ISSN: 0974-5823 Vol. 7 No. 5 May, 2022

International Journal of Mechanical Engineering

# Robust Regression using Least Absolute Deviations Method

**Eakambaram. S[1] and Rex Irudhaya Raj. A[2]**

[1]Department of Statistics, Periyar Government Arts College, Cuddalore- 1.

[2]Research Scholar, Department of Statistics, Annamalai University, Annamalai, Nagar -608 002.

**Abstract**

The least squares regression is optimal and the maximum likelihood estimators of the unknown parameters of the model if the errors are independent will follow a normal distribution with mean zero and a common (though unknown) variance. Robust estimation refers to the ability of a procedure to produce highly insensitive estimates to model misspecifications. Robust methods are known as resistant of abnormal values and other violations of model assumptions and appropriate for a broad category of distributions. A comparative study has been made between Least Absolute Deviations (LAD) method and Least Squares (LS) method. It has been made by using the empirical part which was the generation of experimental data depending on comparison criteria. Finally, it found that the LAD method is more efficient in estimating the parameters in all cases the distribution of errors for the model.

**Keywords:** Least Absolute Deviations, Ordinary Least Square, Maximum Likelihood Estimator, Iteratively Reweighted 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 square regression can be imputed, 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 the maximum likelihood estimators of the unknown parameters of the model if the errors are independent will follow a normal distribution with mean zero and a common (though unknown) variance.

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) stated that "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 confusing. Andrews (1974) 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.

The least absolute deviation errors regression overcomes the 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 double exponential distribution.

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 dangers posed by the occurrence of outliers in the OLS estimates (Rousseuw and Leroy, 2003). The robustness method is considered as an alternative to a LS method, especially if the regression model does not meet the fundamental assumptions.

## 2. Ordinary Least Square Method or $L_2$-norm Method

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

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

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

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

The OLS estimate is optimal when the error distribution is assumed to be normal in the presence of influential observations, robust regression is a suitable alternative to the OLS. Robust procedures have been focussed many studies recently. For a detailed study refer Hampel 1974.

### 2.1. Mean Square Error (MSE) for the model

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

### 2.2. Mean Square Error for Estimator

$$MSE\left(\hat{\beta}\right)_{OLS} = (\sigma^2)_{OLS} \;\; tr\,(X'X)^{-1} \qquad \dots (3)$$

### 2.3. Mean Absolute Error (MAE)

$$MAE = \frac{\sum_{i=1}^{n} \left| U_i \right|}{n} \qquad \dots (4)$$

## 3. Least Absolute Deviation Method or L1-norm Method

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

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

Once LAD estimation is justified over the OLS estimation, an efficient algorithm to obtain LAD estimates has a practical significance. For a detailed study refer Abdelmalek (1971, 1974), Fair (1974), Schlossmacher (1973) and Spyropoulos et.al (1973). They also proposed an improved algorithm for $L_1$ estimation that is very similar to iterative weighted least squares.

Robert (2001) used an iterative procedure when properly initialized, converges to the solution of the $L_1$-regression problem and is called Iteratively Reweighted Least Squares (IRWLS).

The objective function for $L_1$-regression,

$$f(\beta) = \left\| Y - \beta X \right\|_1 \qquad \dots (5)$$

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

Differentiating the above 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. Also, the following formula can be used for the derivative, where it exists

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

Using this, differentiate f with respect to each variable and setting the derivatives to zero, we have

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

where r =1, 2, …., m

Rewriting eqn. (7), we have

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

Let $W$ denote the diagonal matrix, (Das Gupta & Mishra, 2004), where

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

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

We can write these equations in matrix notation as follows:

$(X'WY) = X'WX\beta$

This equation cannot be solved for x in $L_2$-regression because of the dependence of the diagonal matrix on $(\beta)$. Then

$$(X'WX)^{-1}$$

$$\hat{\beta} = (X'WX)^{-1} X'WY \qquad \ldots (9)$$

This formula suggests an iterative scheme that converges to a solution. Indeed, we start by initializing $\beta^{(0)}$ arbitrarily and then use the above formula to successive computation of new approximations. Let $\beta^{(k)}$ denote the approximation at the $k^{th}$ iteration, then formula can be expressed

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

Assuming only that the matrix inverse exists at every iteration, one can show that this iteration scheme converges to a solution to the $L_1$-regression problem.

**3.1. Mean Square Error for the Model**

$$(MSE)_{IRWLS} = (\sigma^2)_{IRWLS}$$

$$= SST - SSR = Y'WY - (\hat{\beta})'X'WY \qquad \ldots (11)$$

**3.2. Mean Square Error for Estimator**

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

**3.3. Mean Absolute Error**

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

**4. Results**

The following model was used

$$y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad \ldots (14)$$

Now, $(X)$ was obtained from set numbers (1, 1.1, 1.2, 1.3 ...), and $\underline{\beta} = [1, 1]$ where the distribution of errors is Normal, Extreme Value, Double Exponential, Cauchy and random size is n = 30, 60, 125 and 250.

