ISSN 0971-0388

Aligarh Journal of Statistics
Vol. 43(2023) 63-86

Comparative Study on Robust Estimators and Evaluating their Performance in Multiple Regression

Lakshmi R. and Sajesh T. A.
[Received on January, 2022. Accepted on March, 2023]

ABSTRACT

Regression analysis is a statistical technique to model data. But the presence of outliers and influential points affects data modelling and its interpretation. Robust regression analysis is an alternative choice to this. Here we make an attempt to study six different robust estimators and their performance on multiple linear regression data. Using Monte Carlo simulation, data is generated and modelled. R software is used for simulation and study. If the fundamental assumptions are true, robust approaches operate as effectively as the OLS estimator. When outliers and leverage points are present, OLS estimators completely fail to work efficiently. Thus, robust estimators are better than OLS estimator. Among the robust estimators, the MM estimator is the best method to rely on and outperform in all situations.

1. Introduction

One of the most essential statistical methods in data modeling is regression analysis. It aids in the prediction of a relationship between the response variable and the predictors. It is often applied in all fields of study including, social science, health science, engineering, physical science, and more. Regression Analysis mainly relies on the ordinary least squares method, which is very vulnerable in the presence of outliers. Informally outlier can be defined as those observations which lie out of place with respect to other observations in the data set. Thus, Robust Regression developed as an upgraded and efficient version of least squares in the presence of contaminated points in the data set. There are numerous robust regression techniques; among them, some are resistant too. The

 : Lakshmi R.
Email: lakshmi.nss19@gmail.com

Extended authors information available after reference list of the article.

most common robust regression method is Huber M estimator. MM estimator developed by Yohai (1987), GM-estimator, Siegal's Repeated Median estimator,

Least Median Square estimator, Least Trimmed Square estimator, S-estimator (Rousseeuw and Yohai 1984), Minimum volume ellipsoid estimator (Rousseeuw and Leroy 1987). Schumacker *et al*. (2002), c0mpared estimators such as OLSE, LTSE, and MME using coefficient of determination as the criterion in a study on robust regression. Muthukrishnan and Radha (2011) compared Huber M and redescending M estimator with the ordinary least squares estimator with the help of the coefficient of determination. All these studies made use of a simple linear regression model and concluded in their papers. In this paper, we discuss about some of the mainstream and efficient robust regression techniques for contaminated data in multiple linear regression models. The main inherent idea is to compare the techniques using simulated data sets, and determine which method is better in multiple regression with respect to different situations. Simulation is done and evaluated by using the Monte Carlo technique. Section 2 briefly describes about the Ordinary Least Squares method, the necessity of robust regression, and important robust regression estimators developed over the years. Section 3 of this paper describes the Monte Carlo simulation scenario. Section 4 provides the comparison and summary of the results obtained in section 3, and Section 5 provides the conclusion of the study.

2. Ordinary Least Squares

Linear regression model is about estimating the parameter $\beta \epsilon R^p$ where:

$$y_i = x_i\beta + \varepsilon_i, \quad i = 1,2,3.\ldots n, \tag{2.1}$$

where $(x_i, y_i)\ \epsilon\ (R^p, R)$ comprise the data and β is the p- dimensional unknown vector and ε_i are unknown errors. The best-known estimator of β is the least square estimators obtained by:

$$min \sum_{i=1}^{n} (y_i - x_i\beta)^2. \tag{2.2}$$

The least square estimators are very popular because of Gauss Markov theorem and very easy to use. These classical least estimators are the best when their assumptions are met by the data. Ordinary least square estimators assume that residuals are normally distributed and that the independent variables have equal variances at all levels. The central limit theorem provides the foundation for the normality of the error distributions, according to Huber (1972), Hampel (2001). The assumption of normality is violated whenever there are outliers in the data,

resulting in unstable estimate prediction by **OLS**. Consequently, if there are outliers in the data, classical estimators are familiar for misbehaving.
The data may contain outliers for a number of reasons, including incorrect data entry, incorrect scoring, and unusual sample data. In regression, outliers can be classified according to their location and effect. Observations would be unusual with respect to *y* values or *x* values. They are categorized as outliers, leverages and influential points based on how they affect the model. The impact of these observations depends on the location where they occur. Outliers in the response variable are observations having significant standardized residuals. They lie far from the best-fit line in the *y* direction. Outliers are often defined as points having standardized residuals that are more than three standard deviations from the mean.
Outliers can also occur in the predictor variables. They might affect the regression results. Extreme values in the predicted variables are called as leverages. Leverages measure how far an independent variable deviates from its mean (Rousseeuw, 1984). The direction of the distance between the remaining data points is not taken into account by leverages. Leverages do not affect the estimates of the regression coefficients. It affects the model summary statistics, standard errors of regression coefficient etc.

Influential points are those data points that unduly influence the regression analysis or a point whose exclusion cause major changes in the fitting of a regression equation. Influential points are those points with unusual *x* coordinate and the unusual *y* value. The regression coefficients are noticeably affected by influential points. Influential points pull the regression model in its direction. A regression model has to be representative of all the sample observations. It is always necessary to assess these points impact on the model. If these points are bad values, then they should be avoided from the sample.

Outliers in either the *x* or *y* directions constitute a significant hazard to least square estimators. Statistical or graphical methods can be used to identify outliers. Mahalanobis distance is a statistical procedure used to locate the outliers in the *x* direction. We cannot say Mahalanobis distance as a perfect method, as it fails to detect the outliers in *y* direction. Other statistical outlier diagnostics works on the idea of erasing one observation at a time and recalculates the regression coefficients; they are called as regression diagnostics, in which diagnostics quantities are obtained using the data with aim of identifying influential points. Following the identification, they are either eliminated or corrected, and then the least squares analysis is performed. As a result, such statistics estimate the change in regression coefficients that would occur if a single observation were

removed following analysis. These statistics are sometimes known as deletion statistics, useful for pinpointing influential points. Cook distance, Studentised residuals, DFFITS, DFBETAS and Jacknife residuals are some of such deletion statistics. Calculation of these diagnostic statistics become complicated when there are multiple unusual observations.

Robust regression estimation is alternative strategy for handling outliers. Robust methods aim to create estimators that are immune to outliers. Diagnostic tools remove outliers before fitting the data using the least square approach, which is the main difference between them and robust regression estimates. Robust regression, on the other hand, fits a regression model to the great majority of the data before identifying outliers as regions with substantial residuals from that reliable response.

