

A ROBUST ESTIMATION PROCEDURE IN LINEAR REGRESSION IN THE PRESENCE OF OUTLIERS

ALANAMU, TAOHEEDAT
(00/55EG007)

B.Sc. (ILORIN) 2004, M.Sc. (ILORIN) 2008

A THESIS SUBMITTED TO THE
DEPARTMENT OF STATISTICS, FACULTY OF PHYSICAL SCIENCES, UNIVERSITY OF
ILORIN, ILORIN, NIGERIA,
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE
DEGREE OF DOCTOR OF PHILOSOPHY (Ph.D.) IN
STATISTICS

JUNE, 2018

CERTIFICATION

This is to certify that this thesis has been read and approved as meeting the requirements of the Department of Statistics, University of Ilorin, Ilorin, Nigeria, for the Award of Doctor of Philosophy (Ph.D.) degree in Statistics.

Dr. G. M. OYEYEMI
(Supervisor)

Date

Prof. B.L. ADELEKE
(Head of Department)

Date

Dr. K. RAUF
(Internal/External Examiner)

Date

(External Examiner)

Date

DEDICATION

To my dear husband, to my dear parents, and to my beloved children

DECLARATION

I, Taoheedat Alanamu, a Ph.D. student in the Department of Statistics, University of Ilorin, Ilorin, hereby declare that this thesis entitled “A robust estimation procedure in linear regression in the presence of outliers”, submitted by me is based on my actual and original work. Any materials obtained from other sources or work done by any other persons or institutions have been duly acknowledged. In addition, the research has been approved by the University of Ilorin Ethical Review Committee.

Taoheedat Alanamu

(00/55EG007)

Date

ACKNOWLEDGEMENTS

All praise is due to **ALLAH**, the Lord of the World. To Him belongs absolute control of everything. He is free from imperfections. My most gratitude goes to Him for making everything possible.

I am sincerely grateful to my supervisor, Dr. G. M. Oyeyemi, for his guidance, encouragement, understanding and beneficial suggestion. May Allah bless you and reward you abundantly sir.

I am most grateful to my husband, Kamaldeen Ahmed for his patience, encouragement and support. He is indeed one in a million. Also, with gratitude I acknowledge my parents, Mr. and Mrs. AbdulAzeez Alanamu for their love and support throughout my life. They have provided me with character that made me to persevere. My children are blessing and wonderful distraction that helped me learns lots in life, I appreciate and value them most.

This will be incomplete without acknowledging Dr. M. K. Garba, he encouraged me to apply for the program and gave me the support and strength when I was about losing hope. You are really a brother and God sent, may Allah take care of you and all that concern you. Also, I will like to appreciate Mr. R. O. Olaniran, for sacrificing out of his limited time to see to mine, God bless you.

And to my siblings, friends, colleagues, and loved ones who have provided me support and assistance during the period of the research, I appreciate you all and say JAZAKUM U-LLAHU KHAYRAN.

Table of Contents

TITLE PAGE.....	i
CERTIFICATION.....	ii
DEDICATION	iii
DECLARATION	iv
ACKNOWLEDGEMENTS	v
LIST OF TABLES	ix
LISTS OF FIGURES.....	xi
ABSTRACT.....	xii
CHAPTER ONE	1
INTRODUCTION.....	1
1.0 Background to the study	1
1.1 Statement of the problem.....	3
1.2 Aim and objectives of the study.....	4
1.3 Justification for the study	4
1.4 Organization of thesis	4
CHAPTER TWO	6
LITERATURE REVIEW	6
2.0 Introduction	6
2.1 Outliers.....	6
2.2. Robust regression.....	7
2.3 High breakdown point estimators	8
2.4. Studies on robust estimators	9
2.5 Properties of estimators.....	16
2.6 Outliers in multiple linear regression.....	17
2.6.1 Effects of Outliers on a Statistic.....	19
2.6.2 Effects of Non-normality on a Statistic	20
2.6.3 Effects of contaminants on Statistic	20
2.7 Comparison of robust estimators	21
2.7.1 Studies on Criteria for comparing Estimators.....	21
2.7.2 Merits of Robust estimators over the Ordinary Least Squares Estimator.....	21

CHAPTER THREE	23
RESEARCH METHODOLOGY	23
3.0 Introduction	23
3.1 Outlier detection methods.....	24
3.2 Some robust regression methods.....	26
3.3 Robust regression based on regularization of case- specific parameters.....	29
3.4 Modified robust regression method.....	30
3.5 Ridge regression.....	30
3.5.1 Methods for selecting the value of parameter λ In Ridge regression modelling.....	31
3.5.2 Derivation of the modified Robust Estimator:	31
CHAPTER FOUR.....	36
SIMULATION STUDY AND PRESENTATION OF RESULTS	36
4.0 Introduction	36
4.1 Dataset simulation.....	36
4.2 Criteria for assessing the estimators performance	37
4.3 Results	39
4.4 Real life data illustration	59
4.4.1 Data analysis and result	60
4.4.2 Discussion of results for the real life data	69
CHAPTER FIVE.....	73
DISCUSSION AND CONCLUSION	73
5.0 Introduction	73
5.1 Summary	73
5.2 Research contribution and findings.....	74
5.3 Conclusion.....	74
5.4 Recommendations	75
Suggestion for further studies	75
REFERENCES.....	76
APPENDICES	79
Appendix A1: Simulation code for CASE I.....	79
Appendix A2: Simulation code for CASE II	82
Appendix A3: Simulation code for CASE III	85

Appendix B: Analysis for Real life data 88

Appendix C: Outlier Detection function 99

LIST OF TABLES

Table 4.1: MSE of various methods over sample size for case I at $\pi = 0.1$	39
Table 4.2: Relative efficiency of various methods over sample size for case I at $\pi = 0.1$	39
Table 4.3: MSE of various methods over sample size for case I at $\pi = 0.2$	39
Table 4.4: Relative efficiency of various methods over sample size for case I at $\pi = 0.2$	40
Table 4.5: MSE of various methods over sample size for case I at $\pi = 0.3$	40
Table 4.6: Relative efficiency of various methods over sample size for case I at $\pi = 0.3$	40
Table 4.7: MSE of various methods over sample size for case I at $\pi = 0.4$	41
Table 4.8: Relative efficiency of various methods over sample size for case I at $\pi = 0.4$	41
Table 4.9: MSE of various methods over sample size for case I at $\pi = 0.5$	41
Table 4.10: Relative efficiency of various methods over sample size for case I at $\pi = 0.5$	42
Table 4.11: MSE of various methods over sample size for case II at $\pi = 0.1$	45
Table 4.12: Relative efficiency of various methods over sample size for case II at $\pi = 0.1$	45
Table 4.13: MSE of various methods over sample size for case II at $\pi = 0.2$	45
Table 4.14: Relative efficiency of various methods over sample size for case II at $\pi = 0.2$	46
Table 4.15: MSE of various methods over sample size for case II at $\pi = 0.3$	46
Table 4.16: Relative efficiency of various methods over sample size for case II at $\pi = 0.3$	46
Table 4.17: MSE of various methods over sample size for case II at $\pi = 0.4$	47
Table 4.18: Relative efficiency of various methods over sample size for case II at $\pi = 0.4$	47
Table 4.19: MSE of various methods over sample size for case II at $\pi = 0.5$	47
Table 4.20: Relative efficiency of various methods over sample size for case II at $\pi = 0.5$	48
Table 4.21: MSE of various methods over sample size for case III at $\pi = 0.1$	52
Table 4.22: Relative efficiency of various methods over sample size for case III at $\pi = 0.1$	52
Table 4.23: MSE of various methods over sample size for case III at $\pi = 0.2$	52
Table 4.24: Relative efficiency of various methods over sample size for case III at $\pi = 0.2$	53
Table 4.25: MSE of various methods over sample size for case III at $\pi = 0.3$	53
Table 4.26: Relative efficiency of various methods over sample size for case III at $\pi = 0.3$	53
Table 4.27: MSE of various methods over sample size for case III at $\pi = 0.4$	54
Table 4.28: Relative efficiency of various methods over sample size for case III at $\pi = 0.4$	54
Table 4.29: Performance of various methods over sample size for case III at $\pi = 0.5$	54
Table 4.30: Performance of various methods over sample size for case III at $\pi = 0.5$	55
Table 4.31: Data set and outlier detection results	61
Table 4.32: Parameter estimates of the Growth study data set with and without the 5% outlying points.	61
Table 4.33: Absolute bias of estimates for the Growth study dataset with and without the 5% outlying points. The true parameter is assumed to be LS without the outlying point.	62
Table 4.34: Standard error (SE) of estimates for the Growth study data set with and without the 5% outlying points	62

Table 4.35: Error variance and coefficient of determination for the Growth study data set with and without the 5% outlying points	63
Table 4.36: Parameter estimates of the Scottish race data set with and without the 5.7% outlying points	63
Table 4.37: Absolute bias of estimates for the Scottish race data set with and without the outlying 5.7% points	64
Table 4.38: Standard error (SE) of estimates for the Scottish race data set with and without the 5.7% outlying points.	64
Table 4.39: Error variance and coefficient of determination for the Scottish race data set with and without the 5.7% outlying points.	65
Table 4.40: Parameter estimates of the Water salinity data set with and without the 7.1% outlying points.	65
Table 4.41: Absolute bias of estimates for the Water salinity data set with and without the 7.1% outlying points.	66
Table 4.42: Standard error (SE) of estimates for the Water salinity data set with and without the 7.1% outlying points.	66
Table 4.43: Error variance and coefficient of determination for the Water salinity data set with and without the 7.1% outlying points.	67
Table 4.44: Parameter estimates of the stackloss data set with and without the 23.8% outlying points.	67
Table 4.46: Standard error (SE) of estimates of the stackloss data set with and without the 23.8% outlying points.	68
Table 4.47: Error variance and coefficient of determination for the stackloss data set with and without the 23.8% outlying points.	69

LISTS OF FIGURES

Figure 4.1: Plot of Average MSE against Outlier percentages for case I..	43
Figure 4.2:Plot of Average RE against Outlier percentages for case I..	44
Figure 4.3: Plot of Average MSE against Outlier percentages for case II..	50
Figure 4.4: Plot of Average RE against Outlier percentages for case II..	51
Figure 4.5 Plot of Average MSE against outlier percentages for case III.....	57
Figure 4.6 Plot of Average RE against outlier percentages for case III.....	58

ABSTRACT

The best method that can be used in regression analysis under some assumptions is Ordinary Least Squares (OLS) method. Such assumptions include; constant variance, normality and uncorrelated error terms among others. Violation of these assumptions such as the presence of outliers may have a large effect on the ordinary least square estimates. The existing robust techniques; Least Absolute (LA) estimator, Huber Maximum (H-M) estimator, Bi-Square Maximum (B-M) estimator, M-Maximum (MM) estimator, Least Median Square (LMS) estimator, Least Trimmed Square (LTS) estimator and S estimator (S) cope with these unusual observations. However, they may be sensitive to the presence of outliers especially when they are in both directions of X and Y. The aim of this study therefore was to obtain a procedure in estimating parameters in linear regression model that is robust in the presence of outliers. The objectives were to: (i) obtain the best method in linear regression analysis among the existing robust methods using efficiency and breakdown point; (ii) propose an efficient procedure in estimation of parameters in Linear Regression in the presence of outliers with high breakdown points; and (iii) validate and compare the proposed procedure with the existing ones.

An estimation method for the classical linear regression was proposed using regularization methods that allow one to handle a variety of inferential problems. Specifically, each outlying point in the data is estimated using case-specific parameter through Ridge regression approach and penalized estimators were suggested when the number of parameters in the model is more than the number of observed data points. The proposed estimation method was compared with the existing methods using simulated data with varying proportion and magnitude of outliers in all directions. The results from simulation were validated using real-life datasets. The performances were assessed using breakdown point and efficiency.

The findings of the study were that:

- (i) the strength and weakness of any robust estimator strongly depend on the percentage, magnitude and direction of the outlier;
- (ii) among the existing robust estimators, MM is the best in terms of unidirectional outlier (X or Y direction);
- (iii) the proposed method was found to be better in dealing with outliers in all directions (X or Y or both); and
- (iv) robust estimators cannot improve the adequacy property of linear regression model.

The study concluded that the proposed method is a better procedure in estimation of parameters in linear regression in the presence of outliers due to its high efficiency and breakdown point of up to 50%. The proposed method is therefore preferred and recommended whenever there is high proportion of outliers and the direction of outliers is unknown.

CHAPTER ONE

INTRODUCTION

1.0 Background to the study

In classical statistical theory, a lot of statistical estimation methods are based on the model with certain assumptions, like the variables are from normal distribution and mutually independent. However, this is not often met in practice. It is hard to describe most of random phenomena with simple statistical models, and furthermore, the optimal estimators could be gotten via these methods only if all the assumptions are fulfilled. In a word, one would find that these statistical models are just approximations to the actual events at a certain degree and when the actual data departure from the model assumptions, the estimators would no longer be the best. A contemporary and important statistical topic named robust statistics was developed under this background. The statistical use of the word "Robust" was first proposed by George Box in 1953. But robustness was accepted by most of people as a sub discipline in statistics was primarily the creation of John W. Turkey and Peter J. Huber in 1960s (Stigler, 2010). In Huber (1964), the M-estimator, was introduced and recognized as a significant contribution to the issues of robustness. After some decades of continued research, Hampel (2001) proposed the concepts of breakdown point and influence curve. This was followed by introduction of R and L estimates, which, together with the M-estimator of Huber (1964) formed the three basic classes of estimates when considering robustness. Stigler (2010) wrote an article where the developments of the robust statistics were summarized. Note that there are other non-parametric methods of estimation, such as jackknife and bootstrap that are outside the scope of this study. These robust methods seek to

provide ways with good performance when the fundamental assumptions for data sets are not fully fulfilled. According to previous studies, it could be concluded that a good robust statistical method should satisfy the following theoretical requirements:

1. When the actual observations are consistent with the model assumptions, the method is near-optimal;
2. When the actual observations have a small departure from the model assumptions, the robust method still has a good performance;
3. When the actual observations have a serious departure from the model assumptions, the corresponding method is still working;

So this study with respect to the comparison of the various robust statistics will take these properties into account.

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 (OLS) estimator 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), 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 OLS estimator 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 nonnormal errors.” And it should be robust with respect to outliers, inliers, leverage points, and contaminations. Robust methods are also used in heteroscedastic models, where variances depend on the independent variables (X), and datasets full of outliers (Tiku and Akkaya, 2004). Robust statistics have not been seriously studied by mathematical statisticians, until recently, although it’s been in use for hundreds of years (Stigler, 1973). It is worthy to 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 statistical tools now have their robust counterparts.

1.1 Statement of the problem

In linear regression analysis, the application of ordinary least squares method works well if the classical 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. In addition, in situations where the existing methods fail to achieve a desirable level of tolerance in terms of breakdown point and efficiency, a modified method is presented.

1.2 Aim and objectives of the study

The aim of this study therefore was to obtain a procedure in estimating parameters in linear regression model that is robust in the presence of outliers. The specific objectives were to:

- i. obtain the best method in linear regression analysis among the existing robust methods using efficiency and breakdown point;
- ii. propose an efficient procedure in estimation of parameters in Linear Regression in the presence of outliers with high breakdown points; and
- iii. validate and compare the proposed procedure with the existing ones using simulated and real life datasets.

1.3 Justification for the study

Despite the fact that many robust techniques have been proposed, each technique suffers from one limitation to the other, varying from low breakdown point to low efficiency. Some of the techniques are not robust to x-axis outliers, while some have troubles with high leverage outliers. Efforts have been made by many in order to make the classical non robust methods to be robust. The aim is to arrive at a procedure that will have high breakdown point, high efficiency and also robust to outliers in all directions.

1.4 Organization of thesis

A clearer introduction of the background, problem statement and specific objective to this study were achieved in chapter one. Chapter two presents studies around outlier in linear model as well as dilemma and strengths of robust methods. The study methodological

framework is presented in chapter three, chapter four presents the simulation studies and results. Chapter five presents the concluding remark.

CHAPTER TWO

LITERATURE REVIEW

2.0 Introduction

In this chapter, the past and recent studies around the effects of outlier in linear regression model will be reviewed. The strength and weakness of Ordinary Least Square (OLS) when the assumption of no outlying observation is violated is discussed in detail. In addition, the strengths, weaknesses and dilemma of recent methods were also reviewed.

2.1 Outliers

Outliers in statistics don't have a rigorous definition; determining whether or not a data point is an outlier is ultimately a subjective exercise. Virtually, how people make the choice of the data model will decide the assumption about the "normal" behaviour of the data. Given this, how the "outliers" are characterized by statisticians from different perspectives are reviewed. Hawkins (1980) defined "outlier" as "the outlying observation appears to deviate markedly from other members of the sample in which it occurs", but this is deemed not enough because it is susceptible by people's anticipation. Some data may act as outliers under one distribution but not outliers any more under another distribution. Barnett and Lewis (1984) gave outliers amore broad but detail definition. Barnett and Lewis (1984) distinguished between extreme observations, outliers, and contaminants and revised the definition of outliers to be the "observations that appear inconsistent with the rest of the data". Extreme observations could be the smallest and largest value of a data set. Contaminants are data points from another distribution "slipped" upward relative to the original distribution. Both extreme observations

and contaminants may or may not be outliers and vice versa. Long after that, Aggarwal (2013) divided outliers, from another point of view, into the noise and anomalies based on analyst interest. The three main purposes regarding outliers are clearly distinguished by Iglewicz and Hoaglin (1993) thus:

- Outlier labelling : Flag potential outliers for further investigation
- Outlier accommodation : the outlying observations are being accommodated using robust statistical techniques
- Outlier identification: identifying whether the observations are outliers using formal tests.

It should be noted that sometimes, outliers cannot be removed from a dataset but attention should be paid to such outlying points (outliers)

2.2. Robust regression

In multiple linear regressions, 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 (2000), if the data contain outliers, non-normality, and/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 breakdown 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 the 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 (HM), Modified Maximum likelihood Estimator (MM), S-Estimator (S), and Least Trimmed Squares Estimator (LTS).

2.3 High breakdown point estimators

To provide reliable estimates in the presence of a large percentage of outlying observation, high breakdown point regression estimators are developed. The high breakdown point regression estimators can achieve up to 50% breakdown point (Wisnowski, 1999).

Due to the drawbacks of the ordinary least squares (OLS) estimator, there are estimators that possess the property of a high breakdown point. The OLS 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 (1984) proposed to minimize a measure of dispersion of the residuals that is less sensitive to extreme values than the variance. 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 weight to large residuals. The loss function is carefully chosen to enable the estimator possess good robustness properties and a high Gaussian efficiency. According to Stuart (2011), there exists a trade-off between the breakdown point and the asymptotic efficiency of the S-estimator. As the 50% high breakdown point S-estimator has efficiency of 28.8% with a

tuning constant of 1.547. And a breakdown point of 10% and asymptotic efficiency of 96.6% with a tuning constant of 5.182. Due to the lapses of the S-estimator, it is usually used as initial value for Modified maximum likelihood estimator. As MM-estimator inherit the high breakdown property of the S-estimator. The fact that MM-estimator is known for its highly efficient property, S-estimator can give it the high breakdown property when used as its initial values. The independence of breakdown point and efficiency of MM-estimator resulted in its recognition as the estimator with both high breakdown point and high efficiency, for both normal and non-normal errors.

2.4. Studies on robust estimators

Regression analysis is mostly used in all areas of study, and the easiest and convenient method normally used is the ordinary least squares. But, due to the limitations of ordinary least square, 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 Bhar (2014), robust M estimator was compared with the ordinary least squares estimator. Bhar (2014) also discussed robust regression methods such as M-estimator, W-estimators, R-estimators, Least median of squares estimator, Least trimmed of squares estimator, and Reweighted least squares estimator. It's been documented that M estimation happens to be the most common robust regression method which is almost as efficient as the least square estimator (Huber, 1973, 1981). This could be as a result of the fact that, rather than minimize the sum of squared residuals as the objective, the M-estimator tends to minimize a function ρ 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. Bhar (2014) 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. Results obtained in Bhar (2014) showed that Huber M estimators performed better than the ordinary least squares estimators when both standard error and coefficient of determination are used as the performance criteria

Study by Fox and Weisberg (2010) on robust regression considered estimators such as the M-estimators, bounded-influence estimators, MM-estimator, and Least median squares estimator. Applying all these methods on Duncan's occupational-prestige data, Fox and Weisberg (2010) revealed that the least squares estimator broke down as a result of vertical outliers while the other robust methods were able to bound the influence of unusual observations. The robust methods serve as formidable statistical tools for identifying unusual observations in the data. In the result by Fox and Weisberg (2010), the least trimmed squares performed better than the estimators based on their individual standard error values, while the ordinary least squares estimator performed poorly when no outlier values were removed.

AL-Noor and Mohammad (2013) performed a simulation study with a view to compare Ordinary Least Squares Method, Least Absolute Deviations method, M-Estimators, Trimmed Least Squares estimators and Nonparametric Regression. In AL- Noor and Mohammad (2013) these estimators were compared for vertical outliers, horizontal outliers and both vertical and horizontal outliers based on their mean square error and relative efficiency. In AL- Noor and

Muhammad (2013), 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. When outliers were introduced in the dependent and the independent variables, Ordinary least squares method, M-estimators and least absolute deviation estimators broke down. The nonparametric methods performed better when outliers were present in both X and Y-directions. Considering the overall performance, the Least trimmed squares estimators performed better than all other estimators. AL-Noor and Mohammad (2013) recommended that future studies should take into consideration the following:

- i. Before analyzing any data, the presence of outliers should be the first thing to check for and then construct the necessary tests whether to see the underlying assumptions are satisfied. After that, the appropriate estimation techniques should be conducted
- 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. When the outliers are appearing in Y- direction, choose M-estimator and LAD method, or choose a LTS method.
- iv. Choose RMSE or Mean Absolute Error (MAE) as criteria for comparing the estimators to avoid dealing with the large values of MSE, when outlier appears in X-direction or XY direction.

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

Radha (2014) did a study on comparison of robust regression estimators by assessing their coefficient of determination. Muthukrishnan and Radha (2014) considered estimators such as Least trimmed squares, Least median squares, Trimmed mean square, Huber M-estimator, and Least absolute deviation, all in relation to the Ordinary least squares estimator. Muthukrishnan and Radha (2014) 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. Muthukrishnan and Radha (2014) used the R-software to analyze two real life data sets. Results of Muthukrishnan and Radha (2014) 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. Muthukrishnan and Radha (2014) concluded that all robust methods are modification of the traditional methods.

Alma(2011) considered the Least Trimmed Squares estimator, Huber M-estimator, Yohai MM-estimator, and S-estimator with respect to the least squares estimator. In Alma (2011), 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. Alma (2011) concluded that S-estimator performed better than the others in efficiencies and was able to bound the influence of outliers and leverage points. The study also showed that MM-estimator breaks down when dealing with high leverage points in small dimensional data. This study will look

at these four robust estimators in a presence of outliers, leverage points, non-normality and contaminants.

