

11-1-2016

An Alternative Algorithm and R Programming Implementation for Least Absolute Deviation Estimator of the Linear Regression Models

Suraju Olaniyi Ogundele

Federal University of Petroleum Resources, Effurun, Delta State, Nigeria., ogundele.olaniyi@fupre.edu.ng


J. I. Mbegbu

Department of Mathematics, University of Benin, Benin-City, Edo State, Nigeria.

C. R. Nwosu

Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Follow this and additional works at: <http://digitalcommons.wayne.edu/jmasm>

 Part of the [Applied Statistics Commons](#), [Social and Behavioral Sciences Commons](#), and the [Statistical Theory Commons](#)

Recommended Citation

Ogundele, Suraju Olaniyi; Mbegbu, J. I.; and Nwosu, C. R. (2016) "An Alternative Algorithm and R Programming Implementation for Least Absolute Deviation Estimator of the Linear Regression Models," *Journal of Modern Applied Statistical Methods*: Vol. 15: Iss. 2, Article 45.

DOI: 10.22237/jmasm/1478004180

Available at: <http://digitalcommons.wayne.edu/jmasm/vol15/iss2/45>

This Algorithms and Code is brought to you for free and open access by the Open Access Journals at DigitalCommons@WayneState. It has been accepted for inclusion in Journal of Modern Applied Statistical Methods by an authorized administrator of DigitalCommons@WayneState.

An Alternative Algorithm and R Programming Implementation for Least Absolute Deviation Estimator of the Linear Regression Models

Cover Page Footnote

*Correspondent Author E-mail/Phone Number: ogundele.olaniyi@fupre.edu.ng, +2348055678505.

An Alternative Algorithm and R Programming Implementation for Least Absolute Deviation Estimator of the Linear Regression Models

Suraju Olaniyi Ogundele
Federal University of Petroleum Resources
Effurun, Nigeria.

J. I. Mbegbu
University of Benin
Benin-City, Nigeria

C. R. Nwosu
Nnamdi Azikiwe University
Awka, Nigeria

We propose a least absolute deviation estimation method that produced a least absolute deviation estimator of parameter of the linear regression model. The method is as accurate as existing method.

Keywords: Linear regression model, least absolute deviation (LAD), equation of a line, R statistical programming and algorithm

Introduction

Regression is a statistical methodology that is use to relate a variable of interest, which is called the dependent variable or response variable, to one or more predictors (independent/regressors) variables. The objective of regression analysis is to build a regression model or prediction equation that helps us to describe, predict and control the dependent variable on the basis of the independent variable(s). When we predict the dependent variable for a particular set of values of the independent variables, we wish to place a bound on the error of prediction. The goal is to build a regression model that produces an error bound that will be small enough to meet our needs.

In the simple linear regression model, the Population Regression Function (PRF) is given by:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

Mr. Ogundele is an Assistant Lecturer in the Department of Mathematics / Computer Science. Email him at: ogundele.olaniyi@fupre.edu.ng. J. I. Mbegbu is a Professor in the Department of Mathematics. C. R. Nwosu is an Associate Professor in the Department of Statistics.

ALGORITHM AND CODE FOR LAD ESTIMATOR

In this model there is only one factor x to explain y . All the other factors that affect y are jointly captured by the error term denoted by ε . We typically refer to y as the endogenous or dependent variable and x as the exogenous or independent variable.

The idea of the regression model is to estimate the population parameters, β_0 and β_1 from a given sample. The Sample Regression Function (SRF) is the sample counterpart of the population regression function (PRF). Since the SRF is obtained for a given sample, a new sample will generate different estimates. The SRF, which is an estimation of the PRF is given by:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (2)$$

Equation (2) is used to calculate the fitted value (\hat{y}_i) for y when $x = x_i$. In the SRF $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators of the parameters β_0 and β_1 . For each x_i we have an observed value (y_i) and a fitted value (\hat{y}_i). The difference between y_i and \hat{y}_i is called the residual $\hat{\varepsilon}_i$ given by:

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

The ordinary least squares (OLS) method is the most widely used method of parameter estimation. The OLS criteria is to minimize the sum of squared error of prediction

$$\hat{\varepsilon}_i^2 = (y_i - \hat{y}_i)^2 \quad (4)$$

OLS regression yields estimates for the parameters that have the desirable property of being minimum variance unbiased estimators (Chatterjee & Hadi, 2006).

