International Journal of Operations Research and Optimization
July-December 2016, Volume 7, No.2, pp. 83-92
ISSN-(Print) 0975-3737, (Online) 2231-4741

Detecting influential observations using Least Absolute Deviations Regression

B. Ashok[1] and R. Elangovan[2]

[1] Department of Community Medicine

Melmaruvathur Adiparasakthi Institute of Medical Sciences and Research

Melamaruvathur – 603 319

Email:bashokstats@gmail.com

[2] Department of Statistics
Annamalai University
Annamalai Nagar – 608 002
Email:srelangovan@rediffmail.com

ARTICLE INFO

Article history:
Received:10-11-2015
Accepted :20-12-2015

ABSTRACT

Robust regression is a form of regression analysis designed to circumvent some limitation of traditional parametric and Non-parametric methods. Robust methods are known as resistant of abnormal values and other violations of model assumptions and appropriate for a broad category of distributions. It is an alternative to lease squares regression when data are contaminated with outliers or influential observations. It can be also used for the purpose of detecting influential observations. Detecting influential observations using Least Absolute Deviations Regression are designed to be not overly affected by violation of assumption by the underlying data generating process. In this paper it is proposed to compare LAD method with the iteratively reweighted least square and ordinarily least square method by detecting influential observations. It is observed that LAD method is more efficient in estimating the parameters in all cases, the distribution of errors follows heavy tailed distributions and in the case of contamination of the data with abnormal values. The LAD regression methods are not only robust but give consistent results in detecting outliers

Keywords:, Deviations Regression, LAD method, Influential observations

1. INTRODUCTION

In the classical approach to the regression problem, the objective is to minimize the sum of squared deviations from the observed and the predicted values of the dependent variable and this method is known as the least squares method; it uses classical optimization methods and generalized inverses. Another method used is minimizing mean absolute deviation from the predicted and observed values of the dependent variable. This problem is known as L_1 – norm minimization or least Absolute Deviation (LAD) method in literature. Third method considered in the literature is that chebyshev criterion of minimizing the maximum of the absolute deviations from the observed and the predicted values of the dependent variables. 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 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 σ^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 a detailed study refer to Huber (1973).

The outliers occurring with extreme values of the regressor variables can be especially disruptive, refer to Andrews (1974). 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, refer to Meyer & Glauber (1964). The least absolute deviation 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. Hence it is professed discuss Least Absolute Deviations Regression by Detecting influential observations. Especially in the case of Pareto, Log-normal, Weibull and Log-Cauchy distribution. Some interesting results can also be seen in Rousseuw and Leroy (2003).

1. Least Square or L_2-norm Method (OLS)

Utilizing the Ordinary Least Squares (OLS) method, the estimator 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 (β) as:

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

The OLS estimate is optimal when the error distribution is assumed to be normal, refer to Hampel (1974), Mosteller & Tukey (1977) in the presence of influential observations.

1.1. Mean Square Error for Model

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

... (2)

1.2. Mean Square Error for Estimator

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

... (3)

1.3. Mean Absolute Error

$$MAE = \frac{\sum_{i=1}^{n} |U_i|}{n}$$

... (4)

2. Least Absolute Deviation (or L1-norm) Method (LAD)

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

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

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 more detailed discussion on L_1 norm iterative weighted least square can be seen in Abdelmalek (1971, 1974), Fair (1974), Schlossmacher (1973), Spyropoulos, Kiountouzis & Young (1973) and Eakambaram and Elangovan (2009, 2010). Following the approach of Robert (2001), the objective function for L_1 regression is

$$f(\beta) = \| Y - \beta X \|_1 \quad \text{... (5)}$$

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

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

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

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

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

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

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

Where r =1, 2, …., m.

Can rewrite (7) as:

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

Let W denote the diagonal matrix as suggested by DasGupta & Mishra (2004), where:

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

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

In the matrix notation

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

Rearrange these equations by multiplying both sides by

$(X'WX)^{-1}$, we get

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

By iterative method equ. (9) converges to a solution. By initializing $\beta^{(0)}$ arbitrarily, the k^{th} iteration after the approximation expressed as

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

The LAD regression estimates are obtainable from the function L_1 fit using the robust regression package suggested by Levent et.al. (2006).

2.1. Mean Square Error for Model

$$(MSE)_{IRWLS}=(\sigma^2)_{IRWLS}$$
$$=SST-SSR=Y'WY-(\hat{\beta})'X'WY \qquad \text{... (11)}$$

2.2. Mean Square Error for Estimator

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

2.3. Mean Absolute Error

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

3. Results

The model under consideration will be of the form

$Y = X\beta + \varepsilon$,

Where

Y is an n x 1 vector of observations on the regressors and

X is an n x p matrix of values of the p regressors

β is a p x 1 vector of parameters and

ε is an n x 1 vector of random disturbances.

Residuals are defined as

$$e = (e_{1,}\ e_{2,\ldots,} e_n)' = Y - X\hat{\beta},$$

Where $\hat{\beta}$ is the estimators of β.

The L_h estimators of β is the $\hat{\beta}$ that minimizes $\sum_{i=1}^{n} |e_i|^h$.

It is to be noted that LAD (h=1) and OLS (h=2) are special cases of L_h estimation.

Now, by using simulation from a set numbers (2, 2.1, 2.2, 2.3 ...), where, distribution of errors (Weibull, Pareto, Log-normal and Log-Cauchy distribution), n=75, n=150, n=225 and n=300, the results were shown in table 1.

