**KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY**

# COMPARISON OF ROBUST REGRESSION ESTIMATORS

By

DAVID ADEDIA

(B.Sc. ACTUARIAL SCIENCE)

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF MPHIL MATHEMATICAL STATISTICS

May 28, 2014

# Declaration

I hereby declare that this submission is my own work towards the award of the MPhil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.



| | | |
|---|---|---|
| David Adedia (PG8064112) | ..................... | .................. |
| Student | Signature | Date |

Certified by:

| | | |
|---|---|---|
| Dr. A. O. Adebanji | ..................... | .................. |
| Supervisor | Signature | Date |

Certified by:

| | | |
|---|---|---|
| Prof. S. K Amponsah | ..................... | .................. |
| Head of Department | Signature | Date |

## Dedication

*To my*

*Family and*

*loved ones*

# Abstract

This study evaluated the performance of the Ordinary Least Squares Estimator (OLSE) method of estimating regression parameters and some robust regression methods. The Least-Trimmed Squares Estimator (LTSE), Huber Maximum likelihood Estimator (HME), S-Estimator (SE) and Modified Maximum likelihood Estimator (MME) were considered in this study. Criteria for the comparison were: coefficients and their standard errors, relative efficiencies, Root Mean Square Errors, coefficients of determination and the power of the test. The sensitivity of these robust methods were considered using Anthropometric data from Komfo Anokye Teaching Hospital. The dataset was on Total Body fat and Body Mass Index, Triceps skin-fold, Arm Fat as percent composition of the body and Height as predictors. Leverages were introduced first into two variables, and into all predictors. The percentages were 5%, 10% and 15 % leverages. Also, 10%, 20% and 30% outliers were introduced in addition to 20% error contamination and contamination with data from non-normal distribution were considered. Results showed that robust methods are as efficient as the OLSE if the assumptions of OLSE are met. OLSE was affected by leverages, outliers, contaminants and non-normality. HME broke-down with leverages in data, and was slightly affected by outliers, contaminants and non-normality; whilst SE and MME were robust to all aberrations. LTSE was affected by contaminants, non-normality, high outliers perturbation and was slightly affected by leverages and low outliers perturbation.

# Acknowledgement

First and foremost, my most gratitude goes to the Most High God, for His mercy and love that he shown to me during the period of writing this thesis. He has made everything simple and endowed me with understanding in writing this work. Secondly, my sincere gratitude goes to my supervisor, Dr. A. O. Adebanji for her guidance and constructive criticisms. Finally, I want to say thank you to all my course mates and friends for being there for me during this period. I really appreciate every help.

# Contents

# List of Tables

# List of Figures

# Chapter 1

# Introduction

A linear regression analysis is an important statistical tool that helps to model a relationship between response variable and the predictors, and it is often applied in all fields of study. Applying Ordinary Least Squares (OLSE) in simple or multiple linear regression always calls for some assumptions: normality of the error terms; equal variance of the error terms; and absence of outliers, leverage points and multicollinearity. According to Hampel (2001) and Huber (1972), normality of the error distributions finds its basis from the central limit theorem; which is a limit theorem based on approximations. Some ground-breakers of statistics including Hampel and Huber defeated the belief that the OLSE would still be approximately optimal under approximate normal, a belief that many statisticians stick up for. This is because, it had been shown that typical error distributions of high-quality datasets usually deviate slightly from the normality and in most cases clearly longer-tailed. Additionally, outliers in the dependent variable, lead to large residual values which further results in the failure of the normality assumption of the error terms. Therefore, in regression analysis, the ordinary Least Squares estimation is the best method if the assumptions are met. However, if these assumptions are not satisfied, the results can easily be affected, Alma (2011).

According to Tiku and Akkaya (2004), an estimator is considered robust if the estimator is efficient for normal errors and do not breakdown completely for non-normal errors. And it should be robust to outliers, inliers, leverage points, as well as contaminations. Also robust methods are used in heteroscesdastic models, where variances depend on the independent variables (X), and datasets full of outliers. Robust statistics have been in use for hundreds of years but not seriously studied by mathematical statisticians until quite recently (Stigler, 1973). We only note that while statisticians have been long aware of the sensitivity of some statistics to slight changes in the basic assumptions, it is only recently that they have had the tools to describe these problems mathematically. As a result, many cook book statistical tools now have their robust counterparts.

## 1.1 Background of the Study

Linear regression analysis as a statistical tool has been in use for centuries to establish a linear relationship between variables. And it is applicable in all fields of study, including; social science, health science, engineering, physical science and many more. For example, banks use regression analysis to determine profit and as a result, they know variables that positively and negatively affect their profits. Medical doctors also use regression to determine total fat in the body of their patients by considering variables that can affect or cause the increase or decrease in the amount of fats in the body. Regression analysis is also applied by statisticians in the hospitals to check types of life styles that could be causal factors for certain diseases; for example, blood pressure. Therefore, regression analysis is an important statistical tool. In this study, we look at the various methods of estimating regression parameters that are resistant to small deviations and also being distributional robust. According to Maronna et al. (2006), robust methods of regression analysis have been under study as far back as in the nineteenth century. And they went on to explain that even though much knowledge about robust estimation was realized in the nineteenth century, the first great steps forward occurred in the 1960s, and the early 1970s with the fundamental work of John Tukey (1960, 1962), Peter Huber (1964, 1967) and Frank Hampel (1971, 1974). Many studies have been carried out and many influential books have been written by Huber (1981), Hampel, Ronchetti, Rousseeuw and Stahel (1986), Rousseeuw and Leroy (1987) and Staudte and Sheather (1990).

## 1.2 Problem Statement

In regression analysis, the application of ordinary least squares method works well if the assumptions of the regression model, variables and the error terms are met. However, in the presence of outliers and leverage points, or failure of the assumptions renders the ordinary least squares method of estimation unreliable. This is because bad leverage points, vertical outliers and good leverage points can influence the coefficients in the model, the residuals, as well as the standard errors of the model and the coefficients.

As a result, this study seeks to evaluate the performances of some robust methods that counteract the influences of the drawbacks in a dataset.

## 1.3 Objectives of the study

The objective of this study is to compare robust regression estimators to the ordinary least squares estimator, checking how resistant they are to aberrations in a dataset. The various estimators will be compared for different datasets to check their consistency in resisting the drawbacks in the datasets.

### 1.3.1 Specific objectives

1. To determine and compare the Ordinary least squares estimator to the robust estimators using their coefficients and standard errors, the Root Mean Square Error (RMSE), relative efficiencies, coefficients of determination and the Power of the test.

2. Also, to be able to compare these estimators using four different datasets. The datasets to be considered are:

   (a) Dataset with normally distributed errors

   (b) Dataset contaminated with outliers and leverages

(c) An error contaminated dataset

(d) Data from non-normal distribution

3. Finally, to be able to show the estimators that perform well generally among the estimators: Least Trimmed Squares Estimator (LTSE), Huber Estimator (HME), S-Estimator (SE), Modified Maximum likelihood Estimator (MME) and Ordinary Least Square Estimator (OLSE); in estimating regression parameters.
   The overall performances will be scrutinized based on the following criteria:
   (a) The coefficients and their standard errors
   (b) The Root Mean Square Error (RMSE) of the estimators
   (c) Their relative efficiencies
   (d) The coefficients of determination (R-Square)
   (e) The Power of the test

## 1.4 Methodology

This study looked at the effects of outliers, leverage points, non-normality, and contaminations on classical least squares estimation in linear regression analysis. Robust methods: MM-estimator, Huber M-estimator, Least trimmed squares estimator, and the S-estimator were compared with the ordinary least squares estimator. And the performances of the robust estimators were examined based on standard error, relative efficiency, regression coefficients and the coefficient of determination.

A secondary data was collected from Komfo Anokye Teaching Hospital in Kumasi municipality on variables; Body Mass Index, Arm Fat as percent composition of the body, Height, Triceps skin-fold, and Total fats in patients' body. Some of these variables are measured and others derived from an OMRON machine. In contaminating and analyzing the data, we used the R statistical package.

Information for this study was gathered from the Internet and the published articles and books that are related to the study.

## 1.5 Justification of Work

In statistical analyses, especially linear regression, we state some assumptions about the dependent variable; the explanatory variables and the error terms. However, in actual situations or applications, these assumptions are rarely met. Therefore robust estimation is very paramount when it comes to linear regression analysis. And this study is concerned with showing the merits of the robust estimation in linear regression analysis.

## 1.6 Organization of Thesis

The Chapter one of this study will contain Background of study, Problem statement, Objectives of study, Methodology, Justification of the study, and Organization of thesis. Chapter two will contain the review of literature, where studies already carried out and are related to our study, methods and application of robust estimation will be looked at. Some robust methods of interest for regression analysis will be considered under chapter three. Chapter four will contain results of the data analysis, where the four datasets will be analyzed using the methods outlined chapter three. Finally, various findings from the analysis will be discussed in Chapter five to check if the goals of this study are achieved. Recommendations will be given with respect to the results obtained that are based on the methods used in the analysis.

# Chapter 2

# Literature Review

In this section of our study, we review studies which had been conducted and are valid to our work. The review will be done in the areas of robust regression estimation with special attention to the estimators such as: least trimmed squares estimator (LTSE), Huber M-Estimator (HME), S-Estimator (SE) and Modified Maximum likelihood Estimator (MME), and in relation to Ordinary Least Squares Estimator (OLSE).

## 2.1 Robust Regression

In multiple linear regression, if the assumptions hold, the ordinary least squares method of estimation is used in estimating the parameters. The assumptions of the ordinary least squares are that: response variable should be continuous; residual errors should be normally distributed, and have equal variance at all levels of the independent variables (homoscedasticity), and be uncorrelated with both the independent variables and with each other. According to Ho and Naughter (cited by Schumacker,et al., 2002), if the data contain outliers; non-normality; or multicollinearity existing between variables, then the sample estimates can be misleading. In addition, aberrations and contaminants in the dependent variable and leverage points in the predictor variables can also lead to the break down of the ordinary least squares method of estimation. These unusual observations in the data can have adverse effects on the sample estimates of the ordinary least squares method and still remain unnoticed.

To rectify the problem pose by the influence of the aberrations in data, is to use regression techniques which can resist the effects of outliers, leverage points, non-normality and contaminations. Some of these robust regression techniques are Huber maximum likelihood Estimator (HME), Modified Maximum likelihood Estimator (MME), S-Estimator (SE), and Least Trimmed Squares Estimator (LTSE).

## 2.2 High Breakdown point estimators

Due to the drawbacks of the ordinary least squares estimator (OLSE), there are estimators that possess the property of a high breakdown point. The OLSE minimizes the variances of the residuals and as a result, the estimates respond to outliers and influential observations. To solve this problem, Rousseeuw and Yohai (1987) propose to minimize a measure of dispersion of the residuals that is less sensitive to extreme values than the variance (Verardi and Croux, 2009).

This estimator is referred to as the S-estimator. Since the OLS gives a huge importance to large residuals, the S-estimator method replaces the square function by a loss function that gives less weights to large residuals. The loss function is carefully chosen to enable the estimator possess good robustness properties and a high Gaussian efficiency. However, there exists a trade-off between the breakdown point and the asymptotic efficiency of the S-estimator.

The 50% high breakdown point S-estimator has efficiency of 28.8% with a tuning constant of 1.547. And it also has a breakdown point of 10% and asymptotic efficiency of 96.6% with a tuning constant of 5.182. As a result of these lapses of the S-estimator, it is usually used as initial value for Modified maximum likelihood estimator. This is done with the sense that MM-estimator can inherit high breakdown property of the S-estimator. Since MM-estimator is highly efficient, using S-estimator as initial values, gives it the high breakdown property. Therefore, MM-estimator is recognized as the estimator with both high breakdown point and efficiency, for both normal and non-normal errors.

## 2.3 Studies on Robust Estimators

Regression analysis is mostly used in all areas of study, and the most handy method normally used is the ordinary least squares. However, due to the drawbacks of this traditional method, many statisticians have come out with solutions to these lapses of the ordinary least square.

In a study by Bhar (2014), the study looked at the Huber M-estimator as an improvement of the ordinary least squares estimator. In his study, robust M-estimator was compared with the ordinary least squares estimator. He discussed robust regression methods such as; M-estimator, W-estimators, R-estimators, Least median of squares estimator, Least trimmed of squares estimator, and Re-weighted least squares estimator. The most common method of robust regression is M-estimation, introduced by Huber (1973, 1981) that is nearly as efficient as least squares estimator. Rather than minimize the sum of squared residuals as the objective, the M-estimator minimizes a function $\rho$ of the standardized residuals. As a result, this method gives smaller weights to observations that are unusual, and hence performs better than the ordinary least squares estimator in the presence of vertical outliers. W-estimators are alternatives to the M-estimator which has a characteristic weight function. And this weight function shows the importance of each observation to the estimator. R-estimators are the estimators computed based on ranks of the data. His study also highlighted L-estimators, which are the estimators computed from a linear combination of order statistics. And they include: least trimmed squares and least median squares. The results from the analysis shows that the Huber M-estimator performs better than the ordinary least squares estimator using both standard error and the coefficient of determination.

Fox and Weisberg (2010) did a study on Robust regression and considered estimators such as: the M-estimators, bounded-influence estimators, MM-estimator, Least median squares estimator. Applying all these methods on Duncan's occupational-prestige data, revealed that the least squares estimator broke down as a result of vertical outliers. But the various robust methods were able to bound the influence of these unusual observations. The robust methods serve as formidable statistical tools for identifying unusual observations in the data. In their study, the main criterion for the comparison of the estimators was the coefficients, but using the standard errors - Least trimmed squares perform better than the others. However, ordinary least squares estimator performed very poorly, when no outliers are removed.

AL-Noor and Mohammad (2013) researched on Model of Robust Regression with Parametric and Nonparametric Methods. A simulation study was performed to compare Ordinary Least Squares Method; Least Absolute Deviations method; M-Estimators; Trimmed Least Squares estimators and Nonparametric Regression. In their study, these estimators were compared for vertical outliers, horizontal outliers and both vertical and horizontal outliers based on their mean square error and relative efficiency. The results for the analysis with no contamination showed that Ordinary least squares estimator performed better than the other estimators. However, when outliers were introduced in the dependent variable, the Ordinary least squares method broke down. And when outliers were introduced in dependent and the independent variables, Ordinary least squares method; M-estimators; Least absolute deviation broke down. The nonparametric methods perform better when outliers were present in both X-dimension and the Y-dimension. Considering the overall performance, the Least trimmed squares estimators performed better than all other estimators. AL-Noor and Mohammad recommended that future studies should take into consideration the following:

i. The poor performance of OLSE estimators with the presence of outliers confirms our need for alternative methods. Therefore, before analyzing the data, we should first check the presence of outliers and then construct the necessary tests whether to see the underlying assumptions are satisfied. After that, we should conduct the appropriate estimation techniques.

ii. Choose a nonparametric method, especially to estimate slope and model, or choose a LTS method when the outliers are in X-direction or XY-direction.

iii. Choose M-estimator and LAD method, or choose a LTS method when the outliers are appearing in Y-direction.

iv. When the outliers appear in X-direction or XY-direction, choose RMSE or mean absolute error (MAE) as criteria for comparing the estimators to avoid dealing with the large values of MSE.