The breakdown point, concept of bounded influence and relative efficiency are ideas that are pertinent to the study of robust regression.

The presence of single outlier can completely invalidate the OLS estimator. Contrast with; we will see estimators that can handle certain percentage of outliers. This particular concept is called as breakdown point. Hodges (1967) provided the first explanation of a breakdown point, and it'd only evaluate location in a single dimension. Hampel (1971) provided broad description of it, but it was highly mathematical in nature and asymptotic. Donoho and Huber (1983) suggested a limited sample version of breakdown point. For a sample Z of n observations,

$$Z = \{(x_{11},\ldots\ldots x_{1p}, y_1),\ldots\ldots,(x_{n1},\ldots,x_{np}, y_n)\}. \quad (2.3)$$

Let T represents a regression estimator. When T is applied to such a sample, the result is a regression coefficient vector as $T(Z) = \hat{\beta}$. Let j of the data sample data points swapped by arbitrary values and call corrupted sample as Z'. The maximal bias generated by such contamination is then calculated as:

$$bias(j; T, Z) = \sup_{Z'} \|T(Z') - T(Z)\|, \quad (2.4)$$

where the supremum is over all possible Z'. If the bias is infinite, j outliers have a significant impact on the estimator. Thus, breakdown of the estimator T at the sample Z is defined as

$$\varepsilon_n^*(T, Z) = min\{\tfrac{j}{n}; (bias(j; T, Z) is\ infinite)\}. \quad (2.5)$$

Or the least amount of contamination that an estimator can tolerate is known as the breakdown point. The breakdown point of ordinary least estimator is $\varepsilon_n^*(T,Z) = \frac{1}{n}$. That is, even the presence of single outlier in the data set can affect least square estimators.

When **OLS** has normally distributed error, the degree to which a robust regression estimator performs like least squares is the measure of the estimator's relative efficiency. Efficiency is commonly stated as a percentage, with a range of 0 to 1. When the mean square error of robust method is less, it can even reach one. The relative efficiency is calculated as the ratio of the robust method MSE to the **OLS** mean square error.

The concept of a bounded impact function refers to the regression estimator's capacity to limit a particular residuals weighting, hence minimizing the impact of the data points on the regression estimator. Leverages have a significant impact on regression coefficient estimates. As a result, having an estimator with a bounded influence function is always preferable. (Birkes and Dodge, 1993).

In order to deal with the barriers due to least square estimators, robust regression estimators were developed over different periods. Robust regression analysis has been investigated since the eighteenth century, according to Maronna *et al.* (2006). Numerous studies have been conducted, and numerous notable publications have been produced by Huber, Staudte and Sheather (2011), Rousseeuw and Leroy (1987) and Hampel, Ronchetti, Rousseeuw and Stahel (1986).

Least Absolute Value Regression

Least square is simple to compute, but this estimator is criticized for its lack of robustness. Even the presence of one outlier largely affects the least square estimate. In 1757, Boscovich introduced the method of least absolute deviation. Adrien Marie Legendre in France discovered the method of least squares around 1805 (Birkes and Dodge, 1993). Edgeworth in 1887 improved Boscovich idea and develops the Least Absolute Deviation as it is now. Least Absolute Value regression which is also called as L_1 estimator. Least absolute value regression is a type of **L**-estimators, the sum of the absolute values of the residuals is minimized to find this value.

$$min \sum_{i=1}^{n} |e_i| = min \sum_{i=1}^{n} |y_i - \sum x_{ij}\beta_j|. \qquad (2.6)$$

The proposed idea is straight forward, but it is not as easy as least squares to calculate. Thus, it was not much recognized in history for a long time. Charnes, Cooper and Ferguson (1955) simplified the **LAD** method to a linear programming problem and advent of computer technology reduced computational difficulty in it. Portnoy and Koenker (1997) have provided a comprehensive summary on Least Absolute Deviation. Large sample properties of the **LAD** estimates are obtained in Bassett and Koenker (1978) and Pollard (1991). Suppose in the model (2.1) the error terms are independently and identically distributed with a common distribution function F, and satisfy the following two conditions:

i) There exist a $\Delta > 0$, such that $f(u) = F'(u)$ where $|u| \leq \Delta$, f is continuous at 0, $f(0) > 0$ and $F(0) = ½$.

ii) S_n is non singular for some *n*, and

$\lim_{n\to\infty} \max_{1\leq i\leq n} x_i' S_n^{-1} x_i = 0.$ (2.7)

Then the above conditions guarantee the asymptotic normality of least squares estimate of the parameter. The **LAD** estimation technique has grown in popularity as a result of these theoretical and computational advances. **LAV** is very resistant to the presence of outliers. **LAV** is less affected than **OLS** in the presence of unusual *Y* values, but it is unable to determine leverage values (Mosteller and Tukey 1977). Thus, its breakdown point is not better than $\frac{1}{n}$. Also, they have a low efficiency. Combination of less efficiency and least breakdown point makes it less attractive than other estimators.

M estimator

The next step in this direction was the **M** estimator. Huber introduces the class of **M** estimators. They are straight forward extension of **M** estimators for location. **M** estimator T_n is defined as a solution of the minimization problem

$min \sum_{i=1}^{n} \rho\left(y_i - \sum x_{ij}\beta_j\right) = min \sum_{i=1}^{n} (e_i)^2$ (2.8)

Or is a root of

$\sum_{i=1}^{n} \psi(y_i - \sum x_{ij}\beta_j) x_{ij} = 0$ (2.9)

Where the function ρ is a properly chosen arbitrary function and it gives contribution of each residual to the objective function. $\psi \equiv \partial\rho/\partial\beta$ is called as influence function. The following are the characteristics of a reasonable ρ:

i) ρ is continuous

ii) $\rho(e) \geq 0$, ρ must be strictly positive and integrable

iii) $\rho(0) = 0$

iv) ρ is symmetric, $\rho(e) = \rho(-e)$ and

v) ρ is monotonically increasing function, $\rho(ei) \geq \rho(ei')$ for $|ei| \geq |ei'|$.

