Least Absolute Deviation Estimation versus Nonlinear Regression Model With Heteroscedasticity Errors

S.Eakambaram\(^1\), M. Salomi\(^2\), and R. Elangovan\(^3\).
\(^1\) Department of Statistics, M. R. Government Arts College, Mannargudi.
\(^2\)&\(^3\) Department of Statistics, Annamalai University, Annamalai Nagar -608 002.
E-mail:salomim.phill@gmail.com

**ARTICLE INFO**

Article history:
Received:10-02-2016
Accepted:25-05-2016

**ABSTRACT**

Least absolute deviation (LAD), also known as least absolute errors (LAE), Least absolute value (LAV) Least absolute residual (LAR) Sum of absolute deviation, or the L\(_1\) norm condition is a statistical optimality criterion and the statistical optimization technique that relies on it. The least absolute Deviation (LAD) estimation in a field of great interest in regression analysis. The application of this method are available in the recent literature covering various aspects of modeling, computational efficiency, error analysis, gross error identification etc. Nonlinear regression is characterized by the fact that the prediction equation depends nonlinearly on one or more unknown parameters. The inferential aspect of unknown regression models and the error assumptions are usually analogous to those made for linear regression models. The various inferential problems on nonlinear regression models, involving heteroscedastic errors have been studied by several author cited in the recent literature. In this paper its proposed to discuss LAD estimations versus nonlinear regression model with heteroscedastic errors. The problem of heteroscedasticity can also be studied with reference to nonlinear regression models with suitable illustration.

**Keyword:** heteroscedastic errors, nonlinear regression models.
1 INTRODUCTION

In linear regression, we study the relationship between \( p \) explanatory variables, \( X_1, \ldots, X_p \), and a continuous response variable \( Y \). This model takes the form \( Y = X\beta + \varepsilon \) where \( X \) is an \( n \times (p + 1) \) matrix of explanatory variables, \( Y \) is an \( n \times 1 \) vector of responses, \( \beta \) is a \( (p+1) \times 1 \) vector of unknown regression coefficients, and \( \varepsilon \) is an \( n \times 1 \) vector of unobservable errors with a Normal \( (0, \sigma^2 I) \) distribution (Cook and Weisberg, 1982). Homoscedasticity is the assumption that the error variance is equal to \( \sigma^2 \) for all \( n \) observation. The assumption is used for estimating the regression coefficient \( \beta \) through ordinary least square (OLS), leading to the solution \( \hat{\beta} = (X'X)^{-1}X'Y \). By the Gauss-Markov theorem, \( \hat{\beta} \) is the linear, unbiased estimator with the smallest variance. Heteroscedasticity and its impact on linear regression have been extensively studied. Based on this research, it has been shown that using linear regression methods to perform prediction with heteroscedastic data can fail dramatically (Carroll & Ruppert, 1988). This is because regions of low variability end up having significantly less influence setting parameters and making predictions than regions containing high variability. This divergence can result in predictions that significantly misrepresent the true mean of the data, especially in regions of low variability. The linear regression in the presence of heteroscedasticity can prevent the type 1 errors and coverage probabilities of confidence intervals (CIs) for model-based predictions from attaining the nominal level (Carroll and Ruppert 1988; Lim, Sen, & Peddada, 2010; Visek, 2011). This failure can cause practitioners to declare an outcome statistically significant when in fact it is not.

Heteroscedasticity is the violation of the homoscedasticity assumption. When it occurs, the OLS estimates \( \hat{\beta} \) are still unbiased, but become inefficient. The regular standard errors of these estimates are wrong, leading to incorrect inferences, although White’s heteroscedastic corrected standard errors (White, 1980) can be larger effect on minimization the unweighted least square
criterion in OLS and contribute more to predictions. Under certain assumptions, the OLS estimator has a normal asymptotic distribution when properly normalized and centered (even when the data does not come from a normal distribution). This result is used to justify using a normal distribution, or a chi square distribution (depending on how the test statistic is calculated), when conducting a hypothesis test. This holds even under heteroscedasticity. More precisely, the OLS estimator in the presence of heteroscedasticity is asymptotically normal, when properly normalized and centered, with a variance-covariance matrix that differs from the case of homoscedasticity. In 1980, White proposed a consistent estimator for the variance-covariance matrix of the asymptotic distribution of the OLS estimator. This validates the use of hypothesis testing using OLS estimators and White’s variance-covariance estimator under heteroscedasticity.