According to Schumacker, Monahan and Mount (2002), the violation of basic assumptions such as: outliers in datasets, non-normality, or multicollinearity among variables lead to estimates not reflecting the actual parameters. In Schumacker *et al.* (2002), the 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 in Schumacker *et al.* (2002) that the MM-estimator performed better than LTS-estimator and OLS-estimator. Therefore, it was concluded in Schumacker *et al.* (2002) that, the best method is the MM-estimator. Schumacker *et al.* (2002) 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 Stuart (2011). 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 (1987) highlighted 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. Yohai (1987) concluded that the MM-estimator was not influenced by outliers as they did to the ordinary least squares.

Jacoby (2014) 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 Jacoby (2014), M-Estimator was analogous to OLS when outliers are removed. Bisquare estimator was not affected by the influential cases like the Huber estimator. The robust estimators such as LTS and LMS 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 breakdown point as high as 50%, often work very well when the sample size is large. Jacoby (2014) concluded on the fact that no single robust estimator is best for all data sets. It was noted by Jacoby (2014) 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-estimator, Huber-M, 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. E,g 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 are satisfied. By applying STATA statistical package on a set of data, Jann (2012) concluded based on the results that even though the robust methods appear to perform better than the OLS, it should be used with the robust methods serving as model diagnostic tools.

Some robust methods of estimation were compared to the ordinary least squares estimation by Cetin and Toka (2011). In Cetin and Toka (2011), S-estimator; Huber M-estimator and least trimmed squares estimator were examined relative to the ordinary least square. A dataset with no outliers but with weak multi-collinearity was used for the comparison. OLS 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 HM. Simulation study was also conducted with non-normal data, and it was found that the OLS is inefficient when the dataset is contaminated with outliers.

According to a study by Rousseeuw, Zaman and Orhan (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 OLS estimator 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 OLS. 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 researchers 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* (2001) emphasized the advantages of using the LTS as a robust method. Rousseeuw *et al.* (2001) concluded on the note that deleting few outlying points brings about drastic changes in regression results.

In a study by Ruppert, Street and Carroll (1988), much attention was paid to Iteratively Reweighted Least Squares method of estimation. As a result, estimators such as Huber estimator, Tukey bi-weight estimator and Hampel estimator were compared with the ordinary least squares estimator. Results from Ruppert *et al.* (1988) showed that the Huber estimator

has problem with leverage points but OLS estimator performed badly overall among the methods.

2.5 Properties of estimators

A researcher performing an inference about a population parameter needs information on a sample statistic of that population parameter. Moreover, an estimator is any quantity calculated from the sample data which is used to give information about an unknown quantity in the population (Arsham, 2014). According to Glass and Hopkins (1984) when performing parametric statistical tests one should be mindful of the properties of estimators. The properties of estimators are unbiasedness, consistency, efficiency and sufficiency.

- If the mean of the sampling distribution of the statistic equals the population parameter being estimated, the estimator is considered unbiased
- If the sample statistic approaches the population parameter as the sample size increases, the estimator is considered consistent
- If the estimator is the only one among all other unbiased estimators with the smallest variance, the estimator is considered efficient
- If the sample (n) from which it is computed contains all the information contained in the population (N), the estimator is considered sufficient

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

2.6 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 (2014), 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. According to Rousseeuw and Leroy (1987), there are three types of outliers in regression analysis that influence the performance of ordinary least squares. 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 direction 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. Outliers in a dataset go a long way to violate the basic assumptions of the error terms in regression analysis. A study by Jacoby (2014) looked at the various types of outliers including leverage points and their influence on the ordinary least squares estimates.

Jacoby (2014) applied OLS method of estimation on a dataset from R statistical-package car, with and without outliers to see the effects of the outliers. Jacoby (2014) also discussed the various methods of identifying outliers in the dependent variable and the leverage points in the independent variables. It was shown Jacoby (2014) 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. According to Jacoby (2014), 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 the influence is assessed by using Cook's Distance (Jacoby, 2014). Since outliers can come about as a result of 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, and they must be included in the analysis, then 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 OLS when there are no deviations in the data, but down weight the effects of influential observations when they exist (Jacoby, 2014).

2.6.1 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:

- Outliers lead to misinterpretation of general pattern in regression plots.
- Unusual observations can have a strong influence on statistical models, and removing outliers from a regression model can sometimes give completely different results.
- These unusual points can also substantially influence the fit of the ordinary least squares model.
- Both the slopes and intercept of the model are substantially influenced by data points that are both outliers and high leverage.
- 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.

2.6.2 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. Faulty distribution assumption of the data may result in presence of outlier, this can result in non-normality of the error terms (Iglewicz and Hoaglin, 1993). Outliers occurring far away from the majority of the data points make the distribution of the response variable to deviate from the normal. Skewness or heavy-tailedness is sign of non-normality, and is as a result of unusual data points. Non-normality due to outliers can cause a deleterious effect on statistical analyses, by increasing the error variance and reduce the power of statistical test; if non randomly distributed, they decrease normality (multivariate normality and assumption of sphericity in multivariate analyses); affects the intercept of the regression model and renders the regression model inefficient and seriously biased estimates Osborne and Overbay (2004).

2.6.3 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 by Stuart (2011), 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.

2.7 Comparison of robust estimators

Robustness of linear regression parameter estimation originated from the theories of Huber (1964) and Hampel (1973), 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. This study will consider some of these criteria

2.7.1 Studies on Criteria for comparing Estimators

A study by Muthukrishnan and Radha (2014) compared M, L and R-estimators with the ordinary least squares estimator by using the coefficient of determination. Schumacker *et al.* (2002), also studied robust regression by comparing estimators such as: OLS, LTS and MM 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 Muhammad (2013) and Gschwandtner and Filzmoser (2012) in their study, used mean square error as the criterion for comparing their estimators.

2.7.2 Merits of Robust estimators over the Ordinary Least Squares Estimator

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 OLS estimator is likely to breakdown.
2. Also, robust estimators help in addressing the effect of outliers in a dataset that could undermine the least squares estimator's performance.
3. Robust estimators are able to bound the influence of unusual observations in a dataset.
4. Robust methods do not need the assumptions of the classical regression model to be satisfied before its applications.
5. Large sample theories undermine the applicability of robust techniques, which is a myth.
6. 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 THREE

RESEARCH METHODOLOGY

3.0 Introduction

Many robust methods have been proposed to achieve high breakdown point or high efficiency or both. M-estimates (Huber, 1981) are solutions of the normal equation with appropriate weight functions. They are resistant to unusual y observations, but sensitive to high leverage points on x . Hence the breakdown point of an M-estimate is $1/n$ (Alma, 2011). R-estimates (Jaeckel 1972) which minimize the sum of scores of the ranked residuals have relatively high efficiency but their breakdown points are as low as those of OLS estimates. Least Median of Squares (LMS) estimates (Siegel 1982) which minimize the median of squared residuals, Least Trimmed Squares (LTS) estimates (Rousseeuw 1983) which minimize the trimmed sum of squared residuals, and S-estimates (Rousseeuw and Yohai 1984) which minimize the variance of the residuals all have high breakdown point but with low efficiency. Generalized S-estimates (GS-estimates) (Croux et al. 1994) maintain high breakdown point as S-estimates and have slightly higher efficiency. MM-estimates proposed by Yohai (1987) can simultaneously attain high breakdown point and efficiencies. Mallows Generalized M-estimates (Mallows 1975) and Schweppe Generalized M-estimates (Handschin et al. 1975) downweight the high leverage points on x but cannot distinguish “good” and “bad” leverage points, thus resulting in a loss of efficiencies. In addition, these two estimators have low breakdown points when p , the number of explanatory variables, is large. Schweppe one-step (S1S) Generalized M-estimates (Coakley and Hettmansperger 1993) overcome the problems of Schweppe. Generalized M-estimates and are

calculated in one step. They both have high breakdown points and high efficiencies. Gervini and Yohai (2002) proposed a new class of high breakdown point and high efficiency robust estimate called robust and efficient weighted least squares estimator (REWLSE). Lee, MacEachern and Jung (2011) and She and Owen (2011) proposed a new class of robust methods based on the regularization of case specific parameters for each response.

3.1 Outlier detection methods

There are many statistical values computed from the sample data that can be used to identify the existence of outliers. The following are common measures of outliers in multiple linear regression.

1. **Standardized Residuals:** To identify the existence of outliers the standardized residuals

$$d_i = \frac{e_i}{\sqrt{MSE}}, \quad i = 1, 2, \dots, n$$

are computed. Large standardized residuals $d_i > 3$ indicates the existence of outliers (Montgomery, et al. 2003).

2. **Studentized Residuals:** For each residual $e_i = y_i - \hat{y}_i$ compute the studentized residuals

$$r_i = \frac{e_i}{\sqrt{MSE(1 - h_{ii})}}, \quad i = 1, 2, \dots, n$$

Where h_{ii} is the hat matrix, are computed. Large studentized residuals $r_i > 3$ indicates the existence of outliers (Montgomery, et al. 2003).

3. **PRESS Residuals:** For each variable observation $X_{ij}, i = 1, 2, \dots, n, j = 1, 2, \dots, p$ compute the prediction error or the PRESS residuals

$$e_{i_{-1}} = y_i - \hat{y}_{i_{-1}}$$

Where $\hat{y}_{i-1}$ is the fitted value of the i^{th} response based on $n - 1$ observations deleting the i^{th} observed values. The PRESS residuals may be computed from the hat matrix and the residual $e_i = y_i - \hat{y}_i$ as

$$r_{(i)} = \frac{e_i}{(1 - h_{ii})}$$

Where h_{ii} is the diagonal element of $H = X(X'X)^{-1}X'$. If $r_{(i)} > 3$ then the i^{th} observation is identified as outliers (Montgomery, *et al.* 2003).

4. The Hat Matrix: Many authors use the value of h_{ii} the i^{th} diagonal element of $H = X(X'X)^{-1}X'$ to indicate outliers. For $h_{ii} > 2\sqrt{p/n}$ (Rousseeuw, and Leroy, 1987), the i^{th} observation is identified as outlier.

5. Cook's Square Distance: Cook's square distance (Besley et al., 1980) of unit i^{th} is a measure based on the square of the maximum distance between the OLS estimate based on all n points $\hat{\beta}$ and the estimate obtained when the i^{th} point, say $\hat{\beta}_{(i)}$. Cook and Weisberg (1989) suggest examining cases with $CD_i^2 > 0.5$ and that case where $CD_i^2 > 1$ should always be studied. This distance measure can be expressed in a general form as;

$$CD_i^2 = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'(X'X)(\hat{\beta}_{(i)} - \hat{\beta})'}{p\hat{\sigma}^2}$$

$i = 1, 2, \dots, n$.

6. DEFFIT Distance: For each observation i compute $(\hat{y}_i - y_{i(i)})$ or $h_{ii}e_i/(1 - h_{ii})$ which tells how much the predicted value $\hat{y}_i$, at the design point x_i would be affected if the case i^{th} were deleted. The standardized version of $DEFFIT_i$ is

$$DEFFIT_i = (\sqrt{h_{ii}}e_i)/(\sigma_i(1 - h_{ii}))$$

Belsley, Kuh and Welsch(1980) suggested that any observation for which $|DEFFIT_i| > 2\sqrt{p/n}$ warrants attention for outliers.

3.2 Some robust regression methods

i. Least Absolute Deviation (LAD)

Least absolute deviation (LAD, also called median regression), estimates are achieved by minimizing the sum of the absolute values of the residuals

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n |y_i - x_i^T \beta|. \quad (3.1)$$

The LAD is also called L_1 estimate due to the L_1 norm used. Although LAD is more resistant than OLS to unusual y values, it is sensitive to high leverage outliers, and thus has a breakdown point of $BP = 1/n \rightarrow 0$ (Rousseeuw and Yohai 1984). Moreover, LAD estimates have a low efficiency of 0.64 when the errors are normally distributed.

ii. Least Median Squares (LMS) Estimates

The LMS estimates (Siegel 1982) are found by minimizing the median of the squared residuals

$$\hat{\beta} = \arg \min_{\beta} \text{Med} |y_i - x_i^T \beta|. \quad (3.2)$$

One good property of the LMS estimate is that it possesses a high breakdown point of near 0.5. However, the LMS estimate has at best an efficiency of 0.37 when the assumption of normal errors is met (see Rousseeuw and Croux 1993). Moreover, LMS estimates do not have a well-defined influence function because of its convergence rate of $n^{-\frac{1}{3}}$ (Rousseeuw 1982). Despite these limitations, the LMS estimate can be used as the initial estimate for some other high breakdown point and high efficiency robust methods.

iii. Least Trimmed Squares (LTS) Estimates

The LTS estimate (Rousseeuw 1983) is defined as

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^q r_{(i)}(\beta)^2. \quad (3.3)$$

where $r_{(i)}(\beta) = y_{(i)} - x_{(i)}^T \beta$, $r_{(1)}(\beta)^2 \leq \dots \leq r_{(q)}(\beta)^2$ are ordered squared residuals, $q = [n(1 - \alpha) + 1]$, and α is the proportion of trimming. Using $q = \left(\frac{n}{2}\right) + 1$ ensures that the estimator has a breakdown point of $BP = 0.5$, and the convergence rate of $n^{-\frac{1}{2}}$ (Rousseeuw 1983). Although highly resistant to outliers, LTS suffers badly in terms of very low efficiency, which is about 0.08, relative to OLS estimates (Stromberg, Hawkins and Hossjer 2000). The reason that LTS estimates call attentions to us is that it is traditionally used as the initial estimate for some other high breakdown point and high efficiency robust methods.

iv. M- Estimates

By replacing the least squares criterion (2) with a robust criterion, M-estimate (Huber, 1964) of β is

$$\hat{\beta} = \arg \min_{\beta} \sum_{i=1}^n \rho \left(\frac{y_i - x_i^T \beta}{\hat{\sigma}} \right) \quad (3.4)$$

where $\rho(\cdot)$ is a robust loss function and $\hat{\sigma}$ is an error scale estimate. The derivative of ρ , denoted by $\psi(\cdot) = \rho'(\cdot)$, is called the influence function. In particular, if $\rho(t) = \frac{1}{2}t^2$ then the solution is the OLS estimate. The OLS estimate is very sensitive to outliers. Rousseeuw and Yohai (1984) indicated that OLS estimates have a breakdown point (BP) of $BP = 1/n$, which tends to zero when the sample size n is getting large. Therefore, one single unusual observation can have large impact on the OLS estimate. One of the commonly used robust loss functions is Huber's ψ function (Huber 1981), where $\psi_c(t) = \rho'(t) = \max\{-c, \min(c, t)\}$. Huber (1981) recommends using $c = 1.345$ in practice. This choice produces a relative efficiency of

approximately 95% when the error density is normal. Another possibility for $\psi(\cdot)$ is Tukey's bisquare function $\psi_c(t) = t\{1 - (t/c)^2\}_+^2$. The use of $c = 4.685$ produces 95% efficiency.

v. S-Estimates

S-estimates (Rousseeuw and Yohai 1984) are defined by

$$\hat{\beta} = \arg \min_{\beta} \hat{\sigma}(r_1(\beta), \dots, r_n(\beta)), \quad (3.5)$$

where $r_i(\beta) = y_i - x_i^T \beta$ and $\hat{\sigma}(r_1(\beta), \dots, r_n(\beta))$ is the scale M-estimate which is defined as the solution of

$$\frac{1}{n} \sum_{i=1}^n \left(\frac{r_i(\beta)}{\hat{\sigma}} \right) = \delta \quad (3.6)$$

for any given β , where δ is taken to be $E_{\phi}[\rho(r)]$. For the biweight scale, S-estimates can attain a high breakdown point of $BP = 0.5$ and has an asymptotic efficiency of 0.29 under the assumption of normally distributed errors (Maronna, Martin, and Yahai, 2006).

vi. MM-Estimates

First proposed by Yohai (1987), MM-estimates have become increasingly popular and are one of the most commonly employed robust regression techniques. The MM-estimates can be found by a three-stage procedure. In the first stage, compute an initial consistent estimate $\hat{\beta}_0$ with high breakdown point but possibly low normal efficiency. In the second stage, compute a robust M-estimate of scale $\hat{\sigma}$ of the residuals based on the initial estimate. In the third stage, find an M-estimate $\hat{\beta}$ starting at $\hat{\beta}_0$. In practice, LMS or S-estimate with Huber or bisquare functions is typically used as the initial estimate $\hat{\beta}_0$. Let $\rho_0(r) = \rho_1(r/k_0)$, $\rho(r) = \rho_1(r/k_1)$, and assume that each of the ρ -functions is bounded. The scale estimate $\hat{\sigma}$ satisfies

$$\frac{1}{n} \sum_{i=1}^n \rho_0 \left(\frac{r_i(\hat{\beta})}{\hat{\sigma}} \right) = 0.5 \quad (3.7)$$

If the ρ -function is biweight, then $k_0 = 1.56$ ensures that the estimator has the asymptotic $BP = 0.5$. Note that an M-estimate minimizes

$$L(\beta) = \sum_{i=1}^n \rho\left(\frac{r_i(\hat{\beta})}{\hat{\sigma}}\right) \quad (3.8)$$

Let ρ satisfy $\rho \leq \rho_0$. Yohai (1987) showed that if $\hat{\beta}$ satisfies $L(\beta) \leq L(\hat{\beta}_0)$, then $\hat{\beta}$'s BP is not less than that of $\hat{\beta}_0$. Furthermore, the breakdown point of the MM-estimate depends only on k_0 and the asymptotic variance of the MM-estimate depends only on k_1 . We can choose k_1 in order to attain the desired normal efficiency without affecting its breakdown point. In order to let $\rho \leq \rho_0$, we must have $k_1 \geq k_0$; the larger the k_1 is, the higher efficiency the MM-estimate can attain at the normal distribution. Maronna, Martin, and Yohai (2006) provides the values of k_1 with the corresponding efficiencies of the biweight ρ -function. However, Yohai (1987) indicates that MM-estimates with larger values of k_1 are more sensitive to outliers than the estimates corresponding to smaller values of k_1 . In practice, an MM-estimate with bisquare function and efficiency 0.85 ($k_1 = 3.44$) starting from a bisquare S-estimate is recommended.

3.3 Robust regression based on regularization of case- specific parameters

She and Owen (2011) and Lee *et al.* (2011) proposed a new class of robust regression methods using the case-specific indicators in a mean shift model with regularization method. A mean shift model for the linear regression is

$$y = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma} + \varepsilon, \varepsilon \sim N(0, \sigma^2 I)$$

Where $y = (y_1, y_2, \dots, y_n)^T$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)^T$, and the mean shift parameter γ_i is non-zero when the i th observation is an outlier and zero otherwise. Due to the sparsity of γ_i s, She and Owen (2011) and Lee *et al.* (2011) proposed to estimate $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by minimizing the penalized least squares using L_1 penalty:

$$L(\boldsymbol{\beta}, \boldsymbol{\gamma}) = \frac{1}{2} \{\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma})\}^T \{\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\gamma})\} + \lambda \sum_{i=1}^n |\gamma_i|, \quad (3.9)$$

Where λ are fixed regularization parameters for $\boldsymbol{\gamma}$. Given the estimate of $\hat{\boldsymbol{\gamma}}$, $\hat{\boldsymbol{\beta}}$ is the OLS estimate with $\mathbf{y}$ replaced by $\mathbf{y} - \boldsymbol{\gamma}$. She and Owen (2011) and Lee *et al.* (2011) proved that the above estimate is in fact equivalent to the M-estimate if Huber's ψ function is used. However, their proposed robust estimates are based on different perspective and can be extended to many other likelihood based models.

3.4 Modified robust regression method

The robust regression method proposed here is based on regularization of case specific parameter method originally developed by She and Owen (2011) and Lee *et al.* (2011). We intend to modify the method using the following steps;

1. Identification of influential (outlying) observations,
2. Estimating the mean shift parameter $\boldsymbol{\gamma}$ using Ridge Regression method,
3. Estimating $\boldsymbol{\beta}$ using OLS given $\hat{\boldsymbol{\gamma}}$ from step 2.

3.5 Ridge regression

The Ordinary least square methods tend to produce estimates that are imprecise and unstable, leading to poor prediction. In order to prevent this difficulties of ordinary least squares method (OLS). Hoerl and Kennard (1970) suggested the ridge regression as an alternative procedure to the least square method in regression analysis. The ridge regression which is the linear transformation of the least square method is based on adding a biasing constant λ to the diagonal of $X'X$ matrix before computing β 's. Therefore, ridge regression is given by

$$\hat{\boldsymbol{\beta}}(\lambda) = (\mathbf{X}'\mathbf{X} + \lambda \mathbf{I})^{-1} \mathbf{X}'\mathbf{Y} \quad \lambda \geq 0 \quad (3.10)$$

where λ is the ridge parameter and I is the identity matrix. The value of λ is appropriate between the interval of $(0, 1)$. Note that if $\lambda = 0$ the ridge estimator becomes the ordinary least square. The values of λ will be selected by the analyst. The corresponding values of the ridge parameter produces different regression coefficient. As the value of λ increases from zero the smaller the variance, the greater the biased introduced. It is always difficult to select the optimal value of λ that produces the stable regression coefficients. Some various methods are used in selecting the appropriate λ .

3.5.1 Methods for selecting the value of parameter λ In Ridge regression modelling

The main step in the ridge regression analysis is to select a value of λ and to obtain the corresponding estimates of the regression coefficients. There are several methods of selecting the appropriate value of λ . The two most often used procedure can be summarized as follows:

- i. Ridge trace: Hoerl and Kennard (1970) suggested a graphical method called ridge trace to select the value of the ridge parameter λ . This is a plot of the values of the regression coefficient plotted against the range of the values of λ .
- ii. Cross validation: the cross validation method does not make use of a particular statistic, like the above; rather it considers several models, using ridge parameter between 0 and 1. For this work, an interval of 0.0001 between successive ridge parameter was used. The one which gives the minimum MSE of regression parameter is selected as the best model.

3.5.2 Derivation of the modified Robust Estimator:

Consider the mean shift model;

$$y = X\beta + X_0\gamma + \varepsilon \quad (3.11)$$

where X_0 is the design matrix for the mean shift parameter γ , whose column is either 1 or 0 and sum of each column is 1.