Therefore, it is necessary to find out the aberrations in the dataset and the basic assumptions of the linear regression model before deciding on the type of method to use.

Muthukrishnan and Radha did a study on comparison of robust regression estimators by assessing their coefficient of determination. The study considered the estimators such as Least trimmed squares, Least median squares, Trimmed mean square, Huber M-estimator, Least absolute deviation, all in relation to the Ordinary least squares estimator. Their study looked at the fact that practical applications of linear regression rarely satisfy the basic assumptions imposed on data sets by the regression model. As a result of this, robust methods are very important when applying the linear regression in practice. They used the R-software to analyze two real life data sets. The results of the study showed that, Huber M-estimator is analogous to the ordinary least squares estimator. On the other hand, Huber M-estimator performs better than the least squares estimator if there are only vertical outliers in the data. Robust estimators such as Least trimmed squares, Least median squares, Least absolute deviation also appear to be similar and are different from the least squares estimator and the

Huber M-estimator. The study therefore concluded on the note that all robust methods are modification of the traditional methods.

Alma researched on Comparison of Robust Regression Methods by considering estimators such as: Least Trimmed Squares estimator, Huber M-estimator, Yohai MM-estimator, S-estimator. These estimators were compared in relation to the least squares estimator. In his study, the comparison of the estimators was done in the presence of outliers and leverage points - by varying the percentage of outliers and leverage points in the data. The study concluded that, S-estimator performed better than the others in efficiencies and was able to bound the influence of outliers and leverage points. Also, the study had shown that MM-estimator breaks down when dealing with high leverage points in small dimensional data.
Our study will look at this four robust estimators in a presence of outliers, leverage points, non-normality and contaminants.

Schumacker et al. (2002) did a study on Comparison of Ordinary least squares and robust regression using the S-PLUS statistical package. The violation of basic assumptions such as: outliers in datasets, non-normality, or multicollinearity among variables leads to estimates not reflecting the actual parameters. In their study, robust estimators such as Least trimmed squares and Modified maximum likelihood estimators were compared to the ordinary least squares estimator. By using a dataset in S-PLUS statistical package, it was shown that the MM-estimator performed better than LTS-estimator and the OLS-estimator. Therefore, it was concluded in their analysis that, the best method is the MM-estimator. Their study also considered the fact that a few statistical packages have the robust methods and as a result, statisticians find it difficult to compare these estimators to check the effect of outliers and the other drawbacks of real life data.

Yohai (1987) developed the MM-estimator, which is by far the most efficient with a high breakdown point. MM-estimator makes use of other estimators, but for the MM-estimator to possess the high breakdown point property, it was proposed to use the S-estimator as initial estimates to compute the MM-estimator. Yohai's study highlights the properties of the MM-estimator such as; high breakdown point, efficiency, Exact fit property and scale equivariance. The robust estimators were compared with the Ordinary least squares using asymptotic biases under contamination. In his study, it was concluded that the MM-estimator was not influenced by outliers as they did to the ordinary least squares.

Jacoby did a study on robust estimators that can be used when there are unusual observations in a data set and when the data set is from a skewed distribution. Many robust estimators were considered with much emphasis on M-estimators and the estimators that bound the influence of unusual observations in a data. Estimators such as; Huber estimator, Bisquare estimator, Least-Trimmed Squares in addition to others were compared. In his study, M-Estimator was analogous to OLSE when outliers are removed. Bisquare estimator was not affected by the influential cases like the Huber estimator. The robust estimators such as LTSE and LMSE did not perform well, this was as a result of the fact that the sample size was small and deleting about half of the data changed the actual information in the data. Robust estimators which have a breaking point as high as 50%, often work very well when the sample size is large. The study concluded on the fact that no single robust estimator is best for all data sets. It was noted that, M-Estimators however perform better for small datasets, but provide unreliable standard errors.

A study by Jann (2012), did look at robust estimators such as; MM, HM, S, LTS, and LMS estimators in relation to the ordinary least squares estimator. Fat tail distributions resulting from non normal errors lead to the breakdown of the ordinary least squares estimators. For example, the t-distribution with few degrees of freedom. Also, the properties of the LS estimator only hold if the assumptions imposed on the data by the least square method of the regression model is satisfied. By applying STATA statistical package on a set of data, the study concluded on the results that even though the robust methods appear to perform better than the OLSE, it should be used with the robust methods serving as model diagnostic tools.

Cetin and Toka (2011) compared some robust methods of estimation to the Ordinary least squares estimation. In their investigation, S-estimator; Huber M-estimator; Least trimmed squares estimator were examined relative to the OLSE. A dataset with no outliers but with weak multi-collinearity was used for the comparison. OLSE was discovered to have been affected when outliers were later introduced into the dataset. The S-estimator performed extremely better than the remaining estimators, followed by the HME. Simulation study was also conducted with non-normal data, and it was found that the OLSE is inefficient when the dataset is contaminated with outliers.

According to a study by Rousseeuw et al. (2001), it was discussed that there are some factors that hinder the use of robust regression methods. It was made clear that the belief that the OLSE is efficient for large sample data is not true and small portion of outliers in a large sample data can distort the performance of the OLSE. This is because, in large sample datasets, data points cluster around each other making it easy for the data points to mask each other without being noticed by simple eyeballing. Some people also find it difficult interpreting the results of robust methods as a result they do not know the advantages of the robust methods. In their study, Rousseeuw et al. emphasized the advantages of using the LTSE as a robust method. The study concluded on the note that deleting few outlying points brings about drastic changes in regression results. And this sometimes leads to a significant fit of the regression model.

In a study by Ruppert et al. (1988), much attention was paid to Iteratively Reweighted Least Squares method of estimation. As a result, estimators such as Huber estimator, Tukey biweight estimator and Hampel estimator were compared with the ordinary least squares estimator. The results from the study showed that the Huber estimator has problem with leverage points but OLSE performed badly overall among the methods.

## 2.4 Properties of Estimators

A researcher performing an inference about a population parameter needs an information on a sample statistic of that population parameter. Moreover, an estimator is a value of a sample statistic that gives information on the population parameter of interest. According to Glass and Hopkins (cited by Schumacker et al., 2002), when performing parametric statistical tests one should be mindful of the properties of estimators. The properties of estimators are unbiasedness, consistency, efficiency and sufficiency.

1. An estimator is considered unbiased if the mean of the sampling distribution of the statistic equals the population parameter being estimated.

2. An estimator is considered consistent if the sample statistic approaches the population parameter as the sample size increases.

3. An estimator is considered efficient if it is the only one among all other unbiased estimators with the smallest variance.

4. An estimator is considered sufficient if the sample (n) from which it is computed contains all the information contained in the population (N).

A method is considered robust if it is efficient when compared to the ordinary least squares estimator, when the errors are independent normal. Moreover, it is substantially efficient than least squares estimator when there are outlying observations in either the response variable or predictor variables or both.

## 2.5 Outliers in Multiple Linear Regression

Outliers in a dataset are unusual observations that occur in a data and do not follow the general pattern of the majority of the data points. In a study by Bhar, it was discussed that outliers come by as a result of a simple operational mistake whereby small sample is included in the data from a different distribution, and they do have serious effects on statistical inference. We have three types of outliers in regression analysis that influence the performance of ordinary least squares according to Rousseeuw and Leroy (1987). These are: vertical outliers, bad leverage points and good leverage points.

According to Verardi and Croux (2009), vertical outliers are those observations that have outlying values for the corresponding error term (the y-dimension) but are not outlying in the space of predictor variables (the x-dimension). Their presence affects the ordinary least squares estimation and in particular the estimated intercept.

Good Leverage points are data points that are outlying in the dimension of predictor variables but are close to the regression line. They do not affect the coefficients of the ordinary least squares estimator, but affect the standard errors and hence influence the statistical inference.

Finally, bad Leverage points are observations that are both outlying in the dimension of explanatory variables and are far from the true regression line. These bad leverage points affect both the intercept and the slope of the ordinary least squares estimation. Some of these data points are very influential and can bring drastic changes in the fitted model.

Therefore, since these types of outliers do affect the traditional method of estimation, many studies were carried out on outliers. Some of these studies are discussed as follows: Outliers in a dataset go along way to violate the basic assumptions of the error terms in regression analysis. A study by Jacoby looked at the various types of outliers including leverage points and their influence on the ordinary least squares estimates.

In his study, OLSE method of estimation was applied on a dataset from R statistical-package car, with and without outliers to see the effects of the outliers. The study also discussed the various methods of identifying outliers in the dependent variable and the leverage points in the independent variables.
It was shown that small samples are more vulnerable to outliers than large sample cases. Unusual observations are only influential when they are both unusual in the dependent variable and the independent variables. This makes it clear that observations that are unusual only in either y-axis or the x-axis do not influence the estimates much as does by the regression outliers, since the regression outliers are the outliers that are unusual in the dependent variable given that they are unusual in the independent variables as well.

In testing for data points that are outliers, methods such as studentized residuals and quantile-comparison plots do help in discovering these points. Leverage is assessed by exploring the hat-values and Influence is assessed by using Cook's Distance. Since outliers can come about as a result of a lot errors in the data collection processes, unusual data points may reflect miscoding, and therefore can be rectified by deleting the observation(s) entirely. But if there are no strong reasons to discard the outliers, then they must be included in the analysis and hence the ordinary least squares method ceases to be effective. Therefore, more robust methods could be used to analyze that kind of data as they produce same results as the OLSE when there are no deviations in the data, but down weight the effects of influential observations when they exist.

### 2.5.1 Effects of Outliers on a Statistic

According to Jacoby (2014), outliers have substantial influence on regression models generally, and these effects can be seen in these areas:

1. Outliers lead to misinterpretation of general pattern in regression plots.

2. Also, unusual observations can have a strong influence on statistical models, and removing outliers from a regression model can sometimes give completely different results.

3. This unusual points can also substantially influence the fit of the Ordinary least squares model.

4. Both the slopes and intercept of the model are substantially influenced by data points that are both outliers and high leverage.

5. Furthermore, Outliers in data show that our models derived cannot capture important characteristics of the data.

An inlier is an observation that lies close to the mean. In a univariate setting, a value close to the mean would not raise any eyebrows. However, as Evans (2001) pointed out, it would be unlikely for an observation to lie near the mean for a large number of variables. So while outliers may be problematic, inliers may be more likely to represent observations that are too good to be true or too good to be real. Of course, a Mahalanobis distance of zero would represent a subject that lies on the mean for every variable.

### 2.5.2 Effects of Non-normality on a Statistic

The observations in a sample may actually come from the same population, but not from a normal one. Errors in a form of outliers and inliers in the response variable usually leads to the non-normality of the error terms, which results in the failure of some model assumptions. Outliers occurring far away from the majority of the data points makes the distribution of the response variable to deviate from the normal. Skewed and Heavy tailed distributions come about as a result of these unusual data points. Therefore, skewness or light-tailedness or heavy-tailedness are signs of non-normality. Non-normality due to outliers affects the intercept of the regression model and renders the regression model inefficient and seriously biased standard errors.

### 2.5.3 Effects of Contaminants on Statistic

In a simulation study by Stuart (2011), it was observed that contaminants in a dataset do cause estimators to breakdown completely. In this simulation study, the MM-estimator was slightly affected when the dataset was contaminated with data points from Cauchy distribution, but the Ordinary least squares estimator broke down completely. Therefore, a dataset that is believed to have come from two probability distributions basically does not follow the rules of the classical regression method.

## 2.6 Comparison of Robust Estimators

Robustness of linear regression parameter estimation originated from the theories of Huber and Hampel and these have laid the foundation for finding practical solutions to many problems. Many researchers have worked in this field and described the methods of robust estimators. Those researchers who did comparative study on these estimators did compare these estimators using many criteria. Some of these are Mean square error, coefficient of determination, standard errors, relative efficiencies, coefficients, bias, breakdown point, wald-statistic and etc. In our study we will consider some of these criteria in our work.

### 2.6.1 Studies on Criteria for comparing Estimators

A study by Muthukrishnan and Radha compared M, L and R-estimators with the ordinary least squares estimator by using the coefficient of determination. Schumacker et al. (2002), also did a study on robust regression by comparing estimators such as: OLSE, LTSE and MME using coefficient of determination as the criterion.

Yohai (1987) compared robust estimators using bias of the estimators under contamination and Stuart (2011) investigated on robust estimators using the coefficients, relative efficiency and the standard errors as the criteria for the comparison.

AL-Noor and Asmaa, and Gschwandtner and Filzmoser (2012) in their study, used mean square error as the criterion for comparing their estimators.

### 2.6.2 Merits of Robust estimators over the Ordinary Least Squares Estimator (OLSE)

Robust methods in estimating parameters for the linear regression models serve a lot of purposes, among these are:

1. Robust estimators are applicable when there are no evidences against the unusual observations in a dataset, since the OLSE is likely to breakdown.
2. Also, robust estimators help in identifying unusual observations in a dataset that could undermine the least squares estimator's performance.
3. Again, robust estimators are able to bound the influence of unusual observations in a dataset.

4. Moreover, robust methods do not need the assumptions of the classical regression model to be satisfied before its applications.

5. Furthermore, large sample theories undermine the applicability of robust techniques, which is a myth.

6. Last but not the least, they help in detecting outliers due to Masking and Swamping, which nullifies the belief that outliers can be detected simply by looking for unusual OLS residuals, box-plot of residuals or test of normality.

# Chapter 3

# Methodology

In this chapter, the various robust regression estimators are outlined and discussed. Attention will be on the robust methods considered in this study and their properties. The aim of linear regression analysis is to study how a dependent variable is linearly related to a set of predictors. We start with the classical methods of linear regression.



## 3.1 Regression Analysis Model

The multiple linear regression model can be written in matrices notation as

$$y = X\beta + e$$

where y is an $n \times 1$ vector of observed response values, X is the $n \times p$ matrix of the predictor variables, $\beta$ is the $p \times 1$ vector contains the unknown parameters and needs to be estimated, and $e$ is the $n \times 1$ vector of random error terms. Therefore to fit this model to the data, we have to use a regression estimator to estimate the unknown parameters in $\beta$, to have $\hat{\beta}$ where $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{pmatrix}$.