Ordinary least squares estimation places certain restrictive assumptions on the random component in the model, the errors of prediction. OLS estimation assumes, among others, that the errors of prediction are normally distributed, with a common error variance at all levels of X [$\varepsilon \sim N(0, \sigma^2)$]. The normality assumption is frequently untenable in practice. Violation of this assumption is often manifested by the presence of outliers in the observed data (Nevitt & Tam, 1998). Thus data containing outlying values may reflect non-normal error distributions with heavy tails or normal error distributions containing observations

atypical of the usual normal distribution with larger variance than the assumed σ^2 . It is well demonstrated that outliers in the sample data heavily influence estimates using OLS regression, sometimes even in the presence of one outlier (Rousseeuw & Leroy, 1987).

If the assumption that the uncertainties (i.e., errors) in the data are uncorrelated and normally distributed are valid for the data at hand, then for most quantitative experiments, the method of least squares is the "best" analytical technique for extracting information from a set of data. The method is best in the sense that the parameters determined by the least squares analysis are normally distributed about the true parameters with the least possible standard deviations (Wolberg, 2006).

However, the assumption of the general applicability of the normal law of errors has been under attack from the very beginning of the development of linear regression and, in particular, the least squares analysis hinges critically on the existence of the second moment of the error distribution. Thus, if it must be assumed the error distribution follows, for instance, a Cauchy distribution or any long-tailed distribution having no finite second moment, then the elegant arguments made in favour of the least squares regression estimators become invalid, and thus, it may become mandatory to look for other criteria to find best estimators for the linear regression model (Giloni & Padberg, 2002). For situations in which the underlying assumptions of OLS estimation are not tenable, the choice of method for parameter estimation is not clearly defined. Thus, the choice of estimation method under non-ideal conditions has been a long-standing problem for methodological researchers (Nevitt & Tam, 1998). The history of this problem is lengthy with many alternative estimation methods having been proposed and investigated (Birkes & Dodge, 1993).

Robust estimation refers to the ability of a procedure to produce highly insensitive estimates to model misspecifications. Hence, robust estimates should be good under wide range of possible data generating distributions. In the regression context, under normality with identically and independently distributed errors, the least squares is the most efficient among the unbiased estimation methods. However, when the normality assumption is not feasible, it is frequently possible to find estimation methods that are more efficient than the traditional least squares. This occurs when the data generating process has fat tails resulting to several outliers compared to the normal distribution. In these cases the least squares becomes highly unstable and sample dependent because of the quadratic weighting, which makes the procedure very sensitive to outlying observations (Pynnonen & Salmi, 1994). Examples of this type of robust estimation are Huber

ALGORITHM AND CODE FOR LAD ESTIMATOR

M-estimation, the method of Least Median of Squares, and the method of Least Absolute Deviations (LAD).

The robust LAD estimator is investigated in the present study and so we provide a brief description of the method. LAD was developed by Roger Joseph Boscovich in 1757, nearly 50 years before OLS estimation (see Birkes & Dodge, 1993 for a review and historical citations). In contrast to OLS estimation which defines the loss function on the residuals as $\sum e_i^2$, LAD finds the slope and Y intercept that minimize the sum of the absolute values of the residuals, $\sum |e_i|$.

Although the concept of LAD is not more difficult than the concept of the OLS estimation but due to computational difficulties in obtaining LAD estimates and lack of exact sampling theory based on such estimates, the LAD method lay in the background and the LS method became popular (Rao & Toutenburg, 1999). Since there are no exact formulas for LAD estimates, an algorithm is used to iteratively obtain the estimate of the parameters.

Methodology

Propose Method

Using the LAD criterion, the model $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ should be constructed by two pairs of data points that yield the minimum sum of absolute deviation, our approach is to investigate the sum of absolute deviation generated by all possible different combinations of data points and then select the two data points that produced the least absolute deviation to find the least absolute deviation estimator.

The two points (x_i, y_i) and (x_j, y_j) yield the following system of equations:

$$\begin{aligned} y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_i \\ y_j &= \hat{\beta}_0 + \hat{\beta}_1 x_j \end{aligned} \tag{5}$$

The solution of equation (5) yield the value of $\hat{\beta}_0$ and $\hat{\beta}_1$, for the selected pair of data. Substituting the value of $\hat{\beta}_1$ and $\hat{\beta}_0$ obtained from (5) into (1), we have

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \tag{6}$$

(6) is then used to determine the sum of absolute deviation of all other data points from the line joining (x_i, y_i) and (x_j, y_j) by substituting for $x_i (i = 1, 2, \dots, n)$ of all data points and calculating

$$\sum_{i=1}^n |y_i - \hat{y}_i|. \quad (7)$$

This procedure will be repeated for all other combinations of data points and the two data points that yielded the least absolute deviation determine the least absolute deviation estimator.