Theorem 3.1: Given the model in (3.11), the robust estimator of the parameter β , given the shrinkage parameter λ and the associated mean shift parameter estimator $\hat{\gamma}_\lambda$ is $\hat{\beta}_\lambda^{\hat{\gamma}} = (X'X)^{-1}X'(y - \hat{\gamma})$.

proof:

Assumption 1: If we merge the columns of X and X_0 to obtain X_* and correspondingly merging the parameters β and γ to obtain a new extended vector θ , thus (3.11) becomes;

$$y = X_*\theta + \varepsilon \quad (3.12)$$

Assumption 2: If we define the constrained error ε variance $\varepsilon'\varepsilon$ as $[\varepsilon'\varepsilon]_\lambda$ given as;

$$[\varepsilon'\varepsilon]_\lambda = \min_{\theta} (y - X_*\theta)'(y - X_*\theta) + \lambda\theta'\theta$$

Our interest is to estimate the regularization parameter γ using ridge regression method thus; minimizing $[\varepsilon'\varepsilon]_\lambda$. Since $\gamma \in \theta$, estimating θ leads to estimating γ .

Now, the partial derivative of $[\varepsilon'\varepsilon]_\lambda$ with respect to γ is;

$$\frac{\partial (y - X_*\theta)'(y - X_*\theta)}{\partial \theta} = 2X_*'(y - X_*\theta)$$

and $\frac{\partial \lambda\theta'\theta}{\partial \theta} = 2\lambda\theta$;

Gathering together gives the first order condition;

$$X_*'y = X_*'X_*\theta + \lambda\theta$$

$$X_*'y = [X_*'X_* + \lambda I]\theta$$

$$\hat{\theta} = [X_*'X_* + \lambda I]^{-1}X_*'y$$

Given the estimate of $\hat{\gamma}$, $\hat{\beta}$ is the OLS estimate with y replaced by $y - \hat{\gamma}$, specifically the robust estimator is thus;

$$\hat{\beta}_\lambda^{\hat{\gamma}} = (X'X)^{-1}X'(y - \hat{\gamma}) \quad (3.13)$$

There exist a close relationship between the modified robust method and ridge estimator, therefore provided the ridge parameter in the model is appropriately defined as well as the associated mean shift parameter $\hat{\gamma}_\lambda$, the modified robust estimator will ensure a minimum mean square error (MMSE) estimator when compared to the benchmark OLS estimator.

Theorem 3.2: Given the robust estimator in (3.12), under mild regularity condition, $\hat{\beta}_\lambda^{\hat{\gamma}} = (X'X)^{-1}X'(y - \hat{\gamma})$ will yield a minimum mean square error (MMSE) estimator define by $MSE(\hat{\beta}_\lambda^{\hat{\gamma}}) = W_\lambda(\sigma^2[X_*'X_*]^{-1})W_\lambda' + (W_\lambda - I)\beta\beta'(W_\lambda - I)'$.

Proof:

Assumption 3: If we define $\hat{\beta}_\lambda^{\hat{\gamma}}$ as a sub estimate from the extended vector θ with associated ridge estimate $\hat{\theta} = [X_*'X_* + \lambda I]^{-1}X_*'y$, there exist a one-to-one correspondence between $MSE(\hat{\beta}_\lambda^{\hat{\gamma}})$ and $MSE(\hat{\theta})$. The derivation of $MSE(\hat{\theta})$ follow as:

Assumption 4: If we define a linear function W_λ of $\hat{\theta}$ over the OLS estimator $\hat{\theta}_{OLS} = [X_*'X_*]^{-1}X_*'y$ as; $\hat{\theta} = W_\lambda\hat{\theta}_{OLS}$; where $W_\lambda = [X_*'X_* + \lambda I]^{-1}X_*'X_*$. Thus, we can derive the $MSE(\hat{\theta})$ as:

$$MSE(\hat{\theta}) = E([\hat{\theta} - \theta][\hat{\theta} - \theta]')$$

$$MSE(\hat{\theta}) = E([W_\lambda\hat{\theta}_{OLS} - \theta][W_\lambda\hat{\theta}_{OLS} - \theta]')$$

Assumption 5: Recall that the mean square error of a parameter vector β can be further decompose into variance and square bias such that:

$$MSE(\hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

$$MSE(\hat{\beta}) = E(\hat{\beta} - E(\hat{\beta}) + E(\hat{\beta}) - \beta)(\hat{\beta} - E(\hat{\beta}) + E(\hat{\beta}) - \beta)'$$

$$MSE(\hat{\beta}) = V(\hat{\beta}) + Bias(\hat{\beta})Bias(\hat{\beta})'$$

$$MSE(\hat{\theta}) = V(\hat{\theta}) + Bias(\hat{\theta})Bias(\hat{\theta})'$$

$$Bias(\hat{\theta}) = E(\hat{\theta} - \theta)$$

$$Bias(\hat{\theta}) = E(W_\lambda \hat{\theta}_{OLS} - \theta)$$

$$Bias(\hat{\theta}) = W_\lambda E(\hat{\theta}_{OLS}) - \theta$$

Assumption 6: The OLS estimator is an unbiased estimator of θ , such that $E(\hat{\theta}_{OLS}) = \theta$.

$$Bias(\hat{\theta}) = \theta(W_\lambda - I)$$

$$Bias(\hat{\theta}) = \theta([X_*'X_* + \lambda I]^{-1}X_*'X_* - I)$$

Similarly, the $V(\hat{\theta})$ follows as;

$$V(\hat{\theta}) = V(W_\lambda \hat{\theta}_{OLS})$$

$$V(\hat{\theta}) = W_\lambda V(\hat{\theta}_{OLS}) W_\lambda'$$

Assumption 7: The variance of OLS estimator is defined as $V(\hat{\theta}_{OLS}) = \sigma^2[X_*'X_*]^{-1}$.

Thus,

$$V(\hat{\theta}) = W_\lambda (\sigma^2[X_*'X_*]^{-1}) W_\lambda'$$

Therefore,

$$MSE(\hat{\theta}) = W_\lambda (\sigma^2[X_*'X_*]^{-1}) W_\lambda' + (W_\lambda - I)\theta\theta'(W_\lambda - I)'$$

$$\begin{aligned} MSE(\hat{\theta}) &= [([X_*'X_* + \lambda I]^{-1}X_*'X_* - I)](\sigma^2[X_*'X_*]^{-1})[([X_*'X_* + \lambda I]^{-1}X_*'X_* - I)]' \\ &\quad + (([X_*'X_* + \lambda I]^{-1}X_*'X_* - I) - I)\theta\theta'(([X_*'X_* + \lambda I]^{-1}X_*'X_* - I) - I)' \end{aligned}$$

Proposition 3.1: (Farebrother, 1976)

Let $\mathbf{A}$ be a $(p \times p)$ -dimensional, positive definite matrix, $\mathbf{b}$ be a nonzero p dimensional vector,

and $c \in \mathbb{R}^+$. Then, $c\mathbf{A} - \mathbf{b}\mathbf{b}' > 0$ if and only if $\mathbf{b}'\mathbf{A}^{-1}\mathbf{b} > c$.

We are now ready to proof the main result, formalized as Theorem 3.2, that for some λ the

modified robust $\hat{\beta}_\lambda^Y$ regression estimator yields a lower MSE than the OLS regression estimator

and likewise other methods defined over it. It follows that the $MSE(\hat{\theta}_{OLS}) = V(\hat{\theta}_{OLS})$.

Thus,

$$\begin{aligned}
MSE(\hat{\theta}_{OLS}) - MSE(\hat{\theta}) &= \sigma^2[X'_*X_*]^{-1} - W_\lambda(\sigma^2[X'_*X_*]^{-1})W'_\lambda + (W_\lambda - I)\theta\theta'(W_\lambda - I)' \\
&= \sigma^2W_\lambda[2\lambda[X'_*X_*]^{-2} + \lambda^2[X'_*X_*]^{-3}]W'_\lambda \\
&\quad - \lambda^2[X'_*X_* + \lambda I]^{-1}\theta\theta'\{[X'_*X_* + \lambda I]^{-1}\}'
\end{aligned}$$

$$MSE(\hat{\theta}_{OLS}) - MSE(\hat{\theta}) = \lambda[X'_*X_* + \lambda I]^{-1}[2\sigma^2I + \lambda\sigma^2[X'_*X_*]^{-1} - \lambda\theta\theta']\{[X'_*X_* + \lambda I]^{-1}\}'$$

The last equation is only positive definite if and only if $2\sigma^2I + \lambda\sigma^2[X'_*X_*]^{-1} - \lambda\theta\theta' > 0$. It is immediately clear that $2\sigma^2I - \lambda\theta\theta' > 0$. Thus, by proposition 3.1, $2\sigma^2(\theta'\theta)^{-1} > \lambda$. If such λ exists, $MSE(\hat{\theta}_{OLS}) - MSE(\hat{\theta}) > 0$.

Now since $\hat{\beta}_\lambda^{\hat{Y}} \subset \hat{\theta}$,

$$MSE(\hat{\beta}_\lambda^{\hat{Y}}) = W_\lambda(\sigma^2[X'_*X_*]^{-1})W'_\lambda + (W_\lambda - I)\beta\beta'(W_\lambda - I)'$$

And thus, $MSE(\hat{\beta}_{OLS}) - MSE(\hat{\beta}_\lambda^{\hat{Y}}) > 0$.

CHAPTER FOUR

SIMULATION STUDY AND PRESENTATION OF RESULTS

4.0 Introduction

In order to compare the robust methods that were discussed in chapter three, two studies are considered; the first one is a simulation study, in which different cases will be used to compare the existing robust regression methods and the proposed method with the existing methods. The second is based on a real life data that contain outliers, and a comparison study will be implemented on the proposed method and the existing robust methods.

4.1 Dataset simulation

We compare different robust methods and report the mean squared errors (MSE) and relative efficiency of the parameter estimates for each estimation method. OLS estimate is compared with seven other commonly used robust regression estimates: the M estimate using Huber's ψ function (M-H), the M estimate using Tukey's bi-square function (M-T), the S estimate, the LTS estimate, the LMS estimate, the MM estimate (using bi-square weights and $k_1 = 4.68$), the LAD and the proposed method (PM). The data generation processes that follow were adapted from Yu, Yao and Bai (2014). We generated n samples from the model

$$Y = X\boldsymbol{\beta} + \varepsilon,$$

where $\mathbf{X} = [X_1, X_2, X_3] \sim N_3(\mathbf{0}, I)$, $\boldsymbol{\beta} = [1, 1, 1]$. In order to compare the performance of different methods, we consider the following three cases for the error density of ε and independent variables $\mathbf{X}$:

Case I (With y direction Outlier): $\varepsilon \sim (1 - \pi)N(0, 1) + \pi N(0, 10^2)$ - contaminated normal mixture with $\pi = 0.1, 0.2, 0.3, 0.4, 0.5$ proportions of contamination.

Case II (With $\mathbf{x}$ direction Outlier): $\mathbf{X} \sim (1 - \pi)N_3(\mathbf{0}, I) + \pi N_3(\mathbf{0}, 10^2 I)$ - contaminated multivariate normal mixture with $\pi = 0.1, 0.2, 0.3, 0.4, 0.5$ proportions of contamination.

Case III (With $\mathbf{x}, y$ direction Outlier): $\mathbf{X} \sim (1 - \pi)N_3(\mathbf{0}, I) + \pi N_3(\mathbf{0}, 10^2 I)$ - contaminated multivariate normal mixture and $\varepsilon \sim (1 - \pi)N(0, 1) + \pi N(0, 10^2)$ with $\pi = 0.1, 0.2, 0.3, 0.4, 0.5$ proportions of contamination, and Overall cases and scenarios, the sample size $n = 10, 30, 100, 200$ and 1000 were used. The replication size was fixed at 1000.

4.2 Criteria for assessing the estimators performance

Mean Square Error of Parameter:

$$MSE_i = \frac{\sum_{j=1}^p (\hat{\beta}_j - \beta_j)^2}{p}$$

$$AMSE = \frac{\sum_{i=1}^I MSE_i}{I}$$

Relative Efficiency of Robust Estimators:

$$RE = \frac{AMSE_{OLS}}{AMSE_{RE}}$$

Where I is the number of iteration and p is the number of parameters in the model.

And for the real life data, the jackknife procedure was used to assess the stability of the results.

The jackknife estimator for each performance metrics are;

$$\begin{aligned}\hat{\theta} &= \frac{\sum_{i=1}^n \hat{\theta}_i}{n} \\ SE(\hat{\theta}) &= \sqrt{\frac{\sum_{i=1}^n [\hat{\theta}_i - \hat{\theta}]^2}{n-1}} \\ \hat{\sigma}^2 &= \frac{\sum_{i=1}^n \hat{\sigma}_i^2}{n} = \frac{\sum_{i=1}^n \left(\frac{\sum_{j=1}^n [\hat{y}_i - y_j]^2}{n-p} \right)_i}{n} \\ \hat{R}^2 &= \frac{\sum_{i=1}^n \hat{R}_i^2}{n}\end{aligned}$$

Where $\hat{\theta}_i$, $\hat{\sigma}_i^2$, and $\hat{R}_i^2$ are i^{th} jackknife estimate based on removing i^{th} observation.

4.3 Results

Table 4.1: MSE of various methods over sample size for case I at $\pi = 0.1$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.9N(0,1) + 0.1N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	1.1388	0.3525	0.4111	0.3858	0.2747	9.6005	0.7657	0.3650	0.1310
$n = 30$	0.4022	0.0689	0.0555	0.0456	0.0443	0.2033	0.1728	0.1120	0.0371
$n = 100$	0.1200	0.0190	0.0150	0.0123	0.0123	0.0620	0.0668	0.0393	0.0110
$n = 200$	0.0691	0.0098	0.0077	0.0062	0.0062	0.0361	0.0397	0.0239	0.0056
$n = 1000$	0.0220	0.0020	0.0017	0.0012	0.0012	0.0126	0.0124	0.0125	0.0011
Average	0.3504	0.0904	0.0982	0.0902	0.0677	1.9829	0.2115	0.1105	0.0372

Table 4.2: Relative efficiency of various methods over sample size for case I at $\pi = 0.1$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.9N(0,1) + 0.1N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	9.6800	31.2600	26.8100	28.5600	40.1200	1.1500	14.3900	30.1900	84.1200
$n = 30$	8.4000	49.0600	60.9000	74.1200	76.3000	16.6300	19.5600	30.1800	91.1100
$n = 100$	8.3300	52.6300	66.6700	81.300	81.300	16.1300	14.9700	25.4500	90.910
$n = 200$	7.2400	51.0200	64.9400	80.6500	80.6500	13.8500	12.5900	20.9200	89.2900
$n = 1000$	4.5500	50.0000	58.8200	83.3300	83.3300	7.9400	8.0600	8.0000	90.9100
Average	7.6400	46.7900	55.6300	69.5900	72.3400	11.1400	13.9100	22.9500	89.2700

Table 4.3: MSE of various methods over sample size for case I at $\pi = 0.2$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.8N(0,1) + 0.2N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	2.2703	0.8241	1.0615	0.9610	0.5032	38.1471	1.1539	0.4902	0.1620
$n = 30$	0.7421	0.0951	0.1005	0.0603	0.0629	0.2206	0.1866	0.1135	0.0429
$n = 100$	0.2519	0.0242	0.0239	0.0154	0.0156	0.0619	0.0634	0.0386	0.0124
$n = 200$	0.1471	0.0126	0.0123	0.0077	0.0078	0.0349	0.0375	0.0252	0.0063
$n = 1000$	0.0621	0.0029	0.0034	0.0016	0.0016	0.0133	0.0123	0.0149	0.0013
Average	0.6947	0.1918	0.2403	0.2092	0.1182	7.6956	0.2907	0.1365	0.0450

Table 4.4: Relative efficiency of various methods over sample size for case I at $\pi = 0.2$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.8N(0,1) + 0.2N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	4.85	13.37	10.38	11.47	21.9	0.29	9.55	22.48	68.02
$n = 30$	4.55	35.54	33.63	56.05	53.74	15.32	18.11	29.78	78.79
$n = 100$	3.97	41.32	41.84	64.94	64.10	16.16	15.77	25.91	80.65
$n = 200$	3.40	39.68	40.65	64.94	64.10	14.33	13.33	19.84	79.37
$n = 1000$	1.61	34.48	29.41	62.50	62.50	7.52	8.13	6.71	76.92
Average	3.68	32.88	31.18	51.98	53.27	10.72	12.98	20.94	76.75

Table 4.5: MSE of various methods over sample size for case I at $\pi = 0.3$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.7N(0,1) + 0.3N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	3.4703	1.9158	2.2627	2.2185	1.9285	56.0916	2.2738	1.7597	0.2250
$n = 30$	1.2107	0.1368	0.2405	0.1224	0.1132	0.2192	0.1804	0.1156	0.0505
$n = 100$	0.3939	0.0320	0.0426	0.0220	0.0237	0.0612	0.0608	0.0410	0.0145
$n = 200$	0.2403	0.0166	0.0219	0.0107	0.0113	0.0345	0.0338	0.0282	0.0072
$n = 1000$	0.1220	0.0047	0.0079	0.0022	0.0024	0.0136	0.0120	0.0179	0.0014
Average	1.0874	0.4212	0.5151	0.4752	0.4158	11.284	0.5122	0.3925	0.0597

Table 4.6: Relative efficiency of various methods over sample size for case I at $\pi = 0.3$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.7N(0,1) + 0.3N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	3.18	5.75	4.87	4.97	5.71	0.20	4.85	6.26	48.98
$n = 30$	2.79	24.71	14.05	27.61	29.86	15.42	18.74	29.24	66.93
$n = 100$	2.54	31.25	23.47	45.45	42.19	16.34	16.45	24.39	68.97
$n = 200$	2.08	30.12	22.83	46.73	44.25	14.49	14.79	17.73	69.44
$n = 1000$	0.82	21.28	12.66	45.45	41.67	7.35	8.33	5.59	71.43
Average	2.28	22.62	15.58	34.04	32.74	10.76	12.63	16.64	65.15

Table 4.7: MSE of various methods over *sample size* for case I at $\pi = 0.4$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.6N(0,1) + 0.4N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	4.4881	3.4259	3.4521	3.6731	4.1416	66.5670	6.9445	4.5967	0.5664
$n = 30$	1.6808	0.2601	0.6035	0.4380	0.3554	0.2745	0.1923	0.1515	0.0640
$n = 100$	0.5720	0.0473	0.0990	0.0441	0.0503	0.0651	0.0599	0.0481	0.0175
$n = 200$	0.3607	0.0246	0.0502	0.0203	0.0239	0.0374	0.0321	0.0341	0.0084
$n = 1000$	0.2039	0.0079	0.0200	0.0041	0.0050	0.0162	0.0138	0.0230	0.0017
Average	1.4611	0.7532	0.8450	0.8359	0.9152	13.392	1.4485	0.9707	0.1316

Table 4.8: Relative efficiency of various *methods* over sample size for case I at $\pi = 0.4$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.6N(0,1) + 0.4N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	2.46	3.22	3.19	3.00	2.66	0.17	1.59	2.40	19.46
$n = 30$	2.01	13.00	5.60	7.72	9.51	12.31	17.58	22.31	52.81
$n = 100$	1.75	21.14	10.10	22.68	19.88	15.36	16.69	20.79	57.14
$n = 200$	1.39	20.33	9.96	24.63	20.92	13.37	15.58	14.66	59.52
$n = 1000$	0.49	12.66	5.00	24.39	20.00	6.17	7.25	4.35	58.82
Average	1.62	14.07	6.77	16.48	14.59	9.48	11.74	12.9	49.55

Table 4.9: MSE of various methods over *sample size* for case I at $\pi = 0.5$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.5N(0,1) + 0.5N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	5.4475	4.8763	4.6195	5.1074	5.8492	86.4832	13.0509	8.1364	1.0992
$n = 30$	1.9811	0.5086	1.0710	1.0378	0.9602	0.4118	0.3238	0.3831	0.0838
$n = 100$	0.7611	0.0769	0.2565	0.1572	0.1639	0.0808	0.0591	0.0671	0.0215
$n = 200$	0.5102	0.0383	0.1341	0.0607	0.0738	0.0481	0.0347	0.0454	0.0103
$n = 1000$	0.3021	0.0141	0.0567	0.0127	0.0187	0.0218	0.0169	0.0333	0.0020
Average	1.8004	1.1028	1.2276	1.2752	1.4132	17.4091	2.6971	1.7331	0.2434

Table 4.10: Relative efficiency of various methods over sample size for case I at $\pi = 0.5$.

Sample Size	Case I (With y direction outlier): $\varepsilon \sim 0.5N(0,1) + 0.5N(0,10^2)$								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	2.02	2.26	2.39	2.16	1.88	0.13	0.84	1.35	10.03
$n = 30$	1.71	6.65	3.16	3.26	3.52	8.21	10.44	8.82	40.33
$n = 100$	1.31	13.00	3.90	6.36	6.10	12.38	16.92	14.90	46.51
$n = 200$	0.98	13.05	3.73	8.24	6.78	10.40	14.41	11.01	48.54
$n = 1000$	0.33	7.09	1.76	7.87	5.35	4.59	5.92	3.00	50.00
Average	1.27	8.41	2.99	5.58	4.73	7.14	9.71	7.82	39.08

The results presented in Table 4.1 through 4.10 cover mean squared errors (MSE) and relative efficiency (RE) of the parameter estimates for each estimation method with sample size $n = 10, 30, 100, 200$ and 1000 respectively. The number of replicates in each design situation is 1000 . Table 4.1 gives the result for the simulated linear regression model with normal distributed error and 0.1 proportion of outlier in the response y direction (Case I), the proposed method (PM) has the smallest MSE over all the sample size used, H-M, B-M and MM have similar lower MSE but they are not as efficient as PM. LS, LMS, LTS, and S have relatively larger MSE due to outlier effect. Increasing the sample size consistently improves the performances of the estimators in terms of reduction of MSE. The Tables also present the average MSE as well as RE which show the overall performance of the method for each level of contaminant outlier factor. Table 4.2 presents the corresponding relative efficiency of the methods in terms of comparing their MSE to the MSE of LS fitting for model without outlier. The result shows that proposed PM method was still able to achieve 90% efficiency at 0.1 proportion of outlier. The next estimator with similar performance is MM which has about 72% relative efficiency at 0.1 proportion of outlier. This result indicates that at 0.1 proportion of outlier in y direction, the proposed PM method is far better than any of the existing methods considered. Similar results

were observed when the proportion of outlier is increased to 0.2, 0.3, 0.4 and 0.5 proportion of outlier in y direction. In general, for y direction outlier, the proposed PM method was found to be relatively far better than any of the existing method in terms of efficiency. Figure 4.1 and 4.2 were used to assess the breakdown of the methods, it could be observed that the breakdown of the proposed PM method is 0.5 compared to other methods considered.

