Available Online at http://www.recentscientific.com

International Journal of Recent Scientific Research
Vol. 7, Issue, 10, pp. 13858-13861, October, 2016

ISSN: 0976-3031

International Journal of Recent Scientific Research

Research Article

LEAST ABSOLUTE DEVIATION METHODS USING NONLINEAR WITH ROBUST REGRESSION MODELS

Eakambaram, S[1]., Salomi M[2] and Elangovan R[3]

[1]Department of Statistics, M. R. Government Arts College, Mannargudi
[2,3]Department of Statistics, Annamalai University, Annamalai Nagar -608 002

ARTICLE INFO

Article History:

Received 05th July, 2016
Received in revised form 08th August, 2016
Accepted 10th September, 2016
Published online 28st October, 2016

Key Words:

Least Absolute Deviation, Nonlinear Regression Models, Ordinary Least Square, Iterative Weighted Least Square.

ABSTRACT

In this paper the research deals with appropriate methods of estimation and the important of using techniques estimating the parameter of nonlinear regression of linear regression with nonlinear regression models using robust methods .The possible of non-normal distribution and infinite variance in particular, has led to development of alternative estimation methods to the least square. Provided that one knows the generating distribution a well established procedure is the method of maximum likelihood, which has several optimal properties. Robust methods are known as resistant of abnormal values and other valuation of models assumption and appropriate for aboard category of distributor. A large number of other estimation methods aimed at achieving robustness have been suggested and a considerable body of literature has also been developed. Gonin and Money (1987), and the reference therein. Generally the robust estimators in the literature can be classified as M-estimators, L-estimators, or R- estimators. Probably most attention has been paid to the lestimators for other type estimators see judge et.al (1985) in the recent past. In this paper least absolute deviation methods using nonlinear with robust regression models has be studied Numerical illustration are also provided.

INTRODUCTION

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 . 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 "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) 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 2σ very satisfactory if the quadraticloss 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 amore 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. 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).

Robust estimation refers to the ability of a procedure to produce highly insensitive estimates to model

*Corresponding author: Eakambaram, S
Department of Statistics, M. R. Government Arts College, Mannargudi

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 one gives up the normality assumption it is frequently possible to find estimation methods that are more efficient than the traditional least squares. Specifically this is true 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. For example in finance and accounting research the idea is not at all uncommon that the underlying distributions could have infinite variances. For example, it has been for long observed that speculative price series tend to have volatility clusterings resulting to kurtic and thick tailed unconditional distributions [see e.g., Mandelbrot (1967), Fama and Roll (1968)]. If so then the least squares approach becomes totally inappropriate, for it minimizes the squared deviations that heavily weight the outlying observations, typical to thick tailed distributions. A Finnish example of the application of the LAD estimation for growth estimation in long-run IRR assessment is Luoma (1983). Also Luoma and Pynnönen (1993) have found the LAD method useful in certain applications of firm' s steady state growth estimation. The possibility of non-normal distributions, and infinite variance in particular, has led to development of alternative estimation methods to the least squares. Provided that one knows the generating distribution, a well established procedure is the method of maximum likelihood, which has several optimal properties. However, this method strongly relies on the knowledge of the distributional form, and hence by construction is not necessarily a robust method, except for the case where the underlying distribution itself is robust.A large number of other estimation methods aimed at achieving robustness have been suggested and a considerable body of literature has developed. See for example, Gonin and Money (1989), Dodge (1987) and the references therein. Generally the robust estimators in the literature can be classified as M-estimators, L-estimators, or R-estimators. Probably most attention has been paid to the *L*estimators. For other type estimators, see e.g. Judge *et al. (1985).*

Nonlinear Regression Models

The basic idea of nonlinear regression, namely to relate a response Y to a vector of Predictor variable X=(x_1, x_2, ….. , x_k $)^T$ (See linear models). Nonlinear regression is characterized by the fact that the prediction equation depends nonlinearly on one or more unknown parameters. Whereas linear regression is often used for building a purely empirical models, nonlinear regression usually arises when there are physical reasons for believing that the relationship between the response and the predictors follows a particular functional form.

A nonlinear regression models has the form

$$Y_i = f(X_i, \theta) + \varepsilon_i \qquad \text{i= 1,2 …. N} \qquad \dots (1)$$

Where Y_i are response f is a known function of the covariate vector X_i =(X; 1, … , $X_{ik})^T$ and the parameter vector $\theta = (\theta_1, \dots \dots, \theta_p)^T$, and ε_i are random error . The ε_i are usually assumed to be uncorrelation with mean zero and constant Variance.