The presence of personal computers make it possible to evaluate repeated process using any programming language of choice. The R program is employed for all calculations, and the program yields the least absolute deviation estimate for the data.

Algorithm for Simple Linear Regression

INPUT: Observations of x and y as vectors X and Y .

OUTPUT: Slope and intercept.

- Step 1. Set $i = 1$ and $j = 2$
- Step 2. While $i \leq (n - 1); j \leq n; i \neq j$, do steps 3 to 4
- Step 3. Select the pairs (x_i, y_i) and (x_j, y_j) from the data and calculate the values of $\hat{\beta}_1$ and $\hat{\beta}_0$ from the system of equations given by (5).
- Step 4. For $i = 1$ to n , determine the estimated value of $y(\hat{y})$ by substituting for (x_i, y_i) in (6) and calculate

$$AbsDev(i, j) = \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Step 5. Determine the minimum value among all $AbsDev(i, j)$ and select the two data points that produced the minimum value.
- Step 6. Print out the values of $\hat{\beta}_1$ and $\hat{\beta}_0$ that correspond to the two selected data points.
- Step 7. If $i > (n - 1)$ and $j > n$, then OUTPUT; stop.

ALGORITHM AND CODE FOR LAD ESTIMATOR

- Step 8. Set $i = i + 1; j + 1$ and go to step 2.
 Step 9. OUTPUT "Method failed after $i > (n - 1)$ and $j > n$ ".

This algorithm is improved upon for multiple linear regression and can be scaled to accommodate the numbers of independent variables present in the data. We provide the algorithm and the R program for the simplest form of the multiple regression with two independent variables

Algorithm for Multiple Linear Regression

INPUT: Observations of x_1, x_2 and y as vectors X1, X2 and Y.

OUTPUT: Intercept, first parameter and second parameter.

- Step 1. Set $i = 1, j = 2$ and $k = 3$
 Step 2. While $i \leq (n - 2); j \leq (n - 1); k \leq n; i \neq j \neq k$, do steps 3 to 4
 Step 3. Select the pairs $(x_i, y_i), (x_j, y_j)$ and (x_k, y_k) from the data and calculate the values of $\hat{\beta}_2, \hat{\beta}_1$ and $\hat{\beta}_0$ from the system of equations

$$\begin{aligned} y_i &= \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} \\ y_j &= \hat{\beta}_0 + \hat{\beta}_1 x_{1j} + \hat{\beta}_2 x_{2j} \\ y_k &= \hat{\beta}_0 + \hat{\beta}_1 x_{1k} + \hat{\beta}_2 x_{2k} \end{aligned} \quad (8)$$

- Step 4. For $i = 1$ to n , determine the estimated value of $y(\hat{y})$ by substituting for (x_{1i}, x_{2i}, y_i) in

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1i} + \hat{\beta}_2 x_{2i} \quad (9)$$

and calculate

$$AbsDev(i, j, k) = \sum_{i=1}^n |y_i - \hat{y}_i|$$

- Step 5. Determine the minimum value among all $AbsDev(i, j, k)$ and select the three data points that produced the minimum value.

- Step 6. Print out the values of $\hat{\beta}_2$, $\hat{\beta}_1$ and $\hat{\beta}_0$ that correspond to the three selected data points.
- Step 7. If $i > (n - 2)$, $j > (n - 1)$ and $k > n$, then OUTPUT; stop.
- Step 8. Set $i = i + 1$; $j = j + 1$; $k = k + 1$ and go to step 2.
- Step 9. OUTPUT "Method failed after $i \leq (n - 2)$, $j \leq (n - 1)$ and $k \leq n$ ".

Results

The R program designed by applying this algorithm is presented in the [Appendix](#). Application of the R program written for the algorithm to the data in [Table 1](#) yielded the same results with the iterative method by Birkes and Dodge (1993). The best two data points are given to be at (x_5, y_5) and (x_{14}, y_{14}) . The least absolute deviation estimate of the model parameters are

$$\hat{\beta}_{0(LAD)} = 46.38444.$$

$$\hat{\beta}_{1(LAD)} = -0.53778$$

Then, the LAD regression line is

$$\hat{y}_i = 46.38444 - 0.53778x_i$$

Table 1. Birth Rate Data

Country	Birth Rate (y_i)	Urban Percentage (x_i)
Canada	16.2	55
Costa Rica	30.5	27.3
Cuba	16.9	33.3
Dominican Republic	33.1	37.1
El Salvador	40.2	11.5
Guatemala	38.4	14.2
Haiti	41.3	13.9
Honduras	43.9	19
Jamaica	28.3	33.1
Mexico	33.9	43.2
Nicaragua	44.2	28.5
Panama	28	37.7
Trinidad-Tobago	24.6	6.8
United States	16	56.5

ALGORITHM AND CODE FOR LAD ESTIMATOR

The R program written for the multiple regression implementation of the proposed method was applied to find the least absolute deviation estimate of the parameter of subset of the supervisor data (Chatterjee & Hadi, 2006) which includes Y, X₁ and X₂. The data is presented in Table 2. The best three data points are given to be at (x₈, y₈), (x₉, y₉) and (x₂₁, y₂₁).