Thus, to define **M** estimator it is necessary to specify the function ρ or ψ. Solving the above set of non-linear equations yields the **M** estimate. However, the problem is that the solutions produced are not scale equivariant. Standardizing residuals by means of estimate of σ helps to overcome this, so that system becomes,

$$\sum_{i=1}^{n} \psi((y_i - \sum x_{ij}\beta_j)/\hat{\sigma})x_{ij} = 0. \quad (2.10)$$

As in the case of **M** estimates of location, the median absolute deviation of σ is often used as the estimate. i.e.,

$$\hat{\sigma} = \frac{median|e_i - median(e_i)|}{0.6745}. \quad (2.11)$$

The choice of a good ψ function is based on the choice of how much weight to assign outliers. Large outliers are not as heavily weighted by a monotone ψ function as they are by least squares. Even when the outlying distance rises, a redescending ψ function increases the weight of an assigned outlier up to a certain distance and then drops the weight to zero. An iterative method is necessary to find **M** estimates because residual cannot be found without the model is fitted. Two methods to solve **M** estimates nonlinear normal equations are Newton Raphson and Iteratively Reweighted Least Squares. Susanti and Pratiwi (2014) has discussed about the algorithm steps of **M** estimator.

The steps involved are:

i) For I=0, OLS method is employed to obtain the initial estimates of regression coefficients, $\hat{\beta}^{(0)}$.

ii) Based on the obtained initial estimates residuals, $e_i^{(0)}$,are calculated and is used to calculate initial weight estimates.

iii) To obtain initial weights a weight function is chosen and applied to the initial OLS estimate of residuals, $w(e_i^{(0)})$.

iv) For I=1 use weighted least squares to obtain $\hat{\beta}^{(1)}$. i.e.,

$$\hat{\beta}^{(1)} = \left(X^T W X\right)^{-1} X^T W y. \quad (2.12)$$

v) The process is continued by calculating new weights using the residuals obtained from the initial weighted least squares. These new weights are used in the next iteration step to obtain the estimate $\hat{\beta}^{(2)}$.

vi) Step 4 and 5 are repeated until the estimate $\hat{\beta}$ converges.

It is easy to study the asymptotic properties of **M** estimators when $\psi(x, \beta)$ is monotone in β. Huber provided complete proof on asymptotic distribution of **M** estimators. i.e., M estimators have asymptotic variance and are normally

distributed asymptotically. (Hampel *et al*., 1986, p. 103), (Shevlyakov *et al*., 2008):

$$V(\psi(\varepsilon), f(\varepsilon)) = \frac{\int_{-\infty}^{\infty} \psi^2(\varepsilon) f(\varepsilon) d\varepsilon}{[\int_{-\infty}^{\infty} \psi'(\varepsilon) f(\varepsilon) d\varepsilon]^2}, \tag{2.13}$$

where f is the density of true error. Also, **M** estimators are asymptotically consistent. Hampel (1968) introduced the breakdown point (BP) concept and is another measure of robustness. The **BP** of an estimator is the proportion of spurious values in the data that the estimator can tolerate and still works properly. The more breakdown, more robust the estimator is. The **BP** of **M** estimator does not depend on its probability density. **M** estimators are highly resistant against *Y* outliers with breakdown point of 0.5 as they are robust against non-constant error variance and heavy tailed error distribution. **M** estimator is the simplest method and is not robust against the leverage points. The method is extensively used in analyzing data. **M** estimates are more efficient than **OLS** estimates. The major drawback of these iterative methods is we can never be sure a root exists, until we find it.

Bickel (1975) suggest an alternative to **M** estimator referred to as one step **M** estimators. In this method, to estimate the parameter with the presence of nuisance parameter, first choose initial parameter for both estimates. Then nuisance parameter would be considered fixed and equal to its initial estimate. Suppose we have an initial estimator $\hat{\beta}_0$ of β in (2.1). Let

$$r = y - X\hat{\beta}_0, \tag{2.14}$$

denote the residuals from $\hat{\beta}_0$. Let $\hat{\sigma}$ be the standard deviation of r. Then a one step **M** estimator of β is written in the form:

$$\hat{\beta} = \hat{\beta}_0 + \widehat{H}_0^{-1} \hat{g}_0, \tag{2.15}$$

where $\hat{g}_0 = \sum_{i=1}^{n} x_i w_i \psi(v_i r_i / \hat{\sigma})$ for some odd function ψ, $w_i = w(x_i)$ and $v_i = v(x_i)$ for weight functions w, v, and an appropriate matrix $\widehat{H}_0$. One step Huber **M** estimator has $w_i = v_i = 1$. One step Mallow's estimator has $v_i = 1$. One step Andrew's estimator has $w_i = 1$, one step Hill and Ryan estimator has $v_i = w_i$.The common choices for the matrix $\widehat{H}_0$ are the Newton Raphson form or iteratively reweighted least squares form. The major advantage of one step **M** estimator is easy to compute and they have asymptotic properties.

Redescending **M** estimators are famous ψ type **M** estimators, in which ψ functions are non-decreasing at the origin, but decreases to zero in the region far

from the origin. According to Holland and Welsch (1977), redescending estimators can be characterized as soft-redescending (pseudo-convex) and hard-redescending (quasi-convex), depending whether the corresponding influence functions are nearly null and exactly null for polluted values of high magnitude. Hard redescending estimators influence function, i.e., ψ function is usually built with discontinuous segments, such as Hampel's three-part estimator (1968).**M** estimators can classify based on the mathematical features of the influence function as: Non robust, quasi robust, robust monotonous, robust soft redescending estimators and robust hard redescending estimators. D.Q.F. de Menezes *et al*., (2021) has reviewed 48 different types of **M** estimators in his paper.

GM estimators

Traditional **M** estimators are not "qualitatively robust" (Hampel 1971 and 1974). This is because they exhibit a large asymptotic bias when the joint distribution of (x_i, y_i) follows the model (2.1) only approximately. This is as a result of the fact that they have unbounded influence function. Also, **M** estimators are vulnerable to the leverage points. To overcome this generalized **M** estimator were introduced. The influence of extreme x_i values were bounded by weight functions. It was Mallows (1975) proposed $\sum_{i=1}^{n} w(x_i)\psi(r_i/\hat{\sigma})x_i = 0$. In this method outliers are handled using the usual **M** estimator and leverage points are down weighted by using appropriate weight function.

