A COMPARISON OF SOME ROBUST REGRESSION TECHNIQUES

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

EZGİ AVCI

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
INDUSTRIAL ENGINEERING

SEPTEMBER 2009

Approval of the thesis:

A COMPARISON OF SOME ROBUST REGRESSION TECHNIQUES

submitted by **EZGİ AVCI** in partial fulfillment of the requirements for the degree of **Master of Science in Industrial Engineering Department, Middle East Technical University** by,

Prof. Dr. Canan Özgen
Dean, Graduate School of **Natural and Applied Sciences** _____________

Prof. Dr. Nur Evin Özdemirel
Head of Department, **Industrial Engineering** _____________

Prof. Dr. Gülser Köksal
Supervisor, **Industrial Engineering Dept, METU** _____________

Examining Committee Members:
Assoc. Prof. Dr. İnci Batmaz
Statistics Department, METU _____________

Prof. Dr. Gülser Köksal
Industrial Engineering Dept., METU _____________

Assist. Prof. Dr. Seçil Savaşaneril
Industrial Engineering Dept., METU

Dr. Özlem İlk
Statistics Department, METU _____________

Dr. Özlem Türker Bayrak
Industrial Engineering Dept, Çankaya University _____________

Date: 11.09.2009

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last name: EZGİ AVCI

Signature:

ABSTRACT

A COMPARISON OF SOME ROBUST REGRESSION TECHNIQUES

Avcı, Ezgi

M.S., Department of Industrial Engineering

Supervisor: Prof. Dr. Gülser Köksal

September 2009, 141 Pages

Robust regression is a commonly required approach in industrial studies like data mining, quality control and improvement, and finance areas. Among the robust regression methods; Least Median Squares, Least Trimmed Squares, M-regression, MM-method, Least Absolute Deviations, Locally Weighted Scatter Plot Smoothing and Multivariate Adaptive Regression Splines are compared under contaminated normal distributions with each other and Ordinary Least Squares with respect to the multiple outlier detection performance measures. In this comparison; a simulation study is performed by changing some of the parameters such as outlier density, outlier locations in the x-axis, sample size and number of independent variables. In the comparison of the methods, multiple outlier detection is carried out with respect to the performance measures detection capability, false alarm rate and improved mean square error and ratio of improved mean square error. As a result of this simulation study, the three most competitive methods are compared on an industrial data set with respect to the coefficient of multiple determination and mean square error.

Keywords: Robust Regression, Multiple Outlier Detection, Multiple Regression

ÖZ

BAZI SAĞLAM REGRESYON YÖNTEMLERİNİN
BİR KARŞILAŞTIRMASI

Avcı, Ezgi

Yüksek Lisans, Endüstri Mühendisliği Bölümü

Tez Yöneticisi: Prof. Dr. Gülser Köksal

Eylül 2009, 141 Sayfa

Sağlam(robust) regresyon endüstriyel çalışmalarda; örneğin veri madenciliği, kalite kontrol ve iyileştirme ve finans alanlarında sıklıkla gerek duyulan bir yaklaşımdır. Sağlam regresyon yöntemlerinden En Küçük Medyan Kareler, En Küçük Budanmış Kareler, M-regresyon, MM, En Küçük Mutlak Değerler, LOWESS ve MARS metotları yaklaşık normal dağılıma göre aykırı değerleri işleme başarımı bakımından En Küçük Kareler metodu ile karşılaştırılmıştır. Bu karşılaştırmada; aykırı değerlerin oranı, x uzayındaki yeri, örneklem büyüklüğü ve bağımsız değişken sayısı gibi parametreler değiştirilerek bir benzetim çalışması gerçekleştirilmiştir. Metotların karşılaştırılmasında aykırı değerleri belirlemede tip 1 ve 2 hata oranı, iyileştirilmiş hata kareler ortalaması ve iyileştirilmiş hata kareler ortalaması oranı gibi başarım ölçüleri kullanılmıştır. Bu benzetim çalışması sonucunda en iyi başarım gösteren üç metot bir endüstriyel veri seti üzerinde uygulanmış olup belirleme katsayısı ve ortalama karesel hata başarım ölçülerine göre bu metotların başarımı tartışılmıştır.

Anahtar Kelimeler: Sağlam Regresyon, Çoklu Aykırı Değer Bulma, Çoklu Regresyon

To my family

ACKNOWLEDGEMENTS

I would like to present my special thanks to my supervisor Prof. Dr. Gülser Köksal for her invaluable support, patience and encouragement throughout this study.

I also would like to thank Assoc. Prof. Dr. İnci Batmaz for her detailed and constructive support during the study. I also wish to thank Elçin Kartal not only for her support but also for her friendship.

I want to express my deepest appreciation to my dear friend Didem Pekkurnaz for her full support, patience and perfect friendship.

I would like to thank my grandfather Mehmet Sarıbaş, my grandmother Nurten Sarıbaş and my uncles, Murat and Ömer. I always feel their love and belief with me. Every moment that we share is invaluable for me.

Most special thanks go to my parents Pervin and Ömer Avcı for their everlasting love, moral support and valuable guidance. I feel so lucky that I have such great parents. I would like to thank them for always being there whenever I need. I would never succeed without their love and support.

TABLE OF CONTENTS

LIST OF TABLES

TABLES

LIST OF FIGURES

FIGURES

CHAPTER 1

INTRODUCTION

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. There are many application areas of regression, including engineering, physical sciences, economics, management, life and biological sciences, and the social sciences. Regression can be said to be the most widely used statistical technique (Montgomery and Peck, 1991).

In industrial applications, regression analysis is commonly used in quality control and improvement, and data mining. Ordinary Least Squares (OLS) method is usually preferred in these studies because it is a well established method and most of the computer packages are capable of making regression analysis with OLS. However, this method has some assumptions. Without satisfying these assumptions, the results will not be valid and should not be used. The most important assumption of OLS is normality. Briefly, normality assumption indicates that the regression errors should be normally distributed. However, the errors are frequently not distributed normally because of some *outliers* in the data.

The concept of an outlier has attracted the attention of the researchers since the earliest attempts to interpret the data. It is more important to decide whether we should delete the observations, which are “unrepresentative” or “mavericks” of the data, than the development of the statistical method. These observations are often seen as contamination to the data. They reduce and affect the information that we may get from the source or generating mechanism. Exploring the interpretation and categorization of outliers is not straightforward (Barnett and Lewis, 1994)

Therefore regression analysis becomes troublesome and requires using techniques like data transformation which may not always solve the problem. Moreover the analysis may take time and typically needs expertise.

To overcome these problems new statistical techniques, which are called robust (resistant), have been introduced that are not so easily affected by outliers (Rousseeuw and Leroy, 1987). Robust regression methods aim to minimize the impact of the outliers on regression estimators, but still invoke parametric assumptions after smoothing the influence of outliers on the regression line (Lane, 2002).

Robust regression methods are gaining more and more importance, but there are few comparative studies about the performances of these approaches. People carrying out applications in various industries appreciate suggestions about which robust methods they should prefer to use.

In this thesis, our aim is to extend the scope of the few comparative studies and guide the people making applications in industry. For this purpose, we extend the simulation comparison of M-regression, MM-method, Least Trimmed Squares (LTS) and Least Median Squares (LMS) performed by Wisnowski (1999). We use the same simulation approach with the addition of Least Absolute Deviations (LAD), Locally Weighted Scatter Plot Smoothing (LOWESS) and Multivariate Adaptive Regression Splines (MARS). The design parameters are selected as the number of independent variables k=2 for n=40 observations or k=6 with n=60 observations. The outlier density is either 10% or 20%, that is; 10% or 20% of the outliers can be outlying.The magnitude of the outliers is between three and five standard deviation units. The outliers are either generated randomly or in groups. These groups are called *clouds*. The number of multiple point clouds is either one or two. The results are compared by performance measures which are detection capability (PP), false alarm rate (PO), improved mean square error (IMSE) and ratio of IMSE. Furthermore from the simulation the most promising robust methods, which are M-regression, LAD and LTS are specified and applied on an industrial data set. A cross-validation approach is used to compare the methods' performances on the industrial data. Criteria on which the robust regression methods compared are coefficient of multiple determination (R^2) and mean square error (MSE). Detection capability and false alarm rate cannot be used since in real life data outliers are not known for sure as in the simulation study.

This thesis is organized into five chapters. In the second chapter, some background information about robustness, outliers, OLS assumptions, robust regression techniques and previous comparative studies are given. In the third chapter, a Monte Carlo simulation study is performed on three different Y-space outlier scenarios, which are adapted from Wisnowski (1999). From the simulation study, the three most promising robust regression methods are determined and their performances are compared on an industrial data set in chapter four. Conclusions and suggestions for future studies are mentioned in chapter five.

CHAPTER 2

LITERATURE REVIEW

In this chapter, first the definition and types of robustness are presented. Then, OLS assumptions are discussed. Since the most basic assumption of OLS is usually violated by outliers; outlier types, which are interior Y-space and interior X-space, are investigated. Robust regression methods, which are especially developed for detecting outliers in the data, are explained next. Their algorithms and basic properties are analyzed. At the end of the chapter, previous comparative studies are briefly discussed.

2.1 Robustness

Box (1953) used the term *robustness* for the first time (Hogg, 1979). Portnoy and He (2000) state that there are more than 3000 entries with "robust" and "robustness" as key words in the Current Index to Statistics. These findings can be classified as "density", "rank", "bootstrap", "censored" or "smoothing" but most of them sees robustness as alternatives to OLS method and normality theory. As indicated by Portnoy and He (2000), the importance of robust estimators was recognized in a study called Princeton which was conducted by Pearson, Student, Box and Tukey. Huber's (1964) classic minimax approach and Hampel's (1968, 1974) introduction of influence functions was the first milestones of the modern theory of robust statistics. After these studies the importance of robustness was understood. Huber (1964) might be considered as the first research that defines robustness as "approximate validity of a parametric model". Hampel (1974) studied the properties of statistical functions and introduced three important robustness concepts: qualitative robustness (continuity), influence function (derivative), and breakdown point (Portnoy and He, 2000).

There are many perspectives of robustness. Box and Tiao (1962) mentioned about two types of robustness: criterion robustness and inference robustness. Criterion robustness selects a criterion for statistical optimality and then investigates its variation as the parent distribution deviates from the form assumed. Inference robustness is about the changes in quantities leading to inference (significance levels, coverage probability, etc.). Then some studies have been performed about efficiency robustness, qualitative robustness, bias robustness, Bayesian robustness, and so on (Portnoy and He, 2000).