Heteroscedasticity is also a major practical issue encountered in ANOVA problems. The $F$ test can still be used in some circumstances. However, it has been said that students in econometrics should not overreact to heteroscedasticity. One author wrote, “unequal error variance is worth correcting only when the problem is severe.” In addition, another word of caution was in the form, “heteroscedasticity has never been a reason to throw out an otherwise good model.” With the advent of heteroscedasticity-consistent standard errors allowing for inference without specifying the conditional second moment of error term, testing conditional homoscedasticity is not as important as in the past. For any non-linear model (for instance Logit and Probit models), however, heteroscedasticity has more severe consequences: the maximum likelihood estimates of the parameters will be biased (in an unknown direction), as well as inconsistent (unless the likelihood function is modified to correctly take into account the precise form of heteroskedasticity). As pointed out by Greene, “simply computing a robust covariance matrix for an otherwise in consistent estimator does not give it redemption. Consequently, the virtue of a robust covariance matrix in this setting is unclear.” Heteroscedasticity can also arise as a result of the presence of outliers. An outlying observation, or outlier, is an observation that is much different (either very small or very large) in relation to the observations in the sample. More precisely, an outlier is an observation from a different population to that generating the remaining sample observations. The inclusion or exclusion of such an observation, especially if the sample size is small, can substantially alter the results of regression analysis.

2. LEAST ABSOLUTE ESTIMATION IN REGRESSION MODEL

Least absolute deviations (LAD), also known as least absolute errors (LAE), least absolute value (LAV), least absolute residual (LAR), sum of absolute deviation, or the $L_1$ norm condition, is a statistical optimality criterion and the statistical optimization technique that relies on it. Similarly to the popular least squares technique, it attempts to find a function which closely approximates a set of data. In the simple case of a set of $(x,y)$ data, the approximation function is
a simple “trend line” in two-dimensional Cartesian coordinates. The method minimizes the sum of absolute errors (SBE) (the sum of the absolute value of the vertical “residuals” between points generated by the function and corresponding points in the data). The least absolute deviations estimate also arises as the maximum likelihood estimate if the errors have a Laplace distribution.

Suppose that the data set consist of the points $(x_i, y_i)$ with $i=1,2,...,n$. We want to find a function $f$ such that $f(x_i) = y_i$. We suppose that the function $f$ is of a particular form containing some parameters which need to be determined. For instance, the simplest form would be linear: $f(x) = bx + c$, where $b$ and $c$ are parameters whose values are not known but which we would like to estimate. Less simply, suppose that $f(x)$ is quadratic, meaning that $f(x) = ax^2 + bx + c$, where $a$, $b$ and $c$ are not yet known. (More generally, there could be not just one explanators, all appearing as arguments of the function $f$. We now seek estimated values of the unknown parameters that minimize the sum of the absolute values of the residuals.

$$s = \sum_{i=1}^{n} |y_i - f(x_i)|$$

The least absolute deviation problem may be extended to include multiple explanatory, constraints and regularization, e.g., a linear model with linear constraints.

$$s = (\beta, b) = \sum_{i} |X_i^\prime \beta + b - y_i|$$

$$x_i^\prime \beta + b - y_i \leq k$$

Where $\beta$ is a column vector of coefficient to be estimated, $b$ is an intercept to be estimated, $X_i$ is a column vector of $i$th observation on the various explanators, $Y_i$ is the $i$th observation on the
dependent variable, and \( k \) is a known constant. Unlike least square regression, least absolute deviation regression does not have an analytical solving method. Therefore, an iterative approach is required. The methods of solving LAD regression listed under the following heads.

1) Simplex-based methods
2) Iteratively re-weighted least square
3) Wesolowsky’s direct descent method
4) Li-Arce’s maximum likelihood approach

For the detailed study refer to Barrodale- Roberts algorithm, year (1973).

3. LEAST ABSOLUTE DEVIATION WITH HETROScedACITY ERROR.

Least Absolute Deviations (LAD) regression is a field of great interests in regression analysis several. The application of this method is available in the recent literature covering various aspects of modeling, computational efficiency, error analysis, gross error identification etc. Linear models, their variants, and extensions are among the most useful and widely used statistical tools for social research. This paper aims to provide an accessible, in-depth, modern treatment of regression analysis, and closely related methods. The linear regression model is given by the equation

\[
y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \epsilon_i
\]