The program gives the least absolute deviation estimate of the parameters as

$$\hat{\beta}_{0(LAD)} = 28.33487$$

$$\hat{\beta}_{1(LAD)} = 0.6835637$$

$$\hat{\beta}_{2(LAD)} = -0.172043$$

Table 2. Subset of Supervisor Data

Y	X1	X2
43	51	30
63	64	51
71	70	68
61	63	45
81	78	56
43	55	49
58	67	42
71	75	50
72	82	72
67	61	45
64	53	53
67	60	47
69	62	57
68	83	83
77	77	54
81	90	50
74	85	64
65	60	65
65	70	46
50	58	68
50	40	33
64	61	52
53	66	52
40	37	42
63	54	42
66	77	66
78	75	58
48	57	44
85	85	71
82	82	39

Conclusion

The proposed method produced a least absolute deviation estimate that is the same as the one provided by the iterative method by Birkes and Dodge (1993) and other existing methods

References

- Birkes, D., & Dodge, Y. (1993). *Alternative methods of regression*. New York, NY: John Wiley & Sons Inc.
- Chatterjee, S., & Hadi, A. S. (2006). *Regression analysis by example*. New York, NY: John Wiley & Sons Inc.
- Giloni A., & Padberg, M. (2002). Alternative methods of linear regression. *Mathematical and Computer Modelling*, 35(3/4), 361-374.
doi: 10.1016/S0895-7177(01)00170-4
- Nevitt, J., & Tam, H. P. (1998). A comparison of robust and nonparametric estimators under the simple linear regression model. *Multiple Linear Regression Viewpoints*, 25, 54-69.
- Pynnonen, S. & Salmi, T. (1994). A report on least absolute deviation regression with ordinary linear programming. *The Finnish Journal of Business Economics*, (1), 33-49.
- R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Australia. Retrieved from <http://www.R-project.org/>.
- Rao, R. C. & Toutenburg, H. (1999). *Linear models: Least squares and alternatives* (2nd ed.). New York, NY: Springer-Verlag, Inc.
- Rousseeuw, P. J. & Leroy, A. M. (1987). *Robust regression and outlier detection*. New York, NY: John Wiley & Sons.
- Wolberg, J. (2006). *Data analysis using the method of least squares*. Berlin Heidelberg: Springer-Verlag.