The word *robust* has many meanings but briefly *robustness* can be defined as "*signifying insensitivity to small deviations from the assumptions"*. Correspondingly *distributional robustness* can be defined as "*the shape of the true underlying distribution deviates slightly from the assumed model (usually the normal distribution")* (Huber, 1981). In our study, we are interested only in the distributional robustness to contaminated normal distributions. Skewed and asymmetric distributions are out of our scope.

A statistical procedure based on the OLS assumptions may be substantially affected by the deviations from normality. In a survey, Tukey (1960) showed that contamination by just two observations from a N(0, 9) distribution among 1000 N(0, 1) observations is enough to make the mean absolute deviation (MAD) estimator more efficient than the sample standard deviation, which is asymptotically optimal for the Gaussian scale parameter (Portnoy and He, 2000).

Huber (1981) indicates that some people may ask whether robust procedures are needed at all, because there is a common two-step approach to deal with outliers:

1. Clean the data by applying some rule for outlier detection.

2. Use classical estimation and testing procedures on the remainder of the data.

Unfortunately, they will not do the same job as the robust estimators. From Huber (1981) the reasons can be summarized as:

1. Separating the two steps accurately is a hard work. Moreover in the multiple regression case, outliers are difficult to detect.
2. Although the sample come from a normal distribution with some gross outliers, the cleaned data may not be normal (there may be both kinds of statistical errors, false rejection and false acceptance). If the sample does not come from a normal distribution, the situation is even worse. As a result, the classical normal theory is not applicable to cleaned samples, so applying the two-step procedure may be more difficult than applying the straight robust procedure.
3. Robust regression methods are a smooth transition between full acceptance and full rejection of an observation; therefore the best rejection procedures are not competitive against the best robust procedures.

Huber (1981) defines a *parametric model* as a "*hopefully good approximation to the underlying situation, but we cannot and do not assume that it is exactly correct".* As a result there should be three properties of any statistical procedure:

1. At the assumed model, it should have a reasonably good efficiency.
2. Small deviations, occurrence of gross errors in a small fraction of the observations, from the model assumptions should affect the performance only slightly.
3. Larger deviations from the assumed model should not affect the parameters substantially.

Therefore robust regression is classified in the family of parametric regression methods. Related with these ideas, *robust estimates* can be specified as *consistent estimates of the unknown parameters at the idealized model*. Since they have robustness property, they will not drift too far away if the model is only approximately true (Huber, 1981).

To illustrate the difference between parametric robustness and nonparametric method, the following example can be given. Suppose that a random sample is drawn from a mixture model with 90% from N (0, 1) and 10% from N (t, 1) for some large value of t. If we use a central model of N (θ, 1), then a robust estimate of location would aim at the center of the majority; that is, $\theta=0$. However, the estimate will be biased from the contamination from N (t, 1). A robust estimate tries to control the bias regardless of the size of t. On the other hand, a nonparametric estimate of location would aim at the center of the mixture distribution, which is not 0 in this case. So, it is important to know what we are searching in the analysis. If we believe in a parametric model that approximates reality and wish to estimate the parameters related with this model, then robust estimates are our best choice. But if we are not sure about the underlying distribution, we consider the data as a sample from an unknown population and are interested in a population quantity, then a nonparametric estimate is more appropriate (see Portney and Welsh 1992) as cited in Portnoy and He (2000).

2.2. Ordinary Least Squares and Assumptions

Regression analysis is a statistical technique for investigating and modeling the relationship between variables. There are lots of application areas of regression including engineering, the physical sciences, economics, management, life and biological sciences, and the social sciences. Regression can be said to be the most widely used statistical technique (Montgomery and Peck, 1991)

Among the various regression methods, OLS is the most commonly used one. It was discovered independently by Carl Friedrich Gauss in Germany around 1975 and by Adrien Marie Legendre in France around 1805. Astronomic and geodetic data were used in the early applications of the method. Its first published appearance was in 1805 in an appendix to a book by Legendre on determining the orbits of comets (Birkes and Dodge, 1993). As it is stated by Rousseeuw and Leroy (1987), among the many possible regression techniques, the OLS method has generally been used because of tradition and ease of its computation.

A regression model that has more than one independent variable is called a *multiple linear regression model*. The model can be formulated as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \varepsilon$$

If we take k=1, the model is called *simple linear regression model*. The parameters β_j are called the regression coefficients. This model illustrates a k-dimensional hyperplane. The parameter β_j indicates the expected change in the response y per unit change in x_j when all the remaining independent variables are held constant. Therefore, β_j can also be called partial regression coefficient (Montgomery and Peck, 1992).

In multiple linear regression, the real functional relationship between the response and the independent variables is unknown. The regression model is an approximation to the real relationship. But over certain ranges of the independent variables, it is an adequate approximation (Montgomery and Peck, 1991).

To estimate the coefficients, the straight line that “best” fits the data points in the plot is found. To judge how well the estimated regression line fits the data, we can analyze the size of the residuals. The smaller the residuals, the more accurate the fit is (Birkes and Dodge, 1993).

The residuals of a regression model are defined as the n differences

$e_i = y_i - \hat{y}_i$, i= 1, 2,…, n ;

where Y_i is an observation and $\hat{Y}_i$ is the corresponding fitted value calculated from the fitted regression equation.

From the definition of the residuals, it can be said that e_i are the amount of response that the regression equation has not been able to explain. Therefore, we can think of the e_i as the *observed errors if the model is correct* (Draper and Smith, 1981).

As stated by Montgomery and Peck (1991), a OLS regression model is valid if and only if the assumptions below are true;

1. The relationship between y and x is linear, or at least it is well approximated by a straight line.
2. The error term ε has zero mean.
3. The error term ε has constant variance σ^2.
4. The errors are uncorrelated.
5. The errors are normally distributed.

Assumptions 4 and 5 mean that the errors are independent random variables. Assumption 5 is the basic requirement for hypothesis testing and interval estimation. If the model is fitted without validating the assumptions, it can be unstable which means that by using a different sample, a totally different model can be fitted with opposite conclusions (Montgomery and Peck, 1991).

The normality assumption violation is the basis for robust regression. Although small departures from normality do not affect the model greatly, gross vialotions of normality assumption is more dangerous since the t- or F-statistics, confidence and prediction intervals depend on the normality assumption. Moreover if the errors come from a heavy-tailed distribution, the OLS regression line may be sensitive to a small subset of the data since these types of outliers "pull" the least squares regression line too much in their direction (Montgomery and Peck, 1992) or outliers may be a result of misplaced decimal points, recording and transmission errors, exceptional phenomena such as earthquakes and strikes, or members of a different population contaminating into the sample. These undesirable situations may be results in non-optimal solutions. Real data usually have outliers but they cannot be identified by the users since nowadays much data is analyzed by computers, without careful inspection and screening. Outliers may totally spoil an ordinary LS analysis. Often such influential points remain hidden because they do not always show up in the usual OLS residual plots (Rousseeuw and Leroy, 1987).

2.3. Outliers in Regression

An outlying observation, or *outlier*, is "*the observation that appears to deviate markedly from other members of the sample in which it occurs"*, as Grubbs (1969) remarks (Barnett and Lewis, 1994).

As discussed by Barnett and Lewis (1994), there are three different sources of variability:

1. Inherent variability: In this type, observations vary over the population; such variation is a natural property of the population. It is uncontrollable and reflects the distributional properties of a correct basic model.
2. Measurement error: Inadequacies in the measurement instrument, the rounding of obtained values, or recording mistakes.
3. Execution error: If we carelessly choose a biased sample or take observations that do not represent the population we try to get information, this type of variability occurs.

From these three sources of variability, we are nearly interested in outliers that may be a perfectly reasonable reflection of the natural inherent variation and reflects an inadequate basic model. A more appropriate model should be assumed in this case. But again it must be remarked that d*istributional robustness (against inherent alternatives)* are out of our scope, *robustness against contamination* is mentioned in our study.

2.3.1. Interior Y-space Outliers

Regression outliers can seriously violate the standard OLS analysis. Outliers can be in both X and Y directions. Outliers in y-direction have received attention in literature because one usually considers the y_i as observations and the x_i as fixed numbers (which is only true when the design has been given in advance) and because such "vertical" outliers often possess large positive or large negative residuals. Even in multiple regression with a large number of independent variables (p), where it is so difficult to visualize the data, such outliers can often be discovered from the list of residuals or from the residual plots (Rousseeuw and Leroy, 1987).

2.3.2. Interior X-space Outliers

Explanatory variables can also have outliers. In many applications, one receives a list of variables, and then has to choose a response variable and some explanatory variables. So, there is no reason why gross errors would only occur in the response variable. Moreover, the probability of having an outlier in explanatory variables is higher than the probability of having an outlier in the response because usually this probability is greater than one; therefore the probability of having incorrect results is higher. An X-space outlier affects LS regression line greatly because it pulls the LS line towards itself. Therefore it is called a *leverage point*. In general, an observation (x_k , y_k) is called a leverage point whenever x_k lies far away from the bulk of the observed x_i in the sample. Note that this does not take y_k into account, so the point (x_k , y_k) does not necessarily have to be a regression outlier. When (x_k , y_k) is close to the regression line specified by the most of the data, then it is a *good leverage point*. In this case, it may perfectly lie on the regression line and it is even useful because it will narrow certain confidence regions (Rousseeuw and Leroy, 1987). The x and y-axis direction outliers are shown in Figure 2.1. The figure is adapted from Rousseuw and Leroy (1987).