The expected value of $y_i$, that is the fitted value, $E(y_i)$ is given by $\hat{y}_i = X_i^T \hat{\beta}$. As a result the residuals can be computed by using $r_i = y_i - \hat{y}_i$, where $i = 1, 2, ..., n$; and n is the sample size. According to Stuart (2011), if the assumptions of the error terms are met, that is the $e_i \sim N(0, \sigma^2)$, then the least squares regression estimator is the maximum likelihood estimator for $\beta$, maximizing

$$\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{e^2}{2\sigma^2}\right) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \hat{y}_i)^2}{2\sigma^2}\right) \tag{3.1}$$

over $\beta$. Which is analogous to maximizing the logarithm of 3.1 over $\beta$: $\sum_{i=1}^{n}\left(-\frac{1}{2}\ln(2\sigma^2) - \frac{e_i^2}{2\sigma^2}\right)$, which corresponds to $\sum_{i=1}^{n} e_i^2$ since $\sigma$ is a constant. Therefore the minimization of the residual sum of squares $\sum_{i=1}^{n} e_i^2$ is the least squares estimate $\hat{\beta}$, (Stuart, 2011). She went on to explain that this will have much effect on the way in which different types of observations affect the regression estimate, $\hat{\beta}$. Rice (1995) explained that finding $\hat{\beta}$ in simple linear regression corresponds to fitting the regression line by minimizing the sum of the squared vertical distances of observed points from the line.

## 3.2 The Ordinary Least Squares Estimator

We define the design matrix X, and the vectors Y and $e$ as,

$$X = \begin{pmatrix} x_{11} & ..... & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & ...... & x_{np} \end{pmatrix} = \begin{pmatrix} x_1^T \\ \vdots \\ x_n^T \end{pmatrix}$$

$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$, *and* $e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$. And the regression linear model is given by $Y = X\beta + e$.

The least squares estimator aims to minimize the sum of the square residuals as:

$$\begin{aligned} \sum_{i=1}^{n} e_i^2 &= e^T e \\ &= (Y - X\beta)^T (Y - X\beta) \\ &= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta \end{aligned}$$

By minimizing the errors:

$$\frac{\partial}{\partial\beta}\sum_{i=1}^{n} e_i^2 = \frac{\partial}{\partial\beta}(Y^TY - Y^TX\beta - \beta^TX^TY + \beta^TX^TX\beta)$$

$$\Rightarrow 0 - 2X^TY + 2X^TX\beta = 0$$

Therefore one can compute the least squares estimates directly from the dataset when $X^TX$ is non-singular by,

$$\hat{\beta} = (X^TX)^{-1}X^TY \tag{3.2}$$



### 3.2.1 The limitations of the Ordinary Least Squares Estimator (OLSE)

The least squares estimator performs well when the error terms are well-behaved or the underlying assumptions hold (Adebanji, 2013). However, failure of these assumptions leads to high sensitivity of the OLSE. Outliers and leverage points do affect the performance of the classic method of estimation. Identifying outliers in the data set can be very difficult without careful investigations. This is because there are good leverage points for example that can be very difficult to discover as they fall on the regression line and can result in inflated standard error which will hence affect the possible inferences. Also, masking and swamping can also make it difficult for outliers to be discovered in the dataset by mere observation. Using Scatter plot to detect outliers in a data for simple linear regression is usually easy, however, when it comes to multiple linear regression it is limited. As a result many robust methods have been formulated to take care of these limitations of the OLSE.

## 3.3 Robust Estimators

Let $z_i = \{(y_i, x_i)\}_{i=1}^{n}$, be a random sample that follows the linear model $Y = X\beta + e$, where the $e_i$ are i.i.d. random variables independent of the $x_i$, with unknown distribution $F_0$. Assuming $G_0(x)$ is the distribution of $x_i$, then the distribution of $z_i$ is given by:

$$H_0(z) = G_0(x)F_0(e) \tag{3.3}$$

Robust estimators should be resistant to a certain degree of data contamination. The general types of departures considered for robustness are $\epsilon$-contamination models with $z_i \sim G_\epsilon$, where $G_\epsilon = (1-\epsilon)G_0 + \epsilon G$ a mixture distribution, where G is a secondary distribution that contaminates the data. Thus the distribution $G_\epsilon$ produces a fraction, $\epsilon$, of outliers coming from G. According to Bondell and Stefanski (2013), the idea is to estimate $\beta$, in the presence of the contamination.

### 3.3.1 Least-Trimmed-Squares Estimator (LTSE)

Rousseeuw (1984) developed the least trimmed squares estimator (LTSE) given by,

$$\hat{\beta} = min \sum_{i=1}^{h} (e_i^2) \tag{3.4}$$

where $e_1^2 \le e_2^2 \le, ..., \le e_n^2$ are the ordered squared residuals, from smallest to largest. LTSE is computed by minimizing the h ordered squared residuals, where $h = \left(\left[\frac{n}{2}\right] + 1\right)$, where n and h are the sample size and the trimming constant, respectively, (AL-Noor & Mohammad, 2013). Using the h trimmed dataset ensures that estimates have a high breakdown point of 50% but a low efficiency of 7.13%, (Matias and Yohai, 2006) .

In a study conducted by Rousseeuw and Leroy (1987), they suggested a trimming constant of $h = [n[1-\alpha] + 1]$ where $\alpha$ is the trimmed percentage. The largest squared residuals are deleted and the least squares method is applied on the trimmed dataset. Moreover, this helps to discard outliers and their influence on the regression estimators. LTSE can be fairly efficient if the trimmed percentage is chosen carefully to discard only the outliers in the data. It can be very efficient based on the size of the trimmed dataset (h) and the level of outliers in the data. Similarly, when the data trimming is done well, this method is computationally equivalent to OLSE. However, in situations where there are more outliers and only some are trimmed this method can also perform as poorly as the ordinary least squares method of estimation. On the contrary, if more observations are deleted where there are only few outliers, good data points will be discarded from the dataset. Least-trimmed-squares has a break-down point of 50%, hence makes the LTSE a high break-down method of estimation. This implies that half of the data has to be influential points before estimates of the least trimmed estimator be affected when the method of the ordinary least squares is applied. LTSE essentially proceeds with OLSE after the deletion of the most extreme positive or negative residuals. LTSE on the other hand, can misrepresent the trend in the data if it is characterized by clusters of extreme cases or if the data set is relatively small. The breakdown value is $\frac{n-h}{n}$ for the LTSE estimate.

**Maximum likelihood type estimators**

The Ordinary least squares estimator is derived by minimizing a function of the residuals. Whilst this is obtained from the maximum likelihood function of the assumed normal distribution of the errors, M-estimators are the estimators that result from the maximum likelihood function of a distribution of the errors which might not be normal. The distribution of the errors might be represented by a different, heavier-tailed, distribution. Assuming this probability distribution function is $f(e_i)$, then the maximum likelihood estimator for $\beta$ is that which maximizes the likelihood function

$$\prod_{i=1}^{n} f(e_i) = \prod_{i=1}^{n} f(y_i - x^T \beta)$$

And this implies it also minimizes the function

$$-\sum_{i=1}^{n} \ln f(e_i) = \sum_{i=1}^{n} -\ln f(y_i - x^T\beta)$$

If the errors terms are well-behaved or normally distributed, then it is just the minimization of the function $\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n}(y_i - x^T\beta)^2$. So if we assume that the errors terms are not well-behaved or distributed differently, then it is now maximum likelihood estimator minimizing a different function. As determined by Bhar and Garner, let $\rho = -\ln f$, then an M-estimator, $\hat{\beta}$, minimizes

$$\sum_{i=1}^{n} \rho(e_i) = \sum_{i=1}^{n} \rho(y_i - x^T\beta) = \sum_{i=1}^{n} \rho(u) \tag{3.5}$$

where $\rho(u)$ is called an objective function, which is continuous, symmetric, positive-definite function with a unique minimum at 0. On the whole, the least squares estimator is in fact a special less robust case of M-estimators. Since getting the actual distribution of the errors for real life quality datasets are normally not easy, the choice of $\rho(u)$ depends on the robustness that is required. In robust estimation, these objective functions are carefully chosen so that the resulting estimator will not be affected by outliers, by down-weighting very large residuals. A robust M-estimator achieves this by minimizing the sum of a less rapidly increasing objective function than that of the least squares estimator $\rho(u) = u^2$, (Andersen, 2008). However the solution to 3.5 is not scale equivariant, and thus the residuals must be standardized by a robust M-estimate of scale, s.

### Desirable property of Robust regression estimates

The desirable property of an estimate is that the estimate be equivariant with respect to affine, regression, and scale transformations, (Matias and Yohai, 2006). This property implies that the estimate transforms properly when the dataset undergo some transformations. Rousseeuw and Leroy discussed this property is explicitly as follows:

1. Regression equivariance

   An estimator T is considered regression equivariant if:

   $T(\{(x_i^T, y_i + x_i^T v);\ i = 1, ..., n\}) = T(\{(x_i^T, y_i);\ i = 1, ..., n\}) + v$

   This implies that any additional linear dependence $y \to y + Xv$ is reflected in the coefficients accordingly $\hat{\beta} \to \beta + \beta v$. When considering the various estimators in regression, this property is used routinely; in proofs of asymptotic properties it allows the fixing of $\beta = 0$ without loss of generality.

2. Affine equivariance

   An estimator is affine equivariant if:

   $T(\{(x_i^T A, y_i);\ i = 1, ..., n\}) = A^{-1}T(\{(x_i^T, y_i);\ i = 1, ..., n\})$, so when the predictor variables, $X_i$, are linearly transformed, $X \to XA$, the estimator is also transformed accordingly, $\hat{\beta} \to A^{-1}\hat{\beta}$. This is useful because it means that contaminating the explanatory variables will not affect the estimate:

   $\hat{y} = X\hat{\beta} = (XA)(A^{-1}\hat{\beta})$

3. Scale equivariance

   An estimator is scale equivariant if the fit produced by it is independent of whether the response variable is contaminated or not. Hence an estimator is scale equivariant if:

   $T(\{(x_i^T, cy_i);\ i = 1, ..., n\}) = cT(\{(x_i^T, y_i);\ i = 1, ..., n\})$

   This implies if $y \to cy$ then $\hat{\beta} \to c\hat{\beta}$

**The scale equivariant estimator**

An estimator is scale equivariant if the estimate from a data do not change even after the error terms are contaminated and hence deviate from the normality assumption. Therefore a scale equivariant estimator is not affected by outliers in a dataset. Moreover, the problem of non scale equivariance is solved by standardizing the error terms by the scale of the residuals. As a result the M-estimators minimize the function:

$$\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) \tag{3.6}$$

The scale or the standard deviation of the residuals is affected by outliers, therefore the scale of the residuals cannot be used here. Andersen (2008) (cited by Stuart, 2011) gave this estimator as the Median Absolute Deviation (MAD), which is given by:

$$s = \frac{MAD}{0.6745} \Rightarrow s = 1.4826 \times MAD, \text{ where MAD is computed as}$$

$$MAD = median\,|e_i - median(e_i)|$$

This estimator for s is highly robust to outliers since it uses the median instead of the mean and has a breakdown point of 50%.

**Proof 3.3.1** *Assuming the sample is large and the error terms $\epsilon_i$ are normally distributed with mean 0 and $\sigma^2$, then*

$$\begin{aligned}
P(|e_i| < MAD) &\approx 0.5 \\
P(\left|\frac{e_i - 0}{\sigma}\right| < \frac{MAD}{\sigma}) &\approx 0.5 \\
P(|Z| < \frac{MAD}{\sigma}) &\approx 0.5 \\
\frac{MAD}{\sigma} &\approx \phi^{-1}(0.75) \\
\frac{MAD}{\phi^{-1}(0.75)} &\approx \sigma \\
\sigma &\approx 1.4826 \times MAD
\end{aligned}$$

*Therefore when the sample is large and the error terms $e_i$ are normally distributed with mean 0 and $\sigma^2$, s estimates the population standard deviation (Stuart, 2011).*

### 3.3.2 The Huber Maximum likelihood Estimator (HME)

The class of M-estimator models contains all models that are derived to be maximum likelihood models. The most common method of robust regression is M-estimation, developed by Huber (1973) that is almost as efficient as OLSE. In estimating this, instead of minimize the sum of squared errors as the objective, the M-estimate minimizes a function $\rho$ of the errors. The M-estimate objective function is,

$$\sum_{i=1}^{n} \rho\left(\frac{e_i}{s}\right) = \sum_{i=1}^{n} \rho\left(\frac{y_i - x^T\beta}{s}\right), \tag{3.7}$$

where "s" is an estimate of scale from a linear combination of the residuals. The function $\rho$ gives the contribution of each residual to the objective function. According to Alma (2011), a reasonable $\rho$ should have the following properties:

$$\begin{aligned} &\rho(e) \geq 0 \\ &\rho(0) = 0 \\ &\rho(e) = \rho(-e) \\ &\rho(e_i) \geq \rho(e_i^{'}) \text{ for } |e_i| \geq \left|e_i^{'}\right|, \text{ and } \rho \text{ is continuous} \end{aligned} \tag{3.8}$$

An example is given for least squares estimation, $\rho(e_i) = e_i^2$. And the system of normal equations to solve this minimization problem is found by taking partial derivatives with respect to $\beta$ and setting them equal to 0. So we minimize equations 3.7 with respect to each of the p parameters, $\beta_1, .., \beta_p$ and this resulted in a system of p equations:

$$\begin{aligned} \sum_{i=1}^{n} x_{ij}\psi\left(\frac{e_i}{s}\right) = \sum_{i=1}^{n} x_{ij}\psi\left(\frac{y_i - x_i^T\beta}{s}\right) = 0;\ j = 1, 2, .....p \\ \text{and } i = 1, 2, .....n \end{aligned} \tag{3.9}$$

where $\psi(u) = \frac{\partial\rho}{\partial u}$ is the score function. We define a weight function as

$$w(u) = \frac{\psi(u)}{u}$$

$$w_i = \frac{\psi\left(\dfrac{y_i - x_i^T}{s}\right)}{\left(\dfrac{y_i - x_i^T}{s}\right)}$$

which results in $w_i = w\left(\frac{e_i}{s}\right)$ for $i = 1, 2, ....n$ with $w_i = 1$ if $e_i = 0$. Substituting this into 3.9 results in:

$$\begin{aligned}
\sum_{i=1}^{n} x_{ij} w_i \left(\frac{e_i}{s}\right) = \sum_{i=1}^{n} x_{ij} w_i \left(\frac{y_i - x_i^T \beta}{s}\right) &= 0,\ j = 1, 2, .....p \\
\Rightarrow \sum_{i=1}^{n} x_{ij} w_i (y_i - x_i^T \beta) &= 0,\ j = 1, 2, .....p \\
\Rightarrow \sum_{i=1}^{n} x_{ij} w_i y_i = \sum_{i=1}^{n} x_{ij} w_i x_i^T \beta &= 0,\ j = 1, 2, .....p
\end{aligned} \tag{3.10}$$

Since $s \neq 0$, we define the weight matrix $W = dig(w_i : i = 1........n)$ as:

$W = \begin{pmatrix} w_1 & & & 0 \\ & w_2 & & \\ & & \ddots & \\ 0 & & & w_n \end{pmatrix}$. This results in the matrix form of the 3.10 as,

$$\begin{aligned}
&\Rightarrow X^T W X \beta = X^T W Y \\
&\Rightarrow \hat{\beta} = (X^T W X)^{-1} X^T W Y
\end{aligned} \tag{3.11}$$

In selecting the type of the $\psi$ function to use, one does this based on how much weight to assign to outliers. As a result the $\psi$ function does not weight large outliers as much as least squares method does. A redescending $\psi$ function increases the weight assigned to an outlier until a specified distance ($e.g. 3\sigma$) and then decreases the weight to 0 as the outlying distance gets larger. The equation above is similar to that of the ordinary least squares estimator but with a weight matrix that down-weight the influence of outliers in the data. Unlike least squares, the least squares method cannot be used to calculate an M-estimate directly from data. According to Alma (2011), the weights however depend on the residuals, the residuals depend on the estimated coefficients, and the estimated coefficients depend on the weights. As a result initial estimates and iterations are required, to eventually converge on W and an M-estimate for $\beta$. According to Draper and Smith (1998), M-estimates of regression are found using the iteratively re-weighted least squares procedure (IRLS), and it is given as:

1. With the iteration counter I set to 0, the least squares method is used to fit an initial model to the data, yielding the initial estimates of the coeffcients $\hat{\beta^{(0)}}$

2. Initial residuals $e_i^{(0)}$ are found using $\hat{\beta^{(0)}}$ and used to calculate $s^{(0)}$.

3. A weight function w(u) is chosen and applied to $\frac{e_i^{(0)}}{s^{(0)}}$ to obtain preliminary weights $w_i^{(0)}$. These give the value of $W^{(0)}$ for $\hat{\beta^{(0)}}$

4. Set I = 1. Using $W^{(0)}$ in 3.11, one obtains the estimate
$\hat{\beta}^{(1)} = (X^T W^{(0)} X)^{-1} X^T W^{(0)} Y$

5. Using $\hat{\beta}^{(1)}$ new residuals, $e_i^{(1)}$ can be found, which, through calculation of $s^{(1)}$ and application of the weight function yield $W^{(1)}$.