The general class of **GM** estimators is

$$\sum_{i=1}^{n} w_i(x_i)\psi\left\{\frac{e_i}{v(x_i)\hat{\sigma}_e}\right\}x_i=0, \qquad (2.16)$$

where ψ the score function generally used is Huber or biweight function. This type of **GM** estimator is called as Schweppe type. The model matrix $\boldsymbol{X}$ determines the weights w_i and v_i initially. The initial estimate is **OLS** and the scale estimate is found by scaling the median of the absolute value of the **OLS** residuals. Final estimates are obtained by iterative procedure. Mallow's proposed another **GM** estimator. Mallow's method down weight both tiny and large residuals which is the main difference between the two estimators. Schweppe's down weight minor residuals only. The current generation of **GM** estimators has a breakdown point of at most 1/p+1 where p is the dimension of x_i. Various **GM** estimators developed over different periods, more frequently discussed are Krasker (1980), Krasker and Welsch (1982) and Marazzi (1993).

After the development of robust **M** estimators and their generalized version, a question raised whether high breakdown point estimators could develop, as an answer to the question, repeated median estimator proposed by Siegal, with a

50% breakdown point. Repeated median estimator is a variation of Theil Sen estimator. Theil (1950) advocated using the median of pairwise slopes as a slope estimator. Sen (1968) improved this estimator by adding the ability to handle ties. TSE is a robust estimator with a breakdown point of 29.3% and a bounded influence function. In addition, they have a high asymptotic efficiency. Theil Sen estimator is only formulated for simple linear model. Many authors like Oja and Niinima (1984), Zhou and Serfling (2008) have made their efforts to extend **TSE** to multiple regression model. However, it is technically difficult and causes the analysis of the properties to be delayed. Thus, Siegal in 1982 proposed highly efficient repeated median estimator and is defined as:
For any p observations $(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \ldots . (x_{ip}, y_{ip})$ the objective is to find the parameter that fits these points exactly. The j[th] coordinate of parameter vector is represented by $\beta_j(i_1, i_2, \ldots, i_p)$. Then the estimator is described as:

$$\hat{\beta}_j = \text{med}_{i_1}(\ldots.(\text{med}_{i_{p-1}}(\text{med}_{i_p}\beta_j(i_1, \ldots\ldots, i_p))).\ldots\ldots). \quad (2.17)$$

The estimator requires the consideration of complete subset of *p* observations and requires a lot of time to compute. It is easy to apply in problems with small *p*. For linear *x*-value transformations, the approach is not equivariant. When the contamination in the data is considerable, the repeated median estimator fails to distinguish between good and bad points of the data, despite having a breakdown point of 50%. Following the failure of repeated median estimators, Rousseeuw develop two high breakdown simple estimators Least Median Square estimator and Least Trimmed Square estimator.

Least Median Square Estimator

Rousseeuw (1984) develop Least Median of Squares, an **L** estimator which is calculated by replacing the objective squared residuals sum in **OLS** by median of the squared residuals. Rousseeuw's this proposal was based on the proposal of Hampel (1975, p.380). The estimates are found by

$$min\ Med\left(y_i - \sum x_{ij}\beta_j\right)^2 = min\ \text{Med}\left(e_i^2\right), \quad (2.18)$$

where Med denotes the median. **LMS** always has a unique solution. Rousseeuw (1984) in his paper has stated that, if p>1, and the observation are in general position, then the breakdown point of **LMS** method is ([n/2]-p+2). The main concept of **LMS** is to reduce the dispersion of the residuals. It has a breakdown point of 0.5. Even though **LMS** has a high breakdown point, it has low efficiency

because of its slow convergence rate. Rousseeuw and Croux (1993) in their paper have shown that **LMS** has a relative efficiency of 37%. Rousseeuw (1984) in his paper itself proposed methods to overcome the slow convergence rate of **LMS** estimator by one step **M** estimators in which initial estimates are obtained by **LMS** method or by making use of an objective function other than that of **LMS**. Thus, least trimmed square estimators developed as an alternative to **LMS** estimator.

Least Trimmed Square Estimator

Least Trimmed Square is yet another *L* estimator method developed by Rousseeuw (1985). The method is extended from trimmed mean. The method minimizes the sum of the trimmed squared residuals. i.e.,

$$min \sum_{i=1}^{q} e_{(i)}^2, \qquad (2.19)$$

where q have to satisfy $\frac{n}{2} \leq q \leq n$. The constant q determines the breakdown point of the **LTS** estimator. **LTS** estimator definition implies that *n-q* observations with largest residual will not affect the estimator. The minimization of the objective function (2.19) implies, choose a subsample of size q observations and find β minimizing the sum of squared residuals of the subsample. Repeating this for every subsample, we get nC_h candidates for **LTS** estimate and among them one that gives the smallest value of the objective function is the final estimate. **LTS** regression consists of finding β belongs to R^p such that sum of q smallest squared residuals is minimum. As mentioned above the value of the constant q determines the breakdown point of the estimator, the choice depends mainly on the purpose for which we use **LTS** estimator. The breakdown point of **LTS** reach the upper bound $([(n-p)/2]+1)/n$ for regression equivariant estimators if the trimming constant $q = [n/2] + [(p+1)/2]$. Setting $q = (n/2) + 1$ ensures the estimator a breakdown of 0.5. It is always better to evaluate **LTS** estimator for a wide range of trimming constant values and compare how the estimator change with increasing value of the constant. Such an analysis provides us an insight on amount of contamination and structure of the data. **LTS** regression is scale and affine equivariant. Visek (1999b) in his paper has shown the asymptotic normality and $\sqrt{n}$ consistency of **LTS** estimator in a general linear regression model with continuously distributed disturbances. One of the drawbacks of **LTS** estimator is non-continuity of **LTS** objective function. Due to this **LTS** estimator sometime might possess highly sensitive to a change of one or several observations. LTS efficiency varies depending on the trimming constant value and the outliers in the data. Stromberg,

Hossjer, and Hawkins (2000) in their paper made clear that **LTS** has a low efficiency of 8%. Its low efficiency makes it non desirable estimator. Though the method has low efficiency, the method is used in other methods as initial estimates. For example, **LTS** is used as the initial estimate in **GM** estimator proposed by Coakley and Hettmansperger (1993).