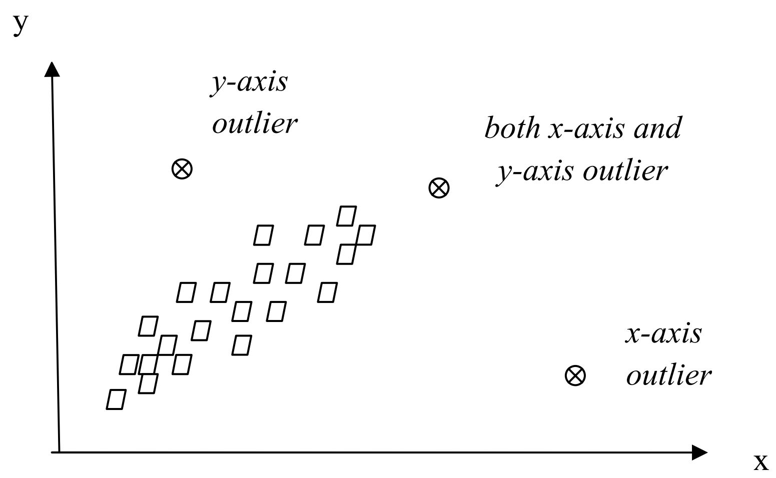

Figure 2.1: x-axis, y-axis and both x-axis and y-axis outliers

2.3.3. Detection of Multiple Outliers

There are two common ways to deal with outliers. The first one is *regression diagnostics* which are computed from the data to identify influential observations. Then these observations are corrected or deleted, followed by an LS analysis on the remaining cases. At the case of a single outlier, the effect of deleting one point at a time is easy to calculate. However, if there are multiple outliers, it is not straightforward to analyze their affects on the regression line and extensive computations are required. The second approach to deal with outliers is to use robust methods whose results are still valid even if a certain amount of data is contaminated. It is thought that robust regression techniques hide the outliers, but the opposite is true because the outliers are far away from the robust fit so they can be detected by their large residuals from it (Rousseeuw and Leroy, 1987).

In fact, diagnostics and robust regression have the same goal, but they proceed in the opposite order. In the diagnostics, first the outliers are identified and then a regression line is fitted to the data without outliers. But in robust regression, first a robust regression line is fitted to the majority of the data and then the outliers are discovered by looking at the large residuals from the robust regression. Sometimes, both methods give exactly the same results, and then the difference is mostly subjective. There are almost as many robust methods as there are diagnostics, and to differentiate between them it is important to compare the robust methods' performances with respect to the OLS method's performance (Rousseeuw and Leroy, 1987).

2.4. Robust Regression Methods

Robust regression methods are developed for situations in which the distribution is close to normal. Many ways have been discussed in literature to make an estimator robust. Changing the minimization criteria of the errors which means using different weight functions; and trimming specified proportions of the data are the most common ways. M-regression and LAD can be cited for the first group; LTS can be cited for the second group.

Still none of them has been shown to be truly superior in all outlier data configurations (Anderson, 2001). The aim of this study is to compare the performances of seven of the robust regression methods which are LAD, M, MM, LTS, LMS, MARS and LOWESS. In fact, in literature MARS and

LOWESS methods are classified in the nonparametric regression methods. However; the aim of this study is to measure the methods' performances against the outliers, not against the model complexity. For this reason, these two methods are taken in the group "robust".

These seven regression methods have been studied for two reasons. The first reason is that they are the most mentioned methods in the alternative regression methods literature. The second one is that these methods are all available in a statistical package called S-PLUS and evaluation procedures are easy to understand for the people carrying out applications in industry.

There are some fundamental concepts of the robust regression methods which are breakdown point, asymptotic efficiency and bounded influence.

Breakdown Point: Let Z be any sample of n data points and T be a regression estimator. If we apply T to such a sample Z, we will get a vector of regression coefficients T (Z). The breakdown point of the estimator at the sample Z is defined as the smallest fraction of contamination that can cause the estimator T to take on values arbitrarily far from T (Z). It is important that this definition includes no probability distributions. For example, OLS has a breakdown point of 1/n which goes to zero for increasing n. This again shows the OLS method's high sensitivity to outliers (Rousseeuw and Leroy, 1987).

Asymptotic Efficiency: Asymptotic can be defined as concerning the limiting behavior of a procedure as the sample size goes to infinity. Asymptotic properties can also be called *large-sample* properties and not directly related to real life data which occurs in finite sample sizes. However, asymptotic properties are often more readily obtainable.

The asymptotic efficiency of an x% means that for an infinitely large sample, the reciprocal of the ratio of the variance of the estimator to the smallest variance, which is the variance of the OLS estimate when the error distribution is normal.

Bounded Influence: An estimator is called bounded influence if it can bound the influence of X-space outliers by means of some weight function w. That is, bounded influence estimators are robust to X-space outliers which have large effect on the OLS regression line because they can pull the line towards themselves.

In the following, the robust regression methods studied in this thesis are introduced briefly.

2.4.1. Least Absolute Deviations

The method of LAD is developed by Joseph Boscovich in 1757, 50 years before the method of OLS, to accommodate inconsistent measurements for the purpose of estimating the shape of the earth. 30 years later Pierre Simon Laplace favored to use the method; it saw occasional use until the OLS method overshadowed it. OLS method is computationally simpler and also Gauss and Laplace developed a supporting theory for it. Computers are now able to do complex calculations and lots of theoretical foundations have been laid for a variety of alternative methods, including the method of LAD (Birkes and Dodge, 1993).

Regression line estimation algorithm is discussed in Birkes and Dodge (1993) very clearly. Assume that the proposed simple linear regression model is $Y = \propto + \beta X + e$. Scatter plot of the data can be checked whether the model is appropriate for representing the relations between X and Y. In the method of OLS, the estimates of the regression coefficients $\hat{\alpha}$ *and* $\hat{\beta}$ are found by minimizing the sum of the squares of the residuals, $\sum \hat{e}_i^2$. Whereas in the LAD method, the estimates are found by minimizing the sum of the absolute values of the residuals, $\sum|\hat{e}_i|$. That is, the LAD estimates $\hat{\alpha}$ *and* $\hat{\beta}$ are the values of a and b that minimize

$$\sum|y_i - (a + bx_i)|.$$

The difference is called the deviation of the point (x_i, y_i) from the line $\hat{Y} = a + bX$. LAD regression is sometimes called L_1-regression because L_1 norm of the vector of deviations is used.

The concept of LAD estimation is simpler than the OLS estimate concept but the LAD method's calculation of the estimates is more complicated. Instead of formulas, there are algorithms for calculating the LAD estimates. Birkes and Dodge (1993) explain this algorithm for simple and multiple regression cases. The aim of the algorithm is to find the line which has the least sum of absolute deviations and best fits the data. The algorithm starts with finding the best line among all the lines passing through for any given point (x_0, y_0). Let's say the best line passing through the initial point is (x_1, y_1), now we should find the best line passing through this point. This line also passes through another data point (x_2, y_2). The algorithm continues like this until the most recent line obtained will be the same as the previous line. This is the best line, called the LAD regression line, among all the lines without regard to what points they pass through.

The procedure is described as follows:

For each point (x_i, y_i) the slope of the line, which is formulated as $(y_i - y_0)/(x_i - x_0)$, passing through the two points (x_0, y_0) and (x_i, y_i) is calculated. If $x_i = x_0$ for some i, the slope is not defined but such points can be ignored. The points are reindexed as,

$$(y_1 - y_0)/(x_1 - x_0) \leq (y_2 - y_0)/(x_2 - x_0) \leq \cdots \leq (y_n - y_0)/(x_n - x_0).$$

Let

$$T = \sum |x_i - x_0|.$$

The index k is found by satisfying the conditions,

$$|x_1 - x_0| + \ldots + |x_{k-1} - x_0| < \frac{1}{2} T$$

$|x_1 - x_0| + \ldots + |x_{k-1} - x_0| + |x_k - x_0| > \frac{1}{2} T$.

The best line passing through (x_0, y_0) is the line,

$\hat{y} = \propto^* + \beta^* x$, where

$$\beta_1^* = \frac{y_k - y_0}{x_k - x_0}$$

$$\beta_0^* = y_0 - \beta_1^* x_0.$$

For the multiple regression case, the regression coefficient estimates are chosen to minimize

$\sum y_i - (|b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3}|)$.

As in the simple regression case, there are no formulas for the minimizing values, but an algorithm is used to obtain the values.

Let,

$$b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} \quad \text{and} \quad x_i = \begin{bmatrix} 1 \\ x_{i1} \\ x_{i2} \\ x_{i3} \end{bmatrix}$$

Then the sum of absolute deviations can be written as

$$\sum |y_i - b' x_i|$$

The vector b will be found by the algorithm which minimizes the above equation.

The algorithm for multiple LAD regression is iterative as in the simple case. The algorithm starts with a vector b, and finds a better vector, and goes on until the best vector $\hat{\beta}$ is obtained. At each step, having a vector of estimates b, we find a better vector b* by first finding a suitable "direction" vector d and then finding the value of t for which b* = b + td is best.

In order to find the best vector of estimates in direction d, we should find the value of t that minimizes

$$\sum |y_i - (b+td)'x_i|$$

If we write $z_i = y_i - b'x_i$ and $w_i = d'x_i$, then the procedure must find the value of t that minimizes

$$\sum |z_i - tw_i|$$

We put the ratios z_i/w_i in increasing order, and reindex the z and w values according to this order, and the index k is found by

$$|w_1| + |w_2| ... + |w_{k-1}| < \frac{1}{2}T$$

$$|w_1| + |w_2| ... |w_{k-1}| + |w_k| > \frac{1}{2}T$$

where $T = \sum |w_i|$. z_k/w_k is the minimum value of t

(Birkes and Dodges, 1993).

The strength of LAD method comes from its robustness to Y-space outliers. For this reason, LAD estimates can sometimes be used as starting values for iterative regression algorithms. This method is especially suitable when the error distribution is heavy-tailed or asymmetric or when the sample size is very large (Birkes and Dodge, 1993). The breakdown point of the LAD estimator is no better than 0% because it is highly sensitive to outliers in the X-space (Rousseeuw and Leroy, 1987). That is; LAD is neither a high breakdown nor a bounded influence estimator.

Even for a simple linear regression model, it is not easy to calculate LAD estimates by using a hand calculator and paper and pencil because it involves the construction of a series of tables. LAD regression estimates are available in the S-PLUS 2000 Robust Library with the function *l1fit*. This function uses the Barrodale-Roberts (1974) algorithm which is based on the simplex algorithm for solving linear programming models. The algorithm for the multiple cases forms an initial vector of estimates; on the other hand the Barrodale-Roberts algorithm includes a special start-up phase (Birkes and Dodges, 1993).

2.4.2. M-regression

The M-estimate is constructed by Huber (1964) to be optimal if the error distribution is contaminated normal. His criterion for the optimality is minimization of the maximum possible variance for infinitely large samples. The M in the M-regression was chosen because there is a relationship between M-estimation and maximum likelihood estimation. If the population of errors were assumed to have a particular distribution, some of the M-estimates would be maximum likelihood estimates. But the main goal of M-regression is to perform well for a wide range of distributions (Birkes and Dodge, 1993).