ALGORITHM AND CODE FOR LAD ESTIMATOR

Appendix

```
# Simple Linear Regression
# Least Absolute Deviation Estimator (LAD)
Y<-c(16.2,30.5,16.9,33.1,40.2,38.4,41.3,43.9,28.3,33.9,44.2,28,24.6,16)
X<-c(55,27.3,33.3,37.1,11.5,14.2,13.9,19,33.1,43.2,28.5,37.7,6.8,56.5)
p<-1
n<-length(Y)
Const<-rep(1,(p+1))
l<-0;AbsError<-0;EstError<-0;VecLAD<-c();Vecj<-c();Veck<-c();Vecx<-c();Vecy<-
c();VecB0<-c();VecB1<-c()
TrialModel1<-function(Xi,Yi,B0,B1){(abs(Yi-(B0+(B1*Xi))))}
LAD1<-function(n,X,Y,B0,B1){
  for (i in 1:n){Xi<-X[i];Yi<-Y[i]
    EstError<-TrialModel1(Xi,Yi,B0,B1)
    AbsError<-AbsError+EstError
  }
  EstLAD<-AbsError
}
Kstart<-2
for (j in 1:(n-1)){xj<-X[j];yj<-Y[j];
  for (k in Kstart:n){xk<-X[k];yk<-Y[k]
    if(k==n){Kstart<-Kstart+1}
    Vecx<-c(xj,xk)
    Vecy<-c(yj,yk)
    A<-cbind(Const,Vecx)
    Det<-round(det(A),4)
    if(Det!=0){
      Beta<-solve(A,Vecy)
      B0<-Beta[1]
      B1<-Beta[2]
      LADEst<-LAD1(n,X,Y,B0,B1)
      l<-l+1
      VecLAD[l]<-c(LADEst)
      Vecj[l]<-c(j)
      Veck[l]<-c(k)
      VecB0[l]<-c(B0)
      VecB1[l]<-c(B1)
    }
  }
}
```

```

    }
  }
}
LADLoc<-sort.int(VecLAD,index.return=TRUE)
PairLoc<-LADLoc$ix[1]
Pair1<-Vecj[PairLoc]
Pair2<-Veck[PairLoc]
Intercept<-round(VecB0[PairLoc],5)
Slope<-round(VecB1[PairLoc],5)
Label1<-"Best two data points for LAD estimate are";Label2<-"and"
Label3<-"The intercept of the LAD regression is"
Label4<-"The slope of LAD regression is"
Label1;Pair1;Label2;Pair2;Label3;Intercept;Label4;Slope

# Multiple Linear Regression
# Least Absolute Deviation Estimator (LAD)
Y<-
c(43,63,71,61,81,43,58,71,72,67,64,67,69,68,77,81,74,65,65,50,50,64,53,40,63,66,
78,48,
85,82)
X1<-
c(51,64,70,63,78,55,67,75,82,61,53,60,62,83,77,90,85,60,70,58,40,61,66,37,54,77,
75,57,
85,82)
X2<-
c(30,51,68,45,56,49,42,50,72,45,53,47,57,83,54,50,64,65,46,68,33,52,52,42,42,66,
58,44,
71,39)
p<-2
n<-length(Y)
Const<-rep(1,(p+1))
l<-0;AbsError<-0;EstError<-0;VecLAD<-c();VecB0<-c();VecB1<-c();VecB2<-
c();Vecx1<-c();Vecx2<-c();Vecy<-c();Vecj<-c()
Veck<-c();Vecv<-c()
TrialModel2<-function(Yi,X1i,X2i,B0,B1,B2){(abs(Yi-(B0+(B1*X1i)+(B2*X2i))))}
LAD2<-function(n,Yi,X1i,X2i,B0,B1,B2){
  for (i in 1:n){X1i<-X1[i];X2i<-X2[i];Yi<-Y[i]

```

ALGORITHM AND CODE FOR LAD ESTIMATOR

```

        EstError<-TrialModel2(Yi,X1i,X2i,B0,B1,B2)
        AbsError<-AbsError+EstError
    }
    EstLAD<-AbsError
}
Kstart<-1;Vstart<-2;EndCount<-3
for (j in 1:(n-2)){x1j<-X1[j];x2j<-X2[j];yj<-Y[j];
    Kstart<-Kstart+1
    for (k in Kstart:(n-1)){x1k<-X1[k];x2k<-X2[k];yk<-Y[k]
        if(Vstart<n){
            Vstart<-Vstart+1
        }else{
            EndCount<-(EndCount+1)
            Vstart<-EndCount
        }
        for (v in Vstart:n){x1v<-X1[v];x2v<-X2[v];yv<-Y[v]
            Vecx1<-c(x1j,x1k,x1v)
            Vecx2<-c(x2j,x2k,x2v)
            Vecy<-c(yj,yk,yv)
            A<-cbind(Const,Vecx1,Vecx2)
            Det<-round(det(A),4)
            if(Det!=0){
                Beta<-solve(A,Vecy)
                B0<-Beta[1]
                B1<-Beta[2]
                B2<-Beta[3]
                LADEst<-LAD2(n,Yi,X1i,X2i,B0,B1,B2)
                l<-l+1
                VecLAD[l]<-c(LADEst)
                VecB0[l]<-c(B0)
                VecB1[l]<-c(B1)
                VecB2[l]<-c(B2)
                Vecj[l]<-c(j)
                Veck[l]<-c(k)
                Vecv[l]<-c(v)
            }
        }
    }
}

```

```

}
LADLoc<-sort.int(VecLAD,index.return=TRUE)
PairLoc<-LADLoc$ix[1]
FirstPoint<-Vecj[PairLoc]
SecondPoint<-Veck[PairLoc]
ThirdPoint<-Vecv[PairLoc]
Constant<-VecB0[PairLoc]
Beta1<-VecB1[PairLoc]
Beta2<-VecB2[PairLoc]
Label1<-"The first LAD data point is";Label2<-"The second data point
is";Label3<-"The third data point is"
Label4<-"The constant of LAD regression model is";Label5<-"The first parameter
is"
Label6<-"The second parameter is"
Label1;FirstPoint;Label2;SecondPoint;Label3;ThirdPoint;Label4;Constant;Label5;Be
ta1;Label6;Beta2

```