6. Set I = 2. A new estimate for $\beta$ is found using $W^{(1)}$. This is $\hat{\beta}^{(2)}$, $e_i^{(2)}$ and $s^{(2)}$ , and in turn next weight, $W^{(2)}$ are then computed.

7. This process is now iterated such that at I = q
$\hat{\beta} = (X^T W^{(q-1)} X)^{-1} X^T W^{(q-1)} Y$
until the estimates of $\beta$ converge, at the point the final M-estimate has been found. The procedure is usually stopped if the criterion is reached. The criterion can be set as an estimate changes by less than a selected percentage between iterations, or after a fixed number of iterations have been carried out. When iterating until a set convergence criterion has been met, the criterion should be of the form
$\frac{\|\hat{\beta}^{(q+1)} - \hat{\beta}^{(q)}\|}{\hat{\beta}^{(q+1)}} < \varepsilon$
where $\varepsilon$ is a small positive number, usually fixed at 0.0001. Another way of setting the convergence criterion is given by iterating until the percentage change in the size of the residuals between iterations is smaller than $\varepsilon$. Moreover, this is the convergence criterion being used in R statistical

software: $\frac{\|e^{(q+1)} - e^{(q)}\|}{e^{(q+1)}} < \varepsilon$, where $e = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix}$.

For all M-estimators (including MM-estimators), when the errors are really normally distributed, then the asymptotic variance is given by:

$$V(\psi, \Phi) = \hat{\sigma}^2 \frac{A(\psi, \Phi)}{B^2(\psi, \Phi)} \tag{3.12}$$

and the variance-covariance matrix of the estimated regression coefficients as:

$$V(\hat{\beta}) = \hat{\sigma}^2 \frac{A(\psi, \Phi)}{B^2(\psi, \Phi)} (X^T X)^{-1} \tag{3.13}$$

from the final IWLS fit. Where $A(\psi, \Phi) = E(\psi^2, \Phi)$ and $B^2(\psi, \Phi) = (E(\psi', \Phi))^2$

**Winsor's Principle**

This principle states that all probability distributions are normal in the middle. This implies that the $\psi$- function of m-estimators should be like the one that is optimal for Gaussian data in the middle. According to Bhar, a $\psi$-function that is linear in the middle is better efficient at the Gaussian distribution.

**Implications of Weight Functions**

In robust regression, making a choice of a weight function to apply to the scaled residuals is analogous to choosing a probability distribution function for the errors. In selecting functions that result in robust estimators, it is not necessary to assume distribution for the error terms. Hence in practice, weight functions are just chosen to apply without considering the associated probability of the errors, $f(e)$. Some statisticians suggested several weight functions, with their corresponding score function and objective function. For every function there is a tuning constant which helps to alter the shape of the function slightly, (Draper & Smith, 1998)(cited by Stuart, 2011).

**Properties of M-estimators**

The weight functions assign weights to the scaled residuals depending on the size of the scaled residuals. M-estimators assign reduced weights to unusual observations but the OLSE assigns weight one to all observations. As a result, OLSE is highly influenced by observations with large residuals than they affect the M-estimators. Because of this, M-estimators are more resistant to heavy-tailed error distributions and non-constant error variance. The type of weight function chosen determines how robust an estimator will be to outliers, and many studies were carried out on finding functions that make the associated M-estimator as robust as possible and still remains fairly efficient (Rousseeuw & Leroy 1987). An increase in tuning constants leads to an improvement in relative efficiency but reduces breakdown point of an estimator. Studies have come out with standard values for tuning constants, resulting in estimators with 95% asymptotic relative efficiency. The table below shows the three weight functions that will be used in this study, with their corresponding score functions and objective functions. The weight function of the OLSE, assigns weight of one (1) to every observation.

The Huber function is analogous to a probability distribution for the errors which is normal in the centre but appears like a double exponential distribution in the tails (Hogg, 1979). This function assigns weight of one (1) to those observations whose scaled residuals are within the central bound, whilst scaled residuals outside that region are given smaller weights. The Huber weight function considers every data point, as a result it does not give weight of zero to any scaled residuals. An estimator is referred to as "Redescending M-estimator" if it assigns weight zero (0) to scaled residuals, such that $\psi(u) = 0$ if $|u| > a$.

An example of redescending function is the Tukey bisquare weight function, also known as biweight function. It is used to compute an M-estimator that is more resistant to regression outliers than the Huber M-estimator and OLSE (Andersen, 2008). The figure 3.1 illustrates how the estimator assigns a smaller proportion of the scaled residuals weight of one (1) than the Huber M-estimator. The M-estimators are more resistant to vertical outliers than the least squares estimator, even though they have varying level of resistance to outliers.

Moreover, when the assumptions of the errors are met, M-estimators are 95% highly efficient as OLSE. On the Other hand, M-estimators are sensitive to leverage points like the least squares estimator. Therefore, M-estimators can breakdown completely in the presence of one bad leverage point, with a breakdown point of 0%. In solving the problem of low breakdown point, Rousseeuw and Yohai (1984) developed the S-estimator, which has a breakdown point of 50%.

Table 3.1: Some Popular functions for M-estimators

| | Objective Function $\rho(u)$ | Score Function $\psi(u)$ | Weight Function $w(u)$ |
|---|---|---|---|
| Least squares | $\frac{1}{2}u^2 \quad -\infty \leq u \leq \infty$ | u | 1 |
| Huber | $\begin{cases} \frac{1}{2}u^2 & \text{if } \lvert u\rvert < a \\ a\lvert u\rvert - \frac{1}{2}u^2 & \text{if } \lvert u\rvert \geq a \end{cases}$ | $\begin{cases} u & \text{if } \lvert u\rvert < a \\ a(sign\ u) & \text{if } \lvert u\rvert \geq a \end{cases}$ | $\begin{cases} 1 & \text{if } \lvert u\rvert < a \\ \frac{a}{\lvert u\rvert} & \text{if } \lvert u\rvert \geq a \end{cases}$ |
| Tukey bisquare | $\begin{cases} \frac{a^2}{6}(1-(1-(\frac{u}{a})^2)^2) & \text{if } \lvert u\rvert \leq a \\ \frac{1}{6}a^2 & \text{if } \lvert u\rvert > a \end{cases}$ | $\begin{cases} u(1-(\frac{u}{a})^2)^2 & \text{if } \lvert u\rvert \leq a \\ 0 & \text{if } \lvert u\rvert > a \end{cases}$ | $\begin{cases} (1-(\frac{u}{a})^2)^2 & \text{if } \lvert u\rvert \leq a \\ 0 & \text{if } \lvert u\rvert > a \end{cases}$ |



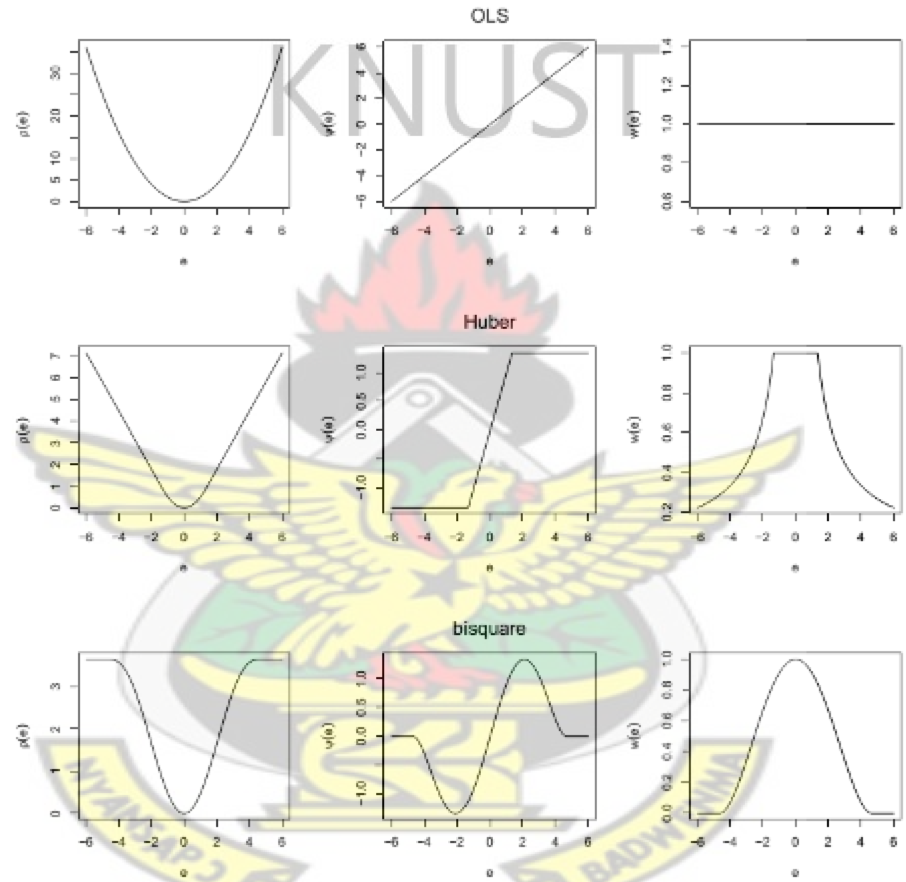


Figure 3.1: Objective, Influence and Weight Functions for Various Estimators

### 3.3.3 The S-Estimator (SE)

The S-estimation is a high breakdown method introduced by Rousseeuw and Yohai (1984) that minimizes the dispersion of the residuals. The S-estimator was introduced to take care of the low breakdown point of the M-estimators. The high breakdown S-estimator possesses a desirable property, that is it is affine, scale and regression equivariant, (Matias and Yohai, 2006). As the least squares estimator minimizes the variance of the residuals, S-estimator minimizes the dispersion of the scaled residuals or S-estimates are the solution that finds the smallest possible dispersion of the residuals $s(r(\beta_1), .....r(\beta_n)$. The robust S-estimation minimizes a robust M-estimate of the residual scale

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{s}\right) = k \tag{3.14}$$

Differentiating 3.14 we obtain the estimating equations for S-estimator:

$$\frac{1}{n}\sum_{i=1}^{n}x_i\psi\left(\frac{e_i}{s}\right) = 0 \tag{3.15}$$

Therefore the value of $\beta$ that minimizes s is the S-estimator. where $\psi$ is replaced with an appropriate weight function. Huber weight function or the biweight function is usually used as with most M-estimation procedures. Although S-estimates have a breakdown point of Break Down Point (BDP)=0.5, it comes at the cost of a very low relative efficiency (Verardi and Croux, 2009).

The choice of the tuning constant is a=1.548 and k=0.1995 for 50% breakdown and about 29% asymptotic efficiency. To increase the efficiency of the S-estimator, if a = 5.182, the Gaussian efficiency rises to 96.6% and unfortunately the breakdown point drops to 10%. Tradeoffs breakdown and efficiency are based on the selection of tuning constant a, and k. The final scale estimate, s, is the standard deviation of the residuals from the fit that minimized the dispersion of the residuals. where k is a constant and the objective function $\rho$ satisfies the following conditions:

1. $\rho$ is symmetric, continuously differentiable and $\rho(0) = 0$.
2. There exists $a > 0$ such that $\rho$ is strictly increasing on [0, a] and constant on $[a, \infty)$.
3. $\dfrac{k}{\rho(a)} = \dfrac{1}{2}$

Therefore an S-estimator is the estimator $\hat{\beta}$ that is the solution of 3.14, with s being the smallest. The second condition on the objective function means that the associated score function will be redescending. The Tukey bisquare weight function in Table 3.1 is usually used to achieve optimal weights assignments. To obtain a breakdown point of 50%, the third condition is required even though it is not strictly necessary, (Stuart, 2011). The choice of k is done so that the resulting s is can estimate the $\sigma$ when the errors are normally distributed. To do this, we set k such that $k = E_{\phi}(\rho(u))$, which is the expected value of the objective function if it is assumed that u has a standard normal distribution (Rousseeuw & Leroy, 1987). To use the Tukey bisquare objective function, Rousseeuw and Yohai (1984) stated that if we set the tuning constant a = 1.547, the third condition is satisfied, and hence makes the S-estimator has 50% BDP. The Tukey bisquare objective function is given by:

$$\rho(u) = \begin{cases} \dfrac{u^2}{2} - \dfrac{u^4}{2a^2} + \dfrac{u^6}{6a^4} & \text{if } |u| \leq a \\ \dfrac{a^6}{6} & \text{if } |u| > a \end{cases}$$

For $u \sim N(0, 1)$ the probability density function is

$$f(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right).$$

Hence with $a = 1.547$, and $k = E_{\phi}(\rho(u))$, we demonstrate the third condition as follows:

**Proof 3.3.2**

$$
\begin{aligned}
k &= E_\phi(\rho(u)) \\
&= \int_{-\infty}^{\infty} \rho(u) f(u) dx \\
&= 2\int_{0}^{\infty} \rho(u) f(u) dx \\
&= 2\int_{0}^{\infty} \rho(u) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) dx \\
&= 2\int_{0}^{1.547} (\frac{u^2}{2} - \frac{u^4}{2(1.547)^2} + \frac{u^6}{6(1.547)^4}) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) dx \\
&+ 2\int_{1.547}^{\infty} \frac{1.547^2}{6} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) dx \\
&= \sqrt{\frac{2}{\pi}} \int_{0}^{1.547} (\frac{u^2}{2} - \frac{u^4}{4.786} + \frac{u^6}{34.365}) \exp\left(-\frac{u^2}{2}\right) dx \\
&+ \int_{1.547}^{\infty} \frac{1.547^2}{3} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) dx \\
&= \sqrt{\frac{2}{\pi}} \int_{0}^{1.547} (\frac{u^2}{2} - \frac{u^4}{4.786} + \frac{u^6}{34.365}) \exp\left(-\frac{u^2}{2}\right) dx + 0.798 \times (1 - \phi(1.547)) \\
&= \sqrt{\frac{2}{\pi}} \times 0.189 + 0.798 \times (1 - 0.939) \\
&= 0.199
\end{aligned}
\tag{3.16}
$$

The $\rho(1.547) = \dfrac{1.547^2}{6} = 0.399$. This implies that $\dfrac{k}{\rho(a)} = \dfrac{E_\phi(\rho(u))}{\rho(1.547)} = \dfrac{1}{2}$. Rousseeuw and Leroy (1987), elaborated more on the breakdown point (BDP) of S-estimator. Consequently, if the objective function of S-estimator satisfies the three conditions, then the S-estimator has BDP of:

$$\frac{\frac{n}{2} - p + 2}{n}$$

which only tends to 0.5 as $n \to \infty$. According to Rousseeuw and Leroy (1987), if the third condition is rewritten as

$$\frac{k}{\rho(a)} = \alpha$$

where $0 < \alpha \leq 0.5$, then the S-estimator would have a breakdown point approaching $\alpha$ as $n \to \infty$. Since the asymptotic efficiency of the S-estimator depends on the objective function, the tuning constants of this function cannot be chosen to give the estimator high breakdown point and high asymptotic efficiency simultaneously.