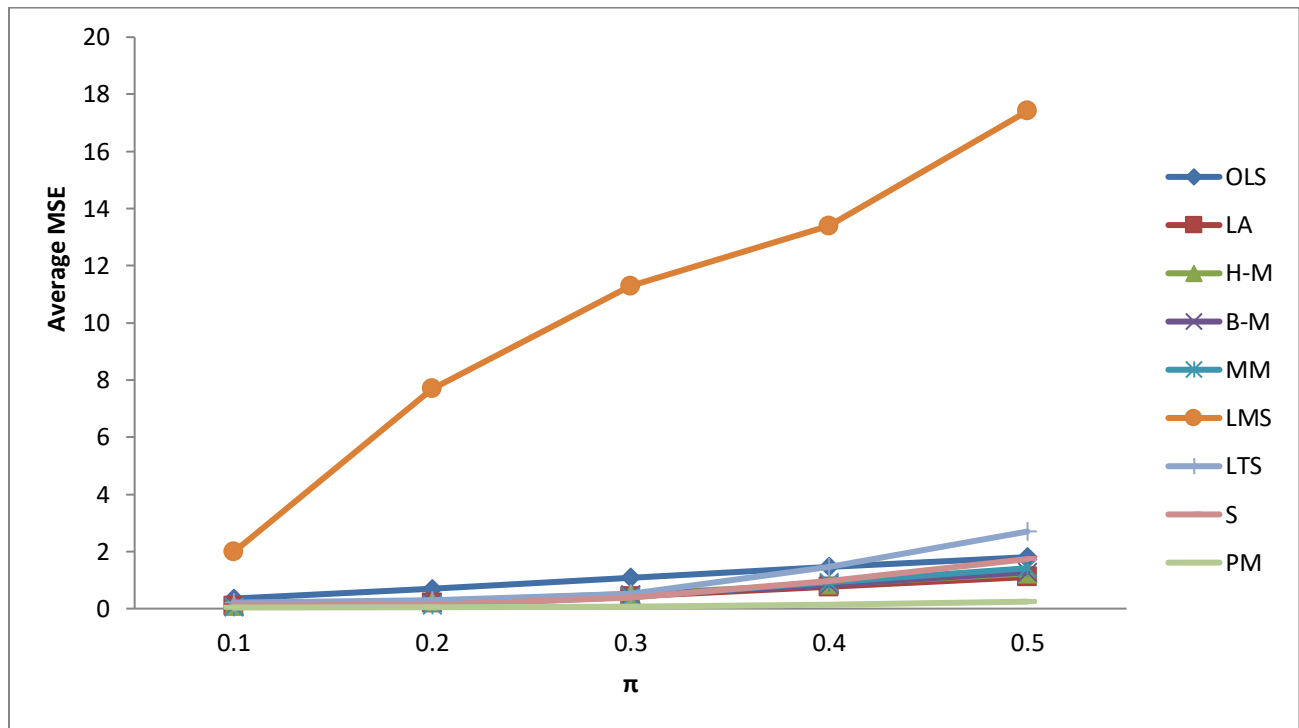


Figure 4.1: Plot of Average MSE against Outlier percentages for case I.

The MSE of the estimators except for PM (grey line) increases with increase in π . The significant increase occurs with LMS (green line). The robust estimators MM, H-M, B-M exhibits similar behavior while S, LTS and LA results are not different.

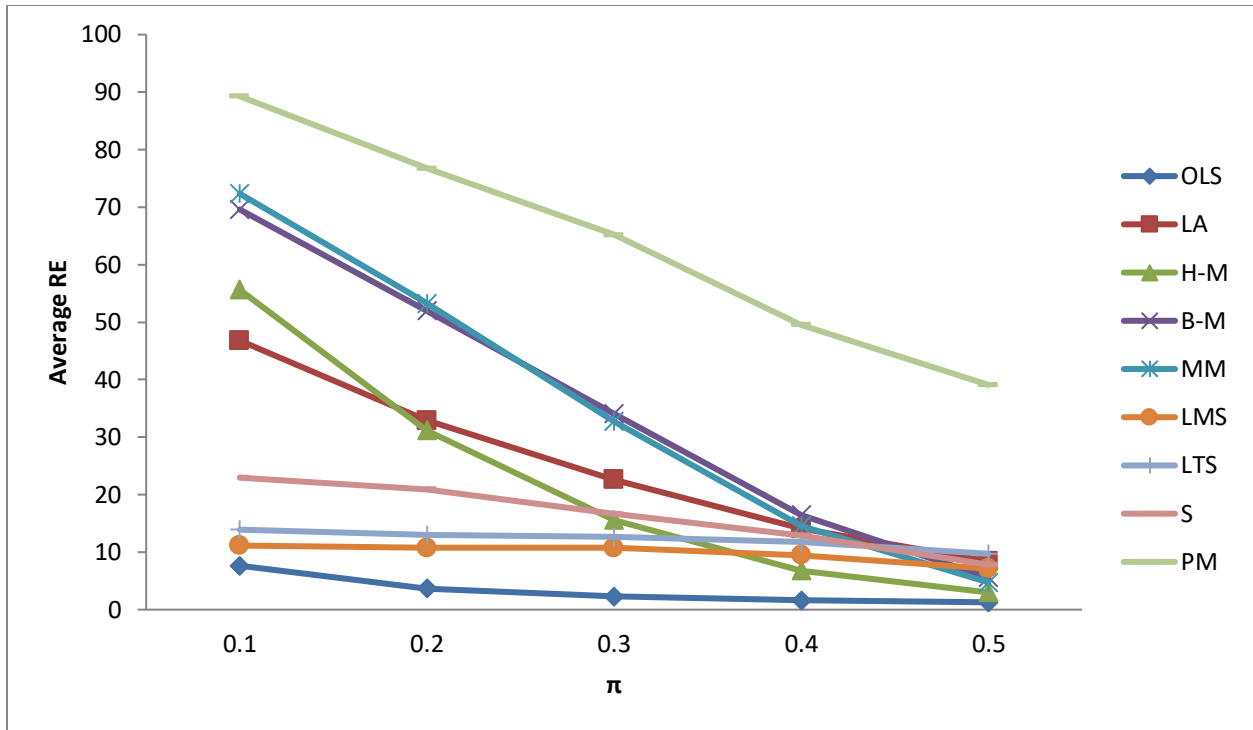


Figure 4.2: Plot of Average RE against Outlier percentages for case I.

The RE of the estimators decreases with increase in π . As expected, OLS has the least RE. The performance of RE for LMS and LTS are not different and likewise S. LA and H-M exhibit similar behavior and they are the only estimators that compete considerably with the proposed estimator (PM). The PM still achieves 50% relative efficiency when $\pi = 0.5$. However, other estimators RE converges around 10% at the same contaminant level 0.5.

Table 4.11: MSE of various methods over sample size for case II at $\pi = 0.1$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	0.3559	0.4787	0.3802	0.4156	0.2512	3.0490	0.6482	0.3709	0.1429
$n = 30$	0.4769	0.4524	0.4522	0.4385	0.0659	0.2117	0.1863	0.1255	0.0384
$n = 100$	0.5946	0.4964	0.5518	0.4417	0.0231	0.0660	0.0692	0.0423	0.0114
$n = 200$	0.6123	0.5236	0.5780	0.4350	0.0128	0.0358	0.0397	0.0255	0.0055
$n = 1000$	0.6277	0.5553	0.6002	0.5265	0.0038	0.0123	0.0121	0.0123	0.0011
Average	0.5335	0.5013	0.5125	0.4515	0.0714	0.6750	0.1911	0.1153	0.0399

Table 4.12: Relative efficiency of various methods over sample size for case II at $\pi = 0.1$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	30.57	22.73	28.62	26.18	43.31	3.57	16.78	29.33	76.14
$n = 30$	7.00	7.38	7.39	7.62	50.68	15.78	17.93	26.61	86.98
$n = 100$	1.66	1.99	1.79	2.24	42.86	15.00	14.31	23.40	86.84
$n = 200$	0.83	0.97	0.88	1.17	39.84	14.25	12.85	20.00	92.73
$n = 1000$	0.16	0.18	0.17	0.19	26.32	8.13	8.26	8.13	90.91
Average	8.04	6.65	7.77	7.48	40.60	11.35	14.03	21.49	86.72

Table 4.13: MSE of various methods over sample size for case II at $\pi = 0.2$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	0.5436	0.7035	0.5748	0.6129	0.3722	4.6178	0.7457	0.4973	0.1794
$n = 30$	0.6625	0.6900	0.6676	0.6663	0.1167	0.2379	0.2175	0.1402	0.0451
$n = 100$	0.6863	0.6865	0.6846	0.6813	0.0515	0.0640	0.0668	0.0445	0.0124
$n = 200$	0.6901	0.6878	0.6878	0.6869	0.0334	0.0373	0.0404	0.0277	0.0064
$n = 1000$	0.6929	0.6885	0.6904	0.6901	0.0141	0.0126	0.0120	0.0134	0.0012
Average	0.6551	0.6913	0.6610	0.6675	0.1176	0.9939	0.2165	0.1446	0.0489

Table 4.14: Relative efficiency of various methods over sample size for case II at $\pi = 0.2$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	19.90	15.38	18.82	17.65	29.07	2.34	14.51	21.76	60.31
$n = 30$	5.04	4.84	5.00	5.01	28.62	14.04	15.36	23.82	74.06
$n = 100$	1.43	1.43	1.43	1.44	19.03	15.31	14.67	22.02	79.03
$n = 200$	0.71	0.71	0.71	0.71	14.67	13.14	12.13	17.69	76.56
$n = 1000$	0.14	0.15	0.14	0.14	7.09	7.94	8.33	7.46	83.33
Average	5.44	4.50	5.22	4.99	19.70	10.55	13.00	18.55	74.66

Table 4.15 MSE of various methods over sample size for case II at $\pi = 0.3$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	0.6994	0.8562	0.7360	0.7825	0.5147	3.3218	0.8010	0.6186	0.2522
$n = 30$	0.7121	0.7359	0.7156	0.7157	0.2839	0.3037	0.2810	0.2483	0.0550
$n = 100$	0.7168	0.7220	0.7175	0.7169	0.3954	0.0793	0.0796	0.0745	0.0144
$n = 200$	0.7145	0.7174	0.7146	0.7145	0.4969	0.0421	0.0417	0.0386	0.0075
$n = 1000$	0.7162	0.7155	0.7156	0.7156	0.6997	0.0133	0.0120	0.0166	0.0014
Average	0.7118	0.7494	0.7199	0.7290	0.4781	0.7520	0.2431	0.1993	0.0661

Table 4.16: Relative efficiency of various methods over sample size for case II at $\pi = 0.3$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	15.38	12.57	14.62	13.75	20.91	3.24	13.43	17.39	42.66
$n = 30$	4.96	4.80	4.93	4.93	12.43	11.62	12.56	14.22	64.18
$n = 100$	1.40	1.39	1.39	1.39	2.53	12.61	12.56	13.42	69.44
$n = 200$	0.71	0.71	0.71	0.71	1.03	12.11	12.23	13.21	68.00
$n = 1000$	0.14	0.14	0.14	0.14	0.14	7.52	8.33	6.02	71.43
Average	4.52	3.92	4.36	4.18	7.41	9.42	11.82	12.85	63.14

Table 4.17: MSE of various methods over *sample size* for case II at $\pi = 0.4$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	0.7614	0.8766	0.7969	0.8400	0.6951	3.3454	0.9759	0.8163	0.4396
$n = 30$	0.7353	0.7595	0.7391	0.7400	0.6137	0.4766	0.4684	0.5353	0.0686
$n = 100$	0.7304	0.7359	0.7309	0.7306	0.7256	0.2775	0.2417	0.5320	0.0181
$n = 200$	0.7281	0.7312	0.7284	0.7283	0.7283	0.2041	0.1439	0.5884	0.0085
$n = 1000$	0.7287	0.7291	0.7286	0.7286	0.7286	0.0688	0.0147	0.6878	0.0017
Average	0.7368	0.7665	0.7448	0.7535	0.6983	0.8745	0.3689	0.6320	0.1073

Table 4.18: Relative efficiency of various methods over *sample size* for case II at $\pi = 0.4$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	14.12	12.26	13.49	12.80	15.47	3.21	11.02	13.17	24.45
$n = 30$	4.46	4.32	4.44	4.43	5.34	6.88	7.00	6.13	47.81
$n = 100$	1.40	1.39	1.40	1.40	1.41	3.68	4.22	1.92	56.35
$n = 200$	0.67	0.67	0.67	0.67	0.67	2.40	3.41	0.83	57.65
$n = 1000$	0.14	0.14	0.14	0.14	0.14	1.45	6.80	0.15	58.82
Average	4.16	3.76	4.03	3.89	4.61	3.52	6.49	4.44	49.02

Table 4.19: MSE of various methods over *sample size* for case II at $\pi = 0.5$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	0.7993	0.8986	0.8218	0.8499	0.8097	6.0614	1.0386	0.9442	1.0106
$n = 30$	0.7408	0.7638	0.7446	0.7462	0.7394	0.7334	0.7303	0.7894	0.0893
$n = 100$	0.7376	0.7423	0.7378	0.7378	0.7378	0.7379	0.6829	0.7649	0.0209
$n = 200$	0.7359	0.7389	0.7361	0.7361	0.7361	0.7553	0.7030	0.7559	0.0101
$n = 1000$	0.7357	0.7365	0.7357	0.7357	0.7357	0.7477	0.7384	0.7472	0.0020
Average	0.7499	0.7760	0.7552	0.7611	0.7517	1.8071	0.7786	0.8003	0.2266

Table 4.20: Relative efficiency of various methods over sample size for case II at $\pi = 0.5$.

Sample Size	Case II (With X direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	13.47	11.99	13.11	12.67	13.30	1.78	10.37	11.41	10.66
$n = 30$	4.66	4.52	4.63	4.62	4.67	4.70	4.72	4.37	38.63
$n = 100$	1.33	1.32	1.33	1.33	1.33	1.33	1.44	1.28	46.89
$n = 200$	0.68	0.68	0.68	0.68	0.68	0.66	0.71	0.66	49.50
$n = 1000$	0.14	0.14	0.14	0.14	0.14	0.13	0.14	0.13	50.00
Average	4.06	3.73	3.98	3.89	4.02	1.72	3.48	3.57	39.14

The results presented in Table 4.11 through 4.20 cover mean squared errors (MSE) and relative efficiency (RE) of the parameter estimates for each estimation method with sample size $n = 10, 30, 100, 200$ and 1000 respectively. The number of replicates in each design situation is 1000 . Table 4.11 gives the result for the simulated linear regression model with normal distributed error and 0.1 proportion of outlier in the response X direction (Case II), the proposed method (PM) has the smallest MSE over all the sample size used, LA, H-M, B-M and OLS have similar large MSE. S, LTS and MM have relatively lower MSE compared to the OLS group due to outlier effect in X direction. Increasing the sample size consistently improves the performances of the estimators in terms of reduction of MSE. The Tables also present the average MSE as well as RE which show the overall performance of the method for each level of contaminant outlier factor. Table 4.12 presents the corresponding relative efficiency of the methods in terms of comparing their MSE to the MSE of LS fitting for model without outlier. It was observed that leverage (X) outlier has more effect than the y direction outlier earlier discussed. The result shows that proposed PM method was still able to achieve 86% efficiency at 0.1 proportion of outlier in X direction. The next estimator with considerable performance is MM which has about 41% relative efficiency at 0.1 proportion of outlier in X direction. This result indicates that at 0.1 proportion of outlier in X direction, the proposed PM method is far

better than any of the existing methods considered. Similar results were observed when the proportion of outlier is increased to 0.2, 0.3, 0.4 and 0.5 proportion of outlier in X direction. In general, for X direction outlier, the proposed PM method was found to be relatively far better than any of the existing method in terms of efficiency. Figure 4.3 and 4.4 were used to assess the breakdown of the methods, it could be observed that the breakdown of the proposed PM method is 0.5 compared to other methods considered.

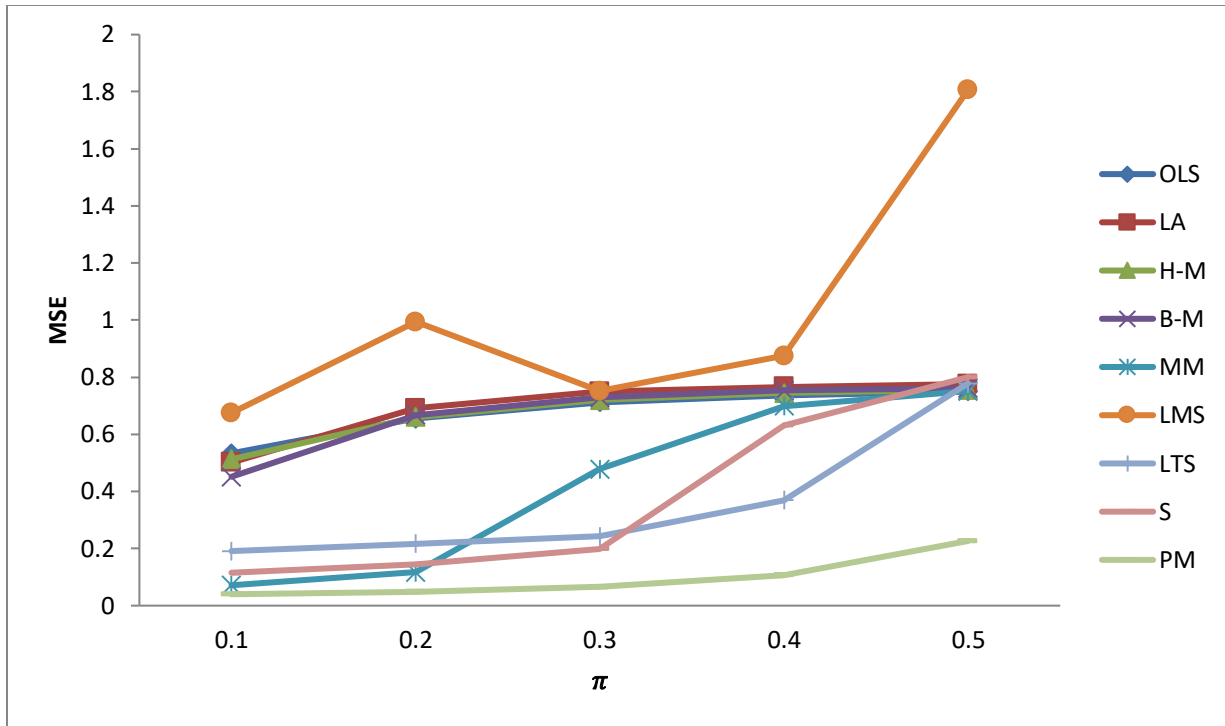


Figure 4.3: Plot of Average MSE against Outlier percentages for case II.

The MSE of the estimators except for PM (grey line) increases significantly with increase in π . The significant increase occurs with LMS (green line). The robust estimators H-M, B-M, LA exhibits similar behavior with OLS while S, LTS and MM results are also similar.

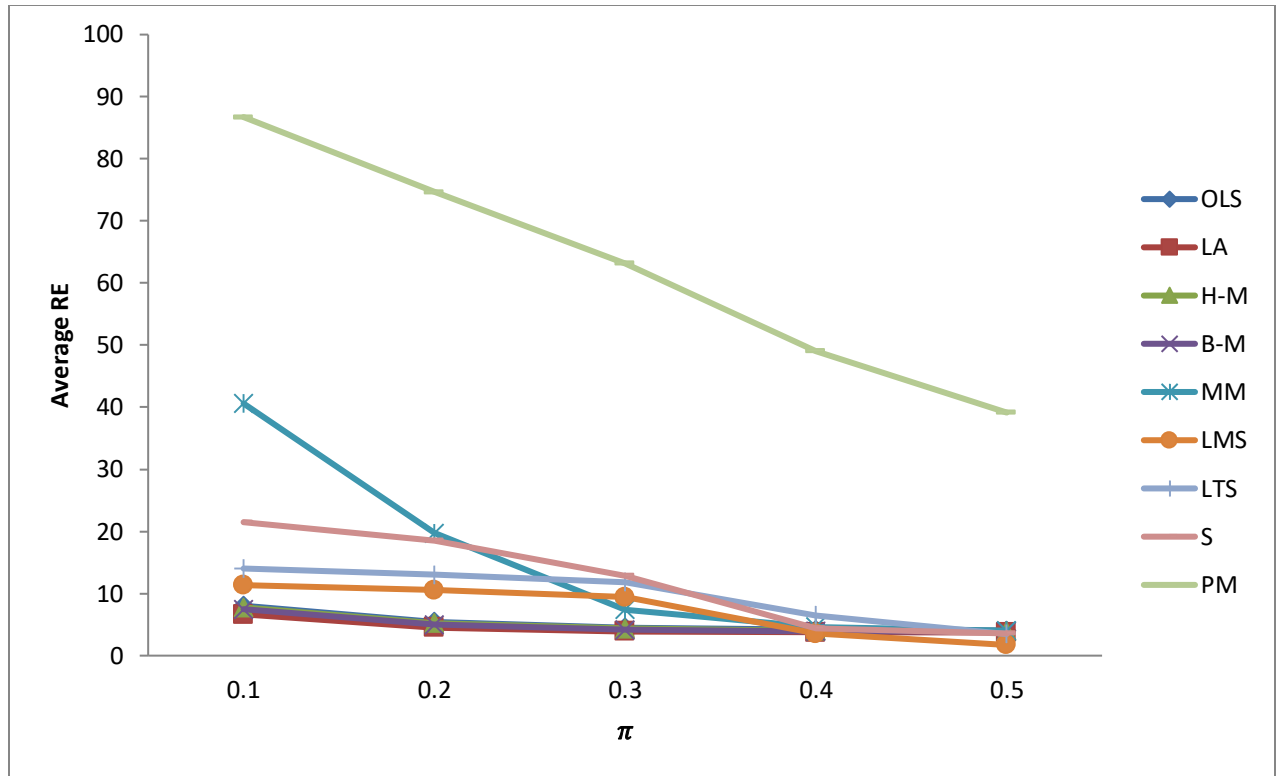


Figure 4.4: Plot of Average RE against Outlier percentages for case II.

The RE of the estimators decreases with increase in π . Here, OLS, B-M, H-M, LA has the least RE. Similar poor performances were observed for LTS, LMS and S. MM estimator withstand the contamination up to 20% of outlier but the performance degrades faster towards 0% efficiency as π climbs 0.3. The PM stood at a distance higher efficiency and still maintains about 40% relative efficiency when $\pi = 0.5$. However, other estimators RE converges around 4% at the same contaminant level 0.5.