S Estimators

Since **LTS** and **LMS** estimators has slow convergence rate, Rousseeuw and Yohai (1984) introduced high breakdown value method that minimizes the dispersion of the residuals with high asymptotic efficiency and better convergence rate of objective function than **LTS**. **S** estimate is the solution to the smallest possible dispersion of the residuals. **S** estimate is not simply minimizing the variance of the residuals; it minimizes the robust **M** estimate of residual scale. That is **S** estimator is obtained by minimization of the dispersion of the residuals:
Minimise $s(r_1(\beta), r_2(\beta), \ldots . r_n(\beta))$ with respect to β,
With final scale estimate
$\hat{\beta} = s(r_1(\hat{\beta}), r_2(\hat{\beta}), \ldots\ldots, r_n(\hat{\beta}))$. (2.20)
The dispersion $s(r_1(\hat{\beta}), r_2(\hat{\beta}), \ldots\ldots, r_n(\hat{\beta}))$ is obtained as an answer of
$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{r_i}{s}\right) =$, (2.21)
where b is a constant defined as $b = E_{\Phi}[\rho(e)]$ and Φ represents the standard normal distribution. The ρ function should satisfy the conditions:

i) ρ is symmetric function and is continuously differentiable with ρ (0) =0.
ii) for c > 0, ρ is strictly increasing on [0, c] and constant otherwise.

S estimators are regression, scale and affine equivariant. For any ρ function satisfying the above conditions along with the condition $\frac{b}{\rho(c)} = 1/2$ has a breakdown of 50%. The major drawback of **S** estimator is low efficient.

MM Estimators

MM estimator was proposed by Yohai (1987) which is most commonly and widely used regression estimator. The **MM** estimation technique involves first estimating the regression parameter with **S** estimation which minimize the scale of the residual from **M** estimation and then proceed with **M** estimation. **MM** estimation is a combination of high breakdown value estimator with high efficiency of approximately 95% relative to ordinary least squares. The name itself hints that more than one **M** estimation is used to calculate the final estimates. An **MM** estimates procedure is as follows:

i) Initially the regression coefficient estimate and residuals are obtained using a high resistant regression method of 50% breakdown point; **S** estimate is commonly used for this purpose.
ii) Using residuals got from step 1, compute **M** estimate of the scale of the residuals.
iii) In the first iteration of weighted least squares, the initial estimate of the residuals from step 1 and the residual scale estimate from step 2 are utilized to produce the M estimates of the regression coefficients. $\sum_{i=1}^{n} w_i \left(\frac{e_i^{(1)}}{\hat{\sigma}_e}\right) x_i = 0.$
iv) New weights, $w_i^{(2)}$, are calculated based on the residuals obtained from initial weighted least squares in step 3.
v) The above steps are reiterated until convergence.

MM estimators are affine equivariant. They have a breakdown point of 0.5. **MM** estimators attain exact fit property from the initial estimate that is if the initial estimators used have an exact fit property, then **MM** estimator inherits the exact fit property. **MM** estimators are consistent and asymptotic normal too. Mathematical proof of all these properties is provided by Yohai (1987). **MM** estimators need a high breakdown start, for that below discussed estimators developed.

Generalised S Estimators

Generalised **S** estimators are proposed by Croux *et al.* (1994) to overcome the low efficiency of **S** estimators. The proposed estimator is calculated by finding a **GM** estimator of the scale of the residuals. Least quartile difference estimator is a special case of **GS** estimator which is defined as:

$$\hat{\beta} = arg\min_{\beta} s_n(\beta), \tag{2.22}$$

where $s_n(\beta)$ is based on the residual $r_i = y_i - \beta^t x_i$ through the equation

$$\binom{n}{2}^{-1} \sum_{i<j} \rho\left(\frac{r_i - r_j}{s_n(\beta)}\right) = k_{n,p}. \tag{2.23}$$

Croux *et al*. (1994:1271) in his paper stated that although the method has high efficiency, but it possesses "slightly increased worst case bias" which makes this method less acceptable. The fact that the objective function of **GS** estimators is independent of the intercept term is a key feature. For models with an asymmetric error distribution, **GS** estimators are ideal, since the objective function only depends on r_i through $|r_i|$, as a result, positive residuals are given the same weight as negative residuals of the same size. Yohai and Zamar (1988) proposed tau estimator obtained by minimizing estimate for the scale of the residuals.

The weight function used in this method is adaptive depending on the error distribution. The estimator has high breakdown and high efficiency, but it does not consider the leverage points, which question the efficiency of the estimator.

3. Simulation

The simulation study is carried out in three phases. Consider the following regression model.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i, i = 1,2,\ldots,n. \quad (3.1)$$

where ε_i follows standard normal distribution. Let the generated explanatory variables be from multivariate normal with mean vector [5 5] and variance-covariance equal to $\begin{bmatrix} 9 & 0 \\ 0 & 9 \end{bmatrix}$ using R software. The values of the regression coefficient are fixed as $\beta_0 = 0, \beta_1 = 3$, and $\beta_2 = 7$. Then y_i values are observed from the model defined in equation (3.1) with the specified values of $\beta_0 = 0, \beta_1 = 3$, and $\beta_2 = 7$. The process is repeated for samples of different size of n=40, 60 and 100.Thus, we obtain data set of explanatory and observed variables. After generating the data, regression model is fitted in the data and the value of the parameters in the model are computed using various robust methods and classical **OLS** in R software and crosscheck the fitted parameter values obtained with the values specified above.

In the second phase, outliers are introduced in the generated dependent variables in different percentages (ε=10% and 20%). That is, mentioned percentage of y_i observations was replaced by values generated from a normal distribution with mean 100 and variance 25. This is repeated for samples of different size (n=40, 60 and 100). After that, regression model is fitted in the data and the parameter values are computed using different robust methods and the classic **OLS** method also. In the third phase, 20% of generated x_i observations in phase one are replaced by the values generated from multivariate normal with mean [100 100] and variance $\begin{bmatrix} 0.5 & 0 \\ 0 & 1 \end{bmatrix}$. And we have already outlier mixed y_i variables. Based on the data thus generated, a regression model is fitted, and the parameter values are computed using different robust methods and classical **OLS.**

4. Discussion on Simulation Results and Real Life Application

After the simulation, the results are tabulated below (table 4.1 – 4.9). Estimators and results are obtained by using R software. We can see for 10% outlier in the data with sample size 40, the **MM** estimator and **LTS** estimator identifies the parameter value approximately. The parameter value computed by **OLS** is distorted from the original value. As the sample size increased to n=60, **MM**, **S**, and **LTS** methods worked properly and obtained the parameter values approximately near to the original value. Also, the relative efficiency of all the robust methods is greater than one. The coefficient of determination of **MM** and **LTS** estimator is 99%, whereas for the **OLS** method, the coefficient of determination is deteriorating from 57% to 2% as the sample size increases from 40 to 60 due to the presence of outliers. From the tables, it is clear that the **MM** estimator performs outstandingly even when leverages are present in the data. However the percentage of leverage points in data increases, the **MM** estimator determines the coefficients approximately. When we try to find out the mean square error of the models with different percentages of leverages, as in tables 4.6 to 4.9 it grows rapidly like, greater than 300. Comparing such mean square values cannot provide any practical information. Thus we don't tabulate the relative efficiency.