Common models

The functional relationship of Y and X of the following form

$$Y_i = f(X_i, \theta) + \varepsilon_i$$

$$Y_i = \theta_1 e^{\theta_2 x_i} + \theta_2 e^{\theta_3 x_i} + \varepsilon_i \qquad \dots (1)$$

Is not linear in the unknown parameters θ_1 and θ_2. We will use the symbol θ to represent a parameter in a nonlinear model where θ is p× 1 vector of unknown parameter. We can liinearize the expectation function by taking logarithms

In E(Y_i) = Inθ_1+ Inθ_2+$\theta_2 x + \theta_3 x \dots \dots$

Tempting to consider rewriting models

InY_i = Inθ_1+ Inθ_2+$\theta_2 x_i + \theta_3 x_i + \varepsilon_i$

β_0, β_1 and β_2 are multiple linear regression to estimate to find β_0, β_1 and β_2 form $[x_i, y_i]$ i=1,2,…,n is called the estimation of the parameter

$$\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i \qquad \dots (2)$$

Using multiple linear regression to estimate β_0, β_1 and β_2 . The linear least square of the parameter in equ.(2) will not in general be equivalent to the nonlinear parameter estimates in the original model equ.(1) the error structure is additive the model in equ.(2).If the error structure is multiplicative

$$y_i = \theta_1 e^{\theta_2 x_1} \varepsilon_i + \theta_2 e^{\theta_3 x_2} \varepsilon_i$$

Taking log on both sides

In Y_i = Inθ_1 + Inθ_2 + In ε

ε is normal distribution.

Models Michaelis menton

$$Y_i = \frac{\theta_1 x_i}{x_i + \theta_2} + \frac{\theta_2 x_i}{x_i + \theta_3} + \varepsilon_i \qquad \dots (3)$$

Expectation function can be linearized

$$\frac{1}{f(x_i, \theta)} = \frac{x_1 + \theta_2}{x_1 \theta_1} + \frac{\theta_2}{x_2 \theta_2} \qquad i = 1,2 \dots n$$

$$\left[\frac{1}{\theta_1} + \frac{\theta_2}{x_1 \theta_1}\right] + \left[\frac{1}{\theta_2} + \frac{\theta_3}{x_2 \theta_2}\right]$$

$$Y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

Where $\beta_0 = \left[\frac{1}{\theta_1} + \frac{1}{\theta_2} + \dots\right]$

$$\beta_1 x_1 = \frac{\theta_2}{\theta_1} \frac{1}{x_1}$$

$$\beta_1 x_1 = \frac{\theta_2}{\theta_1} \frac{1}{x_1}$$

Linear models

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i$$

Where Y= $\frac{1}{y}$ X= $\frac{1}{x}$

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i \qquad \dots (4)$$

$$\varepsilon_i = y_i \quad \beta_0 \quad \beta_1 x_1 \quad \beta_2 x_2$$

1.1 Minimizing Sum of Square Deviation Estimate the Parameters

$$\sum e_i^2 = \sum_{i=1}^{n} [y_i \quad \beta_0 \quad \beta_1 x_1 \quad \beta_2 x_2]^2$$

To estimate β_0, β_1 and β_2 must satisfies and partially differentiate $\sum e_i^2$ with respect $\hat{\beta}_0, \hat{\beta}_1$ and $\hat{\beta}_2$ and set these partial deviation equal to zero .

$$\sum y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_1 + \hat{\beta}_2 \sum x_2 \quad \ldots (5)$$

$$\sum y_i x_1 = \hat{\beta}_0 \sum x_1 + \hat{\beta}_1 \sum {x_1}^2 + \hat{\beta}_2 \sum x_1 \sum x_2 \quad \ldots (6)$$

$$\sum y_i x_2 = \hat{\beta}_0 \sum x_2 + \hat{\beta}_1 \sum x_1 \sum x_2 + \hat{\beta}_2 \sum {x_2}^2 \quad \ldots (7)$$