for \( i=1,2,\ldots,n \) sampled observations. In this equation, \( y_i \) is the dependent variable, the \( x_{ij} \) are regressors, \( \epsilon_i \) and is an unobservable error. The \( \beta_j \) are unknown parameters to be estimated from the data. It is assume that the errors are normally and independently distributed with zero expectation (mean) and constant variance \( \sigma^2: \epsilon_i \sim \text{NID}(0, \sigma^2) \). If the assumptions that errors are normally distributed appears to be violated, and if any prior knowledge is available about error distribution, the maximum likelihood argument could be used to obtain estimates based on the criterion other than least squares namely minimization of \( \sum_{i=1}^{n} |\epsilon_i| \), sum of absolute errors, the linear regression model stated in equ. (3.1) is a particular method of minimizing L_p norm. Taking the linear model of the form stated in equation (3.1), \( Y_i = X_i' \beta + \epsilon_i \), the L_p norm is \( \left[ \sum |\epsilon_i|^p \right]^{1/p} \) with \( p \geq 1 \). Minimizing this, one get estimate of \( \beta \) and also use this for other inferential purposes. Put \( p = 2 \), we get the classical approach to regression problems, namely least square estimates, though they possess properties such as simplicity, they are not well suited if there are violation in the basic assumptions as stated above. When \( p = \infty \) one get the well known
Least Absolute Deviation Estimation versus Nonlinear

When $p = 1$ the $L_1$ norm method is obtained and is known as the Least Absolute Deviation method in literature. The methods of minimizing the sum of absolute and squared deviations from hypothesized linear models have vied for statistical favour for more than 250 years. This venerable method which Laplace called the ‘method of situations’ had a bewildering variety of names. Recently it has accumulated a large array of acronyms: LAE, LAD, MAD, MSAE, and Others. It is also frequently referred to as $L_1$-regression, and less frequently as median regression. A more detailed discussion on LS procedures and its estimates can be seen in Searle (1971). Heteroscedasticity occurs when the variance of the errors varies across observations. If the errors are heteroscedastic, the OLS estimator remains unbiased, but becomes inefficient. More importantly, estimates of the standard errors are inconsistent. The estimated standard errors can be either too large or too small, in either case resulting in incorrect inferences. Given that heteroscedasticity is a common problem in cross-sectional data analysis, methods that correct for heteroscedasticity are important for prudent data analysis. When the form of heteroscedasticity is unknown. In matrix notation.

$$y = X\beta + \varepsilon$$

Where $y$ and $\varepsilon$ are $N \times 1$ matrices, $X$ is $N \times k$ and $\beta$ is $k \times 1$. For the ith row of $X$, we can write

$$y_i = x_i\beta + \varepsilon_i$$

The following are the assumptions of complete the model:
1) Linearity: $y$ is linearly related to the $x$'s through the $\beta$ parameters.
2) Collinearity: The $x$'s are not linearly dependent.
3) Expectation of $\varepsilon$: $E(\varepsilon_i / x_i) = 0$ for all $i$.
4) Homoscedasticity: For a given $x_i$ the errors have a constant variance: $\operatorname{Var}(\varepsilon_i | x_i) = \sigma^2$ for all $i$. 

\[\text{YACS}\]
5) Uncorrelated errors: For two observations i and j, the covariance between $\epsilon_i$ and $\epsilon_j$ is zero.

With these assumption, the OLS estimates $\hat{\beta} = (X'X)^{-1}X'y$ has the covariance matrix:

$$Var(\hat{\beta}) = (X'X)^{-1}X'\Phi X(X'X)^{-1} \quad \ldots (3.2)$$

Where $\Phi$ is the diagonal matrix with $\phi_{ii} = Var(\epsilon_{ii})$. When the errors are homocedastic $\Phi$ can be written as $\Phi = \sigma^2 I$. With this assumption, equ.(3.2) can be simplified.

$$Var(\hat{\beta}) = (X'X)^{-1}X'(\sigma^2 I)X(X'X)^{-1} \quad \ldots (3.4)$$

$$= \sigma^2 (X'X)^{-1}X'X(X'X)^{-1}$$

$$= \sigma^2 (X'X)^{-1}$$

Definition of residuals $e_t = y_t - x_i \hat{\beta}$, we can estimate the usual OLS covariance matrix, hereafter referred to as OLSCM, as:

$$OLSCM = \frac{\sum e_t^2}{N-K}(X'X)^{-1} = s^2 (X'X)^{-1}$$

If the errors are heteroscedastic, then OLSCM is biased and the usual tests of statistical significance are inconsistent. When $\Phi$ is known, equ.(3.2) can be used to correct for heteroscedasticity. Unfortunately, the form of heteroscedasticity is rarely known. When $\Phi$ is
unknown, we need a consistent estimator of $\Phi$ in order to apply equ (3.2). The HCCM is based on the idea that $e_i^2$ can be used to estimate $\phi_{ii}$. This can be thought of as estimating the variance of the error using a single observation: 

$$\hat{\phi}_{ii} = \frac{(e_i - \bar{e})^2}{1} = e_i^2.$$ 

MacKinnom and White (1985) consider alternative estimators designed to improve on the small sample properties of HCO. The equ. (3.2) is based on the OLS residuals $e$ will not be. Specially, if we define $h_{ii} = X_i(X'X)^{-1}X'_i$, then:

$$\text{Var}(e_i) = \sigma^2(1 - h_{ii}) \neq \sigma^2 \quad \cdots (3.4)$$

$\text{Var}(e_i)$ will understand $\sigma^2$. Further, Since $\text{Var}(e_i)$ varies across observation when the errors are homoscedastistic, the OLS residual must be heterscedastic.