From the tables below having the results of a dataset having outliers and leverages in different proportion, we can say **OLS**, **M**, and **LAD** completely fail to work due to the presence of outliers and leverage points. Even though **M** and **LAD** are robust techniques, they fail to work in the presence of leverage points. Thus, **M** and **LAD** are not resistant to leverage points. The **MM**, **S**, and **LTS** estimators computed parameter values approximately equal to the assigned values for the samples of size 40 and 60. According to a study by Alma (2011), the **S** estimator performed better than the **MM** estimator in simple linear case because the **MM** estimator has problems with high leverages in small sample datasets. However, in this multiple linear study, **MM** estimators perform well against the effects of leverages for a small sample too. In this simulation study, all the methods fail to work for the sample size of 100, having both outliers and leverages in different proportions, which is the future work, to investigate the performance of the estimators in multiple regression models with large contaminated samples.

Table 4.1: With n=40,10% outlier.				
Method	β_0	β_1	β_2	RE
OLS	13.9545	2.9566	5.5559	1.0000
M	0.4897	2.9770	6.9521	1.1591
S	-0.2256	3.0243	7.019	1.1591
MM	0.1629	2.9877	6.9729	1.1591
LTS	0.201	2.988	6.9682	1.1591
LMS	-.2029	3.0246	7.0112	1.1591
LAD	0.3875	2.9907	6.9721	1.1591

Table 4.2: With n=60,10% outlier.				
Method	β_0	β_1	β_2	RE
OLS	30.2761	1.8972	3.2752	1.0000
M	15.3908	2.7224	4.9304	1.0696
S	0.1398	2.9832	6.9660	1.3123
MM	0.1198	3.0306	6.9576	1.3107
LTS	0.1592	2.9899	6.9590	1.3114
LMS	-0.1973	3.0264	7.0054	1.3177
LAD	1.1927	3.0226	6.7755	1.2866

Table 4.3: With n=100,10% outlier.				
Method	β_0	β_1	β_2	RE
OLS	47.4302	0.9685	1.0968	1
M	45.9079	1.0302	1.2779	1.000
S	9.5726	2.3333	6.1621	1.4329
MM	44.523	1.0644	1.427	1.0022
LTS	22.1867	1.2867	4.8457	1.2371
LMS	9.3210	2.1759	7.071	1.6590
LAD	37.9894	1.0739	2.1223	1.0310

Table 4.4: With n=40,20% outlier.				
Method	β_0	β_1	β_2	RE
OLS	17.8577	1.7082	6.5452	1
M	0.6700	2.8986	7.0425	1.2404
S	-0.2097	2.9989	7.0437	1.2676
MM	0.0060	2.9810	7.0185	1.2656
LTS	0.0462	2,9815	7.013	1.2649
LMS	-.1973	3.0264	7.0054	1.2725
LAD	0.3376	2.9286	7.0395	1.2522

Table 4.5: With n=60,20% outlier.				
Method	β_0	β_1	β_2	RE
OLS	17.8577	1.7082	6.5452	1
M	0.6700	2.8986	7.0425	1.2404
S	-0.2097	2.9989	7.0437	1.2676
MM	0.0060	2.9810	7.0185	1.2656
LTS	0.0462	2.9815	7.013	1.2649
LMS	-0.1973	3.0264	7.0054	1.2725
LAD	0.3376	2.9286	7.0395	1.2522

Table 4.6: With n=100, 20% outlier.			
Method	β_0	β_1	β_2
OLS	0.0065	3.0149	6.9866
M	-0.0241	3.0137	6.9880
MM	-0.0142	3.0133	6.9882
S	-0.1294	2.9917	7.0087
LTS	-0.0171	3.0110	6.9896
LMS	-0.0017	2.9698	7.0294
LAD	0.0146	2.9945	7.0046

Table 4.7: With n=40,10% outlier and 20 % leverage			
Method	β_0	β_1	β_2
OLS	50.0905	-0.6984	1.0493
M	48.9407	-0.6730	1.0685
S	-0.1607	2.9892	7.0366
MM	0.05747	2.9786	7.0093
LTS	0.0883	2.9794	7.0050
LMS	-0.4112	3.0378	7.0393
LAD	52.0468	0.1116	0.3638

Table 4.8: With n=60,10% outlier and 20% leverage.			
Method	β_0	β_1	β_2
OLS	48.9147	-0.7815	1.0540
M	48.7177	-0.7391	1.0578
S	0.3039	2.9546	6.9742
MM	0.3453	2.9927	6.9668
LTS	0.3956	2.9499	6.9681

LMS	-0.2108	3.0302	6.9793
LAD	48.9218	-1.0636	1.4896

Table 4.9: With n=100,10% outlier and 20 % leverage.			
Method	β_0	β_1	β_2
OLS	52.6665	-0.6825	0.8871
M	52.7578	-0.7128	0.9327
S	53.4607	0.4349	0.0128
MM	52.5612	-0.6619	0.9094
LTS	52.8266	-0.5049	0.9267
LMS	56.8409	0.9371	-0.5680
LAD	53.3020	0.2748	0.1439

Heart Catherization Data (X Direction Outliers)

Connors *et al.* (1996), Chambers *et al.* (1983) analyzed this data set in their respective study. At the femoral area, a catheter is inserted into a significant vein or artery and advanced toward the heart. The doctor must estimate the correct catheter length before inserting it. The purpose of the data set is to describe how the catheter length (Y) relates to the patient's height (X1) and weight (X2). Twelve observations on three variables are included in this data collection. The data is mainly used to explore the relationship between Y and two independent variables. A Strong correlation is established between the independent variables. Along with the association, four outliers in the X direction are detected in the dataset. Thus, we can use the data set to understand how different estimators perform in such situations. From table 4.10, we can understand OLS fails to perform in the presence of anomalies and give a poor estimate of the data. Only MM estimator performs fairly in this situation and detects the outlier point as (3^{rd}, 8^{th}, 9^{th}, and 10^{th}). Even though other estimators like the M estimator, and S estimator estimate the parameters, they don't have the precision in output as that of the MM estimator because they don't detect the outliers in the X direction.