Table 3.2: Tradeoff between BDP and Efficiency as a result of Choice of Tuning constant for the S-Estimator

| Tuning Constant (a) | $k = E_\phi(\rho(u))$ | Breakdown point | Efficiency |
|---|---|---|---|
| 1.547 | 0.1995 | 50% | 28.7% |
| 1.756 | 0.2312 | 45% | 37.0% |
| 1.988 | 0.2634 | 40% | 46.2% |
| 2.251 | 0.2957 | 35% | 56.0% |
| 2.560 | 0.3278 | 30% | 66.1% |
| 2.973 | 0.3593 | 25% | 75.9% |
| 3.420 | 0.3899 | 20% | 84.7% |
| 4.096 | 0.4194 | 15% | 91.7% |
| 5.182 | 0.4475 | 10% | 96.6% |

Table 3.2 shows a summary of the effect of the choice of tuning constant on the BDP and efficiency of an S-estimator defined using the Tukey bisquare weight function, (Stuart, 2011). The low efficiency of the S-estimator when it achieves a high breakdown point makes it unfit for robust regression estimator. However, the high breakdown estimator is needed as an initial estimate for more robust regression estimation processes, since the resulting estimators inherit its high breakdown point (Rousseeuw & Leroy, 1987). Since the S-estimator is less efficient as compared to many other estimators, the use of S-estimator as a standalone estimator is limited. However, because it is highly resistant to outliers, the S-estimator plays a key role in computing the MM-estimator, which is far more efficient. A study by Pena and Yohai (1996) showed that this estimator has a very high computational complexity and therefore there are algorithms that compute only approximate solutions.

### 3.3.4 The Modified Maximum likelihood Estimator (MME)

The MM estimation is a special type of M-estimation developed by Yohai (1987). MM-estimation is a combination of high breakdown value estimation and efficient estimation. Yohai's MM estimator was the first estimator with a high breakdown point and high efficiency under normal error. MM-estimator has three stage procedures. Yohai (1987) describes the three stages that define an MM-estimator:

1. A high breakdown estimator is used to find an initial estimate, which we denote $\tilde{\beta}$. The estimator need not be efficient. Using this estimate the residuals, $r_i(\tilde{\beta}) = y_i - x_i^T\tilde{\beta}$, are computed.

2. Using these residuals from the robust fit and 3.14, an M-estimate of scale with 50% BDP is computed. This $s(r_1(\tilde{\beta})...r_n(\tilde{\beta}))$ is denoted $s_n$. The objective function used in this stage is labeled $\rho_0$.

3. The MM-estimator is now defined as an M-estimator of $\beta$ using a redescending score function, $\psi_1(u) = \dfrac{\partial \rho_1(u)}{\partial u}$, and the scale estimate $s_n$ obtained from Stage 2. So an MM-estimator $\hat{\beta}$ is defined as a solution to

$$\sum_{i=1}^{n} x_{ij}\psi_1\left(\frac{y_i - x_i^T\beta}{s_n}\right) = 0, \text{ where } j = 1, ..., p \tag{3.17}$$

The objective function $\rho_1$ associated with this score function does not have to be the same as $\rho_0$ but it must satisfy the following conditions:

(a) $\rho$ is symmetric and continuously differentiable, and $\rho(0) = 0$.

(b) There exists $a > 0$ such that $\rho$ is strictly increasing on $[0, a]$ and constant on $[a, \infty)$.

(c) $\rho_1(u) \leq \rho_0(u)$

A final condition that must be satisfied by the solution to 3.17 is that

$$\sum_{i=1}^{n} x_{ij}\psi_1\left(\frac{y_i - x_i^T\hat{\beta}}{s_n}\right) \leq \sum_{i=1}^{n} x_{ij}\psi_1\left(\frac{y_i - x_i^T\tilde{\beta}}{s_n}\right) \tag{3.18}$$

### Properties of MM-estimators

The first two stages of the MM-estimation process are responsible for the estimator having high breakdown point, whilst the third stage aims for high asymptotic relative efficiency. This is why $\rho_0$ and $\rho_1$ need not be the same, and why the estimator chosen in stage 2 can be inefficient. Yohai (1987) showed that when estimating MM-estimator, using an estimator with 50% BDP at the first stage will result in the final MM-estimator has 50% BDP. The MM-estimator is very resistant to multiple leverage points and vertical outliers. The MME is also equivariant and hence it transforms 'properly' in some sense (Rousseeuw & Leroy, 1987).

### Finite sample breakdown point of an estimator

The Breakdown point (BDP) of an estimator is a measure of how much an estimator is able to withstand the presence of outliers or errors in a dataset, Andersen (2008) (cited by Stuart, 2011) . Moreover, the BDP of a regression estimator is the smallest proportion of contamination that results in an estimator no longer depicts the general trend in the bulk of the data. An estimator that breaks down due to contamination produces estimates that differ highly from the estimates from the uncontaminated data. Let T be a regression estimator, Z as a sample of size n, and $T(Z) = \hat{\beta}$. Also, let $Z'$ be a contaminated sample where m of the original data points are replaced with arbitrary values. The maximum effect that could be caused by such contamination is

$$\text{effect}(m; T, Z) = sup_{Z'} \parallel T(Z') - T(Z) \parallel \tag{3.19}$$

When 3.19 is infinite, m outliers can have an arbitrarily large effect on T. Then the breakdown point of T at Z is defined as:

$$\text{BDP}(T, Z) = min\{\frac{m}{n} : \text{effect}(m; T, Z) = \infty\} \tag{3.20}$$

Therefore, it is the maximum fraction of contamination that is allowed before $\hat{\theta}$ can take on any value depending on G (Jann, 2012).

### Exact fit property (EFP)

One important robustness property of MM-estimator used by Rousseeuw(1984) is called exact fit property(EFP). An estimator $T_n$ has EFP if given any sample of size n, $(y_1, x_1, ....y_n, x_n)$ for which there exist $\beta$ such that $\#\{i : y_i = \beta^T x_i\} > \dfrac{n}{2}$, then $\#\{i : y_i = T_n^T x_i\} > \dfrac{n}{2}$ too. According to Yohai (1985), the following theorem shows MME inherits the EFP from the initial estimate.

**Theorem 3.3.1** *Assume $\rho_0$ and $\rho_1$ satisfy all their assumptions. Suppose $T_0$ has the EFP and let $T_1$ be any estimate satisfying 3.18. Then $T_1$ has EFP too.*

### Remark

The S-estimates has the EFP and therefore if S-estimates are chosen to be $T_0$, the MM-estimates $T_1$ will also have the EFP.

### Consistency

Theorems 3.2 and 3.3 establish the consistency of the scale estimate $s_n$ defined in stage 2 and of the MM-sequence of estimates $T_1$ of $\theta_0$. In order to prove consistency of MM-estimator we need the following assumptions.

$$\begin{gathered}\text{The function } g(a) = E_{F_0}(\rho_1((u-a)/\sigma)),\ \text{where } \sigma_0 \text{ is defined} \\ E_{F_0}(\rho_1((u)/\sigma)) = K,\ \text{has a unique minimum at } a = 0\end{gathered} \tag{3.21}$$

$$P_{G_0}(\theta' x = 0) < 0.5 \text{ for all } \theta \epsilon R^p \tag{3.22}$$

The error distribution $F_0$ has density $f_0$ with the following properties:

$$\begin{gathered}(i)\ f_0 \text{ is even, } (ii)\ f_0(u) \text{ is monotone non-increasing in } |u|\,,\ \text{and} \\ (iii)\ f_0(u) \text{ is strictly decreasing in } |u| \text{ in the neighborhood } 0.\end{gathered} \tag{3.23}$$

$$\frac{k}{b} = 0.5, \text{where } k = E_\phi(\rho(u)) \text{ and } b = max\ \rho_0(u) \tag{3.24}$$

**Theorem 3.3.2** *Let $(y_1, x_1), ...., (y_1, x_1)$ be i.i.d. observations with distribution F given by 3.3. Assume that $\rho_0$ satisfies 3.8, $T_0$ is a sequence of estimates which is strongly consistent for $\theta_0$. Then $s_n$ is strongly consistent for $\sigma_0$ defined by*

$$E_F(\rho(u/\sigma_0)) = b \tag{3.25}$$

$$\rho_1(u) \leq \rho_0(u) \tag{3.26}$$

$$sup\ \rho_1(u) = sup\ \rho_0(u) \tag{3.27}$$

**Theorem 3.3.3** *Let $(y_1, x_1), ...., (y_1, x_1)$ be i.i.d. observations with distribution F given by 3.3. Assume that $\rho_0$ and $\rho_1$ satisfies 3.8, that 3.24, 3.26, 3.27 hold, and F has satisfies 3.23 and $G_0$ 3.22. Assume also that a sequence $T_0$ is strongly consistent for $\theta_0$, then any other sequence $\{T_1\}$ which satisfies 3.18 is strongly consistent too.*

$$\begin{aligned} &\rho_1 \text{ is odd, twice continuously differentiable and there exists m such that} \\ &\qquad\qquad\qquad\qquad\qquad |u| \geq m \text{ implies that } \rho_1 = b \end{aligned} \tag{3.28}$$

$$G_0 \text{ has second moments and } V = E_{G_0}(x_i x_i^{'}) \text{ is non singular.} \tag{3.29}$$

The following theorem gives the asymptotic normality of M-estimates with scale estimated separately, which include MM-estimates.

**Theorem 3.3.4** *Let $(y_1, x_1), ...., (y_1, x_1)$ be i.i.d. observations with distribution F given by 3.3. Assume that $\rho_1$ satisfies 3.28 and G satisfies 3.29. Let $s_n$ be an estimate of the error scale which converges strongly to $\sigma_0$, let $\{T_n\}$ be the sequence which satisfies 3.17 and which is strongly consistent to the true value $\sigma_0$. Then*

$$n^{1/2}(T_n - \theta_0) \rightarrow_d N\left(0, \sigma_0^2 \left[\frac{A(\psi_1, F_0)}{[B^2(\psi_1, F_0)}\right] V^{-1}\right), \tag{3.30}$$

*where $\rightarrow_d$ denotes convergence in distribution,*

$$A(\psi_1, F_0) = E_F(\psi^2(u)) \tag{3.31}$$

$$B(\psi_1, F_0) = E_F(\psi'(u)) \tag{3.32}$$

**Relative efficiency of an estimator**

The efficiency of an estimator is its minimum possible variance to its actual variance, and this estimator is efficient when the ratio gives one (1). An estimator of a parameter is said to be efficient if it has the minimum variance among other estimators. And an estimator that reaches an acceptable level of efficiency with larger samples is called asymptotically efficient (Andersen, 2008). Moreover, it is important for an estimator to be highly efficient if the idea is to use the estimator to make an inferences about the whole population. Relative efficiency compares the efficiency of an estimator to that of a well known method. Given two estimators $T_1$ and $T_2$ for the parameter $\beta$ where $T_1$ is the most efficient and $T_2$ the less efficient then the relative efficiency of $T_2$ is computed as the ratio of its mean squared error to the mean squared error of $T_1$, (Andersen, 2008). It is given as: $\text{Efficiency}(T_1, T_2) = \dfrac{E(T_1 - \beta)^2}{E(T_2 - \beta)^2}$.
Usually, estimators are compared to the OLSE since it provides the minimum variance estimate, when its assumptions are met, hence it is the most efficient estimator known. According to Andersen (2008), relative efficiency is usually calculated in terms of asymptotic efficiency, so for small samples it is not necessarily a relevant property to consider. Indeed, the relative efficiency of a regression estimator is given as the ratio between its mean squared error and that of the least squares estimator, for an infinite sample with normally distributed errors.

**The Power of a test**

The Power of the test is the probability or the chance of correctly rejecting the Null hypothesis. According to Cohen (1992) the Power analysis for simple or multiple linear regression requires parameters such as Probability of making type

I error ($\alpha$), the numerator and the denominator degrees of freedom, and the effect size, $f^2 = \dfrac{R^2}{1-R^2}$ as inputs. Where the coefficient of determination ($R^2$) is given by;

$$R^2 = \frac{\text{Regression Sum of Squares (SSR)}}{\text{Total Sum of Squares (TSS)}} \tag{3.33}$$

for the ordinary least squares estimator and according to Aelst (2014),

$$R^2 = 1 - \frac{Median(residual^2)}{MAD^2(y)} \tag{3.34}$$

for the robust methods. The effect size is defined as the extent to which the Null hypothesis is false (Cohen, 1992). Post hoc Power analysis is very important and helpful when we fail to reject the Null hypothesis. This is helpful because, it is done to evaluate or to check if failing to reject Null hypothesis could result in a type II error. The Power of the test is affected by: type I error, type II error, sample size and the effect size. Moreover, type I error, sample size and effect size are positively associated to Power, whilst type II error has an inverse relationship with Power of the test.