Now solving equ.,(5) we get $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}_1 - \hat{\beta}_2 \bar{x}_2$

$$\hat{\beta}_1 = \frac{\sum x_1 \, y_i \sum {x_2}^2 \quad \sum x_1 x_2 \sum y_i x_2}{\sum {x_1}^2 \sum {x_2}^2 \quad (\sum x_1 \sum x_2)^2}$$

$$\hat{\beta}_2 = \frac{\sum x_2 \, y_i \sum {x_2}^2 \quad \sum x_1 x_2 \sum y_i x_1}{\sum {x_1}^2 \sum {x_2}^2 \quad (\sum x_1 \sum x_2)^2}$$

Least Square or L2-norm Method (OLS)

Utilizing the Ordinary Least Squares (OLS) method, the estimators found by minimizing the sum of squared residuals:

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

This gives the OLS estimator for (β) as

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

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)

2.1 Mean Square Error for Model

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

SS_T-SS_R = $Y'Y \quad (\beta)'X'Y$

$$= \frac{\sum_{i=1}^{n} {e_i}^2}{d.f(error)} \quad \ldots (2.1)$$

Mean Square Error for Estimator

$$\text{MSE } (\hat{\beta}_1)_{OLS} = (\sigma^2)_{OLS} \operatorname{tr} (X'X)^{-1} \quad \ldots (2.2)$$

Mean Absolute Error

$$\text{MAE} = \frac{\sum_{i=1}^{n} |e_i|}{n} \quad \ldots (2.3)$$

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

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) Fair (1974), Schlossmacher (1973) and Spyropoulos, Kiountouzis & Young (1973), 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), 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)

$$f(\beta) = \| y \quad \beta X \| \quad \ldots (1.1)$$

$$f(\beta) = \sum_{i=1}^{n} \left| y_i \quad \sum_{j=1}^{m} \beta_j X_{ij} \right| \quad \ldots (1.2)$$

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.

$$\frac{\partial f}{\partial \beta_r} = \sum_{i=1}^{n} \frac{y_i \quad \sum_{j=1}^{m} \beta_j X_{ij}}{\left| y_i \quad \sum_{j=1}^{m} \beta_j X_{ij} \right|} (-X_{ir}) = 0 \quad \ldots (1.3)$$

Where r= 1,2,….,m

Can rewrite (5.3)

$$\sum_{i=1}^{n} \frac{\beta_j X_{ij}}{e_i} = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\beta_j X_{ir} X_{ij}}{e_i} \quad \ldots (1.4)$$

Let Wdenote the diagonal matrix, DasGupta & Mishra (2004) where

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

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

We can write these equations in matrix notation as follows:

$$(X'wY) = X'wX\beta$$

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 (β). But let us rearrange this system of equations by multiplying both sides by of:

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

$$\hat{\beta} = (X'WX)^{-1} X'wY \quad \ldots (1.5)$$

3.1 Iteration Weighted Least Square

Let $\hat{\beta}^{(k)}$ denote the approximation at the k[th] iteration, the formula can be expressed as:

$$\hat{\beta}^{(k)}{}_{IRWLS} = (X'X)^{-1} X'WY \quad \ldots (3.1)$$

Mean Square Error For Model

$$(MSE)_{IRWLS} = (\sigma^2)_{IRWLS} \quad \ldots (3.2)$$

$$SS_T\text{-}SS_R = Y'WY \quad \hat{\beta}'X'WY \quad \ldots (3.3)$$

Table 1

Distribution Of error	Estimator	OLS				IRWLS			
		n = 50	**n =100**	**n =150**	**n = 200**	**n = 50**	**n =100**	**n =150**	**n = 200**
Normal Distribution	$\hat{\beta}_0$	1.2532	1.0886	1.2976	1.0658	0.8672	0.3651	1.8324	0.9652
	$\hat{\beta}_1$	0.9672	1.0444	0.8965	1.2346	1.1660	1.1671	0.8751	1.1023
	$\hat{\beta}_2$	1.2432	0.4536	0.9674	1.3343	1.2346	1.2346	1.5674	1.3248
	MSE	1.3789	2.3467	0.763	0.5986	0.6899	1.9823	1.5899	0.5688
	MSE ($\hat{\beta}$)	1.0392	0.9688	1.0000	0.9876	1.324	1.321	0.5834	0.8664
	MAE	0.7684	0.4647	0.7608	0.8743	0.7807	0.9801	0.8231	1.3248
Contaminated Distribution	$\hat{\beta}_0$	14.432	2.0781	10.00	-0.8119	2.8107	0.6576	1.3579	1.1870
	$\hat{\beta}_1$	1.0993	-2.131	-3.413	-0.3524	-2.312	1.2120	0.819	0.5938
	$\hat{\beta}_2$	0.8765	0.8402	0.8161	0.9859	0.8743	1.4435	1.0987	0.2367
	MSE	643.2	924.5	923.3	2.4365	7,0665	6.5832	8.4256	17.56
	MSE ($\hat{\beta}$)	14.653	92.621	32.891	21.4816	0.3312	0.3312	0.0587	0.0165
	MAE	3.2435	9.3320	9.5142	8.4186	2.2160	5.7910	6.814	8.2669
Laplace Distribution	$\hat{\beta}_0$	1.2345	1.9290	1.9380	0.8263	1.2346	1.9823	1.9360	0.8523
	$\hat{\beta}_1$	0.3876	0.1876	0.2347	0.4532	0.1887	1.2356	0.3456	0.5647
	$\hat{\beta}_2$	1.9837	12.384	1.8437	1.8938	1.9860	13.766	1.7654	1.8895
	MSE	0.2987	12.344	0.8829	0.2934	0.9876	12.458	0.9858	0.3456
	MSE ($\hat{\beta}$)	13.567	19.332	14.736	14.244	19.567	19.567	15.456	14.660
	MAE	678.08	84.234	56.344	682.00	685.00	85.432	57.765	691.00

Mean Square Error for Estimation

MSE

$$(\hat{\beta})_{IRWLS} = (\sigma^2)_{IRWLS}\ tr\ (X'WX)^{-1} \quad \text{... (2.4)}$$

Mean Absolute Error

$$\text{MAE}=\frac{\sum_{i=1}^{n}|e_i|}{n} \quad \text{… (2.5)}$$

The Simulation Study

The following model are used

$Y_i = \beta_0 + \beta_1 X_i + \beta_2 X_I + e_i$

Where distribution of the error Normal distribution, Weibull distribution, Laplace distribution. The size of the random n= 50,100, 150 and 200.

CONCLUSION

It is observed that the iterative weighted least square methods provides Robust estimator comparing to OLS. From the table it's observed that for the heavy tailed distribution gives better estimates then the Normal and Laplace distribution. The estimates are provided in this paper are not only Robust but gives consistent results.

Reference

Absolute Deviation Estimator". *J. of Japan Statist. Society*, 31-1 (39-51).

Andrews, D.F. (1974). A robust method for multiple linear regression. Technometrics 16, 523-531.

DasGupta, M & Mishra, SK, (2004), "Least Absolute Deviation Estimation of Linear Econometric Models: A Literature Review", http://mpra.ub.uni-muenchen.de/1781.

Frias H.Thanoon (2015). "Robust Regression by Least Absolute Deviation Method", *International Journal Of Statistics And Application*, 5(3):109-112 DOI: 10.5923/j.statistics.20150503.02

Gordon k.Smyth (2002). "Nonlinear Regression ", John Wiley & Sons Ltd. Chichester, Encylopedia of Enivormetrics Volume 3.pp 1405-1411.ISBN 0471899976.

Hamilton, D.C., Watts, D.G. & Bates, D.M. (1982) Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions, *The Annals of Statistics* 10, 386–393.

Hill, RW & PW Holland (1977). "Two Robust Alternatives to Least Square Regression", *Journal of American Statistical Association*, 72 (828-833).

Hitomi, K & M Kagihara (2001). "Calculation Method for Non-Linear Dynamic Least

Hood, WC & TC, Koopmans (Eds) (1953). Studies in Econometric Method, Cowles Foundation Monograph 14, John Wiley, New York.

Huber, PJ (1964). "Robust Estimation of a Location Parameter", Annals of Mathematical Statistics, 35 (73-101).

Huber, PJ (1972). "Robust Statistics: A Review", *Annals of Mathematical Statistics*, 43 (1041-1067).

Huber, PJ (1973). "Robust Regression: Asymptotics, Conjectures, and Monte Carlo" The Annals of Statistics, 1 (799-821).

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.

Madhuchanda Dasgupta and Sk Mishra (2007). "Least Absolute Deviation Estimation of Linear Econometric Models: A Literature review", MPRA paper No.1781.

Rousseuw P. J., & Leroy A. M., (2003), "Robust Regression and Outlier Detection", New York, John Willy.

Seber, G.A.F. & Wild, C.J. (1989). *Nonlinear Regression*, Wiley, New York.

****