Table 4.10: Showing the result of Heart data.				
Method	Coefficients			MSE
	Intercept	Coefficient of X_1	Coefficient of X_2	
OLS	20.3758	0.2107	0.1911	14.2754
MM	30.3033	-0.1372	0.3136	17.0764
S	63.4433	-1.2279	0.6886	78.1368
LTS	63.3528	-1.2265	0.6884	77.5385
LMS	13.1598	0.5228	0.0415	15.8338
M	26.3711	-0.0134	0.27577	15.1476

5. Conclusion

From our study, we can say that robust methods work with the same efficiency as that of the **OLS** estimator if the basic assumptions in the data are met. **OLS** estimators completely break down in the presence of outliers and leverage points. **M**, **LAD**, **LTS** and, **S** estimators perform consistently if there are only outliers. Whereas the **MM** estimator resists both outliers and leverages. Thus, the study concludes that robust estimators are better than **OLS** estimators. Among the robust estimators, our study concludes **MM** estimator is the best method to rely on and even **LTS** performs well in the presence of both the outlier and leverage in multiple linear regression models. Also, OLS is based on distributional assumptions, whereas robust methods do not have such constraints. Thus it is always better to use robust methods, especially the **MM** method, which helps to identify those observations which are outliers but behaving as normal or those observations which are swamped. Thus, the study recommends it is always better to use robust **MM** estimators, as they are resistant to the observations which are outliers or leverages. Many statisticians hesitate to use robust methods because of their unawareness and their belief that these methods are complex to calculate. Many statistical packages are available with the function of these robust methods. So, we recommend the use of robust methods.

References

Alma, O.G. (2011): Comparison of robust regression methods in linear regression. *Int. J. Contemp. Math. Sciences*, *6*(9), 409-421.

Almetwally, E. M., and Almongy, H. (2018): Comparison between m estimation, s estimation, and mm estimation methods of robust estimation with application and simulation. *International Journal of Mathematical Archive*, *9*(11), 1-9.

Anderson, C., and Schumacker, R.E. (2003): A comparison of five robust regression methods with ordinary least squares regression: Relative efficiency, bias, and test of the null hypothesis. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *2*(2), 79-103.

Bai, Z. D., Rao, C. R., and Yin, Y. Q. (1990): Least absolute deviations analysis of variance. *Sankhyā: The Indian Journal of Statistics, Series A*, 166-177.

Bassett Jr, G., and Koenker, R. (1978): Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, *73*(363), 618-622.

Bickel, P. J. (1973): On some analogues to linear combinations of order statistics in the linear model. *The Annals of Statistics*, 597-616.

Bickel, P. J. (1975): One-step Huber estimates in the linear model. *Journal of the American Statistical Association*, *70*(350), 428-434.

Birkes, D., and Dodge, Y. (1993): Nonparametric Regression. *Alternative Methods of Regression, John Wiley and Sons, Inc., Hoboken, NJ, USA, 111-141*.

Chen, C. (2002): Paper 265-27 Robust regression and outlier detection with the ROBUSTREG procedure. In *Proceedings of the Proceedings of the Twenty-Seventh Annual SAS Users Group International Conference*.

Chambers, J. M., Cleveland, W. S., Kleiner, B., and Tukey, P. A. (1983): Graphical Methods for Data Analysis. Belmont, CA: Wadsworth .

Charnes, A., Cooper, W.W., and Ferguson, R.O. (1955): Optimal estimation of executive compensation by linear programming. *Management science*, 1(2), 138-151.

Croux, C., Rousseeuw, P.J., and Hössjer, O. (1994): Generalized S-estimators. *Journal of the American Statistical Association*, *89*(428), 1271-1281.

Coakley, C. W., and Hettmansperger, T. P. (1993): A bounded influence, high breakdown, efficient regression estimator. *Journal of the American Statistical Association*, *88*(423), 872-880.

Connors, A.F., Speroff, T., Dawson, N.V., Thomas, C., Harell, F.E., Wagner, D., and Knaus, W.A., (1996): The effectiveness of right heart catherization in the initial care of critically III patients. *Jama*, *276*(11),889-897.

De Menezes, D. Q. F., Prata, D. M., Secchi, A. R., and Pinto, J. C. (2021): A review on robust M-estimators for regression analysis. *Computers & Chemical Engineering*, *147*, 107254.

Donoho, D. L., and Huber, P. J. (1983): The notion of breakdown point. *A festschrift for Erich L. Lehmann*, *157184*.

Hadi, A. S. (1992): Identifying multiple outliers in multivariate data. *Journal of the Royal Statistical Society: Series B (Methodological)*, *54*(3), 761-771.

Hampel, F. R. (1968): *Contributions to the theory of robust estimation*. University of California, Berkeley.

Hampel, F. R. (1971): A general qualitative definition of robustness. *The annals of mathematical statistics*, *42*(6), 1887-1896.

Hampel, F. R. (1974): The influence curve and its role in robust estimation. *Journal of the american statistical association*, *69*(346), 383-393.

Hampel, F.R. (1975): Beyond location parameters: Robust concepts and methods.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., Stahel, W. A., and Statistics, R. (1986): The approach based on influence functions. *Robust Statistics. Wiley*.

Hampel, F. R. (1992): Introduction to Huber (1964) robust estimation of a location parameter. *Breakthroughs in Statistics: Methodology and Distribution*, 479-491.

Hampel, F.R. (2001): Robust Statistics: A brief introduction and overview. In *Research Report/Seminar fur Statistik, Eidgenossische Technische Hoshschule (ETH) (Vol. 94).* Seminar fur Statistik, Eidgenossische Technische Hochschule.

Hodges Jr, J. L. (1967, January): Efficiency in normal samples and tolerance of extreme values for some estimates of location. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, pp. 163-186).