The algorithm for estimating the regression line for both the simple and multiple cases is discussed in Birkes and Dodge (1993) very explicitly. M-regression is a generalization of OLS and LAD methods by choosing the minimization criterion as $\sum \rho(\hat{e}_i)$, where ρ(e) is some function of e. OLS and LAD estimation can be regarded as the particular cases of M-estimation in which $\rho(e)= e^2$ and $\rho(e)=|e|$.

The M-estimates mentioned here are Huber M-estimates. In M-estimation, the advantages of LAD and OLS estimation are tried to be combined by defining the ρ(e) function as:

$$\rho(e) = \begin{cases} e^2 & if -k \leq e \leq k \\ 2k|e| - k^2 & if\ e < k\ or\ k < e \end{cases} .$$

Huber suggested that k = 1.5$\hat{\sigma}$, where $\hat{\sigma}$ is an estimate of the standard deviation of the population of random errors. $\hat{\sigma}$ = 1.483 MAD is used to estimate σ, where MAD is the median of the absolute deviations. The constant 1.483 is chosen to guarantee that if the normality assumption is valid, and then $\hat{\sigma}$ is still a good estimate. LAD estimates' advantage is being not as sensitive to outliers as OLS estimates, on the other hand, OLS estimates perform better than the LAD estimates when there no outliers.

The minimization criterion is to minimize:

$$\sum \rho(y_i - (a + bx_i))$$

The values of a and b that minimize this equation will be the Huber M-estimates $\hat{\alpha}$ and $\hat{\beta}$. The algorithm for minimizing the above equation is started by finding the initial estimates of α and β by OLS estimates. These are used to calculate the deviations and an estimate of σ. These will be used to get improved $\hat{\propto}$ *and* $\hat{\beta}$. Let a^0 and b^0 be the current estimates of α and β. The deviations are calculated as $y_i - (a^0 + b^0 x_i)$ and from them $\hat{\sigma}^0 = 1.483MAD$ is calculated. Response values can be defined as $y_i = a^0 + b^0 xi + ei^0$. To get rid of large deviations, an adjustment of y-values is done by $y_i{*} = a^0 + b^0 x_i + e_i{*}$ where $e_i{*}$ is the adjusted error vector obtained by truncating e_i^0, so that none of the deviations is larger than $1.5\hat{\sigma}^0$ in absolute value. If e_i^0 is between $-1.5\hat{\sigma}^0$ and $1.5\hat{\sigma}^0$, $e_i{*} = e_i^0$; if e_i^0 is less than $-1.5\hat{\sigma}^0$, $e_i{*} = -1.5\hat{\sigma}^0$ and $e_i{*} = 1.5\hat{\sigma}^0$ if e_i^0 is greater than $1.5\hat{\sigma}^0$.

For the multiple regression case, the minimization criteria is

$$\sum \rho(y_i - (b_o + b_1 x_{i1} + ... + b_p x_{ip}))$$

where the Huber M-estimates $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_p$ are the values of $b_0, b_1, ..., b_p$.

Let

$$b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_p \end{bmatrix} \text{ and } x_i = \begin{bmatrix} 1 \\ x_{i1} \\ \vdots \\ x_{ip} \end{bmatrix}$$

The vector $\hat{\beta}$ of Huber M-estimates is defined to be the vector b that minimizes

$\sum \rho(y_i - (a + b'x_i))$.

β, the vector of regression coefficients, is first estimated by the vector of LS estimates to calculate deviations and an estimate of σ. Then the deviations and the $\hat{\sigma}$ will be used to obtain an improved estimate of β. The algorithm goes on until the improved estimate of β is at least approximately the same as the previous estimate.

To describe the algorithm, let b^0 the current estimate of β. The deviations of the y_i values from the estimated regression line and the estimate of σ is calculated by $y_i - (b^0)'x_i$ and $\hat{\sigma} = 1.483\ MAD$.

As in the simple regression case; to get rid of large deviations, some adjustments are made to the response values. $e_i^0 = y_i - (b^0)'x_i$ is the deviation of y_i from the current estimated regression line, and if we put y on the right hand side $y_i = (b^0)'x +_i e_i^0$. If we define a new y value as $y_i^* = (b^0)'x + e_i^*$, where $e_i *$ is the adjusted error vector obtained by truncating e_i^0 , so that none of the deviations is larger than $1.5\hat{\sigma}^0$ in absolute value. As a result, we get the improved estimate of β from the adjusted data (Birkes and Dodge, 1993).

M-estimator is statistically more efficient (at a model with Gaussian errors) than LAD method. However its breakdown point is 1/n because of the outliers in X-space which means that it not bounded influence (Rousseeuw and Leroy, 1987).

S-plus *rreg* function can calculate Huber M-estimates by using an algorithm which applies Iteratively Reweighted Least Squares Procedure (S-PLUS Robust Manual).

2.4.3. Least Median of Squares

M-estimator's breakdown point is not high unless they have redescending ψ functions, in which case they need a good starting point (Venables&Ripley, 1999). A regression estimator with a high breakdown point was developed by Rouseeuw (1984) which is given by

$$Minimize\ med\ e_i^2$$

The residual e_i of the ith case is the difference between actual and the estimated value. The square is necessary if n is even, when the central median is taken. This method is very resistant and needs no scale estimate (Venables and Ripley, 1999). This estimator is robust to both x and y-axis outliers with breakdown point 50%. If p>1 and the observations are in general position, then the breakdown point of the LMS method is

$([n/2] - p + 2)/n$.

Unfortunately, LMS has a low asymptotic efficiency. Furthermore, it gives too much sensitivity to central data values (Rousseuw and Leroy, 1987).

S-PLUS has different functions to calculate LMS and LTS estimates like *lmsreg, ltsreg* and *lqs*. But since *lmsreg* and *ltsreg* are not fully documented, and give different results in different releases, we will use *lqs* by choosing the method option as LMS and LTS (Venables and Ripley, 1999).

2.4.4. Least Trimmed Squares

To improve the slow rate of convergence of the LMS, Rousseeuw (1983, 1984) introduced LTS by the following minimization criteria

$$Minimize \sum_{i=1}^{h} (e^2)_{i:n}$$

where $(e^2)_{1:n} \leq \ldots \leq (e^2)_{n:n}$ are the ordered squared residuals from smallest to largest. In OLS estimation, we minimize the sum of all the residuals but here we can limit one's attention to a "trimmed" sum of squares and minimize h of the residuals, so the important thing is to find the number h. LTS estimator breakdown point is 50%, that is; it can cope with several outliers at the same time up to n/2 of the data. LTS is also a reliable data analysis tool because its robustness is not affected by the number of independent variables, so can be used in multiple regression case (Rousseeuw and Leroy, 1987).

Putting $h = [n/2]$+1, the LTS estimator has an breakdown point of $([n/2] - p + 2)/n$. Moreover for $h = [n/2] + [(p + 1)/2]$, the LTS estimator takes its possible maximum value for the breakdown point (Rousseeuw and Leroy, 1987).

Rousseeuw and Van Driessen (2006) stated that the computation time of existing LTS algorithm grows too much with the size of the data set and they developed a new algorithm called Fast-LTS. For small data sets, Fast-LTS finds the same results with LTS; but for larger data sets, it gives more accurate results than existing algorithms for LTS. But since there is no program code available for this new algorithm, we used the existing LTS algorithm.

2.4.5. MM-estimation

To combine the high breakdown property of LMS and LTS with the efficiency of M-estimation; Yohai et al. (1991) introduced MM-estimation (Venables and Ripley, 1999).

Yohai's estimators are defined in three stages. The algorithm starts with calculating a high breakdown point estimate of θ^*. For this purpose, the robust estimator does not need to be efficient. At the second stage, an M-estimate of scale s_n with 50% breakdown is computed on the residuals r_i (θ^*) from the robust fit. At the third stage, equation below is minimized and the solution gives the MM-estimator $\hat{\theta}$.

$$\sum_{i=1}^{n} \psi(r_i\,(\theta)/s_n)x_i = 0,$$

which satisfies

$$S(\theta) \leq S(\theta^*),$$

where

$$S(\theta) = \sum_{i=1}^{n} \rho(r_i\,(\theta)/s_n)\,.$$

The function must satisfy the following conditions:

1. ρ is symmetric and continuously differentiable, and ρ(0)=0.
2. There exists c>0 such that ρ is strictly increasing on $[0, c]$ and constant on $[c, \infty)$.

These two conditions imply that ψ = ρ' has to be properly redescending. The important point about this algorithm is that ρ may be quite different from the second stage scale estimate s_n , because the first and the second stage estimates must have the high breakdown property, on the other hand, the third stage estimate's goal is to have high effiency. Yohai showed that MM-estimators inherit the 50% breakdown point of the first stage and, also has the exact fit property. Also, he proved that if the normality assumption is valid, then MM-estimators are still highly efficient (Rousseeuw and Leroy, 1987).

S-PLUS *rlm* function with the MM option can handle the calculations. Also *lmRobMM* function uses a slightly different M-estimatior with similar properties and gives approximately the same results as *rlm* function (Venables and Ripley, 1999).

Anderson (2001) compared the different options for MM-method in S-PLUS; which are MM1, MM2 and MM3. MM1 efficiency level is 90% with an optimal loss function, MM2 efficiency level is 80% with an optimal loss function and MM3 efficiency level is 85% with a Tukey's Bisquare loss function. The loss function determines the degree of downweighting which the outliers receive in the regression estimation. But she stated after the study that the differences among the MM-type estimators were small with respect to three performance measures: relative efficiency, bias and significance test of the null hypothesis. As a result she recommends that the default options of MM-estimator be used unless the researcher has reason to believe that changing the efficiency or the weighting function would produce improved results.

Up to now, five of the robust regression methods are discussed. Most of the robust regression methods have both strong and weak properties. To summarize, methods with high breakdown point are LMS, LTS and MM. Also these methods are robust to outliers in X-space (high leverage points). However LMS and LTS are not efficient estimators. On the other hand LAD, M and MM have high efficiency (Rousseeuw and Leroy 1987; Anderson 2001).