# Chapter 4

# Analysis

## 4.1 Introduction

This chapter contains the results of our study on comparison of robust regression estimators. Estimators such as Least trimmed squares Estimator (LTSE), Huber Maximum likelihood Estimator (HME), S-Estimator (SE)and Modified Maximum likelihood Estimator (MME) were compared with the Ordinary Least Squares Estimator(OLSE). The analysis was carried out with Total Body fat (bodyfat) as the response variable and four independent variables: Body Mass Index (BMI), Triceps skin-fold (TS), Arm Fat as percent composition of the body (parmfat) and Height of the respondents. Leverages were introduced first into two variables, and finally into all predictors. The percentages were 5%, 10% and 15 % leverages to observe the robustness of the various estimators. Also, 10%, 20% and 30% outliers were introduced in addition to 20% error contamination and data from non-normal distribution were considered to examine how the robust methods would perform when exposed to these aberrations. On Comparing the estimators, coefficients and their standard errors, Residual standard errors (The Root Mean Square Error (RMSE)), Relative efficiencies, Coefficients of determination and the Power of the test for the estimators were used as criteria. The results were presented in a series of tables, for the coefficients and their standard errors were reported as one set of tables, and Residual standard errors (The Root Mean Square Error (RMSE)), Relative efficiencies, Coefficients of determination and the Power of the test of the estimators were also reported as another set of tables. We begin with the results for the dataset with normal errors.

## 4.2 Original dataset with normal Errors

Table 4.1 below contains the estimate of the model parameters and standard errors when the errors are normally distributed. The Residual standard errors, Relative efficiencies, Coefficients of determination and the Power of the tests are presented in Table 4.2 for the normally distributed residuals.

Table 4.1: The coefficients (standard errors) of the estimators for original dataset with normal errors

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 7.3615(0.9448) | 0.8452(0.0555) | 0.1452(0.0824) | 0.0040(0.0104) | 0.2880(0.0184) |
| LTSE | 7.5438(0.8715) | 0.8298(0.0514) | 0.2038(0.0763) | 0.0021(0.0095) | 0.2811(0.0171) |
| HME | 7.4065(1.0123) | 0.8395(0.0594) | 0.1633(0.0883) | 0.0036(0.0111) | 0.2870(0.0198) |
| SE | 7.4543(1.0070) | 0.8322(0.0591) | 0.1726(0.0878) | 0.0045(0.0110) | 0.2865(0.0197) |
| MME | 7.4192(1.0084) | 0.8360(0.0592) | 0.1652(0.0879) | 0.0043(0.0111) | 0.2869(0.0197) |

Table 4.2: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and the Power of the test for original dataset with normal errors

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 1.0650 | 1.0000 | 0.9696 | 1.0000 |
| LTSE | 0.9641 | 1.2203 | 0.9641 | 1.0000 |
| HME | 1.2050 | 0.7811 | 0.9574 | 1.0000 |
| SE | 1.0720 | 0.9870 | 0.9586 | 1.0000 |
| MME | 1.0670 | 0.9963 | 0.9577 | 1.0000 |

From Table 4.1 it is observed that all the estimators perform well, since the errors are normally distributed. This confirmed the saying that, all estimators perform well under normal errors. The The Root Mean Square Error, Relative efficiencies and the coefficients of determination from Table 4.2, also showed that when the errors are normal, all the estimators do well.

## 4.3 5% leverages in BMI and parmfat

Perturbing the Body Mass Index (BMI) and Arm Fat as percent composition of the body (parmfat) reported the following values in Tables 4.3 and 4.4.

Table 4.3: The coefficients (Standard error) of the estimators for 5% leverages in BMI and parmfat

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 4.0169(2.5430) | -0.0413(0.0379) | 0.1041(0.0296) | 0.1433(0.0233) | 0.4918(0.0440) |
| LTSE | 7.8980(1.4995) | 0.8173(0.0660) | 0.1654(0.0881) | 0.2948(0.0200) | 0.2948(0.0200) |
| HME | -3.3689(2.2848) | -0.0108(0.0341) | 0.0719(0.0266) | 0.2501(0.0210) | 0.4222(0.0396) |
| SE | 6.3643(1.0817) | 0.9199(0.0161) | 0.0031(0.0126) | 0.0035(0.0099) | 0.2879(0.0187) |
| MME | 6.2232(1.0822) | 0.9165(0.0161) | 0.0055(0.0126) | 0.0034(0.0099) | 0.2880(0.0187) |

Table 4.4: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 5% leverages in BMI and parmfat

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 2.8030 | 1.0000 | 0.7893 | 1.0000 |
| LTSE | 1.0770 | 6.7735 | 0.9669 | 1.0000 |
| HME | 2.1810 | 1.6517 | 0.8611 | 1.0000 |
| SE | 1.2860 | 4.7508 | 0.9458 | 1.0000 |
| MME | 1.2860 | 4.7508 | 0.9466 | 1.0000 |

When we perturbed BMI and parmfat with 5% leverages, Tables 4.3 and 4.4 showed that some of the estimators broke-down. Estimators like OLSE and HME assumed values which are quite different from when the errors were normal. Considering all the criteria for the comparison, OLSE and HME were affected with 5% leverages. The residual standard errors of these two estimators were inflated, which led to small relative efficiency of these methods. Also, OLSE and HME assumed negative values for some coefficients.

## 4.4 5% leverages in height and TS

The Tables 4.5 and 4.6 below contain the reported statistics from 5% leverages in height and triceps skin-fold (TS) for comparing the regression estimators.

Table 4.5: The coefficients (Standard error) of the estimators for 5% leverages in height and TS

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 10.0350(1.4123) | 1.0915(0.0800) | 0.1476(0.1283) | -0.0544(0.0134) | 0.1194(0.0207) |
| LTSE | 13.4912(1.2982) | 1.1018(0.0617) | 0.1346(0.0983) | -0.1068(0.0150) | 0.2126(0.0210) |
| HME | 12.3368(1.2395) | 1.0939(0.0702) | 0.1594(0.1126) | -0.0888(0.0117) | 0.1749(0.0181) |
| SE | 7.3397(1.0005) | 0.8457(0.0567) | 0.1515(0.0909) | 0.0042(0.0095) | 0.2854(0.0146) |
| MME | 7.3242(1.0006) | 0.8472(0.0567) | 0.1489(0.0909) | 0.0040(0.0095) | 0.2858(0.0146) |

Table 4.6: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test 5% leverages in height and TS

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 1.6570 | 1.0000 | 0.9238 | 1.0000 |
| LTSE | 1.2610 | 1.7267 | 0.9533 | 1.0000 |
| HME | 1.6620 | 0.9940 | 0.9195 | 1.0000 |
| SE | 1.2210 | 1.8417 | 0.9519 | 1.0000 |
| MME | 1.2210 | 1.8417 | 0.9520 | 1.0000 |

In this section, we examined the effects of 5% leverages in height and TS on the estimators. Leverages in height and TS made some estimators to breakdown slightly. The coefficients of height for these estimators have negative relationship with total body fat. OLSE and HME are the estimators that were really affected, since they had low relative efficiencies and coefficients of determination. The intercepts of OLSE, HME and LTSE were inflated, and the coefficients of height for these estimators have negative signs.

## 4.5 5% leverages in BMI, parmfat, height and TS

The results of perturbing all predictors with 5% leverages are shown in Tables 4.7 and 4.8.

Table 4.7: The coefficients (Standard error) of the estimators for 5% leverages in all predictors

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 11.6368(4.1695) | -0.4314(0.1059) | 0.2459(0.0465) | 0.0849(0.0429) | 0.2541(0.0845) |
| LTSE | 11.9490(3.5366) | -0.4201(0.0921) | 0.2430(0.0392) | 0.0669(0.0363) | 0.3187(0.0737) |
| HME | 11.4088(3.9933) | -0.3654(0.1014) | 0.2300(0.0445) | 0.0789(0.0410) | 0.2802(0.0809) |
| SE | 6.4193(1.0667) | 0.9207(0.0271) | 0.0023(0.0119) | 0.0037(0.0110) | 0.2877(0.0216) |
| MME | 6.2231(1.0672) | 0.9165(0.0271) | 0.0055(0.0119) | 0.0034(0.0110) | 0.2880(0.0216) |

Table 4.8: The Root Mean Square Error (RMSE), Relative Efficiency Coefficient of Determination and Power of the test for 5% leverages in all predictors

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 4.6600 | 1.0000 | 0.4176 | 1.0000 |
| LTSE | 3.9020 | 1.4263 | 0.5603 | 1.0000 |
| HME | 4.6860 | 0.9889 | 0.3597 | 1.0000 |
| SE | 1.2860 | 13.1308 | 0.9458 | 1.0000 |
| MME | 1.2860 | 13.1308 | 0.9466 | 1.0000 |

With 5% leverages in all the independent variables, we can see that, MME and SE resisted the influence of the leverages. However, OLSE and HME performed badly. This is as a result of the fact that HME and OLSE lack the resistance to leverages. The intercepts of OLSE, LTSE and HME were affected as well as their standard errors. These estimators also have the coefficients of BMI to be negative which were positive for normal errors. Again, these estimators have large residual standard errors which had affected their relative efficiencies and the coefficients of determination.

## 4.6 10% leverages in BMI, parmfat

The results in Tables 4.9 and 4.10 below are the estimates of the regression parameters computed for 10% leverages in BMI and parmfat.

Table 4.9: The coefficients (Standard error) of the estimators for 10% leverages in BMI, parmfat

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 4.0121(2.6451) | -0.0619(0.0370) | 0.0996(0.0317) | 0.1502(0.0236) | 0.5066(0.0445) |
| LTSE | -4.2087(2.5373) | -0.0262(0.0287) | 0.0591(0.0250) | 0.2746(0.0302) | 0.4138(0.0353) |
| HME | -4.2540(2.3720) | -0.0250(0.0332) | 0.0582(0.0284) | 0.2736(0.0212) | 0.4302(0.0399) |
| SE | 6.6831(1.1291) | 0.9361(0.0158) | -0.0049(0.0135) | 0.0033(0.0101) | 0.2904(0.0190) |
| MME | 6.3714(1.6287) | 0.9260(0.0589) | 0.0012(0.0207) | 0.0028(0.0258) | 0.2909(0.0243) |

Table 4.10: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 10% leverages in BMI, parmfat

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 2.8610 | 1.0000 | 0.7805 | 1.0000 |
| LTSE | 2.0250 | 1.9961 | 0.8749 | 1.0000 |
| HME | 2.2940 | 1.5554 | 0.8466 | 1.0000 |
| SE | 1.3930 | 4.2183 | 0.9435 | 1.0000 |
| MME | 1.3640 | 4.3995 | 0.9457 | 1.0000 |

10% leverages in BMI and parmfat reduced the efficiency of some estimators. However, the MME and SE were not affected like other estimators. The leverages have affected the intercepts and the coefficients of BMI for OLSE, LTSE and HME. Also, they had high residual standard errors making fitted models provided by these methods unreliable.

## 4.7 10% leverages in height, TS

Tables 4.11 and 4.12 show the results of comparison of estimators when 10% leverages are introduced into height and triceps skin-fold (TS).

Table 4.11: The coefficients (Standard error) of the estimators for 10% leverages in height, TS

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 10.1343(1.4558) | 1.1135(0.0816) | 0.1498(0.1320) | -0.05813(0.0136) | 0.1028(0.0202) |
| LTSE | 10.4124(1.4373) | 1.0648(0.0809) | 0.2377(0.1292) | -0.0543(0.0134) | 0.0947(0.0199) |
| HME | 12.7850(1.3399) | 1.1197(0.0751) | 0.1781(0.1214) | -0.0978(0.0125) | 0.1554(0.0186) |
| SE | 7.2947(1.0260) | 0.8443(0.0575) | 0.1578(0.0930) | 0.0038(0.0096) | 0.2878(0.0142) |
| MME | 7.2565(1.0136) | 0.8494(0.0601) | 0.1464(0.0898) | 0.0034(0.0110) | 0.2891(0.0209) |

Table 4.12: The Root Mean Square Error (RMSE), Relative Efficiency Coefficient of Determination and Power of the test for 10% leverages in height, TS

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 1.7010 | 1.0000 | 0.9224 | 1.0000 |
| LTSE | 1.6150 | 1.1093 | 0.9284 | 1.0000 |
| HME | 1.7840 | 0.9091 | 0.9072 | 1.0000 |
| SE | 1.3590 | 1.5666 | 0.9432 | 1.0000 |
| MME | 1.3460 | 1.5970 | 0.9422 | 1.0000 |

The S-estimator and the modified maximum likelihood estimator have been consistent in being robust to the effects of the leverages in the dataset. Moreover, their results have not differed from when the errors were normal. However, Ordinary least squares estimator and Huber Maximum likelihood Estimator have since been affected by the leverages. Least Trimmed Squares Estimator assumes different values some times because of the trimming.

## 4.8 10% leverages in BMI, parmfat, height and TS

Tables 4.13 and 4.14 display the results for 10% leverages in all predictors.

Table 4.13: The coefficients (Standard error) of the estimators for 10% leverages in BMI, parmfat, height and TS

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 13.0788(4.5332) | -0.4754(0.1087) | 0.2703(0.0518) | 0.0685(0.0450) | 0.2114(0.0902) |
| LTSE | 9.9738(5.1715) | -0.4688(0.1166) | 0.3190(0.0653) | 0.0590(0.0460) | 0.1630(0.0940) |
| HME | 12.8249(4.5043) | -0.4736(0.1080) | 0.2638(0.0515) | 0.0703(0.0447) | 0.2407(0.0897) |
| SE | 6.7512(1.1076) | 0.9377(0.0266) | -0.0063(0.0127) | 0.0036(0.0110) | 0.2903(0.0220) |
| MME | 6.3714(1.6287) | 0.9260(0.0589) | 0.0012(0.0207) | 0.0028(0.0258) | 0.2909(0.0243) |

Table 4.14: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 10% leverages in BMI, parmfat, height and TS

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 4.9930 | 1.0000 | 0.3314 | 0.9900 |
| LTSE | 4.9760 | 1.0068 | 0.3209 | 1.0000 |
| HME | 5.3110 | 0.8838 | 0.1776 | 0.9900 |
| SE | 1.4050 | 12.6290 | 0.9432 | 1.0000 |
| MME | 1.3640 | 13.3997 | 0.9457 | 1.0000 |

The intercept was largely affected by leverages in all the independent variables and the values of coefficients of some estimators have also changed significantly. HME performed like the OLSE because, HME does not bound the effect of leverages. These estimators report results which are the average of good and bad data points. According to a study by Alma (2011), the SE performed better than the MME because, MME has problems with high leverages in small sample datasets. However, in this study, MME counteracts the effects of leverages.

## 4.9 5% leverages in BMI, parmfat and 10% outliers

The results presented in the Tables 4.15 and 4.16 below show the performances of the estimators when there are 5% leverages in BMI and parmfat, and 10% vertical outliers.