Holland, P.W., and Welsch, R.E. (1977): Robust regression using iteratively reweighted least-squares. *Communications in Statistics-theory and Methods,* 6(9), 813-827.

Huber, P.J. (1965): A robust version of the probability ratio test. *The Annals of Mathematical Statistics* 1753-1758.

Huber, P. J. (1973): Robust regression: asymptotics, conjectures and Monte Carlo. *The annals of statistics*, 799-821.

Huber, P. J. (1972): The 1972 wald lecture robust statistics: A review. *The Annals of Mathematical Statistics*, *43*(4), 1041-1067.

Huber, P. J. (2011): Robust statistics. In *International encyclopedia of statistical science* (pp. 1248-1251). Springer, Berlin, Heidelberg.

Krasker, W. S. (1980): Estimation in linear regression models with disparate data points. *Econometrica: Journal of the Econometric Society*, 1333-1346.

Krasker, W. S., and Welsch, R. E. (1982): Efficient bounded-influence regression estimation. *Journal of the American statistical Association*, *77*(379), 595-604.

Mallows, C. L (1975): On Some Topics in Robustness. Murray Hill, NJ: Bell Telephone Laboratories. unpublished memorandum.

Maronna, R. A., and Yohai, V. J. (1981): Asymptotic behavior of general M-estimates for regression and scale with random carriers. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, *58*, 7-20.

Maronna, R., Bustos, O., and Yohai,V. (2006,November): Bias-and efficiency-robustness of general M-estimators for regression with random carriers. In *Smoothing Techniques for Curve Estimation: Proceedings of a Workshop held in Heidelberg, April 2-4. 1979* (pp. 91-116). Berlin, Heidelberg:Springer Berlin Heidelberg.

Marazzi, A. (1993): Algorithms, Routines, and S-functions for robust Statistics. CRC Press.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2021): *Introduction to linear regression analysis*. John Wiley & Sons.

Mosteller, F., and Tukey, J.W. (1977): Data analysis and regression: a second course in statistics. *Addison-Wesley series in behavioral science: quantitative methods.*

Muthukrishnan, R., and Radha, M. (2011): Edge detection techniques for image segmentation. *International Journal of Computer Science & Information Technology*, *3*(6), 259.

Nurunnabi, A., West, G., and Belton, D. (2015): Robust locally weighted regression techniques for ground surface points filtering in mobile laser scanning three dimensional point cloud data. *IEEE Transactions on Geoscience and Remote Sensing*, *54*(4), 2181-2193.

Oja, H., and Niinimaa, A. (1985): Asymptotic properties of the generalized median in the case of multivariate normality. *Journal of the Royal Statistical Society. Series B (Methodological)*, 372-377.

Pollard, D. (1991): Asymptotics for least absolute deviation regression estimators. *Econometric Theory*, *7*(2), 186-199.

Portnoy, S., and Koenker, R. (1997): The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators. *Statistical Science*, 12(4),279-300.

Rousseeuw, P., and Yohai, V. (1984): Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis: Proceedings of a*

Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983 (pp. 256-272). Springer US.

Rousseeuw, P. J. (1984): Least median of squares regression. *Journal of the American statistical association*, *79*(388), 871-880.

Rousseeuw, P.J., and Leroy, A. (1987): *Robust regression and outlier detection*. New York: Wiley.

Rousseeuw, P.J. (1985): Multivariate estimation with high breakdown point. *Mathematical statistics and applications*, *8*(283-297), 37.

Rousseeuw, P., and Yohai, V. (1984): Robust regression by means of S-estimators. In *Robust and Nonlinear Time Series Analysis: Proceedings of a Workshop Organized by the Sonderforschungsbereich 123 "Stochastische Mathematische Modelle", Heidelberg 1983* (pp. 256-272). Springer US.

Rousseeuw, P.J., and Croux, C. (1993): Alternatives to the median absolute deviation. *Journal of the American Statistical association*, 88(424),1273-1283.

Schumacker, R.E., Mohanana, M.P., and Mount, R.E. (2002): A comparison of OLS and robust regression using S-PLUS. *Multiple Linear Regression Viewpoints*, 28(2),10-13.

Sen, P.K. (1968): Estimates of the regression coefficient based on Kendall's tau. *Journal of the American statistical association*, 63(324),1379-1389.

Shevlyakov, G., Morgenthaler, S., and Shurygin, A. (2008): Redescending M-estimators. *Journal of Statistical Planning and Inference, 138*(10), 2906-2917.

Siegel, A.F. (1982): Robust regression using repeated medians. *Biometrika*, *69*(1), 242-244.

Staudte, R. G., and Sheather, S.J. (2011): *Robust estimation and testing*. Vol. 918. John Wiley & Sons.

Stromberg, A.J., Hossjer, O., and Hawkins, D.M. (2000): The least trimmed differences regression estimator and alternatives. *Journal of the American Statistical Association*, *95*(451), 853-864.

Susanti, Y., Pratiwi, H., Sulistijowati, S., and Liana, T. (2014): M estimation, S estimation, and MM estimation in robust regression. *International Journal of Pure and Applied Mathematics*, *91*(3), 349-360.

Theil, H. (1950): A rank-invariant method of linear and polynomial regression analysis. *Indagationes mathematicae*,12(85), 173.

Yohai, V. J. (1987): High breakdown-point and high efficiency robust estimates for regression. *The Annals of statistics*, 642-656.

Yohai, V.J., and Zamar, R. H. (1988): High breakdown-point estimates of regression by means of the minimization of an efficient scale. *Journal of the American statistical association*, *83*(402), 406-413.

Visek, J. (1999): The least trimmed squares–random carriers. *Bulletin of the Czech econometric society*, *6*.
Yazici, M. (2016, July): The weighted least squares ratio (WLSR) method to M-Estimators. In *2016 SAI Computing Conference (SAI)* (pp. 434-445). IEEE.
Zhou,W. and Serfling, R. (2008): Multivariate spatial U-quantiles: a Bahadur-Kiefer representation, a Theil-Sen estimator for multiple regression, and a robust dispersion estimator. *Journal of Statistical Planning and Inference*, *138*(6),1660-1678.

Authors and Affiliations

Lakshmi R.[1] and Sajesh T. A.[2]

Sajesh T. A.
Email: sajesh.t.abraham@gmail.com

[1,2]Department of Statistics, St. Thomas College (Autonomous), University of Calicut, Malappuram, Thrissur, Kerala, India