As mentioned before, the last two regression methods are in the class of nonparametric regression. Huber (1981) discusses the difference between robust and nonparametric methods very clearly. A procedure is called *nonparametric* if it is supposed to be used for a broad, non-parameterized set of underlying distributions. For instance, the sample mean and the sample median are the nonparametric estimates of the population mean and median but the sample mean is not a robust estimate of the population mean because it is highly sensitive to outliers. Nonparametric procedures can also be *robust*. For instance, median is a nonparametric statistic, but it is also a highly robust estimate for estimating the center of a symmetric distribution as a central model.

In this study, the robustness properties of the two nonparametric methods MARS and LOWESS will also be analyzed and discussed. You will find brief description of these methods in the following sections.

2.4.7. Multivariate Adaptive Regression Splines

MARS is an adaptive procedure that can be used for multiple regression cases (Hastie et al., 2001). The aim of the MARS procedure is to combine the recursive partitioning and spline fitting's advantages. The advantage of recursive partitioning is its adaptability through its local variable subset selection strategy which tracks the dependencies associated with a wide variety of complex functional forms. The two disadvantages of recursive partitioning are the lack of continuity of its models and its inability to capture simple relationships such as linear, additive or interactions of low order compared to n. Whereas, spline fitting is a nonadaptive procedure which produces continuous models with continuous derivatives. But it has the disadvantage that very large basis function sets are usually required in high dimensions to capture relatively simple functional relationships (Friedman, 1991).

MARS uses expansions in piecewise linear basis functions of the form $(x\text{-}t)_+$ and $(t\text{-}x)_+$. The "+" means positive part, so

$$(x-t)_+ = \begin{cases} x-t, & \text{if } x>t \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad (t-x)_+ = \begin{cases} t-x, & \text{if } x<t \\ 0, & \text{otherwise.} \end{cases} .$$

Each function is piecewise linear, with a *knot* at the value t. These are linear splines which are called a *reflected pair*. Our aim is to form these reflected pairs for each input X_j with knots at each observed value x_{ij} of that input. So, the collection of basis functions is

$$C = \{(X_j - t)_+, (t - X_j)_+\}_{\substack{t \in \{x_{1j}, x_{2j}, \ldots\ldots, x_{Nj}\} \\ j=1,2,\ldots,p}} .$$

There are totally $2N_p$ basis functions if all of the input values are distinct. Although, each basis function depends only on a single X_j , it is considered as function over the entire input space, for example, $h(X) = (X_j - t)_+$.

The model-building strategy is similar to forward stepwise linear regression, but different from regression, it can use functions from the set C (and their products) instead of using the original input variables. Therefore, the model has the form

$$f(X) = \beta_0 + \sum_{m=1}^{M} \beta_m h_m(X),$$

where each $h_m(X)$ is a function in C, or a product of two or more such functions.

The coefficients β_m are estimated by the minimization criteria of standard linear regression. The trick is finding the functions $h_m(X)$. The algorithm starts with only the constant function $h_0(X) = 1$ in the model and all functions in the set C are candidate functions.

At each stage, we consider all products of a function h_m in the model set *M* with one of the reflected pairs in C as a new basis function pair. We add to the model *M* the term of the form

$$\hat{\beta}_{M+1}h_l(X).(X_j - t)_+ + \hat{\beta}_{M+2}h_l(X).(t - X_j)_+, h_l \in M$$

that produces the largest decrease in training error. Here, $\hat{\beta}_{M+1}$ and $\hat{\beta}_{M+2}$ are coefficients estimated by least squares along with all the other M+1 coefficients in the model. Then the winning products are added to the model and the process is continued until the model set M contains some preset number of terms.

For example, at the first stage a function of the form $\beta_1(X_j - t)_+ + \beta_1(t - X_j)_+; t \in \{x_{ij}\}$ is considered to be added to the model, since multiplication by the constant function just produces the function itself. Suppose the best choice of the function form is $\beta_1(X_2 - x_{11})_+ + \beta_1(x_{11} - X_2)_+$. Then we include this pair of basis functions to the set M. At the second stage, we consider adding a pair of products of the form

$h_m(X).(X_j - t)_+$ and $h_m(X).(t - X_j)_+ , t \in \{x_{ij}\}$

where for h_m, we have the choices

$$h_0(X) = 1,$$
$$h_1(X) = (X_2 - x_{11})_+, or$$
$$h_2(X) = (x_{11} - X_2)_+ .$$

At the end of this process, we have a large model which overfits the data, so a backward deletion procedure is applied. The term whose removal causes the smallest increase in residual square error (RSS) is deleted from the model at each stage. These deletions produce an estimated best model $\widehat{f_\lambda}$ of each number of terms of λ.

Cross-validation can be used to estimate the optimal value of the λ, but MARS procedure uses generalized cross-validation for computational time saving. Generalized cross-validation criterion is computed by

$$GCV(\lambda) = \frac{\sum_{i=1}^{N}\left(y_i - \hat{f}_\lambda(x_i)\right)^2}{\left(1 - M(\lambda)/N\right)^2}$$

where M(λ) is the effective number of parameters in the model. Effective number of parameters is the total number of terms in the model plus the number of parameters used in selecting the optimal positions of the knots. By backward selection, we choose the model that minimizes the GCV (λ) criterion.

Advantage of the MARS method is that the piecewise linear basis functions has the ability to operate locally, that is outside of their range they are zero. But when they are multiplied together, the result is nonzero only over the small part of the feature space where both component functions are nonzero. Therefore, the regression surface is built up by using nonzero components locally-only when they are needed. This advantage is important because parameters should be spent carefully in high dimensions, as they can run out quickly. The second advantage of MARS method is easiness of computation. First, the reflected pair is fit with right most knot. As the knot is moved successively one position at a time to the left; over the left part of the domain area, the basis functions differ by zero and by a constant over the right part. (Hastie et al., 2001).

MARS method is added to the S-PLUS by calling the MDA and MASS libraries from R-language. After adding the libraries, MARS method can be applied on S-PLUS with the function *"mars()"*.

2.4.8. Local Weighted Scatter Plot Smoothing

A small number of outliers can seriously affect the estimates accuracy. So, to decrease the influence of outliers, some smoothing methods have been developed. Among them LOWESS is a well known method which is called a "robust version of LOESS" (Takezawa, 2006). The LOWESS algorithm is quite complex; it uses robust locally linear fits. A window is placed about x, we weight the data points that lie inside the window so that points near to x get the most weight and a robust weighted regression is used to predict the value at x (Venables and Ripley, 1999).

LOWESS algorithm is

1. $\widehat{Y}_i$ is the result of the smoothing data by LOESS.
2. Robustness weights ($\{\sigma_i\}$ $(i = 1, \dots, n)$) are derived as follows

   $$\sigma_i = B(r_i/(6\hat{s})),$$

   where

   $r_i = Y_i - \widehat{Y}_i$ are the residuals

   $\hat{s} = med(|r_i|)$ is the scale estimate

   B(u) is a bisquare weight function which is

   $$B(u) = \begin{cases} \frac{15}{16}(1-u^2)^2 & 0 \leq |u| \leq 1 \\ 0 & 1 < |u| \end{cases}$$

3. Smoothing by LOESS is carried out with weight of σ_i on the ith data. That is, $w\left(\frac{X_i - x^*}{h_k(x)}\right)$ in eq(…) is replaced with $\left(\sigma_i \,.\, w\left(\frac{X_i - x^*}{h_k(x)}\right)\right)$ to obtain estimates. The residuals are calculated as $r_i = Y_i - \widehat{Y}_i$
4. (2) and (3) above are repeated three more times.

As a result, we obtain smoothing that is robust to outliers.

S-PLUS has a function called *loess ()* which implements the loess method. If family= "symmetric" is assigned in loess (), calculation of LOWESS is carried out. If family is not chosen or family= "Gaussian" is assigned, LOESS is executed (Takezawa, 2006). The parameter f controls the size of the window and is the proportion of the data which is included. In S-PLUS, the default value is f=2/3 but it is often too large for scatter plots with appreciable structure (Venables and Ripley, 1999).

2.5. Previous Comparative Studies

Stigler (1977) presents a comparison of the performances of eleven robust estimators by using real data sets. He mentioned that most of the robustness studies have relied upon mathematical theory, computer simulated data or a combination of these and there is a lack of real data studies. His data sets are from the physics like the speed of light or density of the earth. The estimators considered in this study are the mean, median, 10%, 15%, and 25% trimmed means, three versions of M-estimators, Edgeworth, outmean and Hogg's T1 which are taken from the comprehensive Princeton Simulation study. The performance measures are relative error and relative rank. The relative error measures the absolute value of an estimator's error relative to the sizes of the errors achieved by other estimators for the same data set. On the other hand, the relative rank does not take into account the actual errors of estimation. Only their ranks are calculated for each data set. Eventually it is found that light trimming improves the sample mean, but that the sample mean is also competitive among the many recently proposed methods.

Nevit and Tam (1997) investigate some nonparametric and robust regression methods' performance for situations in which the underlying assumptions of OLS are violated by the presence of outliers in the observed data. This study is only carried out for simple linear regression. A program called GAUSS is used for the simulation study. Design parameters included are sample size and types of the distribution. Unit normal, contaminated normal, lognormal and t-5 df distributions are considered. Variance, bias, mean squared error and relative mean square error are chosen as the performance measures. LAD, 10% and 20% Winsorized least squares, 10% trimmed least squares are the robust methods and monotonic regression, weighted median estimator, Theil median estimator are the nonparametric methods included in this study. The results of the study demonstrated that under mild data contamination (10%), none of the methods outperform the others. But when the outlier density is increased to 30%, the LAD slope estimator is the most competitive one. Moreover, it can be said that for the nonnormal distributions, the symmetry of the error distribution dramatically affects the estimator performance. LAD estimator is no more desirable but the Winsorized least squares and the nonparametric method Theil are preferable.

Wisnowski (1999) carried out a comprehensive multiple outlier detection study by using some robust methods and Monte Carlo Simulation. He mentions about the two reasons why robust regression is not widely used. The first mentioned reason is that an extra effort should be spent to get the appropriate software which is capable of making robust analyses. The second is that performance analyses are required. Therefore, the aim of his study is to compare the performance of the leading multiple outlier detection procedures for the linear regression model. Both the direct and indirect procedures are taken in this study. The indirect procedures include the robust regression methods which are LMS, LTS, M and MM. Detection capability and false alarm rate are used as performance measures. The factorial design parameters are the sample size, percentage of outliers, number of independent variables, outlier location, number of multiple point clouds and the proportion of independent variables with outliers. The factorial design parameters are selected to be convenient with the literature. The simulated data were analyzed in S-PLUS 4.5.His simulation results demonstrate that OLS, M and MM methods outperform the other two robust methods LTS and LMS. OLS method's strong performance is unexpected here. One lack of this comprehensive study is that, he did not compare the methods on real life data.