### 4. NON LINEAR MODELS WITH HETROSCEDASTCITY ERRORS

Consider the standard nonlinear regression model with usual assumptions in vector notation as

$$Y_{n \times n} = f_{n \times 1}(\beta) + \varepsilon_{n \times 1} \quad \cdots (4.1)$$

and $\beta$ is $(p \times 1)$ vector of unknown parameters. Suppose that $\hat{\beta}$ is the nonlinear least squares estimator of $\beta$.

For large samples, the NLLS residuals vector is given by
\( e = [Y - \hat{Y}] = [Y - f(\hat{\beta})] \) \quad \ldots (4.2)

Where \( \hat{\beta} = \beta + (F'F)^{-1} F' \varepsilon \) \quad \ldots (4.3)

And \( F = F(\beta) = \left( \frac{\partial}{\partial \beta} f(X_i, \beta) \right)_{n \times p} \) \quad \ldots (4.4)

Here \( \frac{\partial}{\partial \beta} f(X_i, \beta) \) is the \((i,j)^{th}\) element of \((n \times p)\) matrix \( F(\beta) \).

An approximate relationship between \( e \) and \( \varepsilon \) is given by

\( e \cong M\varepsilon \), \quad \text{Where } M = \left[ I - F(F'F)^{-1} F' \right] \)

Or \( e \cong [I - V] \varepsilon \), \quad \text{where } v = (v_{ij}) = F(F'F)^{-1} F' \) is symmetric idempotent matrix known as ‘HAT’ matrix in scalar form,

\( e_i = [\varepsilon_i - \sum_{j=1}^{n} v_{ij} \varepsilon_j], i = 1,2, ..., n \) \quad \ldots (4.5)

Since, \( v \) is symmetric idempotent matrix, it follows that

\( \text{Trace}(v) = \text{rank}(v) = p \)

And \( \sum_{j=1}^{n} v_{ij}^2 = v_{i\mu}, \quad l = 1,2, ..., n \)

If \( \varepsilon \) follows \( N(0, \sigma^2 I) \) then \( e \) follows a singular normal distribution with zero mean vector and variance \( \sigma^2 \). Here \( v \) controls the variation in \( e \). Since, the variance of each \( e_i \) is a function of both.
\( \sigma^2 \) and \( v_{ii} \); the NLLS residuals have a probability distribution that is scale dependent.

The nonlinear studentized residuals do not depend on either of these quantities and they have probability distribution that is free of the nuisance scale parameters. One can make a further distinction between internal studentization and external studentization.

5(i) NONLINEAR STUDENTIZED RESIDUALS

In NLLS regression, the internally nonlinear studentized residuals are defined by,

\[
e_i^* = \frac{e_i}{\hat{\sigma}_{(i)} \sqrt{1-v_{ii}}}, \quad i=1,2,...,n
\]  

Where \( \hat{\sigma}^2 = \left[ \frac{\sum e_i^2}{n-p} \right] \beta_{\alpha \gamma} \) is the beta distribution with parameter \( \frac{1}{2} \) and \( \frac{(n-p-1)}{2} \).

It follows that \( E(e_i) = 0 \) and \( \text{var}(e_i) = 1, \forall \ i = 1,2,...,n \)

Also, \( \text{Cov}(e_i^*, e_j^*) = \frac{-v_{ij}}{[(1-v_{ii})(1-v_{jj})]^{\frac{1}{2}}}, \forall, \ i \neq j = 1,2,...,n \)

(ii) EXTERNALLY NONLINEAR STUDENTIZED RESIDUALS

The externally nonlinear studentized residuals are defined by,

\[
e_i^{**} = \frac{e_i}{\hat{\sigma}_{(i)} \sqrt{1-v_{ii}}} , \forall, \ i = 1,2,...,n
\]  

Where

\[
\hat{\sigma}^2_{(i)} = \frac{(n-p)\hat{\sigma}^2 - \left[ \frac{\sum e_i^2}{(1-v_{ii})} \right]}{n-p-1}
\]
Or
\[ \hat{\sigma}^2(i) = \hat{\sigma}^2 \left[ \frac{n - p - e_i^2}{n - p - 1} \right] \]

Under normality \( \hat{\sigma}^2(i) \) and \( e_i \) are independent.