Table 4.21: MSE of various methods over sample size for case III at $\pi = 0.1$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	0.5015	0.6446	0.4988	0.5133	0.2785	2.2208	0.5415	0.3743	0.1352
$n = 30$	0.8460	0.4879	0.5959	0.5146	0.0651	0.1944	0.1761	0.1323	0.0400
$n = 100$	0.6597	0.3305	0.4157	0.1785	0.0248	0.0750	0.0720	0.0448	0.0116
$n = 200$	0.6635	0.2785	0.3845	0.0313	0.0142	0.0358	0.0367	0.0245	0.0057
$n = 1000$	0.6329	0.2265	0.3340	0.0035	0.0034	0.0118	0.0119	0.0117	0.0011
Average	0.6607	0.3936	0.4458	0.2482	0.0772	0.5076	0.1676	0.1175	0.0387

Table 4.22: Relative efficiency of various *methods* over sample size for case III at $\pi = 0.1$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	21.75	16.93	21.87	21.25	39.17	4.91	20.15	29.15	80.70
$n = 30$	4.07	7.05	5.77	6.68	52.84	17.70	19.53	26.00	86.00
$n = 100$	1.58	3.15	2.50	5.83	41.94	13.87	14.44	23.21	89.66
$n = 200$	0.77	1.83	1.33	16.29	35.92	14.25	13.90	20.82	89.47
$n = 1000$	0.16	0.44	0.30	28.57	29.41	8.47	8.40	8.55	90.91
Average	5.67	5.88	6.35	15.72	39.86	11.84	15.28	21.55	87.35

Table 4.23: MSE of various methods over sample size for case III at $\pi = 0.2$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	1.2498	1.4292	1.1845	1.2202	0.5248	4.3297	1.0246	0.6505	0.2130
$n = 30$	0.9546	0.6919	0.7464	0.7363	0.0909	0.2350	0.2040	0.1337	0.0425
$n = 100$	0.7686	0.5141	0.6030	0.4230	0.0464	0.0701	0.0705	0.0512	0.0123
$n = 200$	0.7578	0.5213	0.5918	0.2970	0.0294	0.0351	0.0415	0.0269	0.0070
$n = 1000$	0.7030	0.4862	0.5594	0.0538	0.0094	0.0121	0.0116	0.0128	0.0012
Average	0.8868	0.7285	0.7370	0.5461	0.1402	0.9364	0.2704	0.1750	0.0552

Table 4.24: Relative efficiency of various methods over sample size for case III at $\pi = 0.2$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	9.27	8.10	9.78	9.49	22.07	2.67	11.30	17.80	54.37
$n = 30$	3.57	4.93	4.57	4.63	37.51	14.51	16.72	25.50	80.24
$n = 100$	1.22	1.83	1.56	2.22	20.26	13.41	13.33	18.36	76.42
$n = 200$	0.71	1.04	0.91	1.82	18.37	15.38	13.01	20.07	77.14
$n = 1000$	0.14	0.21	0.18	1.86	10.64	8.26	8.62	7.81	83.33
Average	2.98	3.22	3.40	4.00	21.77	10.85	12.60	17.91	74.30

Table 4.25: MSE of various methods over sample size for case III at $\pi = 0.3$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	1.6959	1.7156	1.5168	1.5755	0.7239	3.3111	0.9174	0.8095	0.2983
$n = 30$	1.0967	0.8502	0.8821	0.8562	0.1826	0.2674	0.2161	0.2010	0.0539
$n = 100$	0.8132	0.6204	0.6705	0.5988	0.0804	0.0616	0.0624	0.0528	0.0129
$n = 200$	0.7637	0.6005	0.6575	0.5061	0.0645	0.0386	0.0424	0.0318	0.0076
$n = 1000$	0.7541	0.5949	0.6529	0.4476	0.0409	0.0138	0.0126	0.0177	0.0014
Average	1.0247	0.8763	0.8760	0.7968	0.2185	0.7385	0.2502	0.2226	0.0748

Table 4.26: Relative efficiency of various methods over sample size for case III at $\pi = 0.3$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	7.22	7.13	8.07	7.77	16.91	3.70	13.34	15.12	41.03
$n = 30$	3.09	3.99	3.84	3.96	18.57	12.68	15.69	16.87	62.89
$n = 100$	1.14	1.50	1.39	1.55	11.57	15.10	14.90	17.61	72.09
$n = 200$	0.72	0.92	0.84	1.09	8.53	14.25	12.97	17.30	72.37
$n = 1000$	0.13	0.17	0.15	0.22	2.44	7.25	7.94	5.65	71.43
Average	2.46	2.74	2.86	2.92	11.60	10.60	12.97	14.51	63.96

Table 4.27: MSE of various methods over sample size for case III at $\pi = 0.4$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	2.3510	2.3353	2.0652	2.1073	1.3803	5.3698	1.5573	1.4358	0.6050
$n = 30$	1.3164	0.8986	0.9771	0.9121	0.3402	0.3124	0.2887	0.3175	0.0801
$n = 100$	0.9063	0.7088	0.7681	0.7017	0.1954	0.0713	0.0677	0.0754	0.0183
$n = 200$	0.8311	0.6883	0.7343	0.6833	0.2429	0.0380	0.0350	0.0439	0.0082
$n = 1000$	0.7763	0.6484	0.7012	0.6206	0.2875	0.0153	0.0121	0.0191	0.0016
Average	1.2362	1.0559	1.0492	1.0050	0.4893	1.1614	0.3922	0.3783	0.1426

Table 4.28: Relative efficiency of various methods over sample size for case III at $\pi = 0.4$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	4.86	4.89	5.53	5.42	8.27	2.13	7.33	7.95	18.88
$n = 30$	2.67	3.91	3.59	3.85	10.32	11.24	12.16	11.06	43.82
$n = 100$	1.15	1.47	1.35	1.48	5.32	14.59	15.36	13.79	56.83
$n = 200$	0.63	0.76	0.71	0.76	2.14	13.68	14.86	11.85	63.41
$n = 1000$	0.13	0.15	0.14	0.16	0.35	6.54	8.26	5.24	62.50
Average	1.89	2.24	2.26	2.33	5.28	9.64	11.59	9.98	49.09

Table 4.29: Performance of various methods over sample size for case III at $\pi = 0.5$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	2.2092	1.6587	1.8399	1.8448	1.1602	4.0268	0.9721	1.1182	1.1079
$n = 30$	1.4134	0.9107	1.0899	1.0558	0.5741	0.3638	0.3444	0.4368	0.0876
$n = 100$	0.8969	0.6819	0.7468	0.7054	0.5402	0.1034	0.0821	0.1953	0.0214
$n = 200$	0.8943	0.7195	0.7798	0.7290	0.6487	0.0644	0.0467	0.1497	0.0108
$n = 1000$	0.8056	0.6929	0.7397	0.7029	0.6783	0.0290	0.0138	0.0634	0.0020
Average	1.2439	0.9327	1.0392	1.0076	0.7203	0.9175	0.2918	0.3927	0.2459

Table 4.30: Performance of various methods over sample size for case III at $\pi = 0.5$.

Sample Size	Case III (With X, y direction outlier)								
	METHODS								
	OLS	LA	H-M	B-M	MM	LMS	LTS	S	PM
$n = 10$	5.25	6.99	6.30	6.29	10.00	2.88	11.93	10.37	10.47
$n = 30$	2.54	3.94	3.29	3.40	6.25	9.87	10.42	8.22	40.98
$n = 100$	1.10	1.45	1.33	1.40	1.83	9.57	12.06	5.07	46.26
$n = 200$	0.62	0.76	0.71	0.75	0.85	8.54	11.78	3.67	50.93
$n = 1000$	0.12	0.14	0.14	0.14	0.15	3.45	7.25	1.58	50.00
Average	1.93	2.66	2.35	2.40	3.82	6.86	10.69	5.78	39.73

The results presented in Table 4.21 through 4.30 cover mean squared errors (MSE) and relative efficiency (RE) of the parameter estimates for each estimation method with sample size $n = 10, 30, 100, 200$ and 1000 respectively. The number of replicates in each design situation is 1000 . Table 4.21 gives the result for the simulated linear regression model with normal distributed error and 0.1 proportion of outlier in both X, y direction (Case III), the proposed method (PM) has the smallest MSE over all the sample size used, LA, H-M, B-M and OLS have similar large MSE. S, LTS and MM have relatively lower MSE compared to the OLS group due to outlier effect in X, y direction. Increasing the sample size consistently improves the performances of the estimators in terms of reduction of MSE. The Tables also present the average MSE as well as RE which show the overall performance of the method for each level of contaminant outlier factor. Table 4.22 presents the corresponding relative efficiency of the methods in terms of comparing their MSE to the MSE of LS fitting for model without outlier. It was observed that the X, y outlier has more effect than the y and X alone direction outliers earlier discussed. The result shows that proposed PM method was still able to achieve 87% efficiency at 0.1 proportion of outlier in X, y direction. The next estimator with considerable performance is MM which has about 40% relative efficiency at 0.1 proportion of outlier in X, y

direction. This result indicates that at 0.1 proportion of outlier in X, y direction, the proposed PM method is far better than any of the existing methods considered. Similar results were observed when the proportion of outlier is increased to 0.2, 0.3, 0.4 and 0.5 proportion of outlier in X, y direction. In general, for X, y direction outlier, the proposed PM method was found to be relatively far better than any of the existing method in terms of efficiency. Figure 4.5 and 4.6 were used to assess the breakdown of the methods, it could be observed that the breakdown of the proposed PM method is 0.5 compared to other methods considered.

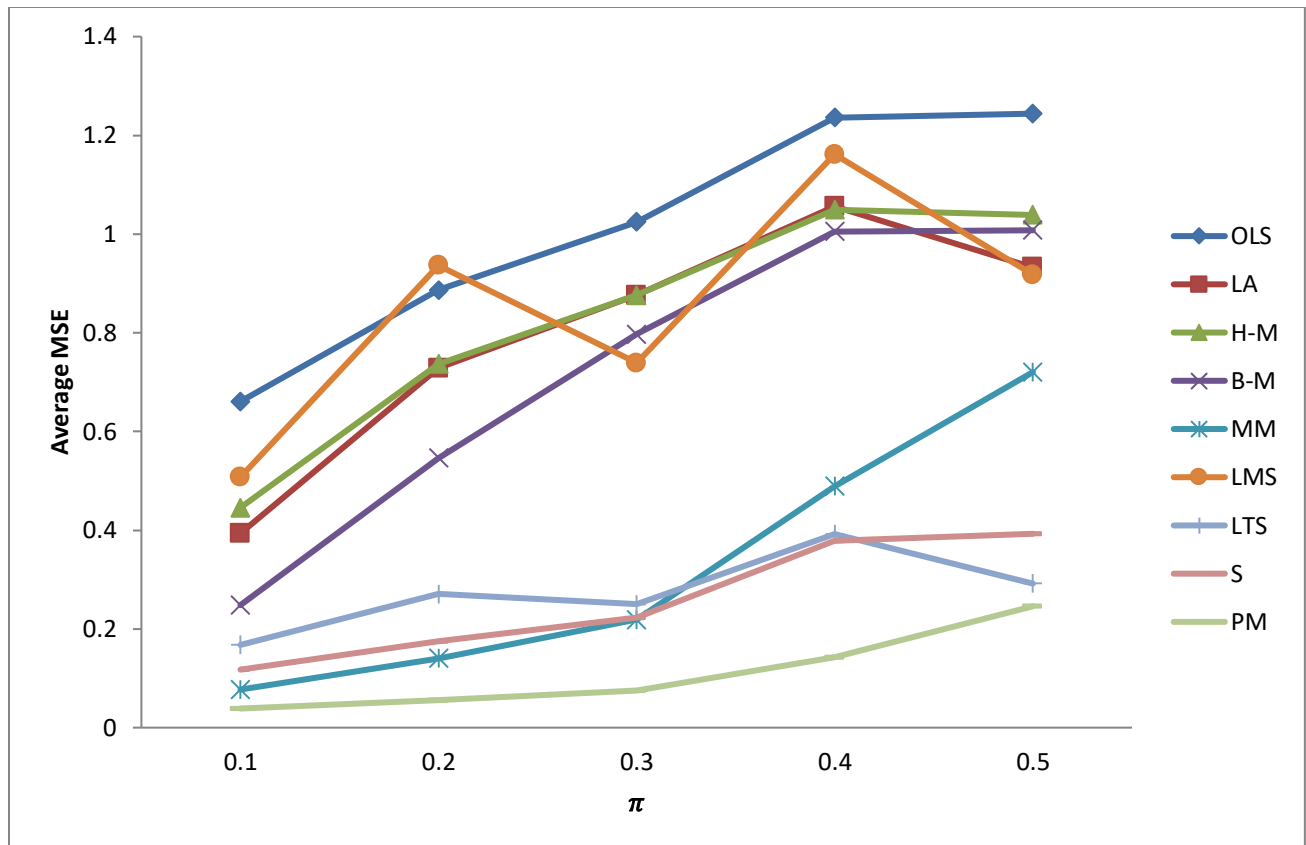


Figure 4.5 Plot of Average MSE against outlier percentages for case III.

The MSE of the estimators except for PM (grey line) increases significantly with increase in π . The significant increase occurs with OLS (blue with triangle line). The robust estimators H-M, B-M, LA and LMS exhibits similar behavior with OLS while S, LTS and MM results are also similar.

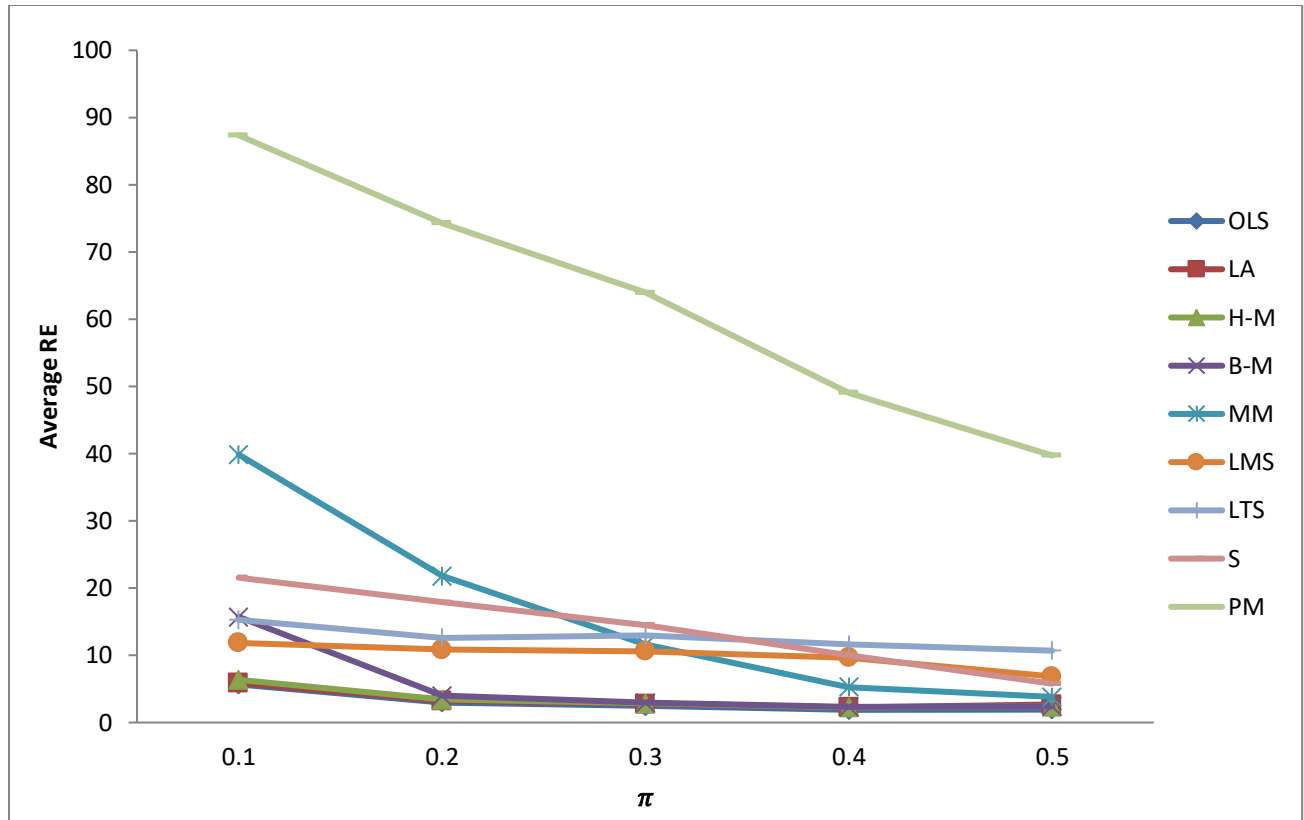


Figure 4.6 Plot of Average RE against outlier percentages for case III.

The RE of the estimators decreases with increase in π . Here, OLS, B-M, H-M, LA has the least RE. Similar poor performances were observed for LTS, LMS and S. MM estimator withstand the contamination up to 20% of outlier but the performance degrades faster towards 0% efficiency as π climbs 0.3. The PM stood at a distance higher efficiency and still maintains about 40% relative efficiency when $\pi = 0.5$. However, other estimators RE converges around 3% at the same contaminant level 0.5.

4.4 Real life data illustration

Datasets used in this research

1. The Stackloss Dataset (Brownlee, 1965; Rousseeuw and Leroy, 1987)

The dataset describes the operation of a plant for the oxidation of ammonia to nitric acid. The $\text{stackloss}(y)$ is a function of the rate(x_1), temperature (x_2) and acid concentration (x_3). The dataset has 21 observations or cases. It is widely acclaimed that the data points (1, 3, 4, 21) and possibly the point (2) are outliers. While the points (1, 3, 4, 21) are considered outliers, Kashyap and Maiyuran (1993) estimate the parameters as (-37.65, 0.80, 0.577, -0.067) of which the first is the y-intercept and the subsequent three are the coefficients associated with x matrix.

2. The Water Salinity Dataset (Rupert and Carrol, 1980; Rousseeuw and Leroy, 1987)

The water salinity (i.e., its salt concentration) dataset (Rupert and Carrol, 1980) comprises data on water salinity (y) as the dependent variable and lagged salinity(x_1), trend (x_2) and river discharge in North Carolina's Pamlico Sound (x_3) as the explanatory variables. The dataset has 28 points. In this dataset, Rousseeuw and Leroy's (1987) method detects the points (5, 16, 23, 24) as outliers whereas Rupert and Carrol's method detects (1, 11, 13, 15, 16, 17) as outliers. Kashyap and Maiyuran's (1993) method detects (5, 8, 15, 16, 17) as outliers for which the regression coefficients are (22.30, 0.724, -0.279, -0.786).

3. The Scottish Hills Races Data (Charterjee and Hadi 2006)

The Scottish Hills Races data consist of a response variable (record times, in minutes) and two explanatory variables (the distance in miles(x_1), and the climb in feet(x_2)) for 35 races in

Scotland in 1984. Charterjee and Hadi (2006) claimed observations (7, 11, 18, 31 and 33) are outliers. They gave the regression estimate as $(-8.9833, 6.2166, 0.0111)$.

4. Growth Study (Zaman, Rousseeuw, and Orhan 2001; De Long and Summers 1991)

Zaman, Rousseeuw, and Orhan (2001) show how robust techniques improve over OLS using results for the growth study of De Long and Summers. De Long and Summers (1991) studied the national growth of 61 countries including Nigeria from 1960 to 1985 using OLS. The response variable is the GDP growth per worker (GDP) and the regressors are the constant term, labor force growth (LFG), relative GDP gap (GAP), equipment investment (EQP), and non-equipment investment (NEQ). Zaman, Rousseeuw, and Orhan (2001) claimed that observation 60 corresponding to Zambia is an outlier. They gave the regression estimates as $(-0.01430, -0.02981, 0.02026, 0.26538, 0.06236)$.

4.4.1 Data analysis and result

The proposed robust regression method strongly depends on outlier detection methods; however different method of outlier detection often yields different result. Thus, in this research we used combined outlier detection method based on all methods described in section 3.1. That's, any observation flagged as an outlier by any of the method is regarded as an outlier in this research.

Table 4.31: Data set and outlier detection results

Data set	Sample size (n)	Percentage of outlier	Outlying points	Direction of outliers
Growth Study	61	5%	[2, 8, 60]	x, y
Scottish Hills Races Data	35	5.7%	[16, 18]	x, y
Water Salinity Dataset	28	7.1%	[4, 16]	X
Stackloss Dataset	21	23.8%	[2, 4, 8, 10, 14]	x, y

Table 4.32: Parameter estimates of the Growth study data set with and without the 5% outlying points.

Method	With outlier					Without outlier				
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
LS	0.0002	-0.1650	0.0910	0.0920	0.0082	-0.0170	0.0436	0.1510	0.1417	0.0148
LA	-0.0134	-0.1060	0.1841	0.1100	0.0157	-0.0238	0.0910	0.2354	0.1294	0.0184
H-M	-0.0128	-0.0740	0.1507	0.1219	0.0152	-0.0204	0.0197	0.1769	0.1433	0.0181
B-M	-0.0209	-0.0015	0.1901	0.1386	0.0195	-0.0209	0.0305	0.1817	0.1430	0.0184
MM	-0.0190	-0.0062	0.1785	0.1356	0.0182	-0.0188	0.0359	0.1656	0.1418	0.0165
LMS	-0.0118	0.0460	0.2433	0.0543	0.0178	-0.0086	0.7136	0.2810	0.0337	0.0018
LTS	-0.0194	-0.0595	0.2310	0.1066	0.0218	-0.0093	0.7574	0.2922	0.0333	0.0014
S	-0.0274	-0.0519	0.2555	0.1411	0.0240	-0.0141	0.6419	0.2730	0.0732	0.0019
PM	-0.0138	0.0417	0.1419	0.1337	0.0125	-0.0146	0.0476	0.1459	0.1353	0.0130

Table 4.33: Absolute bias of estimates for the Growth study dataset with and without the 5% outlying points. The true parameter is assumed to be LS without the outlying point.

