	1.4622
	$\hat{\beta}_1$	4.458	9.826	1.280	0.7562	1.7923	1.4644	1.2373	0.9912
	MSE	46.88	3418	1776	889.3	3.4731	11.4779	3.3222	2.8047
	$MSE(\hat{\beta})$	36.1982	491.637	95.3313	20.5728	0.00005	0.6194	0.1325	0.0247
	MAE	3.8029	20.5249	8.9826	5.5301	3.2332	11.2649	7.3912	4.5529

This formula suggests an iterative scheme that hopefully converges to a solution. Indeed, we start by initializing  $\beta^{(0)}$  arbitrarily and then use the above formula to successively compute new approximations. If we let  $\beta^{(k)}$  denote the approximation at the  $k^{\text{th}}$  iteration, then the update formula can be expressed as:

$$\hat{\beta}^{(k)}_{IRWLS} = (X'WX)^{-1} X'WY \quad (10)$$

Assuming only that the matrix inverse exists at every iteration, one can show that this iteration scheme converges to a solution to the L1-regression problem. LAD regression estimates are obtainable from the function L1 fit in the computer language (S-PLUS) and from the robust regression package ROBSTAT, Levent, Soner, Gokhan & Hulya (2006) [8].

### 3.1. Mean Square Error for Model

$$\begin{aligned} (MSE)_{IRWLS} &= (\sigma^2)_{IRWLS} \\ &= SST - SSR = Y'WY - (\hat{\beta})'X'WY \end{aligned} \quad (11)$$

### 3.2. Mean Square Error for Estimator

$$MSE \left( \hat{\beta} \right)_{IRWLS} = (\sigma^2)_{IRWLS} \text{tr} (X'WX)^{-1} \quad (12)$$

### 3.3. Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^n |U_i|}{n} \quad (13)$$

## 4. The Simulation Study

The following model was used

$$y_i = \beta_0 + \beta_1 x_i + U_i \quad (14)$$

Now,  $(X)$  was obtained from a set numbers (1, 1.1, 1.2, 1.3 ...),  $\beta = [1, 1]$  where, distribution of errors (normal, Contaminant, Laplace, Cauchy). And size of random:  $n=20$ ,  $n=50$ ,  $n=100$  and  $n=200$ .

## 5. Concluding Remarks

Through the above table:

1. It has been shown that the LAD ( $L_1$ -norm) is more efficient in using IRWLS approach to estimate the parameters in all cases of distributions of error for the model, especially in the case of following normal

distribution in the LS method in all sample sizes, and this was confirmed by all standards of comparison used in the research.

2. The method of LAD style is very suitable and efficient for estimating the parameters and regression analysis, for using the absolute deviations of error in the estimate.

## REFERENCES

- [1] Abdelmalek, NN (1971). "Linear L1 Approximation for a Discrete Point Set and L1-Solutions of Overdetermined Linear Equations", Journal of Assoc. Comput. Mach., 18. (47-41).
- [2] Abdelmalek, NN (1974). "On the Discrete Linear L1 Approximation and L1 Solution of Over-determined Linear Equations", Journal of Approximation Theory, 11(35-53).
- [3] Andrews, D.F. (1974). A robust method for multiple linear regression. Technometrics 16, 523-531.
- [4] DasGupta, M & Mishra, SK, (2004), "Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review", <http://mpira.uni-muenchen.de/1781>.
- [5] Fair, RC (1974). "On the Robust Estimation of Econometric Models", Annals of Economic and Social Measurement, 3 (667-677).
- [6] Hampel F R, (1974), "The influence curve and its role in robust estimation", Journal of the American Statistical Association, 69, 383-393.
- [7] Huber, PJ (1973). "Robust Regression: Asymptotics, Conjectures, and Monte Carlo" The Annals of Statistics, 1 (799-821).
- [8] Levent, S; Soner, C; Gokhan, G & Hulya, A, (2006), "A Comparative Study of Least Square Method and Least Absolute Deviation Method for Parameters Estimation in Length-Weight Relationship for White Grouper (Ephinephelus Aeneus)", Pakistan Journal of Biological Sciences 9(15): 2919-2921.
- [9] Meyer, JR & Glauber R.R (1964). "Investment Decisions: Economic Forecasting and Public Policy", Harvard Business School Press, Cambridge, Massachusetts.
- [10] Mosteller and Tukey J W, (1977), "Data Analysis and Regression. Reading", MA: Addison Wesley.
- [11] Rousseeuw P. J., & Leroy A. M., (2003), "Robust Regression and Outlier Detection", New York, John Wiley.
- [12] Robert J. Vanderbei, (2001), "Linear Programming: Foundations and Extensions", Robert J. Vanderbei, USA
- [13] Schlossmacher, EJ (1973). "An Iterative Technique for Absolute Deviations Curve Fitting", Journal of the American Statistical Association, 68 (857-859).
- [14] Spyropoulos, K; Kiountouzis, E & Young, A (1973). "Discrete Approximation in the  $L_1$  Norm", Computer Journal, 16 (180-186).