Here \( e_i^{**} \sim \text{student's } t\text{-distribution with } (n - p - 1) \text{ degrees of freedom.} \)

A relationship between internally and externally nonlinear standardized residual is given by
\[ e_i^{**} = e_i^* \left[ \frac{n - p - 1}{n - p - e_i^2} \right]^{1/2}, \quad i = 1,2, ..., n \quad \text{... (5.3)} \]

Thus, \( e_i^{**} \) is a monotonic transformation of \( e_i^{**2} \).

6. ESTIMATION OF NONLINEAR REGRESSION MODEL WITH HETROSCARDATIC ERRORS BY USING NONLINEAR STUDENTIZED RESIDUALS.

Consider the standard nonlinear regression model.

\[ Y_i = f(X_i; \beta) + \varepsilon_i, \quad i = 1,2, ..., n \quad \text{... (6.1)} \]

This may be written in matrix notation as.
\[ Y_{n \times 1} = f_{n \times 1}(\beta) + \varepsilon_{n \times 1} \quad \text{... (6.2)} \]

Where \( X_i = (X_{i1}, X_{i2}, ..., X_{ik}) \) is a k-component vector denotes the \( i^{th} \) observation on known k-explanatory variables;
\( \beta \) is a \( p \times 1 \) vector of unknown parameters;

\( f(\beta) \) is a known twice continuously differentiable function of \( \beta \);

The usual assumption of the nonlinear regression model are:

1) \( E \left[ \frac{Y_i}{X_i} \right] = f(X_i, \beta), \ i = 1, 2, \ldots, n; \)

2) \( \beta \) is estimable or identified

3) \( E \left[ \frac{\varepsilon_i}{f(X_i, \beta)} \right] = 0 \)

4) \( E \left[ \frac{\varepsilon_i^2}{f(X_i, \beta)} \right], \ j = 1, 2, \ldots, n \) = \( \sigma^2 \) a finite constant and

5) \( E \left[ \frac{\varepsilon_i \varepsilon_j}{f(X_i, \beta), f(X_j, \beta)} \right], \ i, j = 1, 2, \ldots, n \) = 0 \( \forall \ j \neq i \)

That is, the \( \varepsilon_i \)'s are conditional homoscedastic and nonautocorrelated random error variables (or)

\( E(\varepsilon \varepsilon^T) = \sigma^2 I_n \)

6) \( \varepsilon_i \)'s are normally distributed i.e., \( \varepsilon_i \sim i.i.d. N(0, \sigma^2) \), \( i = 1, 2, \ldots, n \)

By minimizing the residual sum of squares
With respect to \( \tilde{\beta} \), for large samples, under iterative process, an iterative nonlinear least square (NLLS) estimator for \( \beta \) is given by

\[
\tilde{\beta}_{n+1} = \tilde{\beta}_n + [F'(\tilde{\beta}_n)F(\tilde{\beta}_n)]^{-1} F'(\tilde{\beta}_n) \left[ Y - f(\tilde{\beta}_n) \right]
\]  

(6.3)

Where \( F'(\tilde{\beta}_n) \) is the regressor matrix.

Here, all the terms on the R.H.S of (4.11) are evaluated to \( \tilde{\beta}_n \) and \( Y - f(\tilde{\beta}_n) \) is the vector of nonlinear least square residuals for an arbitrary value of \( \lambda \).

By violating the assumption of homoscedastic errors in the nonlinear model (6.1) one may assume that

\[
E[\varepsilon \varepsilon'] = \phi = \sigma^2 \psi
\]  

(6.4)

Where \( \phi \) or \( \psi \) is symmetric positive definite matrix.

If the diagonal elements of dispersion matrix \( \phi \) are not all identical and \( \varepsilon \) is free from autocorrelation then \( \phi \) can be considered a diagonal matrix.

\[
\phi = \text{diag}(\sigma_1^2, \sigma_2^2, ..., \sigma_n^2)
\]

And with \( i \)th diagonal element is given by \( \sigma_i^2 \).

Define the proposed iterative (NLLS) residual vector based on \( \tilde{\beta}_n \) as \( e_n = [Y - f(\tilde{\beta}_n)] \).
Least Absolute Deviation Estimation versus Nonlinear

Also, Iterative Nonlinear Internally Studentized Residuals are defined by

$$e_{nl}^* = \frac{\sigma_{nl}}{\sqrt{1 - v_{nl}}}$$, $i = 1, 2, ..., n$  \hspace{1cm} ...(6.5)

Where

$$v_n = \left( v_{nij} \right) = \left[ F(\hat{\beta}_n) F(\hat{\beta}_n) \right]^{-1} F'(\hat{\beta}_n)$$  \hspace{1cm} ...(6.6)

Is symmetric idempotent matrix known as ‘HAT’ matrix.