Table 4.15: The coefficients (Standard error) of the estimators for 5% leverages in BMI, parmfat and 10% outliers

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 1.4201(9.8870) | 0.5970(0.1474) | 0.2169(0.1151) | -0.0217(0.0907) | 0.1670(0.1712) |
| LTSE | 6.4701(1.2791) | 0.8403(0.0194) | 0.0191(0.0153) | 0.0084(0.0116) | 0.2855(0.0219) |
| HME | 5.9785(1.2898) | 0.8588(0.0192) | 0.0224(0.0150) | 0.0066(0.0118) | 0.2831(0.0223) |
| SE | 6.7819(1.1152) | 0.9005(0.0166) | -0.0005(0.0130) | 0.0066(0.0102) | 0.2955(0.0193) |
| MME | 6.3948(1.3226) | 0.8999(0.0233) | 0.0047(0.0181) | 0.0056(0.0112) | 0.2939(0.0213) |

Table 4.16: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 5% leverages in BMI, parmfat and 10% outliers

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 10.9000 | 1.0000 | 0.5432 | 1.0000 |
| LTSE | 1.3780 | 62.5683 | 0.9882 | 1.0000 |
| HME | 1.3440 | 65.7740 | 0.9643 | 1.0000 |
| SE | 1.3180 | 68.3947 | 0.9668 | 1.0000 |
| MME | 1.3030 | 69.9784 | 0.9668 | 1.0000 |

5 % leverages and 10% outliers affected the OLSE more than the other estimators. Methods such SE and MME were able to bound the effects of both leverages and outliers. HME was also able to resist the influence of outliers but was not able to withstand the effects of leverages. Using coefficient of determination for the comparison, the LTSE performed better than other estimators. MME reported the most reliable RMSE and relative efficiency.

## 4.10 5% leverages in height, TS and 10% outliers

The various numerical measures computed for estimators when there are 5 % leverages in height and TS, and 10% outliers are presented in the tables 4.17 and 4.18

Table 4.17: The coefficients (Standard error) of the estimators for 5% leverages in height, TS and 10% outliers

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 0.1911(9.7234) | 0.6262(0.5506) | -0.4570(0.8836) | 0.1469(0.0921) | 0.6012(0.1423) |
| LTSE | 0.3861(2.4603) | 0.2944(0.1396) | 0.0532(0.2242) | 0.1512(0.0232) | 0.6636(0.0360) |
| HME | -7.6135(2.3185) | 0.3118(0.1313) | -0.0205(0.2107) | 0.2585(0.0220) | 0.5246(0.0339) |
| SE | 7.2946(1.0218) | 0.8443(0.0579) | 0.1578(0.0929) | 0.0038(0.0097) | 0.2878(0.0150) |
| MME | 7.2565(1.0136) | 0.8494(0.0601) | 0.1464(0.0898) | 0.0034(0.0110) | 0.2891(0.0209) |

Table 4.18: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 5% leverages in height, TS and 10% outliers

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 11.4100 | 1.0000 | 0.4992 | 1.0000 |
| LTSE | 2.8710 | 15.7945 | 0.9486 | 1.0000 |
| HME | 2.5370 | 20.2269 | 0.8729 | 1.0000 |
| SE | 1.3590 | 70.4908 | 0.9616 | 1.0000 |
| MME | 1.3460 | 71.8590 | 0.9609 | 1.0000 |

MME and SE are still resistant to both outliers and leverages. These robust methods try to get models that fit the majority of the data, whilst OLSE provides models that fit the average of the data. As a result, OLSE is always affected by few unusual observations. The coefficients of OLSE and HME were largely influenced by the aberrations in the data at this level, whilst LTSE was slightly affected. Also, the standard error of the intercept revealed that the fitted model of the OLSE differed a lot from when the residuals were normal.

## 4.11 5% leverages in BMI, parmfat, height and TS and 10% outliers

The tables 4.19 and 4.20 below display how the estimators fair in the presence of 5% perturbations in all predictor variables and also 10% vertical outliers.

Table 4.19: The coefficients (Standard error) of the estimators for 5% leverages in BMI, parmfat, height and TS and 10% outliers

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 0.9480(9.7595) | 0.4336(0.2479) | 0.2387(0.1087) | 0.0047(0.1003) | 0.1697(0.1978) |
| LTSE | 6.8715(1.5254) | 0.5081 (0.0385) | 0.0717 (0.0174) | 0.0406 (0.0154) | 0.2939(0.0307) |
| HME | 6.3526(1.7249) | 0.5091(0.0438) | 0.0848(0.0192) | 0.0365(0.0177) | 0.2821(0.0350) |
| SE | 6.5114(1.0915) | 0.9337(0.0277) | -0.0026(0.0122) | 0.0029(0.0112) | 0.2918(0.0221) |
| MME | 6.3714(1.6287) | 0.9260(0.0589) | 0.0012(0.0207) | 0.0028(0.0258) | 0.2909(0.0243) |

Table 4.20: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 5% leverages in BMI, parmfat, height and TS and 10% outliers

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 10.9100 | 1.0000 | 0.5423 | 1.0000 |
| LTSE | 1.6660 | 42.8844 | 0.9827 | 1.0000 |
| HME | 1.8690 | 34.0746 | 0.9310 | 1.0000 |
| SE | 1.4080 | 60.0405 | 0.9619 | 1.0000 |
| MME | 1.3640 | 63.9765 | 0.9632 | 1.0000 |

From tables 4.19 and 4.20, LTSE in addition to other robust methods performed better. LTSE does perform well when the trimming is done properly. The robust methods also performed well using; relative efficiencies, coefficients of determination, residual standard errors and power of the tests. On the contrary, the OLSE only did well in post hoc power of the test. The intercepts of all robust methods are similar and are close to that of the original data with normal errors. However, the intercept and its standard error for OLSE was largely affected to extent that the standard error assumed a very large value.

## 4.12 5% leverages in BMI and parmfat and 15% outliers

The numerical measures computed for the estimators with percentage of vertical outliers increased to 15% and 5% leverages in predictor variables; BMI and parmfat.

Table 4.21: The coefficients (Standard error) of the estimators for 5% leverages in BMI and parmfat and 15% outliers

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 1.7657(14.1125) | 0.7450(0.2104) | 0.2843(0.1643) | -0.1055(0.1294) | 0.1750(0.2443) |
| LTSE | 10.3727(1.6828) | 1.2738(0.0229) | -0.0026(0.0174) | -0.1070(0.0198) | 0.2131(0.0237) |
| HME | 8.0821(1.6947) | 1.1794(0.0253) | -0.0031(0.0197) | -0.0637(0.0155) | 0.2417(0.0293) |
| SE | 6.7893(1.1339) | 0.9031(0.0169) | -0.0032(0.0132) | 0.0077(0.0104) | 0.3008(0.0196) |
| MME | 6.3755(1.6066) | 0.8955(0.0271) | 0.0041(0.0220) | 0.0065(0.0202) | 0.3008(0.0241) |

Table 4.22: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test 5% leverages in BMI and parmfat and 15% outliers

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 15.5600 | 1.0000 | 0.4699 | 1.0000 |
| LTSE | 1.3880 | 125.6725 | 0.9909 | 1.0000 |
| HME | 1.6410 | 89.9088 | 0.9572 | 1.0000 |
| SE | 1.4510 | 114.9964 | 0.9670 | 1.0000 |
| MME | 1.3930 | 124.7719 | 0.9674 | 1.0000 |

When the percentage of outliers increased, only OLSE is very much affected by using Coefficient of determination from Table 4.22. Moreover, using other criteria, the other estimators are also affected but not as compared to the OLSE. The coefficient of parmfat for LTSE, HME and SE were assumed negative values. Also, the coefficient of height for OLSE, LTSE and HME assumed different values due to the perturbations in the dataset. By comparing the estimators using coefficients of determination, relative efficiencies, standard error and power of the tests, LTSE performed better than the other estimators.

## 4.13 5% leverages in height and TS and 15% outliers

Tables 4.23 and 4.24 below showed the results for the comparison of the estimators, when there are 5% leverages in height and TS and 15% outliers.

Table 4.23: The coefficients (Standard error) of the estimators for leverages in height and TS and 15% outliers

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 1.6230(13.9801) | 0.7005(0.7916) | -0.4677(1.2705) | 0.0912(0.1324) | 0.8494(0.2046) |
| LTSE | -2.6700(3.1505) | 0.6442(0.1890) | -0.5709(0.3047) | 0.1188(0.0317) | 0.7916(0.0525) |
| HME | -6.5925(2.8160) | 0.3711(0.1595) | -0.0761(0.2559) | 0.2299(0.0267) | 0.5803(0.0412) |
| SE | 7.4203(1.0439) | 0.8350(0.0591) | 0.1896(0.0949) | 0.0024(0.0099) | 0.2883(0.0153) |
| MME | 7.4208(1.0234) | 0.8388(0.0601) | 0.1810(0.0912) | 0.0019(0.0110) | 0.2896(0.0210) |

Table 4.24: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for leverages in height and TS and 15% outliers

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 16.4100 | 1.0000 | 0.4103 | 1.0000 |
| LTSE | 3.6240 | 20.5041 | 0.9206 | 1.0000 |
| HME | 2.8310 | 33.5999 | 0.8726 | 1.0000 |
| SE | 1.4790 | 123.1065 | 0.9643 | 1.0000 |
| MME | 1.4620 | 125.9860 | 0.9637 | 1.0000 |

By using $R^2$, OLSE was affected from Table 4.24. In addition, by using coefficients, only MME and SE were not affected. The coefficients of parmfat for OLSE, LTSE and HME differed a lot from that of the normal errors. Also, vertical outliers in the data has influenced the standard error of the OLSE. MME and SE perform better than the remaining estimators.

## 4.14 5% leverages in BMI, parmfat, height and TS and 15% outliers

The numerical measures (criteria) for comparing the Ordinary least squares estimator to the robust methods, when all predictors contain 5% leverages and the response variable contains 15% vertical outliers are listed in Tables 4.25 and 4.26.

Table 4.25: The coefficients (Standard error) of the estimators for 5% leverages in BMI, parmfat, height and TS and 15% outliers

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 4.2019(13.8930) | 0.8002(0.3529) | 0.2962(0.1548) | -0.1496(0.1428) | 0.1464(0.2816) |
| LTSE | 7.5769(1.5162) | 1.0886(0.0321) | -0.0069(0.0136) | -0.0422(0.0218) | 0.2864(0.0214) |
| HME | 6.3722(1.4537) | 1.0238(0.0369) | 0.0017(0.0162) | -0.0193(0.0149) | 0.2841(0.0295) |
| SE | 6.6606(1.1539) | 1.0724(0.0293) | -0.0290(0.0129) | -0.0093(0.0119) | 0.2901(0.0234) |
| MME | 6.3643(1.2301) | 1.0690 (0.0305) | -0.0224(0.0148) | -0.0114 (0.0118) | 0.2878(0.0233) |

Table 4.26: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 5% leverages in BMI, parmfat, height and TS; and 15% outliers

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 15.5300 | 1.0000 | 0.4717 | 1.0000 |
| LTSE | 1.1450 | 183.9636 | 0.9938 | 1.0000 |
| HME | 1.6450 | 89.1274 | 0.9570 | 1.0000 |
| SE | 1.4540 | 114.0812 | 0.9642 | 1.0000 |
| MME | 1.4480 | 115.0286 | 0.9647 | 1.0000 |

From Tables 4.25 and 4.26, the LTSE performed better than all estimators when all independent variables are perturbed with 5% leverages and 15% outliers. The LTSE reported the highest coefficient of determination and the lowest RMSE. It also performed better using the relative efficiency and the coefficients. This is because, some of its coefficients have not differed from when the errors were normally distributed.

## 4.15 10% outliers Perturbation

Presented below are Tables 4.27 and 4.28 which contain the measures for comparing the estimators for 10% outliers in the response variable and no leverages.

Table 4.27: The coefficients (Standard error) of the estimators for 10% outliers Perturbation

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 5.6421(13.0350) | 1.5801(0.7652) | -0.5968(1.1366) | -0.0480(0.1429) | 0.1941(0.2545) |
| LTSE | 7.2565(0.9772) | 0.8494(0.0579) | 0.1464(0.0865) | 0.0034(0.0106) | 0.2891(0.0201) |
| HME | 6.9244(1.3156) | 0.8952(0.0772) | 0.1247(0.1147) | 0.0003(0.0144) | 0.2791(0.0257) |
| SE | 7.2947(1.0641) | 0.8443(0.0625) | 0.1578(0.0928) | 0.0038(0.0117) | 0.2878(0.0208) |
| MME | 7.2565(1.0136) | 0.8494(0.0601) | 0.1464(0.0898) | 0.0034(0.0110) | 0.2891(0.0209) |

Table 4.28: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 10% outliers Perturbation

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 14.6900 | 1.0000 | 0.1702 | 0.9850 |
| LTSE | 1.0810 | 184.6682 | 0.9681 | 1.0000 |
| HME | 1.2850 | 130.6885 | 0.9674 | 1.0000 |
| SE | 1.3590 | 116.8435 | 0.9616 | 1.0000 |
| MME | 1.3460 | 119.1114 | 0.9609 | 1.0000 |

From Tables 4.27 and 4.28, it is evident that only OLSE was affected when the dependent variable has 10% unusual observations. The coefficients of the OLSE have values which differ much from when the errors were normally distributed. It also has the least coefficient of determination among all the estimators. Moreover, OLSE reporting large value for residual standard error and small value for relative efficiency makes it unreliable.

## 4.16 20% outliers Perturbation

The 20% outliers perturbation reported the following results in Tables 4.29 and 4.30 on all criteria for comparing the robust methods with the OLSE.

Table 4.29: The coefficients (Standard error) of the estimators for 20% outliers Perturbation

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | -3.6776(19.7251) | 2.2783(1.1579) | -1.2675(1.7199) | -0.0091(0.2162) | 0.0859(0.3851) |
| LTSE | 5.9006(3.0205) | 0.8777(0.1837) | 0.2281(0.2663) | 0.0074(0.0306) | 0.2750(0.0606) |
| HME | 6.2893(1.7319) | 0.9272(0.1017) | 0.1368(0.1510) | -0.0001(0.0190) | 0.2804(0.0338) |
| SE | 7.4357(1.1137) | 0.8363(0.0654) | 0.1878(0.0971) | 0.0008(0.0122) | 0.2932(0.0217) |
| MME | 7.4591(1.0328) | 0.8383(0.0609) | 0.1822(0.0912) | 0.0004(0.0109) | 0.2941(0.0210) |

Table 4.30: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for 20% outliers Perturbation

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 22.2300 | 1.0000 | 0.1186 | 0.9000 |
| LTSE | 3.0940 | 51.6224 | 0.8083 | 1.0000 |
| HME | 1.5520 | 205.1613 | 0.9638 | 1.0000 |
| SE | 1.6270 | 186.6826 | 0.9621 | 1.0000 |
| MME | 1.6130 | 189.9373 | 0.9618 | 1.0000 |

20% outliers had rendered the OLSE limited according to tables 4.29 and 4.30. The fit provided by the OLSE is unreliable in the sense that, it has failed by the use of almost all the criteria. It has large residual standard error, low relative efficiency and low coefficient of determination, which has undermine the usefulness of the OLSE for this perturbed dataset. On the contrary, robust methods such as MME and SE perform well, this is because, they reported estimates which are similar to estimates for the normal error dataset.