Method	With outlier					Without outlier				
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
LS	0.0172	0.2086	0.0701	0.0496	0.0066	0.0000	0.0000	0.0000	0.0000	0.0000
LA	0.0066	0.1886	0.0544	0.0317	0.0027	0.0068	0.1252	0.0844	0.0159	0.0036
H-M	0.0045	0.1227	0.0170	0.0197	0.0028	0.0033	0.0337	0.0259	0.0037	0.0033
B-M	0.0046	0.0670	0.0419	0.0039	0.0055	0.0039	0.0210	0.0307	0.0030	0.0036
MM	0.0038	0.0651	0.0350	0.0061	0.0047	0.0018	0.0100	0.0146	0.0011	0.0017
LMS	0.0167	0.4553	0.1764	0.1112	0.0118	0.0129	0.9266	0.2218	0.1107	0.0195
LTS	0.0129	0.3287	0.1260	0.0476	0.0085	0.0135	0.9320	0.2105	0.1159	0.0222
S	0.0116	0.1997	0.1093	0.0264	0.0101	0.0152	0.7014	0.1432	0.0739	0.0221
PM	0.0032	0.0110	0.0091	0.0079	0.0023	0.0024	0.0050	0.0051	0.0063	0.0018

Table 4.34: Standard error (SE) of estimates for the Growth study data set with and without the 5% outlying points

Method	With outlier					Without outlier				
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$
LS	0.0064	0.0868	0.0457	0.0171	0.0035	0.0038	0.1057	0.0251	0.0073	0.0021
LA	0.0076	0.1390	0.0712	0.0135	0.0033	0.0038	0.2121	0.0336	0.0107	0.0052
H-M	0.0060	0.0812	0.0463	0.0108	0.0032	0.0046	0.0859	0.0327	0.0061	0.0022
B-M	0.0069	0.0704	0.0552	0.0101	0.0040	0.0049	0.1013	0.0366	0.0057	0.0020
MM	0.0069	0.0761	0.0539	0.0123	0.0038	0.0040	0.1058	0.0285	0.0063	0.0020
LMS	0.0184	0.4888	0.1548	0.0546	0.0126	0.0193	0.7538	0.1900	0.0452	0.0228
LTS	0.0180	0.4042	0.1406	0.0560	0.0094	0.0167	0.6559	0.1630	0.0350	0.0216
S	0.0080	0.1730	0.0607	0.0302	0.0060	0.0134	0.6531	0.0773	0.0576	0.0226
PM	0.0037	0.0950	0.0218	0.0077	0.0023	0.0037	0.0992	0.0229	0.0074	0.0023

Table 4.35: Error variance and coefficient of determination for the Growth study data set with and without the 5% outlying points

Method	With outlier		Without outlier	
	$\hat{\sigma}^2$	R^2	$\hat{\sigma}^2$	R^2
LS	0.0002	0.2287	0.0001	0.5260
LA	0.0002	0.1913	0.0001	0.5043
H-M	0.0002	0.1953	0.0001	0.5237
B-M	0.0002	0.1394	0.0001	0.5233
MM	0.0002	0.1534	0.0001	0.5254
LMS	0.0002	0.1432	0.0002	0.0318
LTS	0.0002	0.0618	0.0002	0.0139
S	0.0002	0.0907	0.0002	0.2610
PM	0.0002	0.1703	0.0001	0.5254

Table 4.36: Parameter estimates of the Scottish race data set with and without the 5.7% outlying points

Method	With outlier			Without outlier		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
LS	4.5280	5.2177	0.0060	-8.9324	6.6365	0.0079
LA	-3.5836	5.7162	0.0075	-5.9435	5.9509	0.0085
H-M	-4.5724	6.0550	0.0071	-8.3146	6.4702	0.0081
B-M	-5.9087	6.2029	0.0071	-8.2904	6.4837	0.0081
MM	-6.1145	6.2303	0.0072	-7.6358	6.3858	0.0079
LMS	0.2386	4.9088	0.0082	-6.4201	5.6790	0.0105
LTS	-0.3110	4.9714	0.0080	-8.2975	5.9888	0.0102
S	-1.3071	5.2596	0.0076	-1.5309	5.2749	0.0075
PM	-7.8897	6.4492	0.0079	-7.9885	6.4789	0.0078

Table 4.37: Absolute bias of estimates for the Scottish race data set with and without the outlying 5.7% points

Method	With outlier			Without outlier		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
LS	13.4604	1.7499	0.0025	0.0000	0.0000	0.0000
LA	6.3007	1.0544	0.0009	2.9890	0.6856	0.0008
H-M	5.5306	0.8821	0.0010	0.8228	0.1855	0.0003
B-M	3.9898	0.6644	0.0008	0.8195	0.1658	0.0003
MM	3.8479	0.6451	0.0008	1.3523	0.2548	0.0003
LMS	9.1534	1.7304	0.0005	2.9928	0.9575	0.0026
LTS	8.6299	1.6651	0.0006	4.2451	0.6724	0.0023
S	7.5536	1.3590	0.0010	7.4614	1.3617	0.0010
PM	1.0427	0.1873	0.0002	0.9440	0.1576	0.0001

Table 4.38: Standard error (SE) of estimates for the Scottish race data set with and without the 5.7% outlying points.

Method	With outlier			Without outlier		
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$
LS	9.3255	1.3577	0.0018	2.4532	0.4072	0.0004
LA	4.1336	0.6598	0.0008	3.1311	0.3950	0.0012
H-M	3.5793	0.5444	0.0005	3.0488	0.4942	0.0006
B-M	1.6387	0.3135	0.0003	3.0032	0.4893	0.0006
MM	1.8354	0.3333	0.0003	2.7271	0.4523	0.0006
LMS	0.2342	0.0832	0.0002	5.1435	0.6648	0.0020
LTS	0.2899	0.0729	0.0003	6.0768	0.8277	0.0018
S	2.5863	0.4497	0.0008	2.3977	0.3276	0.0009
PM	2.6328	0.4414	0.0003	2.4918	0.4010	0.0004

Table 4.39: Error variance and coefficient of determination for the Scottish race data set with and without the 5.7% outlying points.

Method	With outlier		Without outlier	
	$\hat{\sigma}^2$	R^2	$\hat{\sigma}^2$	R^2
LS	245.3347	0.5888	24.6176	0.9621
LA	271.7340	0.5396	26.7946	0.9593
H-M	268.8480	0.5408	25.0208	0.9618
B-M	282.1541	0.5183	25.3903	0.9617
MM	281.8388	0.5182	27.8589	0.9605
LMS	308.7882	0.5221	38.0292	0.9458
LTS	309.9242	0.5179	35.5466	0.9504
S	295.5963	0.5264	39.4257	0.9485
PM	293.4022	0.491679	24.70937	0.961911

Table 4.40: Parameter estimates of the Water salinity data set with and without the 7.1% outlying points.

Method	With outlier				Without outlier			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	6.3708	0.7895	-0.0188	-0.1590	17.9822	0.7610	-0.1905	-0.6352
LA	8.2229	0.8018	-0.1542	-0.2263	16.8603	0.7738	-0.1408	-0.5989
H-M	7.4668	0.7911	-0.0534	-0.2018	18.2957	0.7345	-0.2004	-0.6330
B-M	8.2582	0.7871	-0.0743	-0.2316	18.4468	0.7262	-0.1972	-0.6347
MM	18.5969	0.7114	-0.2063	-0.6335	18.6581	0.7376	-0.1932	-0.6521
LMS	22.9201	0.5924	-0.1719	-0.7843	39.9935	0.3814	-0.7797	-1.3697
LTS	28.4516	0.5219	-0.1177	-0.9997	32.7395	0.4915	-0.1224	-1.1795
S	20.1284	0.6744	-0.2379	-0.6795	19.5817	0.6497	-0.2214	-0.6502
PM	16.8412	0.7553	-0.1691	-0.5851	17.7716	0.7429	-0.1824	-0.6183

Table 4.41: Absolute bias of estimates for the Water salinity data set with and without the 7.1% outlying points.

Method	With outlier				Without outlier			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	11.6114	0.0454	0.1717	0.4762	0.0000	0.0000	0.0000	0.0000
LA	9.7593	0.0727	0.0598	0.4089	1.4805	0.0403	0.0712	0.0501
H-M	10.5154	0.0475	0.1372	0.4334	0.7965	0.0293	0.0125	0.0233
B-M	9.7239	0.0454	0.1327	0.4037	1.0590	0.0384	0.0274	0.0334
MM	1.4674	0.0504	0.0249	0.0494	0.9535	0.0264	0.0159	0.0317
LMS	7.2600	0.1680	0.0881	0.2825	22.1212	0.3809	0.6996	0.7384
LTS	10.7181	0.2388	0.0759	0.3828	14.5349	0.2788	0.0836	0.5352
S	3.8665	0.0910	0.0751	0.1343	2.4731	0.1096	0.0633	0.1169
PM	1.1410	0.0057	0.0215	0.0501	0.2106	0.0181	0.0081	0.0170

Table 4.42: Standard error (SE) of estimates for the Water salinity data set with and without the 7.1% outlying points.

Method	With outlier				Without outlier			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	1.6928	0.0289	0.0414	0.0617	2.3611	0.0313	0.0301	0.1004
LA	5.4476	0.0583	0.0738	0.2104	2.2681	0.0455	0.0586	0.0678
H-M	3.7154	0.0349	0.0420	0.1410	1.6397	0.0322	0.0193	0.0728
B-M	5.5462	0.0399	0.0850	0.2127	1.5236	0.0351	0.0346	0.0663
MM	3.1435	0.0448	0.0298	0.1219	3.4782	0.0474	0.0347	0.1406
LMS	11.0390	0.1367	0.1588	0.4296	4.8146	0.0758	0.6554	0.1458
LTS	6.7928	0.0929	0.0383	0.2621	3.7504	0.0963	0.0636	0.1337
S	3.7395	0.0593	0.0610	0.1431	5.9438	0.0823	0.0615	0.2376
PM	2.4738	0.0283	0.0315	0.1030	2.1640	0.0293	0.0313	0.0905

Table 4.43: Error variance and coefficient of determination for the Water salinity data set with and without the 7.1% outlying points.

Method	With outlier		Without outlier	
	$\hat{\sigma}^2$	R^2	$\hat{\sigma}^2$	R^2
LS	1.045464	0.885868	0.470029	0.948524
LA	1.389568	0.84747	0.517901	0.943256
H-M	1.103879	0.878561	0.488311	0.946511
B-M	1.241231	0.861153	0.512829	0.943784
MM	3.388016	0.639657	0.532461	0.941039
LMS	6.780952	0.234713	4.543204	0.528788
LTS	10.03148	0.07292	2.176898	0.763477
S	3.57838	0.620607	0.759613	0.914573
PM	2.906391	0.689683	0.474028	0.948098

Table 4.44: Parameter estimates of the stackloss data set with and without the 23.8% outlying points.

Method	With outlier				Without outlier			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	-34.0890	0.6147	1.2894	-0.1467	-26.9565	0.8416	0.4046	-0.1796
LA	-38.0274	0.8144	0.6371	-0.0833	-34.4153	0.7907	0.4192	-0.0633
H-M	-33.8608	0.7382	0.8725	-0.1372	-33.8055	0.7881	0.4430	-0.0748
B-M	-34.2875	0.8241	0.5967	-0.1262	-33.8228	0.7898	0.4351	-0.0740
MM	-33.7745	0.8051	0.6571	-0.1336	-154.7810	0.4596	0.8859	1.4305
LMS	-39.1613	0.8101	0.5249	-0.0417	-7.6271	0.5898	0.8510	-0.3504
LTS	-38.9728	0.7961	0.5721	-0.0446	-31.7617	1.2182	-0.4655	-0.1589
S	-39.0810	0.8190	0.5633	-0.0576	-155.8220	0.4632	0.8744	1.4432
PM	-30.0905	0.7354	0.5264	-0.1008	-29.0123	0.7711	0.4653	-0.1230

Table 4.45: Absolute bias of estimates for the stackloss data set with and without the 23.8% outlying points.

Method	With outlier				Without outlier			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	10.4443	0.2269	0.8848	0.0904	0.0000	0.0000	0.0000	0.0000
LA	14.1534	0.1035	0.2837	0.1574	8.3081	0.0865	0.1148	0.1244
H-M	9.9803	0.1049	0.4825	0.0938	7.0462	0.0791	0.0877	0.1056
B-M	10.0977	0.0976	0.2971	0.0953	7.1058	0.0781	0.0932	0.1077
MM	10.2200	0.0984	0.2853	0.0970	128.5129	0.4015	0.5326	1.6105
LMS	14.5469	0.0907	0.3115	0.1489	26.1384	0.2626	0.5789	0.2809
LTS	14.1662	0.1011	0.3373	0.1459	10.8076	0.4123	1.1271	0.1061
S	14.1344	0.1033	0.3094	0.1446	129.7385	0.4179	0.5971	1.6292
PM	4.0130	0.1061	0.1218	0.0788	2.3840	0.0704	0.0650	0.0566

Table 4.46: Standard error (SE) of estimates of the stackloss data set with and without the 23.8% outlying points.

Method	With outlier				Without outlier			
	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$
LS	4.0204	0.0895	0.1917	0.0576	11.1095	0.1124	0.2981	0.1290
LA	3.8788	0.0388	0.1423	0.0524	8.9963	0.0546	0.2545	0.0565
H-M	3.4481	0.0578	0.1209	0.0515	8.6193	0.0612	0.2603	0.0471
B-M	2.9532	0.0148	0.0298	0.0456	8.4939	0.0609	0.2584	0.0447
MM	3.7563	0.0608	0.1761	0.0529	557.6245	1.4528	2.1015	6.8899
LMS	2.6186	0.0494	0.0893	0.0219	40.4065	0.8650	2.1271	0.4482
LTS	1.8339	0.0900	0.1501	0.0359	22.7939	0.4457	1.2846	0.2498
S	1.3367	0.0289	0.0583	0.0221	557.4060	1.4542	2.1113	6.8871
PM	7.2577	0.0987	0.2452	0.0904	9.4034	0.0916	0.2477	0.1076

Table 4.47: Error variance and coefficient of determination for the stackloss data set with and without the 23.8% outlying points.

Method	With outlier		Without outlier	
	$\hat{\sigma}^2$	R^2	$\hat{\sigma}^2$	R^2
LS	9.1232	0.9114	1.1482	0.9880
LA	12.3116	0.8794	1.5399	0.9843
H-M	10.4492	0.8988	1.4225	0.9857
B-M	12.8572	0.8745	1.4417	0.9855
MM	12.2157	0.8808	11.4455	0.8908
LMS	13.7099	0.8648	18.7645	0.8254
LTS	13.2595	0.8688	7.8402	0.9305
S	13.1065	0.8706	11.5818	0.8806
PM	15.5594	0.8539	1.2001	0.9876

4.4.2 Discussion of results for the real life data

Robust regression methods are often assessed using efficiency and break down point (percentage of outliers the method can tolerate). We have used the efficiency and breakdown point proxy by the standard error of estimates (SE) and parameter estimate respectively. However, to specifically study the breakdown point of each method on the real life data sets used, the results of the method with the outlying point as well as without the outlying point are presented. Also, to further assess the adequacy of each method, error variance and coefficient of determination are presented. All the results presented were based on jackknife estimates. The jackknife procedure was used to assess the stability of the results. The jackknife estimator for each performance metrics were presented in chapter three

Starting with the lowest level (5%) of outlier in the x, y directions, the estimate of LS with the outlying points is 0.000178, -0.16501, 0.091041, 0.092038 and 0.008246 and removing the outlying points leads to an estimate of -0.01702, 0.043629, 0.151029, 0.141665 and 0.014816. The breakdown point of the robust estimator is assessed by comparing the estimate and absolute

biased results of the estimator in the presence and absence of the outlying points. Based on the comparison, it was observed that out of all the robust estimators the PM is the best in terms of maintaining its estimate in the presence and absence of the outlying point. This implies the PM estimate can withstand 5% of outlying points from x, y directions. Similar, estimators that are competing with the PM estimator are HM, BM and MM. Another, influence of outlier on linear regression model is efficiency, outliers reduce the efficiency of the estimator. As expected, the LS estimates with outlying observation are relatively far from when the outlying observations are removed. Again, the best estimate in terms of relative closeness to the true situation is PM followed by HM, BM and MM. Moving to model adequacy property of model with outlier, the result using the residual variance and coefficient of multiple determination shows that the presences of outlier do not affect the residual variance and coefficient of multiple determination. Hence LS still take the leading position in these two criteria. The performance of LTS, LMS and S are not better than the LS estimator. Moving to (5.7%) of outlier in the x,y directions, similar result as observed with 5% of outlier in the x, y directions was also observed. For (7.1%) of outlier in the x direction, the estimate of LS with the outlying points is 6.370766, 0.789466, -0.01884 and -0.159 and removing the outlying points leads to an estimate of 17.98217, 0.761, -0.19054, 0.141665 and -0.63523. The breakdown point of the robust estimator is assessed by comparing the estimate and absolute biased results of the estimator in the presence and absence of the outlying points. Based on the comparison, it was observed that out of all the robust estimators the PM is the best in terms of maintaining its estimate in the presence and absence of the outlying point. This implies the PM estimate can withstand 7.1% of outlying points from x direction. Similar, estimators that are competing with the PM estimator is MM. Moving to efficiency, the LS estimates with outlying observation are relatively far from when the outlying

observations are removed. The best estimate among the robust estimator is BM. However, in terms of breakdown of efficiency the PM estimator is the best. Which implies the efficiency of PM is not sensitive to the outlying observations. For model adequacy property of model with outlier, the result using the residual variance and coefficient of multiple determination shows that the presences of outlier do not affect the residual variance and coefficient of multiple determination. Hence LS still take the leading position in these two criteria. Again, the performance of LTS, LMS and S are not better than the LS estimator.

For (7.1%) of outlier in the x direction, the estimate of LS with the outlying points is 6.370766, 0.789466, -0.01884 and -0.159 and removing the outlying points leads to an estimate of 17.98217, 0.761, -0.19054, 0.141665 and -0.63523. The breakdown point of the robust estimator is assessed by comparing the estimate and absolute biased results of the estimator in the presence and absence of the outlying points. Based on the comparison, it was observed that out of all the robust estimators the PM is the best in terms of maintain its estimate in the presence and absence of the outlying point. This implies the PM estimate can withstand 7.1% of outlying points from x direction. Similar, estimators that are competing with the PM estimator is MM. Moving to efficiency, the LS estimates with outlying observation are relatively far from when the outlying observations are removed. The best estimate among the robust estimator is BM. However, in terms of breakdown of efficiency the PM estimator is the best. Which implies the efficiency of PM is not sensitive to the outlying observations. Moving to model adequacy property of model with outlier, the result using the residual variance and coefficient of multiple determination shows that the presences of outlier do not affect the residual variance and coefficient of multiple determination. Hence LS still take the leading position these two criterion. As earlier observed, the performance of LTS, LMS and S are not better than the LS estimator.

Further increase in the percentage of outlier to 23.8% in the x, y directions increases the severity of outliers on the estimate of LS with outlier. Similar severity effect was observed with MM and likewise the known high breakdown S and LMS estimators. Under this condition, the PM estimator justifiably takes the lead in terms breakdown (absolute bias and estimate) and efficiency.

CHAPTER FIVE

DISCUSSION AND CONCLUSION

5.0 Introduction

This chapter consists of four sections; section 5.1 summarizes the whole results of the study. Section 5.2, research contribution and findings, brings out the main findings in this study. Section 5.3 gives the conclusion, while section 5.4 gives the recommendations.

5.1 Summary

The study is on obtaining a robust procedure in estimating parameters in linear regression model, in the presence of outliers. Chapter one is the introduction, throwing light on multiple regression, linear regression model and how it's been affected by outliers, and properties robust regression must exhibit. Chapter two reviews the literature, and it revealed that efforts have been made to make the classical non robust method to be robust, due to the fact that each method suffers from one limitation to the other. Hence the objective of this study is to develop an efficient procedure in estimating parameters in Linear Regression in the presence of outliers, with high breakdown point. Chapter three introduces the proposed method through Monte Carlo Simulation for various proportions and magnitude of outliers in all direction (X and Y).

Chapter four reports the comparison of the proposed method with the classical Least Square method and seven other existing methods (LA, H-M, B-M, MM, LMS, LTS, and S). The comparison was done based on breakdown points and efficiency (that is, properties of robust regression method).

The proposed method (PM) does not breakdown having the smallest MSE over all the sample sizes at 0.1 proportion of outlier in y-direction; followed by H-M, B-M and MM; while LS, LMS, LTS and S have larger MSE due to the outlier effect. Also, the PM achieved high efficiency at 0.1 proportion of outlier, followed by MM. At 0.2, 0.3 0.4 and 0.5 proportion of

outlier in y-direction, similar results were observed. For Leverage outlier, similar results observed for y- direction outlier were also observed, while none of the existing methods could withstand 0.1 proportion of outlier as PM, that yielded high efficiency at X and y direction of outlier condition.

Studying the performances of each method on real life data set, PM have high efficiency and can withstand 5% of outlying points in X and y direction, followed by H-M, B-M, MM. Similar results was observed for 5.7% and 7.1% of outlier in x and y direction. While increase in percentage of the outlier to 23.8% in X and y direction does not affect the breakdown point and efficiency of PM.

5.2 Research contribution and findings

Comparison of various robust estimators were carried out based on breakdown point and efficiency which help to identify the strength and weakness of the methods

- 1 Based on the weakness of the existing robust estimators reviewed, robust estimator based on mean shift and shrinkage approach is proposed.
- 2 The strength of the proposed robust estimator using published real life data and simulated data was examined.
- 3 The proposed estimator is capable of achieving high breakdown point and high efficiency.

5.3 Conclusion

Based on the various experimental scenarios considered, the following conclusions are drawn;

- 1 The strength and weakness of any robust estimator strongly depends on the sample size, percentage and direction of outlier
- 2 Among the existing robust estimators, MM is the best in terms of unidirectional outlier (either X or Y) using efficiency and breakdown point
- 3 Among the entire robust estimator including the proposed PM estimator, PM is the best with high breakdown point up to 50% as well as high efficiency.
- 4 Robust estimator cannot improve the adequacy property of linear regression model.

5.4 Recommendations

Based on the result of this research, the proposed PM estimator is highly recommended, given its capability to withstand low and high percentage of outlying point from x direction, y direction, and both directions. That is, maintaining high breakdown point and efficiency simultaneously. This property is often rare in robust estimator as it is often required to sacrifice one to achieve the other.

Suggestion for further studies

Further studies can be done on;

A robust estimation Procedure in a nonlinear regression in the presence of outlier

A robust estimation procedure in regression in the presence of missing data.