$$\hat{\sigma}^2 = \left[ e_n^* e_n^* \right] = \frac{\sum_{i=1}^n e_{ni}^2}{n - p}$$  \hspace{1cm} ...(6.7)

Here, $\left[ e_{ni}^* \right] \sim \text{Beta distribution with parameter } \frac{1}{2} \text{ and } (n-p-1) / 2$,

It follows that, $E[e_{ni}^*] = 0$ and $\text{Var}(e_{ni}^*) = 1$, $\forall\ l = 1, 2, ..., n$

Also

$$\text{Cov}(e_{ni}^*, e_{nj}^*) = \frac{-v_{nj}}{\left( (1 - v_{nij})(1 - v_{nij}) \right)^{1/2}}$$, $\forall\ i \neq j = 1, 2, ..., n$  \hspace{1cm} ...(6.8)

Consider an estimate for $\varphi$

$$\hat{\varphi}_{n}^* = \text{diag} \left[ e_{n1}^*, e_{n2}^*, ..., e_{nn}^* \right]$$  \hspace{1cm} ...(6.9)

Now, an iterative Estimated Nonlinear Generalized Least Square (IENLGLS) estimator for $\beta$ is given by

$$\hat{\beta}_{n+1} = \hat{\beta}_n^* + \left[ F'(\hat{\beta}_n^*) \hat{\varphi}_n F(\hat{\beta}_n^*) \right]^{-1} F'(\hat{\beta}_n^*) [Y - \lambda f(\hat{\beta}_n^*)]$$  \hspace{1cm} ...(6.10)
Here $F(\hat{\beta}_n^*) = \left[ \frac{\partial f}{\partial \beta} \right]_{\hat{\beta}_n^*}$ as the regressor matrix and $[Y - \lambda f(\hat{\beta}_n^*)]$ is the vector of IENLGLS residual for an arbitrary value of $\lambda$.

Further Var $\left( \hat{\beta}_n^* \right) = \left[ F' (\hat{\beta}_n^*) \hat{\sigma}_n^* F(\hat{\beta}_n^*) \right]^{-1}$ ... (6.11)

7. A TEST FOR HETEROSCEDASTICITY IN NONLINEAR REGRESSION MODEL

STUDENTIZED RESIDUALS

One of the crucial assumptions of the nonlinear regression model is that the error observations have equal variances. But, in practice, it has been observed that errors are heteroscedastic. If errors are heteroscedastic, the Nonlinear Least Squares (NLLS) estimates of the parameters are inefficient and usual method of inference may produce misleading conclusions. Thus, there is need for testing the existence of the problem of heteroscedasticity in the nonlinear regression model. A wide number of tests have been developed, with a quickening of interest in the last two decades.

With usual notation, consider the nonlinear regression model

$$Y_i = f(X_i, \beta) + \epsilon_i \quad \text{for } i = 1, 2, \ldots, n$$

Where $X_i = (X_{i1}, X_{i2}, \ldots, X_{ik}), i = 1, 2, \ldots, n$ is a k-component vector denoting the $i^{th}$ observation on known independent variables;

$\beta$ is $(p \times 1)$ vector of unknown parameter. $\epsilon_i, i = 1, 2, \ldots, n$ are i.i.d. error random variables with mean zero and unknown unequal variance i.e., the errors are hetroescedatic errors.

In testing the null hypothesis of homoscedasticity of errors, the following procedure may be applied:

**Step-1** The observation on dependent variable $Y$ are arranged according to the ascending order of observation on independent variable $X$, with which the heteroscedasticity might be associated.
**Step-2** Divide the arranged data into $K$ groups of sizes $n_1, n_2, \ldots, n_K$ respectively. Here $n_1, n_2, \ldots, n_K$ should be approximately equal. One may choose $k$ such that the size of each group is reasonable small and it is greater than the number of parameter in the nonlinear regression model. For instance, for a sample of 30 observations, $k$ may be chosen as 3 such that $n_1, n_2 \text{ and } n_3$ may be equal to 10.

**Step-3** Run separate nonlinear regression models on these $k$ groups of observation and obtain the Iterative Nonlinear Least Square Internally Studentized Residual Sum of Square (INLLSISRSS) for each nonlinear regression model and pooled them as (RSS) with degrees of freedom $(n_1 - p) + (n_2 - p) + \ldots + (n_K - p)$.

**Step-4** Obtain the INLLSISRSS for the combined data as (RSS) by estimating a single nonlinear model with degree of freedom $(\sum_{j=1}^{k} n_j - p)$.