Table 1: LAD regression estimates

Error Distribution	Estimator	OLS				IRWLS			
		75	150	225	300	75	150	225	300

Weibull	$\hat{\beta}_0$	1.5074	1.4052	1.5324	1.8326	0.8349	0.6243	1.2932	0.9743
	$\hat{\beta}_1$	0.8396	1.2344	0.9342	0.9783	1.6702	1.2753	0.8760	1.0123
	MSE	1.2340	0.8612	1.0302	1.0102	0.4908	0.8160	0.8731	0.7364
	$MSE(\hat{\beta})$	0.7813	0.1245	0.0463	0.2124	0.0153	0.0412	0.0015	0.0002
	LAD	0.7736	0.1238	0.0362	0.7213	0.7134	0.8354	0.7334	0.7380
Pareto	$\hat{\beta}_0$	8.3810	7.5498	5.2546	3.5478	6.7845	0.2354	0.2015	0.0215
	$\hat{\beta}_1$	-2.1322	-5.2325	6.5421	3.2451	7.0245	7.1554	8.1203	8.3521
	MSE	15.8345	16.5421	18.2152	18.5421	6.2154	5.2124	4.2546	3.2245
	$MSE(\hat{\beta})$	12.7983	8.4235	6.8354	7.0667	2.4956	5.9335	6.3103	6.5643
	LAD	3.8341	0.00012	0.0146	0.0075	0.3329	21.4186	13.8891	18.633
Log-normal	$\hat{\beta}_0$	-0.7432	8.2779	6.8761	6.7910	2.2460	8.4985	8.5245	10.3330
	$\hat{\beta}_1$	1.4382	1.3254	1.3021	0.2457	1.8775	1.5687	1.4265	1.3256
	MSE	5.4732	6.5462	8.2464	5.2156	4.2181	3.0216	2.0315	1.0335
	$MSE(\hat{\beta})$	4.0012	0.8521	0.7725	0.6554	0.5642	0.4578	0.3266	0.2123
	LAD	1.6183	1.4585	1.6224	1.5876	1.4568	1.6548	1.5246	1.4203
Log-Cauchy	$\hat{\beta}_0$	-5.4632	-4.5650	-3.3564	6.2240	5.0213	4.0264	3.2260	2.3256
	$\hat{\beta}_1$	3.3524	8.2321	4.2335	3.0215	2.0315	1.2354	0.2351	0.1254
	MSE	32.8230	125.0	123.5	136.2	2.3245	12.5450	1.0255	2.1542
	$MSE(\hat{\beta})$	28.1392	56.2546	89.3256	98.2215	0.5488	0.6421	0.4578	0.3654
	LAD	3.0342	23.0221	7.2264	6.2254	5.2145	12.6544	8.2566	7.5562

4. Conclusion

It has been shown that the LAD (L_1-norm) is more efficient in estimate the parameters in all cases of distributions of error using IRWLS approach. Especially in the case of following heavy tailed log normal distribution. LAD regression method is very suitable and efficient for detecting influential observations especially in the case of longer tailed distribution.

REFERENCES

1. Abdelmalek, NN (1971). "Linear L_1 Approximation for a Discrete Point Set and L_1-Solutions of Over determined Linear Equations", Journal of Assoc. Comput. Mach., 18, pp.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, pp.35-53.
3. Andrews, D.F. (1974). A robust method for multiple linear Regressions. Technometrics 16, pp.523-531.
4. DasGupta, M & Mishra, SK, (2004), "Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review", http://mpra.ub.uni-muenchen.de/1781.
5. Eakambaram, S and R Elangovan. (2009). "Least square versus Least Absolute Deviations Estimation in Regression Model". Int. J. Agricult. Stat. Sci., Vol.5, No.2, pp.355-372.
6. Eakambaram, S and R Elangovan. (2010). "On the Least Absolute Error Estimation of Linear Regression Models with Auto-Correlated Errors". Ultra Scientist, Vol.21(1)M, pp.213-220.
7. Fair, RC (1974). "On the Robust Estimation of Econometric Models", Annals of Economic and Social Measurement, 3, pp.667-677.
8. Hampel F R, (1974)," The influence curve and its role in robust estimation", Journal of the American Statistical Association, 69, pp.383-393.
9. Huber, PJ (1973). "Robust Regression: Asymptotics, Conjectures, and Monte Carlo" The Annals of Statistics, 1, pp.799-821.
10. 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): pp.2919-2921.
11. Meyer, JR & Glauber R.R (1964). "Investment Decisions: Economic Forecastingand Public Policy", Harvard Business School Press, Cambridge, Massachusetts.
12. Mosteller and Tukey J W, (1977)," Data Analysis and Regression. Reading", MA: Addison Wesley.
13. Rousseuw P. J., & Leroy A. M., (2003), "Robust Regression and Outlier Detection", New York, John Willy.
14. Robert J. Vanderbei, (2001), "Linear Programming: Foundations and Extensions", Robert J. Vanderbei,USA

15. Schlossmacher, EJ (1973). "An Iterative Technique for Absolute Deviations Curve Fitting", Journal of the American Statistical Association, 68, pp.857-859.
16. Spyropoulous, K; Kiountouzis, E & Young, A (1973). "Discrete Approximation in the L_1 Norm", Computer Journal, 16, pp.180-186.