REFERENCES

- Aggarwal C.C. (2013). An Introduction to Outlier Analysis. In Outlier Analysis. Springer, New York, NY
- Alma, O. G. (2011). Comparison of robust regression methods in linear regression. *International Journal for Contemporary Mathematics and Science*, 6,409–421.
- 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.
- Arsham, H. (2014). Estimators and their qualities in Statistical thinking for decision making: revealing facts from figures
- Barnett, V. & Lewis, T. (1984). *Outliers in statistical data*. 3rd Edn., John Wiley: New York.
- Bhar, L. (2014). Robust regression. <http://www.iasri.res.in/ebook/EBADAT/3Diagnostics>.
- Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression diagnostics: identifying influential data and sources of collinearity*. New York: Wiley.
- Brownlee, K. A. (1965), *Statistical Theory and Methodology in Science and Engineering*, 2nd edn. John Wiley & Sons:New York.
- Cetin, M. & Toka, O. (2011). The comparing of s-estimator and m-estimators in linear regression. *Gazi University Journal of Science*, 24(4), 747–752.
- Chatterjee, S & Hadi, A.S. (2006). Regression analysis by example. 4th edn. Hoboken, NJ: Wiley-Interscience.
- Coakley, C. W. & Hettmansperger, T. P. (1993). A Bounded Influence, High Breakdown, Efficient Regression Estimator. *Journal of American Statistical Association*, 88, 872-880.
- Cook, R. D.& Weisberg, S. (1989). Regression diagnostics with dynamic graphics (with discussion). *Technometrics*, 31, 277-311.
- Croux, C., Rousseeuw, P. J.& H'ossjer O. (1994). Generalized S-estimators. *Journal of American Statistical Association*, 89, 1271-1281.
- De Long, J.B. & Summers, L.H. (1991). Equipment investment and economic growth. *Quarterly Journal of Economics*, 106, 445-501.
- Evans, S. (2001). Statistical aspects of the detection of fraud. In: Lock, S, Wells, F & Farthing, M, Edt. *Fraud and Misconduct: in Biomedical Research*, 3rd Edn BMJ Books.
- Freedman, W. L., Wilson, C. D.& Madore, B. F. (1991). New Cepheid Distances to Nearby Galaxies Based on BVRI CCD Photometry. *Astrophysical Journal*, 372,455-470.
- Farebrother, R. W. (1976). Further results on the mean square error of ridge regression. *Journal of the Royal Statistical Society, Series B (Methodological)*, 248–250.
- Fox, J. & Weisberg, S. (2010). An appendix to an r companion to applied regression 2nd edn. 1–17.
- Gervini, D. & Yohai, V. J. (2002). A Class of Robust and Fully Efficient Regression Estimators. *The Annals of Statistics*, 30, 583-616.
- George, E.P. B. (1953), Non-normality and tests on variance. *Biometrika*, 40(3/4), 318-335
- Glass, G.V. & Hopkins, K.D. (1984). *Statistical Methods in Education and Psychology*, 2ndEdn. Englewood Cliffs, NJ: Prentice Hall. In Schumacker, R. E., Monahan, M. P.& Mount, R. E. (2002). A comparison of OLS and robust regression using s-plus.*Multiple Linear Regression View Point*, 28(2), 10–13.
- Gschwandtner, M. & Filzmoser, P. (2012). Computing robust regression estimators:

- Developments since Dutter (1977). *Austrian Journal of Statistics*, 41, 45–58.
- Hampel, F. R. (1973). Robust estimation: A condensed partial survey. *Probability theory and related fields*, 27(2), 87–104
- Hampel, F. (2001). *Robust statistics: A brief introduction and overview*. 1st International Symposium on Robust Statistics and Fuzzy Techniques in Geodesy and GIS, 13–17. Carsio, A. & Kutterer, H. (Eds). Swiss Federal Institute of Technology, Zurich
- Handschin, E., Kohlas, J., Fiechter, A. & Schweppe, F. (1975), Bad Data Analysis for Power System State Estimation. *IEEE Transactions on Power Apparatus and Systems*, 2, 329–337.
- Hawkins, D. M. (1980). *Identification of outliers*. Monographs on Statistics and Applied Probability Chapman and Hall, New York: London.
- Hoerl, A. E. & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometric*, 12(1), 55–67
- Ho, K. & Naugher, J. (2000). Outliers Lie: An Illustrative Example of Identifying Outliers a Applying Robust Models. *Multiple Linear Regression Viewpoints*, 26(2), 2–6
- Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 73–101
- Huber, P. J. (1972). The 1972 Wald lectures: Robust statistics. *The Annals of Mathematical Statistics*, 43, 1041–1067.
- Huber, P. J. (1981). *Robust Statistics*. New York: John Wiley and Sons.
- Iglewicz, B. & Hoaglin, D. (1993). *How to detect and handle outliers*. The ASQC Basic References in Quality Control: Statistical Techniques. ASQC Quality Press, 16
- Jackel, L. A. (1972). Estimating regression coefficients by minimizing the dispersion of the residuals. *Annals of Mathematical Statistics*, 5, 1449–1458.
- Jacoby, B. (2014). Regression III: Advanced methods. Accessed on Dec. 19, 2015 from: <http://polisci.msu.edu/jacoby/icpsr/regress3/lectures/week3/11.Outliers.pdf>.
- Jann, B. (2012). *Robust regression in stata*. German stata user group meeting, Berlin, 1–34.
- Kashyap, R. L. & Maiyuran, S. (1993). Robust procedures in Regression and outlier set estimation using likelihood reasoning. *Electrical and Computer Engineering ECE Technical Reports, TR-EE 93-8*, Purdue University School of Electrical Engineering.
- Lee, Y., MacEachern, S. N., & Jung, Y. (2011). Regularization of case-specific parameters for robustness and efficiency. Submitted to Statistical Science.
- Mallows, C. L. (1975), *On Some Topics in Robustness*. Unpublished memorandum, Bell Tel. Laboratories, Murray Hill.
- Maronna, R. A., Martin, R. D. & Yohai, V. J. (2006). *Robust Statistics*. John Wiley.
- Montgomery, D. C., Peck, E. A., & Vining, G. G. (2003). *Introduction to Linear Regression Analysis*, 3rd edn. New York: John Wiley & Sons.
- Muthukrishnan, R. & Radha, M. (2014). Comparison of robust estimators in regression. <http://interstat.statjournals.net/YEAR/2010/articles/1005003.pdf>.
- Osborne, J. W. & Overbay, A. (2004) The power of outliers (and why researchers should always check for them). *Practical assessment, research and evaluation* 9(6).
- Rousseeuw, P. J. (1982). Least Median of Squares regression. Research Report No. 178, Centre for Statistics and Operations research, VUB Brussels.

- Rousseeuw, P.J. (1983). *Multivariate Estimation with High Breakdown Point*. Research Report No. 192, Center for Statistics and Operations research, VUB Brussels.
- Rousseeuw, P.J. & Croux, C. (1993). Alternatives to the median absolute deviation. *Journal of American Statistical Association*, 94, 388-402.
- Rousseeuw, P.J. & Leroy, A.M. (1987). *Robust Regression and Outlier Detection*, Wiley. New York
- Rousseeuw, P.J. & Yohai, V. J. (1984). Robust Regression by Means of S-estimators.
- Rousseeuw, P. J., Zaman, A. & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Economic Letters*, 71, 1–8.
- Rupert, D. & Carroll, R.J., (1980). Trimmed Least Squares Estimation in the Linear Model, *Journal of American Statistical Association*, 75, 828 – 838.
- Ruppert, D., Street, J. O. & 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.& Mount, R. E. (2002). A comparison of OLS and robust regression using s-plus.*Multiple Linear Regression View Point*, 28(2),10–13.
- She, Y. &Owen, A. B. (2011). Outlier Detection Using Nonconvex Penalized regression. *Journal of American Statistical Association*, 106, 626-639.
- Stigler, S. M. (2010). The changing history of robustness.(author abstract). *The American Statistician*, 64(4),277.
- Stromberg, A. J., Hawkins, D. M., & Hossjer, O. (2000). The least trimmed differences regression estimator and alternatives. *Journal of American Statistical Association*, 95, 853-864.
- Stuart, C. (2011). Robust regression. *Department of Mathematical Sciences*, Durham University, 169.
- Tiku, M. L. & Akkaya, A. D. (2004). *Robust estimation and hypothesis testing*. New Age International (P) Ltd.
- Verardi, V. & Croux, C. (2009). Robust regression in stata. *The Stata Journal*, 3,439–453.
- Wisnowski, J. W. (1999). *Multiple Outliers in linear regression: advances in detection methods, robust estimation, and variable selection*. A doctoral desertation of Arizona State University.
- Yohai, V. J. (1987), High breakdown-point and high efficiency robust estimates for regression. *The Annals of Statistics*, 15, 642-656.
- Yu, C., Yao,W. &Bai, X.(2014). *Robust linear regression: A review and comparison*. Accessed on Dec. 19, 2015 from: <https://arXiv.org>.
- Zaman, A., Rousseeuw, P.J. & Orhan, M. (2001). Econometric applications of high-breakdown robust regression techniques. *Econometrics Letters*, 71, 1-8.

APPENDICES

Appendix A1: Simulation code for CASE I

```
#####  
##### Simulation code for Case I #####  
#####
```

```
rm(list=ls())  
library(MASS)  
library(quantreg)  
robust=function(n,p,ittr)  
{  
  
MSE=matrix(NA,nrow=ittr,ncol=11)  
cat("Repetitions done:", "\n")  
for(i in 1:ittr)  
{  
  
set.seed(i)  
#n=10 ;p=0.1  
X=mvnrm(n,mu=rep(0,3),Sigma=diag(3),empirical=T)  
  
x1=X[,1] ; x2=X[,2] ; x3=X[,3]  
  
  
b0=1 ; b1=b2=b3=1  
e=rnorm(n)  
  
yc=b0+b1*x1+b2*x2+b3*x3+e  
y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))  
#y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))  
#y=b0+b1*x+rt(n,3)  
dat=data.frame(y,x1,x2,x3)  
datc=data.frame(yc,x1,x2,x3)  
  
modc=lm(yc~x1+x2+x3,datc)  
mod1=lm(y~x1+x2+x3,dat)  
mod2=rq(y~x1+x2+x3,data=dat)  
mod3=rlm(y~x1+x2+x3,dat)  
mod4=rlm(y~x1+x2+x3,dat,psi=psi.bisquare)  
mod5=rlm(y~x1+x2+x3,dat,method="MM")  
mod6=lmsreg(y~x1+x2+x3,dat)  
mod7=ltsreg(y~x1+x2+x3,dat)  
mod8=lqs(y~x1+x2+x3,dat,method="S")
```

```

#difffts<- as.vector(dffits(mod1))
#g <- which(abs(difffts) > 0.2828)

g=((n*(1-p))+1):n
Xout=matrix(NA,nrow=n,ncol=length(g))

for(uu in 1:length(g))
{
xout=c()
for(vv in 1:n)
{
xout=c(xout,ifelse(vv==g[uu],1,0))
}
Xout[,uu]=xout
}
mod=lm(y~x1+x2+x3+Xout)

ridg=lm.ridge(y~x1+x2+x3+Xout,lambda=seq(0,1,length=itr))
modr=lm.ridge(y~x1+x2+x3+Xout,lambda=as.numeric(names(which.min(ridg$GCV))))

msec=mean((coef(modc)-c(b0,b1,b2,b3))^2)
mse1=mean((coef(mod1)-c(b0,b1,b2,b3))^2)
mse2=mean((coef(mod2)-c(b0,b1,b2,b3))^2)
mse3=mean((coef(mod3)-c(b0,b1,b2,b3))^2)
mse4=mean((coef(mod4)-c(b0,b1,b2,b3))^2)
mse5=mean((coef(mod5)-c(b0,b1,b2,b3))^2)
mse6=mean((coef(mod6)-c(b0,b1,b2,b3))^2)
mse7=mean((coef(mod7)-c(b0,b1,b2,b3))^2)
mse8=mean((coef(mod8)-c(b0,b1,b2,b3))^2)
mse9=mean((coef(mod)[1:4]-c(b0,b1,b2,b3))^2)
mse10=mean((coef(modr)[1:4]-c(b0,b1,b2,b3))^2)

MSE[i,]=cbind(msec,mse1,mse2,mse3,mse4,mse5,mse6,mse7,mse8,mse9,mse10)
if (i %in% seq(0, itr, 1)) cat(i, "... ")
}
colnames(MSE)=c("C-LS","LS","LA","H-M","B-M","MM","LMS","LTS","S","PM1","PM2")
AMSE=round(colMeans(MSE),4)
RE=round((AMSE[1]/AMSE)*100,2)
return(list(AMSE=AMSE,RE=RE))
}
res1=robust(10,0.1,1000)
res2=robust(30,0.1,1000)
res3=robust(100,0.1,1000)
res4=robust(200,0.1,1000)
res5=robust(1000,0.1,1000)

```

```
Avg1=round(colMeans(rbind(res1$AMSE,res2$AMSE,res3$AMSE,res4$AMSE,res5$AMSE)),
4)
Avg2=round(colMeans(rbind(res1$RE,res2$RE,res3$RE,res4$RE,res5$RE)),2)

ress=rbind(rbind(res1$AMSE,res1$RE),rbind(res2$AMSE,res2$RE),
rbind(res3$AMSE,res3$RE),rbind(res4$AMSE,res4$RE),
rbind(res5$AMSE,res5$RE),rbind(Avg1,Avg2))

#write.csv(ress,"TAOHEEDAT_SEC_0.1.csv")
```

Appendix A2: Simulation code for CASE II

```
#####  
#####Simulation code for Case  
II#####
```

```
rm(list=ls())  
library(MASS)  
library(quantreg)  
robust=function(n,p,ittr)  
{  
  
MSE=matrix(NA,nrow=ittr,ncol=11)  
cat("Repetitions done:", "\n")  
for(i in 1:ittr)  
{  
  
set.seed(i)  
#n=10 ;p=0.1  
X=mvrnorm(n,mu=rep(0,3),Sigma=diag(3),empirical=T)  
  
x1=X[,1] ; x2=X[,2] ; x3=X[,3]  
  
b0=1 ; b1=b2=b3=1  
e=rnorm(n)  
  
yc=b0+b1*x1+b2*x2+b3*x3+e  
y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))  
#y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))  
#y=b0+b1*x+rt(n,3)  
dat=data.frame(y,x1,x2,x3)  
datc=data.frame(yc,x1,x2,x3)  
  
modc=lm(yc~x1+x2+x3,datc)  
mod1=lm(y~x1+x2+x3,dat)  
mod2=rq(y~x1+x2+x3,data=dat)  
mod3=rlm(y~x1+x2+x3,dat)  
mod4=rlm(y~x1+x2+x3,dat,psi=psi.bisquare)  
mod5=rlm(y~x1+x2+x3,dat,method="MM")  
mod6=lmsreg(y~x1+x2+x3,dat)  
mod7=ltsreg(y~x1+x2+x3,dat)  
mod8=lqs(y~x1+x2+x3,dat,method="S")  
  
#dffts<- as.vector(dffits(mod1))  
#g <- which(abs(dffts) > 0.2828)
```

```

g=((n*(1-p))+1):n
Xout=matrix(NA,nrow=n,ncol=length(g))

for(uu in 1:length(g))
{
xout=c()
for(vv in 1:n)
{
xout=c(xout,ifelse(vv==g[uu],1,0))
}
Xout[,uu]=xout
}
mod=lm(y~x1+x2+x3+Xout)

ridg=lm.ridge(y~x1+x2+x3+Xout,lambda=seq(0,1,length=itr))
modr=lm.ridge(y~x1+x2+x3+Xout,lambda=as.numeric(names(which.min(ridg$GCV))))

msec=mean((coef(modc)-c(b0,b1,b2,b3))^2)
mse1=mean((coef(mod1)-c(b0,b1,b2,b3))^2)
mse2=mean((coef(mod2)-c(b0,b1,b2,b3))^2)
mse3=mean((coef(mod3)-c(b0,b1,b2,b3))^2)
mse4=mean((coef(mod4)-c(b0,b1,b2,b3))^2)
mse5=mean((coef(mod5)-c(b0,b1,b2,b3))^2)
mse6=mean((coef(mod6)-c(b0,b1,b2,b3))^2)
mse7=mean((coef(mod7)-c(b0,b1,b2,b3))^2)
mse8=mean((coef(mod8)-c(b0,b1,b2,b3))^2)
mse9=mean((coef(mod)[1:4]-c(b0,b1,b2,b3))^2)
mse10=mean((coef(modr)[1:4]-c(b0,b1,b2,b3))^2)

MSE[i,]=cbind(msec,mse1,mse2,mse3,mse4,mse5,mse6,mse7,mse8,mse9,mse10)
if (i %in% seq(0, itr, 1)) cat(i, "... ")
}
colnames(MSE)=c("C-LS","LS","LA","H-M","B-M","MM","LMS","LTS","S","PM1","PM2")
AMSE=round(colMeans(MSE),4)
RE=round((AMSE[1]/AMSE)*100,2)
return(list(AMSE=AMSE,RE=RE))
}
res1=robust(10,0.1,1000)
res2=robust(30,0.1,1000)
res3=robust(100,0.1,1000)
res4=robust(200,0.1,1000)
res5=robust(1000,0.1,1000)

```

```

Avg1=round(colMeans(rbind(res1$AMSE,res2$AMSE,res3$AMSE,res4$AMSE,res5$AMSE)),
4)
Avg2=round(colMeans(rbind(res1$RE,res2$RE,res3$RE,res4$RE,res5$RE)),2)

ress=rbind(rbind(res1$AMSE,res1$RE),rbind(res2$AMSE,res2$RE),
rbind(res3$AMSE,res3$RE),rbind(res4$AMSE,res4$RE),
rbind(res5$AMSE,res5$RE),rbind(Avg1,Avg2))

#write.csv(ress,"TAOHEEDAT_SEC_0.1.csv")

```

Appendix A3: Simulation code for CASE III

```
#####  
##### Simulation code for Case III #####  
#####  
=====
```

```
rm(list=ls())  
library(MASS)  
library(quantreg)  
robust=function(n,p,ittr)  
{  
  
  MSE=matrix(NA,nrow=ittr,ncol=11)  
  cat("Repetitions done:", "\n")  
  for(i in 1:ittr)  
  {  
  
    set.seed(i)  
    #n=10 ;p=0.1  
  
    Xc=mvnrm(n,mu=rep(0,3),Sigma=diag(3),empirical=T)  
    X=rbind(Xc[(1:round(n*(1-p)))],cbind(rnorm(round(n*p),0,10),rnorm(round(n*p),0,10),  
    rnorm(round(n*p),0,10)))  
  
    x1=X[,1] ; x2=X[,2] ; x3=X[,3]  
    xc1=Xc[,1] ; xc2=Xc[,2] ; xc3=Xc[,3]  
  
    b0=1 ; b1=b2=b3=1  
    e=rnorm(n)  
    yc=b0+b1*xc1+b2*xc2+b3*xc3+e  
    yn=b0+b1*x1+b2*x2+b3*x3+e  
    y=c(yn[1:round(n*(1-p))],rnorm(round(n*p),0,10))  
    #y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))  
    #y=b0+b1*x+rt(n,3)  
    dat=data.frame(y,x1,x2,x3)  
    datc=data.frame(yc,xc1,xc2,xc3)  
  
    modc=lm(yc~x1+x2+x3,datc)  
    mod1=lm(y~x1+x2+x3,dat)  
    mod2=rq(y~x1+x2+x3,data=dat)  
    mod3=rlm(y~x1+x2+x3,dat)  
    mod4=rlm(y~x1+x2+x3,dat,psi=psi.bisquare)  
    mod5=rlm(y~x1+x2+x3,dat,method="MM")  
    mod6=lmsreg(y~x1+x2+x3,dat)  
    mod7=ltsreg(y~x1+x2+x3,dat)  
    mod8=lqs(y~x1+x2+x3,dat,method="S")
```

```

#difffts<- as.vector(dffits(mod1))
#g <- which(abs(difffts) > 0.2828)

g=((n*(1-p))+1):n
Xout=matrix(NA,nrow=n,ncol=length(g))

for(uu in 1:length(g))
{
xout=c()
for(vv in 1:n)
{
xout=c(xout,ifelse(vv==g[uu],1,0))
}
Xout[,uu]=xout
}
mod=lm(y~x1+x2+x3+Xout)

ridg=lm.ridge(y~x1+x2+x3+Xout,lambda=seq(0,1,length=itr))
modr=lm.ridge(y~x1+x2+x3+Xout,lambda=as.numeric(names(which.min(ridg$GCV))))

msec=mean((coef(modc)-c(b0,b1,b2,b3))^2)
mse1=mean((coef(mod1)-c(b0,b1,b2,b3))^2)
mse2=mean((coef(mod2)-c(b0,b1,b2,b3))^2)
mse3=mean((coef(mod3)-c(b0,b1,b2,b3))^2)
mse4=mean((coef(mod4)-c(b0,b1,b2,b3))^2)
mse5=mean((coef(mod5)-c(b0,b1,b2,b3))^2)
mse6=mean((coef(mod6)-c(b0,b1,b2,b3))^2)
mse7=mean((coef(mod7)-c(b0,b1,b2,b3))^2)
mse8=mean((coef(mod8)-c(b0,b1,b2,b3))^2)
mse9=mean((coef(mod)[1:4]-c(b0,b1,b2,b3))^2)
mse10=mean((coef(modr)[1:4]-c(b0,b1,b2,b3))^2)

MSE[i,]=cbind(msec,mse1,mse2,mse3,mse4,mse5,mse6,mse7,mse8,mse9,mse10)
if (i %in% seq(0, itr, 1)) cat(i, "... ")
}
colnames(MSE)=c("C-LS","LS","LA","H-M","B-M","MM","LMS","LTS","S","PM1","PM2")
AMSE=round(colMeans(MSE),4)
RE=round((AMSE[1]/AMSE)*100,2)
return(list(AMSE=AMSE,RE=RE))
}
res1=robust(10,0.1,1000)
res2=robust(30,0.1,1000)
res3=robust(100,0.1,1000)
res4=robust(200,0.1,1000)
res5=robust(1000,0.1,1000)

```

```

Avg1=round(colMeans(rbind(res1$AMSE,res2$AMSE,res3$AMSE,res4$AMSE,res5$AMSE)),
4)
Avg2=round(colMeans(rbind(res1$RE,res2$RE,res3$RE,res4$RE,res5$RE)),2)

ress=rbind(rbind(res1$AMSE,res1$RE),rbind(res2$AMSE,res2$RE),
rbind(res3$AMSE,res3$RE),rbind(res4$AMSE,res4$RE),
rbind(res5$AMSE,res5$RE),rbind(Avg1,Avg2))

#write.csv(ress,"TAOHEEDAT_SEC_0.1.csv")

```

Appendix B: Analysis for Real life data

```
#####  
##### Modified Ridge Regression code: lm.ridge1#####  
#####
```

```
rm(list=ls())  
lm.ridge1 = function (y,x,lambdab)  
{  
  dat=data.frame(y,x)  
  Y <-y  
  X <- model.matrix(y~.,dat)  
  n <- nrow(X)  
  p <- ncol(X)  
  
  #Xscale<- drop(rep(1/n, n) %*% X^2)^0.5  
  #X <- X/rep(Xscale, rep(n, p))  
  Xs<- svd(X)  
  rhs<- t(Xs$u) %*% Y  
  d <- Xs$d  
  lscoef<- Xs$v %*% (rhs/d)  
  lsfit<- X %*% lscoef  
  resid<- Y - lsfit  
  s2 <- sum(resid^2)/(n - p)  
  HKB <- (p - 2) * s2/sum(lscoef^2)  
  LW <- (p - 2) * s2 * n/sum(lsfit^2)  
  k <- length(lambdab)  
  dx <- length(d)  
  div <- d^2 + rep(lambdab, rep(dx, k))  
  a <- drop(d * rhs)/div  
  dim(a) <- c(dx, k)  
  coef<- Xs$v %*% a  
  #dimnames(coef) <- list(names(X), format(lambdab))  
  GCV <- colSums((Y - X %*% coef)^2)/(n - colSums(matrix(d^2/div,  
    dx)))^2  
  #res <- list(coef = drop(coef),  
    #lambdab = lambdab, GCV = GCV, kHKB = HKB,  
    #kLW = LW)  
  res1=rbind(GCV,drop(coef))  
  optimum=res1[,as.numeric(which.min(res1[1,]))][1]  
  optimum  
}
```