**Step-5** Compute the $F$-test statistic for testing the null hypothesis of homoscedastic errors as

$$F = \frac{[RSS - RSS_{SS}]/p}{\sum_{j=1}^{k} n_j - kp} \sim F[p, (\sum_{j=1}^{k} n_j - kp)]$$

And compare the calculated value of $F$-test statistic with its critical value and draw the inference accordingly.

8. **NUMERICAL STUDY**

In this section we present the numerical computation to compare the least squares deviation LAD, NLAD and Heteroscedasticity (Known $\beta$). The Simulation Study used is to generate stable symmetrically distributed errors is given in Chambers et al. The characteristics exponent was varied of 1.0 to 2.5. the heteroscedasticity parameter $\beta$ was set to 0, ± 0.1, ± 0.5, and ± 0.6.
Table 1: Observed MSE when the disturbance Terms in the Heteroscedasticity Process are Stable with index $\lambda$, n=50

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th></th>
<th>-0.6</th>
<th>-0.5</th>
<th>-0.1</th>
<th>0</th>
<th>0.1</th>
<th>0.5</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>LAD</td>
<td>0.524</td>
<td>0.068</td>
<td>0.058</td>
<td>0.091</td>
<td>0.111</td>
<td>0.356</td>
<td>0.910</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>0.051</td>
<td>0.032</td>
<td>0.054</td>
<td>0.076</td>
<td>0.091</td>
<td>0.320</td>
<td>0.791</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>0.270</td>
<td>0.065</td>
<td>0.088</td>
<td>0.148</td>
<td>0.321</td>
<td>0.278</td>
<td>1.333</td>
</tr>
<tr>
<td>2.0</td>
<td>LAD</td>
<td>2.982</td>
<td>0.195</td>
<td>0.112</td>
<td>0.094</td>
<td>0.161</td>
<td>0.441</td>
<td>10.781</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>0.040</td>
<td>0.627</td>
<td>0.349</td>
<td>0.555</td>
<td>0.671</td>
<td>0.685</td>
<td>10.937</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>1.043</td>
<td>0.070</td>
<td>0.099</td>
<td>0.099</td>
<td>0.156</td>
<td>0.471</td>
<td>4.951</td>
</tr>
<tr>
<td>1.5</td>
<td>LAD</td>
<td>1.110^3</td>
<td>0.517</td>
<td>0.139</td>
<td>0.160</td>
<td>0.238</td>
<td>14.214</td>
<td>12.781</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>531.421</td>
<td>92.312</td>
<td>25.251</td>
<td>63.670</td>
<td>52.489</td>
<td>181.566</td>
<td>6.502</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>1.113^2</td>
<td>0.198</td>
<td>0.151</td>
<td>0.175</td>
<td>0.197</td>
<td>7.512</td>
<td>2.410^5</td>
</tr>
<tr>
<td>1.0</td>
<td>LAD</td>
<td>4.310^{10}</td>
<td>5.210^2</td>
<td>2.261</td>
<td>0.631</td>
<td>1.210^2</td>
<td>4.710^7</td>
<td>2.610^6</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>2.110^9</td>
<td>6.810^6</td>
<td>1.10^8</td>
<td>1.1010^8</td>
<td>7.010^{10}</td>
<td>4.310^6</td>
<td>1.410^4</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>3.210^{11}</td>
<td>61.087</td>
<td>0.619</td>
<td>0.521</td>
<td>15.679</td>
<td>2.110^2</td>
<td>4.810^{11}</td>
</tr>
</tbody>
</table>

Table 2: Observed MSE when the disturbance Terms in the Heteroscedasticity Process are stable with Index $\lambda$, n=100