In 2001, Anderson also made a similar simulation study about some robust regression methods using Monte Carlo Simulation. M, MM, MM1, MM2, MM3 and LAV methods are compared on different data configurations. MM1 is the MM method with the tuning parameter changed from 0.85 to 0.80 and MM2 uses the 0.90 tuning constant. MM3 is the MM method with the rho function is Tukey instead of optimal choice. Factorial design parameters are sample size, number of independent variables, outlier density and outlier location. Her performance measures are relative efficiency, bias and test of the null hypothesis. Her simulation results demonstrated that the MM type robust regression methods outperform OLS and LAV with respect to the three performance measures.

Moreover, Lane (2002) made a robust regression study aiming to discuss and compare some robust regression methods which are LTS and MM method. A heuristic data set about the number of international phone calls from Belgium in years 1950-1973 was taken from Rousseeuw and Leroy (1987). In this data set, the outliers occurred by using a different system of measurement in the years 1964-1969 and known. He used S-PLUS 2000 to model the data with outliers. As a performance criterion he compared the regression lines. Both the robust methods' lines fit the data better than the OLS regression line. As a result, he advised to use robust regression to handle with the outliers in the data.

CHAPTER 3

A SIMULATION STUDY FOR COMPARING PERFORMANCES OF ROBUST REGRESSION METHODS

To compare robust regression techniques, simulation has been a commonly used tool ever since Pearson study conducted in 1930 published at Biometrika. The researcher is free to specify the type of the distribution and is able to know what kind of a mechanism produced his data. For example, outlier observations are exactly known. Therefore it is easy to evaluate the performance of the methods for such simulated data. But even if the researcher is very experienced, there is no guarantee that the simulated samples actually represent the data. In fact, the researchers generally focus on a narrow range of alternatives to normality that is independent, identically distributed samples from long tailed symmetric continuous distributions. However the real data can be correlated biased and the outlying observations are not known for sure. As a result the performance of the methods should also be compared on real data sets (Stigler, 1977).

In this study, we used both approaches. In this chapter, a Monte Carlo simulation study is performed with respect to the scenarios indicated by Wisnowski (1999). The seven robust regression methods are compared by the performance measures which are detection capability, false alarm rate and an improved mean square error. After the simulation results are obtained, for each performance measure a Repeated ANOVA study is conducted and analyzed to see whether there is a significant difference between the methods. The results are discussed and the most promising methods which will be used in the real data case are determined.

3.1. Monte Carlo Simulation Study Planning

3.1.1. Data Generation and Outlier Planting

Monte Carlo simulation approach is used to test the performance of the robust regression procedures studied. The S-PLUS code for the simulation and data generation scenarios are adapted from Wisnowski (1999). A fixed percentage of clean observations are generated and then outliers are placed at the specified locations as suggested by the scenario and design parameters.

For the clean observations, the independent variable levels are generated from a multivariate normal distribution with $\mu_x = 7.5$ and $\sigma_x = 4$. Wisnowski (1999) states that these parameter values are selected to be consistent with some of the results in the literature. The dependent variable for the ith clean observation is generated by $y_i = x_i^{'}\beta + \varepsilon_i$ where β is the vector of known regression coefficients, which are all equal to 5 for each of the k independent variables and equal to 0 for the intercept and ε_i is distributed N(0,1).

To generate outliers; the ith independent variable value for the jth observation can be taken as $x_{ij} = \bar{x}_{i,clean} + 4\delta_L + \varepsilon_{ij}^{*}$ where $\bar{x}_{i,clean}$ is the average of the clean values for the ith independent variable, δ_L is the magnitude of the outlying shift distance in X-space in standard deviation units, σ_x. ε_{ij}^{*}, which is used to separate multiple observations in a cloud to protect against singular matrices, is a random variable generated from a Uniform (0, 0.025). If the ith observation is a y-axis outlier, the response value is calculated by $y_i = x_i^{'}\beta + \delta_R$ where δ_R is the magnitude of the outlying distance off the regression plane in standard deviation units.

In our simulation study, outliers are placed at three different locations which are *randomly scattered outliers in the interior of X-space*, *outliers in multiple point clouds at the centroid of X-space* and *outliers in multiple point clouds* *when the independent variables are randomly scattered in the interior of X-space.* These scenarios represent only the X-space outliers since the LAD, M and MM methods are vulnerable in high leverage situations as stated by Wisnowski (1999).

3.1.2. Performance Measures

To compare the robust regression methods' performances, we have chosen three performance measures: detection capability, false alarm rate and mean square error of the clean data. The first two measures are adapted from Hadi and Simonoff (1993) and Wisnowski (1999). The third one is an improved version of relative efficiency used in Anderson (2001).

Detection Capability (PP) = P (at least one planted outlier is detected) = p_1

The complement of detection capability is the masking probability and can be shown as, P (Masking) = P (none of the planted outlier is detected) =1- p_1.

False Alarm Rate (PO) = P (a clean observation is swamped)

Improved Mean Square Error (IMSE): This performance measure is improved based on the idea of relative efficiency measure mentioned in Anderson (2001). Relative efficiency can be defined as the degree to which an estimator performs like OLS, when OLS has normally distributed errors. Relative efficiency is usually expressed as a percentage as in defined by Ryan (1997) as cited by Anderson (2001),

Relative Efficiency = MSE Robust / MSE OLS

This ratio theoretically is between 0 and 1, but can exceed one if the robust MSE is less than the OLS MSE. For this reason, higher percentages are more desirable.

The MSE OLS in the ratio is the mean square error of the data without outliers. However, MSE Robust is the mean square error of the robust regression model calculated by including all the error terms in the summation. But this gives a disadvantage to robust estimators because these methods give outliers very large values so that they can be easily detected.

When we fit a robust regression line to the data, the model includes outliers and as a result the outliers' residuals will be much larger than the normal observations' residuals. When calculating the MSE of robust regression methods, the outliers' residual values should be ignored. By this way, we will prevent penalizing these methods because they fit the regression line based on majority of the data, but not close to outliers. Therefore, in our simulation study *improved MSE (IMSE)* is used as the third performance measure.

Ratio of the IMSE: It is found by dividing the IMSE of the robust method by the IMSE of OLS. It shows how the method performs compared with OLS when there are no outliers in the data.

3.1.3. Factorial Design

The factorial design of the study considers the sample size, percentage of outliers, number of independent variables (dimension of the data), outlier location (δ_L is the magnitude of the unusualness in x-space; δ_R is the magnitude of unusualness in y-space), the number of multiple point clouds and the proportion of independent variables with outliers. The factorial design parameters are selected to be convenient with the literature (Wisnowski, 1999).

The number of independent variables are k=2 for n=40 observations or k=6 with n=60 observations. The outlier density is either 10% or 20%. The magnitude of the outliers is between 3 and 5 standard deviation units. The number of multiple point clouds is either 1 or 2.

To define an observation as an outlier, we should use cut off values for each robust regression procedure. The simulated cutoff values for each procedure is calculated as the 95th percentile of the absolute value of the residuals from the normally distributed data which does not contain outliers. S-PLUS is used to perform the simulations.

The results of the cut-off simulation results are not the same as in the Wisnowski (1999). The difference is not because of the iteration number, if we iterate 10000 times, the results do not change. Also, if analyze the *seed (i)* function in S-PLUS 4.5 which is the S-PLUS version used in Wisnowski, the generated seeds are as the same as the S-PLUS 6 version. Moreover, the random number generator function *rnorm ()* gives the same random numbers both in S-PLUS 4.5 and S-PLUS 6. To validate our cut-off values, Minitab14 is used and we get the same cut-off values as in ours from Minitab.

Our performance measures; detection capability, false alarm rate and improved mean square error are calculated by performing 500 replications.

3.2. Simulation Results and Performance Analysis

In all three scenarios, there is no independent variable that is unusual in the X-space for the interior X-space outliers. That is, no high-leverage point is intentionally located in the samples. The response values for the interior X-space outliers are δ_R sigma away from the regression plane which is obtained from the clean cases. δ_R is the magnitude of unusualness in y-space whose value can be 3σ, 4σ or 5σ.

There are three cases for the interior X-space outliers. In the first case, multiple outliers are randomly scattered in the interior of X-space. In the second case, multiple point clouds or clusters of outliers are located near the centroid of X-space. The third study considers multiple point clouds randomly placed (different for each replication) in the interior of X-space.

3.2.1. Scenario 1: Randomly Scattered Regression Outliers in the Interior of X-space:

In this scenario, the outliers have random levels of the independent variables with the same distribution as the clean observations but the dependent variable values are placed at δ_R sigma away from the regression plane. The response to the ith clean observation is generated by $y_i = x_i'\beta + \varepsilon_i$ where β is the coefficients vector which is known and selected to be 5 for each of the k independent variables and 0 for the intercept. x_i is the vector of k independent variables distributed N (7.5, 4) and ε_i is the random error term distributed N (0, 1). The response to the ith outlying observation is generated by $y_i = x_i'\beta + \delta_R$ where δ_R is the outlying distance off the regression plane in standard deviation units, which equals 1 in our study. The factorial design with the simulation results in Table 3.1 includes the following factors: A (n: sample size and k: number of independent variables in the regression equation), B(dens : outlier density), C(δ_R : is the magnitude of unusualness in y-space), D (cld : number of clouds).

3.2.2. Scenario2: Regression Outliers in Multiple Point Clouds at the Centroid of X-space:

This scenario compares the performance of the robust regression procedures when there are multiple Y-space outliers forming clouds at the centroid of X-space.

The response to the ith clean case is generated by $y_i = x_i'\beta + \varepsilon_i$ where β is the coefficients vector which is known and selected to be 5 for each of the k independent variables and 0 for the intercept. X_i is the vector of k independent variables whose distribution is N (7.5, 4) and ε_i is the random error term distributed N (0, 1).

Table 3.1: Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for Scenario-1.