**Table –** OLS and IRWLS estimators

| **Distribution of Errors** | **Estimator** | **OLS** | | | | **IRWLS** | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **n=30** | **n=60** | **n=125** | **n=250** | **n=30** | **n=60** | **n=125** | **n=250** |
| **Normal Distribution** | $\hat{\beta}_0$ | 1.3023 | 1.1936 | 1.3879 | 1.1758 | 0.8389 | 0.8240 | 1.2928 | 1.0544 |
| | $\hat{\beta}_1$ | 1.0567 | 1.1444 | 1.0643 | 1.0985 | 1.2602 | 1.2750 | 1.0740 | 1.1116 |
| | **MSE** | 1.1290 | 0.9600 | 1.1600 | 2.0000 | 0.6909 | 0.9261 | 0.9632 | 0.8664 |
| | **MSE ($\hat{\beta}$)** | 0.8947 | 0.2245 | 0.1566 | 0.1225 | 0.1157 | 0.10017 | 0.1016 | 0.10015 |
| | **MAE** | 0.8863 | 0.8643 | 0.9340 | 0.8808 | 0.8680 | 0.8574 | 0.9323 | 0.8779 |
| **Extreme Value** | $\hat{\beta}_0$ | 10.7810 | 20.7210 | 10.1000 | -0.9020 | 2.9103 | 0.8576 | 1.8359 | 1.2890 |
| | $\hat{\beta}_1$ | -3.2310 | -3.5190 | -0.3534 | 1.2120 | 0.2918 | 1.2589 | 1.0265 | 1.0895 |
| | **MSE** | 17.9700 | 643.8000 | 631.4000 | 926.0000 | 2.5956 | 7.1667 | 6.9354 | 8.5235 |
| | **MSE ($\hat{\beta}$)** | 13.8952 | 92.7330 | 33.9891 | 21.5186 | 0.4329 | 0.1075 | 0.1146 | 0.1001 |
| | **MAE** | 2.9484 | 10.4330 | 8.6242 | 8.5985 | 2.3460 | 6.8910 | 6.9761 | 8.3779 |
| **Double Exponential Distribution** | $\hat{\beta}_0$ | -0.9430 | -0.6235 | 0.5891 | 0.6647 | 0.8892 | 0.5808 | 0.7186 | 0.8657 |
| | $\hat{\beta}_1$ | 1.6608 | 1.4463 | 1.1393 | 1.1383 | 1.0258 | 1.1492 | 1.1236 | 1.1120 |
| | **MSE** | 6.5770 | 6.2200 | 4.2800 | 3.9000 | 1.9253 | 1.8463 | 1.4321 | 1.1287 |
| | **MSE ($\hat{\beta}$)** | 6.1020 | 0.9812 | 0.3244 | 0.1889 | 0.1003 | 0.1003 | 0.1010 | 0.1051 |
| | **MAE** | 1.8173 | 1.6193 | 1.3943 | 1.3172 | 1.7797 | 1.5578 | 1.3931 | 1.3011 |
| **Log Normal** | $\hat{\beta}_0$ | -7.5580 | -22.7700 | 3.3130 | 5.7940 | -1.1363 | -0.7567 | 1.2605 | 1.5622 |
| | $\hat{\beta}_1$ | 4.5580 | 9.9260 | 1.3800 | 0.8562 | 1.8923 | 1.5644 | 1.3373 | 1.0912 |
| | **MSE** | 46.9800 | 3419 | 1777 | 889.4000 | 3.5731 | 11.5779 | 3.4222 | 2.9047 |
| | **MSE ($\hat{\beta}$)** | 36.2982 | 491.7370 | 95.4313 | 20.6728 | 0.1001 | 0.7194 | 0.2325 | 0.1247 |
| | **MAE** | 3.9029 | 20.6249 | 9.0826 | 5.6301 | 3.3332 | 11.3649 | 7.4912 | 4.6529 |

## 5. Conclusion

From the above table it is observed that the iterative reweighted least square methods provide robust estimator comparing to OLS. It's also seen that the heavy tailed distribution gives better estimates than the normal and double exponential distribution. The estimates are provided in this paper are not only robust but gives consistent results. The method of LAD style is very suitable and efficient for estimating the parameters and regression analysis.

## References


- Abdelmalek, N. N (1971). Linear L1 Approximation for a Discrete Point Set and L1-Solutions of Overdetermined Linear Equations, *Journal of Assoc. Comput. Mach.*, 18, pp. 47-41.
- Abdelmalek, N. N (1974). On the Discrete Linear L1 Approximation and L1 Solution of Over- determined Linear Equations, *Journal of Approximation Theory*, 11, pp. 35-53.
- Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*, 16, pp. 523-531.
- DasGupta, M and Mishra, S. K (2004) Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review, http://mpra.ub.uni-muenchen.de/1781.
- Fair, R. C (1974). On the Robust Estimation of Econometric Models, *Annals of Economic and Social Measurement*, 3, pp. 667-677.
- Hampel, F. R, (1974). The influence curve and its role in robust estimation, *Journal of the American Statistical Association*, 69, pp. 383-393.
- Huber, P. J. (1973). Robust Regression: Asymptotics, Conjectures, and Monte Carlo, *The Annals of Statistics*, Vol. 1, pp. 799-821.
- Robert J. Vanderbei (2001). Linear Programming: Foundations and Extensions, Robert J. Vanderbei, USA
- Rousseuw P. J. and Leroy A. M. (2003). Robust Regression and Outlier Detection, *John Willy*, New York.
- Schlossmacher, E. J (1973). An Iterative Technique for Absolute Deviations Curve Fitting, *Journal of the American Statistical Association*, 68, pp. 857-859.
- Spyropoulous, K., Kiountouzis, E., and Young, A (1973). Discrete Approximation in the $L_1$ Norm, *Computer Journal*, 16, pp. 180-186.