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th></th>
<th>-0.6</th>
<th>-0.5</th>
<th>-0.1</th>
<th>0</th>
<th>0.1</th>
<th>0.5</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n=100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>LAD</td>
<td>0.245</td>
<td>0.071</td>
<td>0.042</td>
<td>0.082</td>
<td>0.212</td>
<td>0.311</td>
<td>0.613</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>0.042</td>
<td>0.052</td>
<td>0.047</td>
<td>0.067</td>
<td>0.071</td>
<td>0.318</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>0.368</td>
<td>0.079</td>
<td>0.091</td>
<td>0.247</td>
<td>0.433</td>
<td>0.465</td>
<td>1.458</td>
</tr>
<tr>
<td>Lamda</td>
<td>LAD</td>
<td>0.235</td>
<td>0.276</td>
<td>0.084</td>
<td>0.264</td>
<td>0.517</td>
<td>9.683</td>
<td></td>
</tr>
<tr>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td>-------</td>
<td></td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>0.459</td>
<td>0.734</td>
<td>0.256</td>
<td>0.635</td>
<td>0.538</td>
<td>0.783</td>
<td>8.973</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>1.563</td>
<td>0.346</td>
<td>0.879</td>
<td>0.739</td>
<td>0.169</td>
<td>0.378</td>
<td>4.952</td>
</tr>
<tr>
<td>2.0</td>
<td>LAD</td>
<td>1.2.10^2</td>
<td>0.176</td>
<td>0.199</td>
<td>0.340</td>
<td>0.738</td>
<td>12.614</td>
<td>10.752</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>430.32</td>
<td>72.302</td>
<td>15.352</td>
<td>73.674</td>
<td>32.423</td>
<td>101.426</td>
<td>3.518</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>1.1.14^1</td>
<td>0.121</td>
<td>0.169</td>
<td>0.235</td>
<td>0.596</td>
<td>6.352</td>
<td>1.3.10^3</td>
</tr>
<tr>
<td>1.5</td>
<td>LAD</td>
<td>3.2.10^9</td>
<td>4.1.10^2</td>
<td>0.245</td>
<td>0.933</td>
<td>1.1.10^4</td>
<td>4.3.10^5</td>
<td>2.2.10^5</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>1.1.10^8</td>
<td>6.2.10^6</td>
<td>1.1.10^5</td>
<td>1.1.10^8</td>
<td>4.0.10^3</td>
<td>2.3.10^7</td>
<td>1.3.10^8</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>3.1.10^10</td>
<td>81.687</td>
<td>0.609</td>
<td>0.215</td>
<td>14.779</td>
<td>2.1.10^4</td>
<td>2.8.10^10</td>
</tr>
<tr>
<td>1.0</td>
<td>LAD</td>
<td>3.2.10^9</td>
<td>4.1.10^2</td>
<td>0.245</td>
<td>0.933</td>
<td>1.1.10^4</td>
<td>4.3.10^5</td>
<td>2.2.10^5</td>
</tr>
<tr>
<td></td>
<td>NLAD</td>
<td>1.1.10^8</td>
<td>6.2.10^6</td>
<td>1.1.10^5</td>
<td>1.1.10^8</td>
<td>4.0.10^3</td>
<td>2.3.10^7</td>
<td>1.3.10^8</td>
</tr>
<tr>
<td></td>
<td>HSD</td>
<td>3.1.10^10</td>
<td>81.687</td>
<td>0.609</td>
<td>0.215</td>
<td>14.779</td>
<td>2.1.10^4</td>
<td>2.8.10^10</td>
</tr>
</tbody>
</table>

Table 3: LAD and NLAD Versus hetrosedascity change in $\lambda$, $\beta$ Number of Replication that an LAD of $\beta_1$.

<table>
<thead>
<tr>
<th>Lamda</th>
<th>$\lambda$</th>
<th>$\beta$</th>
<th>N</th>
<th>-0.6</th>
<th>-0.5</th>
<th>-0.1</th>
<th>0</th>
<th>0.1</th>
<th>0.5</th>
<th>0.6</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>LAD VS HETROSCEDASCITY</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.5</td>
<td>10</td>
<td>31c</td>
<td>61c</td>
<td>60c</td>
<td>73a</td>
<td>77a</td>
<td>76a</td>
<td>79</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>30c</td>
<td>56c</td>
<td>68c</td>
<td>81</td>
<td>70b</td>
<td>80</td>
<td>92</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.0</td>
<td>10</td>
<td>28c</td>
<td>80c</td>
<td>121c</td>
<td>113a</td>
<td>177a</td>
<td>103a</td>
<td>97</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>34c</td>
<td>78c</td>
<td>171</td>
<td>76</td>
<td>89a</td>
<td>56</td>
<td>91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.5</td>
<td>10</td>
<td>37c</td>
<td>80b</td>
<td>127c</td>
<td>143a</td>
<td>167a</td>
<td>113a</td>
<td>70</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>54c</td>
<td>78c</td>
<td>171</td>
<td>87</td>
<td>54a</td>
<td>87b</td>
<td>102</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
It is observed that from table (1) heteroscedasticity process using linear and nonlinear case increase and gives better estimate when the parameter values $\lambda = 2.5, 2.0, 1.5$ and $\beta$ takes the values $\pm 0.62$ from table 2 observed that the estimated value closer to the true value of the parameter when compared with linear and nonlinear with heteroscedasticity error. It’s cleared that the estimated values performs well for both linear and nonlinear cases when compared with LAD estimation based on heteroscedasticity error.

**Conclusion**

It is evidence that in inferential aspects on nonlinear regression models with heteroscedasticity errors gives better estimates. When compared with OLS, and overcome partially some of the short comings the problem of heteroscedasticity in nonlinear regression model Iterative Procedure gives better results when compared with LAD regression methods.

**Reference**

Least Absolute Deviation Estimation versus Nonlinear