A ***n,k***	**B** ***dens***	**C** δ_R	**OLS**	**M**	**MM**	**LTS**	**LMS**	**LAD**	**LWS**	**MARS**
40,2	**10%**	**3σ**	0.970 0.067 1.043 1	0.971 0.057 0.986 0.945	0.972 0.058 0.998 0.957	0.832 0.043 1.270 1.218	0.819 0.037 1.218 1.168	0.960 0.064 1.002 0.961	0.873 0.076 0.903 0.866	0.872 0.074 0.983 0.942
60,6	**10%**	**3σ**	0.902 0.065 1.020 1	0.901 0.055 0.964 0.945	0.902 0.057 0.967 0.948	0.744 0.064 1.472 1.443	0.733 0.059 1.418 1.390	0.901 0.056 0.995 0.975	0.576 0.059 0.775 0.760	0.747 0.069 1.130 1.108
40,2	**20%**	**3σ**	0.885 0.101 1.306 1	0.872 0.087 1.236 0.946	0.876 0.089 1.229 0.941	0.824 0.046 1.283 0.982	0.807 0.046 1.292 0.989	0.889 0.066 1.112 0.851	0.796 0.113 1.135 0.869	0.773 0.113 1.248 0.956
60,6	**20%**	**3σ**	0.744 0.098 1.282 1	0.721 0.086 1.234 0.962	0.727 0.088 1.230 0.959	0.691 0.083 1.726 1.346	0.676 0.073 1.651 1.288	0.773 0.074 1.142 0.891	0.475 0.095 1.009 0.787	0.590 0.103 1.177 0.918
40,2	**10%**	**4σ**	0.997 0.078 1.135 1	0.998 0.054 0.965 0.850	0.998 0.056 0.965 0.850	0.986 0.039 1.229 1.083	0.984 0.035 1.198 1.055	0.998 0.054 1.002 0.883	0.949 0.093 1.004 0.884	0.941 0.091 1.102 0.971
60,6	**10%**	**4σ**	0.989 0.080 1.134 1	0.995 0.052 0.935 0.825	0.995 0.054 0.940 0.829	0.959 0.059 1.391 1.227	0.962 0.053 1.337 1.179	0.994 0.056 0.996 0.878	0.759 0.077 0.889 0.784	0.917 0.090 1.269 1.119
40,2	**20%**	**4σ**	0.984 0.141 1.610 1	0.982 0.096 1.302 0.809	0.983 0.095 1.268 0.787	0.983 0.038 1.171 0.727	0.980 0.037 1.185 0.736	0.990 0.067 1.113 0.691	0.918 0.159 1.423 0.884	0.921 0.159 1.587 0.986

Table 3.1 (cont'd): Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for Scenario-1.

A *n,k*	B *dens*	C δ_R	OLS	M	MM	LTS	LMS	LAD	LWS	MARS
60,6	**20%**	**4σ**	0.951 0.140 1.608 1	0.944 0.103 1.369 0.851	0.945 0.101 1.337 0.831	0.952 0.063 1.427 0.887	0.951 0.049 1.320 0.821	0.973 0.074 1.145 0.712	0.683 0.139 1.309 0.814	0.791 0.150 1.498 0.932
40,2	**10%**	**5σ**	1 0.094 1.253 1	1 0.051 0.941 0.751	1 0.053 0.938 0.749	0.999 0.039 1.220 0.974	0.999 0.039 1.184 0.945	1 0.054 1.002 0.799	0.976 0.115 1.134 0.905	0.957 0.111 1.240 0.990
60,6	**10%**	**5σ**	0.999 0.099 1.280 1	0.999 0.047 0.895 0.699	1 0.048 0.893 0.698	0.096 0.057 1.370 1.070	0.996 0.052 1.321 1.032	0.999 0.056 0.996 0.778	0.853 0.099 1.036 0.809	0.921 0.113 1.236 0.966
40,2	**20%**	**5σ**	0.995 0.195 2 1	0.997 0.090 1.259 0.629	0.998 0.083 1.170 0.585	0.998 0.037 1.158 0.579	0.999 0.035 1.158 0.579	0.999 0.067 1.134 0.567	0.948 0.215 1.793 0.896	0.950 0.218 2.028 1.014
60,6	**20%**	**5σ**	0.985 0.195 2.028 1	0.987 0.106 1.386 0.683	0.988 0.094 1.277 0.630	0.993 0.058 1.347 0.664	0.995 0.044 1.244 0.613	0.998 0.074 1.145 0.564	0.784 0.191 1.694 0.835	0.844 0.204 1.912 0.943
Average probabilities			0.883 0.113 1.334 1	0.947 0.074 1.099 0.825	0.949 0.073 1.085 0.814	0.838 0.052 1.338 1.017	0.908 0.047 1.298 0.983	0.956 0.064 1.058 0.796	0.799 0.119 1.128 0.841	0.852 0.125 1.318 0.987

The response to the i[th] outlying observation is generated by $y_i = x_i'\beta + \delta_R$ where x_i is the vector of k independent variables distributed unif (7.375, 7.625) and δ_R is the outlying distance off the regression plane in standard deviation units, which equals 1 in our study. If the number of clouds is two, the response values for the outliers in the first cloud are generated as above and the second clouds response values are generated by $y_i = x_i'\beta - \delta_R$.

The four factors in this scenario are A (n: sample size and k: number of independent variables in the regression equation), B (dens: outlier density), C (δ_R : is the magnitude of unusualness in y-space), D (cld: number of clouds).

Wisnowski (1999) stated that in this scenario the levels of δ_R are chosen close to one another because initial studies have demonstrated that none of the robust regression procedures have detection capability below $3\sigma_e$ and nearly all had perfect detection capability at $5\sigma_e$ and below.

The factorial design and the simulation results are presented in Table 3.2.

3.2.3. Scenario3: Regression Outliers in Multiple Point Clouds: Independent Variables Randomly Scattered on the Interior of X-Space

In this case, the multiple outlier clouds are not placed at the centroid but at different locations in X-space for each replication. In a single point cloud, the location of the independent variables for outlying observations is determined by using the median of the first three clean observations for each variable. To guarantee the variation of the outlying observations, we add unif (0, 0.25) to this median value. In the second cloud, outliers are placed around the median value of the last three clean observations in each variable. To cover the X-Space adequately, median of the three observations is used. If median of more than three observations is used, then the outlying observations will be placed too close to the centroid of X-space (Wisnowski, 1999).

The factorial design matrix with the simulation results is demonstrated in Table 3.3 where the factors are A (n: sample size and k: number of independent variables in the regression equation), B (dens: outlier density), C (δ_R: is the magnitude of unusualness in y-space) and D (cld: number of clouds).

Table 3.2: Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for Scenario-2.

A *n,k*	B *dens*	C δ_R	D *cld*	OLS	LTS	LMS	M	MM	LAD	LWS	MARS
40,2	10%	3σ	1	1 0.063 1.012 1	0.934 0.037 1.197 1.183	0.928 0.033 1.177 1.164	1 0.053 0.960 0.949	1 0.055 0.962 0.951	1 0.053 0.989 0.977	0.969 0.074 0.894 0.883	0.594 0.081 1.043 1.031
60,6	10%	3σ	1	1 0.058 0.981 1	0.977 0.058 1.387 1.414	0.964 0.052 1.343 1.369	1 0.049 0.925 0.943	1 0.051 0.930 0.948	1 0.055 0.978 0.997	1 0.198 1.734 1.767	0.227 0.080 1.365 1.391
40,2	20%	3σ	1	0.997 0.098 1.279 1	0.902 0.055 1.678 1.312	0.881 0.053 1.600 1.251	0.998 0.081 1.188 0.928	0.996 0.084 1.189 0.929	0.994 0.061 1.076 0.841	0.290 0.105 1.084 0.847	0.038 0.119 1.292 1.010
60,6	20%	3σ	1	1 0.094 1.257 1	0.712 0.162 5.673 4.513	0.787 0.103 3.082 2.452	0.998 0.079 1.171 0.931	0.998 0.081 1.175 0.935	0.998 0.065 1.061 0.844	1 0.452 3.173 2.524	0.002 0.110 1.591 1.266
40,2	10%	4σ	1	1 0.073 1.086 1	1 0.037 1.197 1.102	1 0.033 1.177 1.084	1 0.050 0.941 0.866	1 0.052 0.942 0.867	1 0.053 0.990 0.912	1 0.090 0.990 0.912	0.776 0.109 1.239 1.141
60,6	10%	4σ	1	1 0.068 1.059 1	1 0.058 1.387 1.309	1 0.052 1.343 1.268	1 0.048 1.902 1.796	1 0.049 0.906 0.855	1 0.055 0.978 0.923	1 0.358 2.627 2.481	0.793 0.106 1.203 1.136
40,2	20%	4σ	1	1 0.139 1.569 1	0.994 0.039 1.266 0.807	0.998 0.033 1.153 0.735	1 0.083 1.197 0.763	1 0.086 1.189 0.758	1 0.061 1.076 0.686	0.909 0.142 1.352 0.862	0.243 0.170 1.688 1.076
60,6	20%	4σ	1	1 0.137 1.561 1	0.964 0.073 2.322 1.487	0.994 0.047 1.450 0.929	1 0.081 1.182 0.757	1 0.084 1.190 0.762	1 0.065 1.061 0.680	1 0.701 5.220 3.344	0.011 0.150 1.890 1.211
40,2	10%	3σ	2	1 0.050 0.918 1	0.948 0.038 1.195 1.301	0.944 0.033 1.177 1.282	1 0.049 0.922 1.004	1 0.050 0.921 1.003	1 0.048 0.966 1.052	1 0.053 0.765 0.833	1 0.054 0.854 0.930

Table 3.2 (cont'd): Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for Scenario-2.