## 4.17 30% outliers Perturbation

The Tables 4.31 and 4.32 below present the numerical measures (criteria) for comparing the regression estimators when there are 30% vertical outliers with no leverages.

Table 4.31: The coefficients (Standard error) of the estimators for 30% outliers Perturbation

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | -18.3096(19.2388) | 3.1934(1.1293) | -2.6911(1.6775) | 0.1408(0.2108) | -0.4080(0.3756) |
| LTSE | -16.8367(8.3929) | 2.6412(0.5020) | -1.3608(0.7256) | 0.0340(0.0848) | -0.2171(0.1763) |
| HME | -13.1413(13.9522) | 2.6703(0.8190) | -1.8427(1.2165) | 0.0621(0.1529) | -0.1995(0.2724) |
| SE | 7.6760(1.2056) | 0.8365(0.0708) | 0.1801(0.1051) | -0.0012(0.0132) | 0.2958(0.0235) |
| MME | 7.6827(1.2112) | 0.8385(0.0736) | 0.1762(0.1067) | -0.0015(0.0120) | 0.2960(0.0241) |

Table 4.32: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test 30% outliers Perturbation

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 21.6800 | 1.0000 | 0.1132 | 0.9800 |
| LTSE | 8.4150 | 6.6376 | 0.5075 | 1.0000 |
| HME | 13.3400 | 2.6412 | 0.4934 | 1.0000 |
| SE | 2.0980 | 106.7843 | 0.9690 | 1.0000 |
| MME | 2.0470 | 112.1716 | 0.9682 | 1.0000 |

From Tables 4.31 and 4.32, 30% outliers perturbation led to the breakdown of the OLSE with some robust methods slightly affected. Introducing 30% outliers resulted in some of the estimators having very large standard errors and small coefficients of determination with unreliable coefficients. Tables 4.31 and 4.32 show that the modified maximum likelihood estimator and the S-estimator were robust to the influence of the outliers.

## 4.18 20% error contamination

The Tables 4.33 and 4.34 present the estimates for regression parameters from a dataset with contaminated response variable. The 20% of the observations of the response variable were replaced with observations from the Cauchy distribution.

Table 4.33: The coefficients (Standard error) of the estimators for 20% error contamination

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | -10.3260(122.1340) | 11.7430(7.1690) | -10.7740(10.6490) | -1.3130(1.3380) | -1.0680(2.3840) |
| LTSE | 1.7149(23.4060) | 5.4468(1.1992) | -4.5888(1.5097) | -0.5307(0.3338) | -0.4099(0.3215) |
| HME | 7.5333(1.5790) | 0.9287(0.0927) | -0.0271(0.1377) | 0.0022(0.0173) | 0.2728(0.0308) |
| SE | 7.4882(0.9944) | 0.8452(0.0584) | 0.1359(0.0867) | 0.0021(0.0109) | 0.2918(0.0194) |
| MME | 7.4272(0.9460) | 0.8528(0.0568) | 0.1231(0.0857) | 0.0016(0.0099) | 0.2923(0.0198) |

Table 4.34: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test 20% error contamination

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 137.6000 | 1.0000 | 0.0261 | 0.2500 |
| LTSE | 15.6700 | 77.1079 | 0.2376 | 0.9960 |
| HME | 1.4140 | 9469.7399 | 0.9683 | 1.0000 |
| SE | 1.4160 | 9443.0081 | 0.9695 | 1.0000 |
| MME | 1.4030 | 9618.8140 | 0.9695 | 1.0000 |

In this section, 20% error contamination from Cauchy distribution rendered some estimators limited. The OLSE and LTSE performed poorly. However, estimators such as, MME, SE and HME performed well. The performances of the MME and SE are very impressive in this study. The coefficients of MME and SE are very similar to that of the normal errors. In addition, the residual standard error, coefficients of determination and power of the test are also analogous to the estimates of the original data with normal errors. Moreover, because the residual standard error of OLSE was inflated, it the relative efficiencies of MME and SE have also inflated.

## 4.19 Non-normal distribution(lognormal)

Distributional robustness of the robust methods was assessed by simulating dataset from log-normal distribution. The coefficients of the model for the original data with normal errors were used as the parameters and in conjunction with simulated predictors from log-normal distribution to simulate the response variable. Tables 4.35 and 4.36 below presented the results for the simulated dataset above.

Table 4.35: The coefficients (Standard errors) of the estimators for Non-normal distribution(lognormal)

| Methods | Intercept | BMI | parmfat | height | TS |
|---|---|---|---|---|---|
| OLSE | 207.5770(429.4370) | 10.677(10.6690) | -3.8540(21.1100) | -2.8590(1.9820) | -1.1140(6.2380) |
| LTSE | 47.6217(26.1914) | 0.4520(0.6718) | -0.4530(1.2742) | -0.1289(0.1249) | 0.4783(0.3787) |
| HME | 11.2693(3.3728) | 0.8039(0.0838) | 0.2134(0.1658) | -0.0051(0.0156) | 0.3203(0.0490) |
| SE | 8.5576(1.3716) | 0.8371(0.0341) | 0.1335(0.0674) | 0.0068(0.0063) | 0.2828(0.0199) |
| MME | 7.9685(1.7403) | 0.8431(0.0405) | 0.1519(0.0749) | 0.0075(0.0073) | 0.2883(0.0237) |

Table 4.36: The Root Mean Square Error (RMSE), Relative Efficiency, Coefficient of Determination and Power of the test for Non-normal distribution(lognormal)

| Method of estimation | RMSE | Relative efficiency | Coefficient of determination | Power of the test |
|---|---|---|---|---|
| OLSE | 582.8000 | 1.0000 | 0.0273 | 0.2600 |
| LTSE | 35.1200 | 275.3785 | 0.0235 | 0.2200 |
| HME | 4.2420 | 18875.4693 | 0.5737 | 1.0000 |
| SE | 1.7270 | 113881.8231 | 0.9394 | 1.0000 |
| MME | 1.6550 | 124006.1117 | 0.9469 | 1.0000 |

From the Tables 4.35 and 4.36, by using $R^2$, we see that OLSE and LTSE broke down with HME slightly affected. Moreover, using the coefficients and the relative efficiency, it is also clear that OLSE and LTSE did not do well. This is because the data was simulated from a heavy tailed distribution. On the other hand, MME and SE were still robust to the aberrations in the data. Moreover, robust methods like MME and SE are robust to data from fat tailed distributions, therefore the results in this section have illustrated the distributional robustness of MME and SE.

# Chapter 5

# Conclusion and Recommendations

## 5.1 Introduction

In this study, we compared the Ordinary least squares estimator (OLSE) of multiple linear regression to the following robust regression estimators: Least Trimmed Squares Estimator (LTSE), Huber Maximum Likelihood Estimator (HME), S-estimator and the Modified Maximum Likelihood Estimator (MME). This comparison was carried out to evaluate the robustness of some robust methods in estimating regression parameters. In assessing the performances of these estimators the following datasets were considered: normal error dataset, datasets perturbed with leverages, datasets perturbed with both outliers and leverage, only outliers perturbed datasets, error contaminated dataset, and simulated dataset from log-normal distribution using the fitted model of data with normal errors. To evaluate the overall performances of the robust methods, we used the following criteria: the coefficients and their standard errors, the residual standard errors, relative efficiencies, coefficients of determination and the power of the test. Studies by Alma (2011), Muthukrishnan and Radha (2014) and others compared robust estimators to the Ordinary Least Squares Estimator using coefficient of determination. Also, some studies compared some of these estimators and others using coefficients and their standard errors, the residual standard errors, relative efficiencies as bases for their comparisons. The chapter three of this study discussed the methods listed above. The discussions on these methods included their robustness properties and weaknesses. Above all, applying these methods on the above datasets reported results in chapter four which confirmed the theories of these methods as discussed in chapters two and three above.

## 5.2 Findings and Conclusions

The results reported in chapter four showed that robust methods are as efficient as the OLSE if the basic assumptions are satisfied. In addition, small deviations from normality did not substantially impair robust methods like, LTSE, HME, SE. Moreover, the robust methods such as MME and SE did not breakdown completely even when the errors largely deviate from normality. Perturbing the independent variables with leverages, caused Ordinary least squares estimator and Huber Maximum likelihood estimator to breakdown, whilst the Least trimmed squares estimator was slightly affected. However, the S-estimator and the Modified Maximum Likelihood Estimator did very well. This was as a result of the fact that, these methods were developed as modifications to the OLSE and other robust methods, therefore, they are able to resist the influences of aberrations that limit the performances of OLSE and other methods.
Also, when we perturbed the dependent variable with outliers and independent variables with leverages, OLSE broke down completely but HME and LTSE were slightly affected. This substantiates the claim that, HME performs well if there are only vertical outliers, whilst other robust methods are able to resist both leverages and vertical outliers.

In addition, when we introduced only vertical outliers, HME, SE and MME did very well. This reiterate the point that, when a dataset has only outliers in the space of the response variable, HME can be used to analyze such a dataset.

Also, when we contaminated the data with some data points from Cauchy distribution, it was observed that OLSE and LTSE were affected, but HME was slightly affected. The total breakdown of the OLSE explained the concept that OLSE performs well only when the errors are normal and satisfies the basic rudiment of the classical regression method.

Finally, when the simulated dataset from log-normal distribution was analyzed, the results showed the robustness of MME and SE. Other estimators like, OLSE, LTSE broke-down completely.

Therefore, from the discussion above, the first objective of this study which is to determine all the criteria for comparing the estimators has been achieved, since all the estimators were compared using these standards (criteria) of the comparison. Also, the second objective which is to compare OLSE to the robust methods using four datasets was also achieved, since the study compared the OLSE successfully with the robust methods for the four datasets by using all the stated criteria for the comparison. Moreover, with respect to the last objective of this study which is to come out with some robust methods which perform well for all datasets using the criteria of comparison has been achieved. This is because, robust methods such as MME and SE were able to perform well using all the bases of the comparison. Therefore, we conclude that robust methods are very paramount in estimating regression parameters.

## 5.3 Recommendations

In respect of our findings, the following recommendations are given on the use and application of robust methods in linear regression analysis.

1. We recommend the use of robust methods because of the effects of masking and swamping. Robust methods help to uncover observations which may be outliers but are behaving as usual observations, or the observations which are not outliers but because of other data points they appear as one.

2. The central limit theorem is based on large sample theory and it is not in all situations that this law holds, therefore it is advisable to use robust methods, since they do not impose strict distributional assumptions on the

datasets.

3. Again, robust methods can be used concurrently with the Ordinary Least Squares method of Estimation as diagnostic tools.

4. Finally, many statisticians do not use robust methods because, they believe these methods are computationally complex with less information on how they are used. However, we recommend the use of these methods because, there are statistical packages which now have functions for the application of robust methods.

# REFERENCES


Aelst, S. V. (2014). Robust strategies and model selection. http://www.dcs.bbk.ac.uk/ercim/COMISEFTutorialTalks/VanAelstTutorial.pdf.

AL-Noor, H. N. and Mohammad, A. A. (2013). Model of robust regression with parametric and nonparametric methods. *Mathematical Theory and Modeling*, 3:27–39.

Alma, O. G. (2011). Comparison of robust regression methods in linear regression. *International Journal for contemp Maths and Science*, 6:409–421.

Andersen, R. (2008). Modern methods for robust regression. *Sage Publications*, pages 2–52.

Bhar, L. (2014). Robust regression. http://www.iasri.res.in/ebook/EBADAT/3-Diagnostics

Bondell, H. D. and Stefanski, L. A. (2013). Efficient robust regression via two-stage generalized empirical likelihood. pages 1–42.

Cetin, M. and Toka, O. (2011). The comparing of s-estimator and m-estimators in linear regression. *Gazi University Journal of Science*, 24(4):747–752.

Cohen, J. (1992). Statistical power analysis. *Sage Publications*, 3:98–101.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis Third edition*. John Wiley and Sons Ltd.

Fox, J. and Weisberg, S. (2010). An apendix to an r companion to applied regression second edition. pages 1–17.

Garner, W. (2014). Robust regression. $http : //math.ucsd.edu/\ wgarner/research/pdf/robust_r egression.pdf$.

Gschwandtner, M. and Filzmoser, P. (2012). Computing robust regression estimators: Developments since dutter(1977). *Austrian Journal of Statistics*, 41:45–58.

Hampel, F. (2001). Robust statistics: A brief introduction and overview. pages 1–5.

Huber, P. J. (1972). The 1972 wald lecture: Robust statistics. *The Annals of Mathematical Statistics*, 43:1041–1067.

Jacoby, B. (2014). Regression iii: Advanced methods. http://polisci.msu.edu/jacoby/icpsr/regress3/lectures/week3/11.Outliers.pdf.

Jann, B. (2012). Robust regression in stata. pages 1–34.

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006). *Robust Statistics, Theory and Methods*. John Wiley and Sons Ltd.

Matias, S.-B. and Yohai, V. J. (2006). A fast algorithm for s-regression estimates. *Journal of Computational and Graphical Statistics*, 15(2):414–427.

Muthukrishnan, R. and Radha, M. (2014). Comparison of robust estimators in regression. http://interstat.statjournals.net/YEAR/2010/articles/1005003.pdf.

Pena, D. and Yohai, V. J. (1996). A procedure for robust estimation and diagnostics in regression. *Statistics and Econometrics Series 19*, pages Working paper 96–48.

Rice, J. A. (1995). *Mathematical Statistics And Data Analysis Second Edition*. An Imprint of Wadsworth Publishing Company Belmont, California.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection*. JOHN WILEY and SONS.

Rousseeuw, P. J., Zaman, A., and Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economic Letters*, 71:1–8.

Ruppert, D., Street, J. O., and Carroll, R. J. (1988). A note on computing robust regression estimates via iteratively reweighted least squares. *The American Statistician*, 42:152–154.

Schumacker, R. E., Monahan, M. P., and Mount, R. E. (2002). A comparison of ols and robust regression using s-plus. Vol. 28(2):10–13.

Staudte, R. G. and Sheather, S. J. (1990). *Robust Estimation and Testing*. John Wiley and Sons Ltd.

Stuart, C. (2011). Robust regression.

Tiku, M. L. and Akkaya, A. D. (2004). *Robust estimation and hypothesis testing*. New Age International (P) Ltd.

Verardi, V. and Croux, C. (2009). Robust regression in stata. *The Stata Journal*, 3:439–453.

Yohai, V. J. (1985). High breakdown point and high efficient robust estimates for regression. 66:1–47.

Yohai, V. J. (1987). High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15:642–656.