#dat=stackloss

```
#y= dat[,4]
#x=dat[,-4]
#lambda=seq(0,1,length=1000)
#lm.ridge1(y,x,lambda)
```

```
#####
##### Real life data Analysis #####
#####
```

```
library(MASS)
library(quantreg)
comprobust=function(y,x)
{
  dat=data.frame(y,x)
  p=ncol(dat)
```

```
  betas1=betas2=betas3=betas4=betas5=betas6=betas7=betas8=betas9=matrix(NA,ncol=p,nrow=nrow(dat))
  betas11=betas22=betas33=betas44=betas55=betas66=betas77=betas88=betas99=matrix(NA,ncol=p,nrow=nrow(dat))
```

```
  SR1s=SR2s=SR3s=SR4s=SR5s=SR6s=SR7s=SR8s=SR9s=matrix(NA,ncol=2,nrow=nrow(dat))
  SR11s=SR22s=SR33s=SR44s=SR55s=SR66s=SR77s=SR88s=SR99s=matrix(NA,ncol=2,nrow=nrow(dat))
```

```
  for(i in 1:nrow(dat))
  {
```

```
    dat=dat[-i,]
```

```
    n=nrow(dat)
```

```
    mod1=lm(y~.,dat)
    lminf<- influence.measures(mod1)$is.inf
    lminfs=apply(lminf,1,sum)
    g=which(lminfs>0)
```

```
  Xout=matrix(NA,nrow=n,ncol=length(g))
```

```
  for(uu in 1:length(g))
  {
```

```

xout=c()
for(vv in 1:n)
{
xout=c(xout,ifelse(vv==g[uu],1,0))
}
Xout[,uu]=xout
}

```

with analysis

```

modLA=rq(y~.,data=dat)
modHM=rlm(y~.,dat)
modBM=rlm(y~.,dat,psi=psi.bisquare)
modMM=rlm(y~.,dat,method="MM")
modLMS=lmsreg(y~.,dat)
modLTS=ltsreg(y~.,dat)
modS=lqs(y~.,dat,method="S")

```

```

pmdata=data.frame(dat,Xout)
X=as.matrix(cbind(dat$x,Xout))

```

```

#cv1=cv.glmnet(X,y,alpha=0)

```

```

#modr=lm.ridge1(dat$y, X, lambda=seq(0,1,length=1000))

```

```

ridg=lm.ridge(y~.,data=pmdata,lambda=seq(0,1,length=1000))
modr=lm.ridge(y~.,data=pmdata,lambda=as.numeric(names(which.min(ridg$GCV))))

```

```

#modr=glmnet(X,y,alpha=0,lambda=cv1$lambda.min,standardize=F)

```

without analysis

```

withoutdata=dat[-g,]

```

```

modLA2=rq(y~.,data=withoutdata)
modHM2=rlm(y~.,withoutdata)
modBM2=rlm(y~.,withoutdata,psi=psi.bisquare)
modMM2=rlm(y~.,withoutdata,method="MM")
modLMS2=lmsreg(y~.,withoutdata)
modLTS2=ltsreg(y~.,withoutdata)
modS2=lqs(y~.,withoutdata,method="S")

```

```

xn=as.matrix(cbind(withoutdata[, -1]))
#cv11=cv.glmnet(xn,withoutdata[,1],alpha=0)

```

```

#modr2=glmnet(xn,withoutdata[,1],alpha=0,lambda=cv1$lambda.min,standardize=F)

#modr2=lm.ridge1(withoutdata[,1], xn, lambda=seq(0,1,length=1000))

ridg2=lm.ridge(y~,data=withoutdata,lambda=seq(0,1,length=1000))
modr2=lm.ridge(y~,data=withoutdata,lambda=as.numeric(names(which.min(ridg2$GCV))))

modr3=lm(y~,withoutdata)

##### result

b1=coef(mod1) ; b2=coef(modLA) ; b3=coef(modHM) ; b4=coef(modBM)

b5=coef(modMM) ; b6=coef(modLMS) ; b7=coef(modLTS) ; b8=coef(modS)

b9=as.matrix(coef(modr))[1:p]

b11=coef(modr3) ; b22=coef(modLA2) ; b33=coef(modHM2) ; b44=coef(modBM2)

b55=coef(modMM2) ; b66=coef(modLMS2) ; b77=coef(modLTS2) ; b88=coef(modS2)

b99=as.matrix(coef(modr2))[1:p]

sigma1=sum(resid(mod1)^2)/(n-p) ; sigma2=sum(resid(modLA)^2)/(n-p)

sigma3=sum(resid(modHM)^2)/(n-p) ; sigma4=sum(resid(modBM)^2)/(n-p)

sigma5=sum(resid(modMM)^2)/(n-p) ; sigma6=sum(resid(modLMS)^2)/(n-p)

sigma7=sum(resid(modLTS)^2)/(n-p) ; sigma8=sum(resid(modS)^2)/(n-p)

xm=model.matrix(mod1)
bx=matrix(b9,nrow=p,ncol=1)
sigma9=sum((dat[,1]-xm%*%bx)^2)/(n-p)

sigma11=sum(resid(modr3)^2)/(n-p) ; sigma22=sum(resid(modLA2)^2)/(n-p)

sigma33=sum(resid(modHM2)^2)/(n-p) ; sigma44=sum(resid(modBM2)^2)/(n-p)

sigma55=sum(resid(modMM2)^2)/(n-p) ; sigma66=sum(resid(modLMS2)^2)/(n-p)

sigma77=sum(resid(modLTS2)^2)/(n-p) ; sigma88=sum(resid(modS2)^2)/(n-p)

```

```

xm2=model.matrix(modr3)
bx2=matrix(b99,nrow=p,ncol=1)
sigma99=sum((withoutdata[,1]-xm2%*%bx2)^2)/(n-p)

```

```

sstw=sum((dat[,1]-mean(dat[,1]))^2)
sstwo=sum((withoutdata[,1]-mean(withoutdata[,1]))^2)

```

```

R1=1-(((n-p)*sigma1)/sstw) ; R2=1-(((n-p)*sigma2)/sstw) ; R3=1-(((n-p)*sigma3)/sstw)

```

```

R4=1-(((n-p)*sigma4)/sstw) ; R5=1-(((n-p)*sigma5)/sstw) ;R6=1-(((n-p)*sigma6)/sstw)

```

```

R7=1-(((n-p)*sigma7)/sstw) ; R8=1-(((n-p)*sigma8)/sstw) ; R9=1-(((n-p)*sigma9)/sstw)

```

```

R11=1-(((n-p)*sigma11)/sstw) ; R22=1-(((n-p)*sigma22)/sstw) ; R33=1-(((n-p)*sigma33)/sstw)

```

```

R44=1-(((n-p)*sigma44)/sstw) ; R55=1-(((n-p)*sigma55)/sstw) ;R66=1-(((n-p)*sigma66)/sstw)

```

```

R77=1-(((n-p)*sigma77)/sstw) ; R88=1-(((n-p)*sigma88)/sstw) ; R99=1-(((n-p)*sigma99)/sstw)

```

```

##### jacknifes

```

```

betas1[i,]=b1 ; betas2[i,]=b2 ; betas3[i,]=b3 ; betas4[i,]=b4 ; betas5[i,]=b5

```

```

betas6[i,]=b6 ; betas7[i,]=b7 ; betas8[i,]=b8 ; betas9[i,]=b9 ; betas11[i,]=b11

```

```

betas22[i,]=b22 ; betas33[i,]=b33 ; betas44[i,]=b44 ; betas55[i,]=b55 ; betas66[i,]=b66

```

```

betas77[i,]=b77 ; betas88[i,]=b88 ; betas99[i,]=b99

```

```

SR1s[i,]=c(sigma1,R1) ; SR2s[i,]=c(sigma2,R2) ; SR3s[i,]=c(sigma3,R3)

```

```

SR4s[i,]=c(sigma4,R4) ; SR5s[i,]=c(sigma5,R5) ; SR6s[i,]=c(sigma6,R6)

```

```

SR7s[i,]=c(sigma7,R7) ; SR8s[i,]=c(sigma8,R8) ; SR9s[i,]=c(sigma9,R9)

```

```

SR11s[i,]=c(sigma11,R11) ; SR22s[i,]=c(sigma22,R22) ; SR33s[i,]=c(sigma33,R33)

```

```

SR44s[i,]=c(sigma44,R44) ; SR55s[i,]=c(sigma55,R55) ; SR66s[i,]=c(sigma66,R66)

```

```

SR77s[i,]=c(sigma77,R77) ; SR88s[i,]=c(sigma88,R88) ; SR99s[i,]=c(sigma99,R99)

```

```

cat(i,"...")
}

betasw=rbind(colMeans(betas1),colMeans(betas2),colMeans(betas3),colMeans(betas4),
             colMeans(betas5),colMeans(betas6),colMeans(betas7),colMeans(betas8),colMeans(betas9))

betaswo=rbind(colMeans(betas11),colMeans(betas22),colMeans(betas33),colMeans(betas44),
              colMeans(betas55),colMeans(betas66),colMeans(betas77),colMeans(betas88),colMeans(betas99
              ))

se.betasw=rbind(apply(betas1,2,sd),apply(betas2,2,sd),apply(betas3,2,sd),apply(betas4,2,sd),
                apply(betas5,2,sd),apply(betas6,2,sd),apply(betas7,2,sd),apply(betas8,2,sd),apply(betas9,2,sd))

se.betaswo=rbind(apply(betas11,2,sd),apply(betas22,2,sd),apply(betas33,2,sd),apply(betas44,2,s
                d),

                apply(betas55,2,sd),apply(betas66,2,sd),apply(betas77,2,sd),apply(betas88,2,sd),apply(betas99,2,
                sd))

SRw=rbind(colMeans(SR1s),colMeans(SR2s),colMeans(SR3s),colMeans(SR4s),
          colMeans(SR5s),colMeans(SR6s),colMeans(SR7s),colMeans(SR8s),colMeans(SR9s))

SRwo=rbind(colMeans(SR11s),colMeans(SR22s),colMeans(SR33s),colMeans(SR44s),
           colMeans(SR55s),colMeans(SR66s),colMeans(SR77s),colMeans(SR88s),colMeans(SR99s))

betas=as.matrix((cbind(betasw,betaswo)))
se.betas=as.matrix((cbind(se.betasw,se.betaswo)))
SR=as.matrix((cbind(SRw,SRwo)))

rownames(betas)=rownames(se.betas)=rownames(SR)=c("LS","LA","H-M","B-
M","MM","LMS","LTS","S","PM")

propoutlier=100*(length(g)/n)

absbetasw=rbind(colMeans(abs(betas1-betas11)),colMeans(abs(betas2-
betas11)),colMeans(abs(betas3-betas11)),
colMeans(abs(betas4-betas11)),colMeans(abs(betas5-betas11)),
             colMeans(abs(betas6-betas11)),colMeans(abs(betas7-betas11)),colMeans(abs(betas8-
betas11)),colMeans(abs(betas9-betas11)))

absbetaswo=rbind(colMeans(abs(betas11-betas11)),colMeans(abs(betas22-
betas11)),colMeans(abs(betas33-betas11)),

```

```

colMeans(abs(betas44-betas11)),colMeans(abs(betas55-betas11)),
  colMeans(abs(betas66-betas11)),colMeans(abs(betas77-betas11)),colMeans(abs(betas88-
betas11)),colMeans(abs(betas99-betas11)))

absbias=as.matrix((cbind(absbetasw,absbetaswo)))

return(list(betas=betas,se.betas=se.betas,absbw=absbetasw,absbias=absbias,SR=SR,g,propoutlier=
propoutlier))
}

```

```

dat=stackloss
y= dat[,4]
x=dat[, -4]
lambda=seq(0,1,length=1000)
M1=comprobust(y,x)

```

```

url = "http://www.statsci.org/data/general/hills.txt"
races.table = read.table(url, header=TRUE, sep="\t")
head(races.table)

```

```

y=races.table[,4]
x=cbind(races.table[, -c(1,4)])
M2=comprobust(y,x)

```

```

y=c(7.6, 7.7, 4.3, 5.9, 5, 6.5, 8.3,8.2, 13.2, 12.6,10.4,10.8,13.1,12.3, 10.4,10.5,
  7.7,9.5,12,12.6,13.6,14.1,13.5, 11.5,12,13,14.1,15.1)

```

```

x1=c(8.2, 7.6, 4.6,4.3,5.9,5,6.5,8.3,10.1,13.2,12.6,10.4,10.8,13.1,13.3,10.4,10.5,
  7.7,10,12,12.1,13.6,15,13.5,11.5,12.0,13.0,14.1)

```

```

x2=c(4,5,0,1,2,3,4,5,0,1,2,3,4,5,0,1,2,3,0,1,4,5,0,1,2,3,4,5)

```

```

x3=c(23.005,23.873,26.417,24.868,29.895,24.200,23.215,21.862,22.274,23.830,25.144,
  22.430,21.785,22.380,23.927,33.443,24.859,22.686,21.789,22.041,21.033,21.005,
  25.865, 26.290, 22.932, 21.313,20.769, 21.393)
x=cbind(x1,x2,x3)
M3=comprobust(y,x)

```

```

dat=read.table("growth.r",header=T)
x=dat[, -c(1,2)]
y=dat[,2]
M4=comprobust(y,x)

```

```

write.csv(M1$betas,"Stackloss.beta.csv")
write.csv(M1$absbias,"Stackloss.absbias.csv")
write.csv(M1$se.betas,"Stackloss.se.beta.csv")
write.csv(M1$SR,"Stackloss.SR.csv")

write.csv(M2$betas,"Race.beta.csv")
write.csv(M2$absbias,"Race.absbias.csv")
write.csv(M2$se.betas,"Race.se.beta.csv")
write.csv(M2$SR,"Race.SR.csv")

write.csv(M3$betas,"WaterS.beta.csv")
write.csv(M3$absbias,"WaterS.absbias.csv")
write.csv(M3$se.betas,"WaterS.se.beta.csv")
write.csv(M3$SR,"Water.SR.csv")

write.csv(M4$absbias,"Growth.absbias.csv")

rm(list=ls())
library(MASS)
library(quantreg)
robust=function(n,p,case_,itr)
{

MSE=matrix(NA,nrow=itr,ncol=9)
cat("Repetitions done:", "\n")
for(i in 1:itr)
{
set.seed(i)
# n=10 ;p=0.1
x=rnorm(n)
b0=1 ; b1=1
yc=b0+b1*x+rnorm(n)
if(case_=="I"||case_=="II")
{
y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))
}
else {y=b0+b1*x+rt(n,3)}
dat=data.frame(y,x)
datc=data.frame(yc,x)

modc=lm(yc~x,datc)
mod1=lm(y~x,dat)
mod2=rq(y~x,data=dat)
mod3=rlm(y~x,dat)
mod4=rlm(y~x,dat,psi=psi.bisquare)

```

```

mod5=rlm(y~x,dat,method="MM")
mod6=lmsreg(y~x,dat)
mod7=ltsreg(y~x,dat)
mod8=lqs(y~x,dat,method="S")

msec=mean((coef(modc)-c(b0,b1))^2)
mse1=mean((coef(mod1)-c(b0,b1))^2)
mse2=mean((coef(mod2)-c(b0,b1))^2)
mse3=mean((coef(mod3)-c(b0,b1))^2)
mse4=mean((coef(mod4)-c(b0,b1))^2)
mse5=mean((coef(mod5)-c(b0,b1))^2)
mse6=mean((coef(mod6)-c(b0,b1))^2)
mse7=mean((coef(mod7)-c(b0,b1))^2)
mse8=mean((coef(mod8)-c(b0,b1))^2)

MSE[i,]=cbind(msec,mse1,mse2,mse3,mse4,mse5,mse6,mse7,mse8)
if (i %in% seq(0, itr, 1)) cat(i, "... ")
}
colnames(MSE)=c("C-LS","LS","LA","H-M","B-M","MM","LMS","LTS","S")
AMSE=round(colMeans(MSE),4)
RE=round((AMSE[1]/AMSE)*100,2)
return(list(MSE=MSE,AMSE=AMSE,RE=RE))
}
res1=robust(10,0.5,"II",1000)
res2=robust(30,0.5,"II",1000)
res3=robust(100,0.5,"II",1000)
res4=robust(200,0.5,"II",1000)
res5=robust(1000,0.5,"II",1000)

Avg1=round(colMeans(rbind(res1$AMSE,res2$AMSE,res3$AMSE,res4$AMSE,res5$AMSE)),
4)
Avg2=round(colMeans(rbind(res1$RE,res2$RE,res3$RE,res4$RE,res5$RE)),2)

ress=rbind(rbind(res1$AMSE,res1$RE),rbind(res2$AMSE,res2$RE),
rbind(res3$AMSE,res3$RE),rbind(res4$AMSE,res4$RE),
rbind(res5$AMSE,res5$RE),rbind(Avg1,Avg2))

write.csv(ress,"TAOHEEDAT_SS1_CII_5.csv")

```

```

rm(list=ls())
library(MASS)
library(quantreg)

```

```

robust=function(n,p,itr)
{

MSE=matrix(NA,nrow=itr,ncol=11)
cat("Repetitions done:", "\n")
for(i in 1:itr)
{

set.seed(i)
#n=10 ;p=0.1
X=mvrnorm(n,mu=rep(0,3),Sigma=diag(3),empirical=T)

x1=X[,1] ; x2=X[,2] ; x3=X[,3]


b0=1 ; b1=b2=b3=1
e=rnorm(n)

yc=b0+b1*x1+b2*x2+b3*x3+e
y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))
#y=c(yc[1:round(n*(1-p))],rnorm(round(n*p),0,10))
#y=b0+b1*x+rt(n,3)
dat=data.frame(y,x1,x2,x3)
datc=data.frame(yc,x1,x2,x3)

modc=lm(yc~x1+x2+x3,datc)
mod1=lm(y~x1+x2+x3,dat)
mod2=rq(y~x1+x2+x3,data=dat)
mod3=rlm(y~x1+x2+x3,dat)
mod4=rlm(y~x1+x2+x3,dat,psi=psi.bisquare)
mod5=rlm(y~x1+x2+x3,dat,method="MM")
mod6=lmsreg(y~x1+x2+x3,dat)
mod7=ltsreg(y~x1+x2+x3,dat)
mod8=lqs(y~x1+x2+x3,dat,method="S")

#diffs<- as.vector(dffits(mod1))
#g <- which(abs(diffs) > 0.2828)

g=((n*(1-p))+1):n
Xout=matrix(NA,nrow=n,ncol=length(g))

for(uu in 1:length(g))
{
xout=c()
for(vv in 1:n)
{

```

```

xout=c(xout,ifelse(vv==g[uu],1,0))
}
Xout[,uu]=xout
}
mod=lm(y~x1+x2+x3+Xout)

ridg=lm.ridge(y~x1+x2+x3+Xout,lambda=seq(0,1,length=itr))
modr=lm.ridge(y~x1+x2+x3+Xout,lambda=as.numeric(names(which.min(ridg$GCV))))

msec=mean((coef(modc)-c(b0,b1,b2,b3))^2)
mse1=mean((coef(mod1)-c(b0,b1,b2,b3))^2)
mse2=mean((coef(mod2)-c(b0,b1,b2,b3))^2)
mse3=mean((coef(mod3)-c(b0,b1,b2,b3))^2)
mse4=mean((coef(mod4)-c(b0,b1,b2,b3))^2)
mse5=mean((coef(mod5)-c(b0,b1,b2,b3))^2)
mse6=mean((coef(mod6)-c(b0,b1,b2,b3))^2)
mse7=mean((coef(mod7)-c(b0,b1,b2,b3))^2)
mse8=mean((coef(mod8)-c(b0,b1,b2,b3))^2)
mse9=mean((coef(mod)[1:4]-c(b0,b1,b2,b3))^2)
mse10=mean((coef(modr)[1:4]-c(b0,b1,b2,b3))^2)

MSE[i,]=cbind(msec,mse1,mse2,mse3,mse4,mse5,mse6,mse7,mse8,mse9,mse10)
if (i %in% seq(0, itr, 1)) cat(i, "... ")
}
colnames(MSE)=c("C-LS","LS","LA","H-M","B-M","MM","LMS","LTS","S","PM1","PM2")
AMSE=round(colMeans(MSE),4)
RE=round((AMSE[1]/AMSE)*100,2)
return(list(AMSE=AMSE,RE=RE))
}
res1=robust(10,0.1,1000)
res2=robust(30,0.1,1000)
res3=robust(100,0.1,1000)
res4=robust(200,0.1,1000)
res5=robust(1000,0.1,1000)

Avg1=round(colMeans(rbind(res1$AMSE,res2$AMSE,res3$AMSE,res4$AMSE,res5$AMSE)),
4)
Avg2=round(colMeans(rbind(res1$RE,res2$RE,res3$RE,res4$RE,res5$RE)),2)

ress=rbind(rbind(res1$AMSE,res1$RE),rbind(res2$AMSE,res2$RE),
rbind(res3$AMSE,res3$RE),rbind(res4$AMSE,res4$RE),
rbind(res5$AMSE,res5$RE),rbind(Avg1,Avg2))

write.csv(ress,"TAOHEEDAT_SEC_0.1.csv")

```

Appendix C: Outlier Detection function

outlierKD function

```
outlierKD<- function(dt, var) {
var_name<- eval(substitute(var),eval(dt))
  na1 <- sum(is.na(var_name))
  m1 <- mean(var_name, na.rm = T)
par(mfrow=c(2, 2), oma=c(0,0,3,0))
boxplot(var_name, main="With outliers")
hist(var_name, main="With outliers", xlab=NA, ylab=NA)
  outlier <- boxplot.stats(var_name)$out
mo<- mean(outlier)
var_name<- ifelse(var_name %in% outlier, NA, var_name)
boxplot(var_name, main="Without outliers")
hist(var_name, main="Without outliers", xlab=NA, ylab=NA)
title("Outlier Check", outer=TRUE)
  na2 <- sum(is.na(var_name))
cat("Outliers identified:", na2 - na1, "n")
cat("Propotion (%) of outliers:", round((na2 - na1) / sum(!is.na(var_name))*100, 1), "n")
cat("Mean of the outliers:", round(mo, 2), "n")
  m2 <- mean(var_name, na.rm = T)
cat("Mean without removing outliers:", round(m1, 2), "n")
cat("Mean if we remove outliers:", round(m2, 2), "n")
  response <- readline(prompt="Do you want to remove outliers and to replace with NA?
[yes/no]: ")
if(response == "y" | response == "yes"){
dt[as.character(substitute(var))] <- invisible(var_name)
  assign(as.character(as.list(match.call())$dt), dt, envir = .GlobalEnv)
cat("Outliers successfully removed", "n")
  return(invisible(dt))
} else{
cat("Nothing changed", "n")
  return(invisible(var_name))
}
```