A *n,k*	B *dens*	C δ_R	D *cld*	OLS	LTS	LMS	M	MM	LAD	LWS	MARS
60,6	10%	3σ	2	1 0.046 0.811 1	0.967 0.059 1.384 1.706	0.955 0.052 1.343 1.656	1 0.046 0.888 1.095	1 0.047 0.885 1.091	1 0.051 0.952 1.174	1 0.036 0.600 0.740	1 0.044 0.767 0.946
40,2	20%	3σ	2	1 0.049 0.907 1	0.945 0.372 1.156 1.274	0.925 0.032 1.144 1.261	1 0.047 0.909 1.002	1 0.048 0.907 1	1 0.049 0.958 1.056	1 0.049 0.725 0.799	1 0.055 0.847 0.934
60,6	20%	3σ	2	1 0.044 0.864 1	0.973 0.056 1.346 1.558	0.953 0.048 1.297 1.501	1 0.044 0.867 1.003	1 0.044 0.864 1	1 0.051 0.941 1.089	1 0.028 0.530 0.613	0.953 0.028 0.558 0.646
40,2	10%	4σ	2	1 0.050 0.918 1	1 0.038 1.195 1.302	1 0.033 1.177 1.282	1 0.049 0.922 1.004	1 0.050 0.921 1.003	1 0.048 0.966 1.052	1 0.053 0.765 0.833	1 0.054 0.854 0.930
60,6	10%	4σ	2	1 0.073 1.086 1	1 0.037 1.197 1.102	1 0.033 1.177 1.084	1 0.050 0.941 0.866	1 0.052 0.942 0.867	1 0.053 0.990 0.912	1 0.090 0.990 0.912	0.776 0.109 1.239 1.141
40,2	20%	4σ	2	1 0.049 0.907 1	1 0.036 1.147 1.264	1 0.032 1.135 1.251	1 0.047 0.910 1.003	1 0.048 0.909 1.002	1 0.049 0.959 1.057	1 0.049 0.725 0.799	1 0.055 0.847 0.934
60,6	20%	4σ	2	1 0.045 0.864 1	1 0.056 1.336 1.546	1 0.048 1.297 1.501	1 0.044 0.868 1.005	1 0.044 0.865 1.001	1 0.051 0.941 1.089	1 0.028 0.530 0.613	0.747 0.005 0.247 0.286
Avrg. Prob.				1 0.071 1.067 1	0.957 0.076 1.629 1.511	0.958 0.045 1.380 1.317	1 0.056 1.043 0.995	1 0.058 0.981 0.936	1 0.054 0.993 0.959	0.948 0.157 1.419 1.235	0.635 0.083 1.095 1.001

Table 3.3: Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for Scenario-3

A *n,k*	B *dens*	C δ_R	D *cld*	OLS	LTS	LMS	M	MM	LAD	LWS	MARS
40,2	**10%**	**3σ**	**1**	0.726 0.070 1.063 1	0.672 0.042 1.237 1.163	0.663 0.038 1.229 1.156	0.762 0.057 0.983 0.924	0.763 0.059 0.983 0.924	0.752 0.055 1.006 0.946	0.633 0.081 0.926 0.871	0.550 0.086 1.058 0.995
60,6	**10%**	**3σ**	**1**	0.619 0.072 1.072 1	0.650 0.064 1.459 1.361	0.631 0.054 1.381 1.288	0.683 0.057 0.975 0.909	0.686 0.059 0.976 0.910	0.684 0.061 1.022 0.953	0.739 0.068 0.840 0.784	0.390 0.082 1.542 1.438
40,2	**20%**	**3σ**	**1**	0.567 0.111 1.377 1	0.612 0.055 1.446 1.050	0.596 0.051 1.410 1.024	0.593 0.090 1.258 0.913	0.599 0.091 1.247 0.905	0.628 0.076 1.186 0.861	0.404 0.109 1.105 0.802	0.243 0.116 1.274 0.925
60,6	**20%**	**3σ**	**1**	0.394 0.115 1.402 1	0.456 0.107 2.216 1.581	0.412 0.104 2.204 1.572	0.398 0.102 1.366 0.974	0.406 0.104 1.362 0.971	0.420 0.098 1.371 0.978	0.653 0.100 1.048 0.747	0.170 0.102 1.511 1.078
40,2	**10%**	**4σ**	**1**	0.919 0.083 1.167 1	0.915 0.039 1.215 1.041	0.904 0.036 1.203 1.031	0.950 0.056 0.970 0.831	0.951 0.058 0.970 0.831	0.947 0.056 1.007 0.863	0.837 0.098 1.036 0.888	0.727 0.110 1.244 1.066
60,6	**10%**	**4σ**	**1**	0.836 0.091 1.224 1	0.905 0.059 1.395 1.140	0.903 0.050 1.328 1.085	0.921 0.056 0.962 0.786	0.926 0.057 0.955 0.780	0.915 0.061 1.028 0.840	0.877 0.095 1.025 0.837	0.570 0.107 1.454 1.188
40,2	**20%**	**4σ**	**1**	0.800 0.159 1.735 1	0.906 0.039 1.200 0.692	0.897 0.039 1.223 0.705	0.842 0.110 1.410 0.813	0.857 0.103 1.337 0.771	0.881 0.078 1.207 0.696	0.593 0.148 1.383 0.797	0.315 0.159 1.617 0.932
60,6	**20%**	**4σ**	**1**	0.590 0.166 1.822 1	0.776 0.085 1.986 1.090	0.764 0.081 1.945 1.067	0.602 0.141 1.716 0.942	0.620 0.134 1.664 0.913	0.674 0.115 1.545 0.848	0.782 0.159 1.436 0.788	0.225 0.139 1.742 0.956
40,2	**10%**	**5σ**	**1**	0.984 0.099 1.302 1	0.987 0.039 1.206 0.926	0.984 0.036 1.204 0.925	0.997 0.054 0.948 0.728	0.997 0.056 0.947 0.727	0.995 0.056 1.007 0.773	0.946 0.116 1.178 0.905	0.825 0.139 1.478 1.135
60,6	**10%**	**5σ**	**1**	0.945 0.118 1.420 1	0.986 0.056 1.360 0.958	0.988 0.049 1.308 0.921	0.990 0.050 0.919 0.647	0.990 0.050 0.907 0.639	0.989 0.061 1.028 0.724	0.938 0.132 1.263 0.889	0.679 0.133 1.655 1.165

Table 3.3 (cont'd): Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for for Scenario-3.

A ***n,k***	**B** ***dens***	**C** δ_R	**D** ***cld***	**OLS**	**LTS**	**LMS**	**M**	**MM**	**LAD**	**LWS**	**MARS**
40,2	**20%**	**5σ**	**1**	0.927 0.217 2.196 1	0.987 0.037 1.147 0.522	0.983 0.037 1.179 0.537	0.957 0.121 1.493 0.680	0.970 0.097 1.279 0.582	0.978 0.078 1.211 0.551	0.756 0.188 1.741 0.793	0.384 0.199 1.079 0.491
60,6	**20%**	**5σ**	**1**	0.758 0.224 0.362 1	0.937 0.067 1.609 0.681	0.938 0.058 1.546 0.654	0.771 0.182 2.123 0.899	0.800 0.147 1.848 0.782	0.847 0.123 1.657 0.701	0.864 0.227 1.934 0.819	0.245 0.168 2.665 1.128
40,2	**10%**	**3σ**	**2**	0.802 0.287 4.817 1	0.802 0.041 1.228 0.255	0.794 0.038 1.213 0.252	0.868 0.054 0.944 0.196	0.868 0.055 0.944 0.196	0.862 0.052 0.988 0.205	0.830 0.253 5.602 1.163	0.802 0.363 10.383 2.155
60,6	**10%**	**3σ**	**2**	0.817 0.464 16.223 1	0.820 0.061 1.404 0.086	0.812 0.051 1.328 0.082	0.871 0.050 0.901 0.055	0.872 0.049 0.902 0.056	0.862 0.054 0.977 0.060	0.882 0.551 20.708 1.276	0.598 0.325 24.428 1.508
40,2	**20%**	**3σ**	**2**	0.752 0.461 13.832 1	0.793 0.040 1.192 0.086	0.784 0.039 1.201 0.086	0.827 0.061 1.015 0.073	0.830 0.062 1.017 0.073	0.822 0.061 1.080 0.078	0.748 0.362 15.898 1.149	0.704 0.462 23.012 1.664
60,6	**20%**	**3σ**	**2**	0.789 0.618 56.112 1	0.782 0.066 1.484 0.026	0.778 0.058 1.426 0.025	0.779 0.063 1.038 0.018	0.781 0.064 1.036 0.018	0.792 0.074 1.149 0.020	0.850 0.659 46.080 0.821	0.447 0.369 33.609 0.599
40,2	**10%**	**4σ**	**2**	0.898 0.295 4.913 1	0.922 0.039 1.205 0.245	0.915 0.036 1.204 0.245	0.954 0.053 0.939 0.191	0.954 0.054 0.940 0.191	0.946 0.053 0.989 0.201	0.911 0.258 5.670 1.154	0.862 0.370 10.078 2.051
60,6	**10%**	**4σ**	**2**	0.886 0.465 16.362 1	0.932 0.060 1.395 0.085	0.934 0.051 1.322 0.081	0.961 0.048 0.892 0.054	0.961 0.048 0.893 0.055	0.956 0.054 0.977 0.060	0.911 0.556 20.955 1.281	0.665 0.338 24.545 1.500
40,2	**20%**	**4σ**	**2**	0.813 0.473 14.177 1	0.921 0.036 1.146 0.081	0.918 0.035 1.157 0.082	0.938 0.063 1.023 0.073	0.939 0.064 1.022 0.072	0.936 0.062 1.082 0.076	0.820 0.374 16.100 1.136	0.743 0.478 23.436 1.653
60,6	**20%**	**4σ**	**2**	0.823 0.623 56.501 1	0.922 0.058 1.376 0.024	0.922 0.050 1.328 0.023	0.902 0.069 1.096 0.019	0.907 0.069 1.076 0.019	0.918 0.076 1.164 0.020	0.876 0.662 46.609 0.825	0.477 0.388 35.210 0.623

Table 3.3 (cont'd): Design matrix with detection capability (first), false alarm rates (second), IMSE (third) and ratio of the IMSE (fourth) for Scenario-3

A ***n,k***	**B** ***dens***	**C** δ_R	**D** ***cld***	**OLS**	**LTS**	**LMS**	**M**	**MM**	**LAD**	**LWS**	**MARS**
40,2	**10 %**	**4σ**	**2**	0.898 0.295 4.913 1	0.922 0.039 1.205 0.245	0.915 0.036 1.204 0.245	0.954 0.053 0.939 0.191	0.954 0.054 0.940 0.191	0.946 0.053 0.989 0.201	0.911 0.258 5.670 1.154	0.862 0.370 10.078 2.051
60,6	**10 %**	**4σ**	**2**	0.886 0.465 16.362 1	0.932 0.060 1.395 0.085	0.934 0.051 1.322 0.081	0.961 0.048 0.892 0.054	0.961 0.048 0.893 0.055	0.956 0.054 0.977 0.060	0.911 0.556 20.955 1.281	0.665 0.338 24.545 1.500
40,2	**20 %**	**4σ**	**2**	0.813 0.473 14.177 1	0.921 0.036 1.146 0.081	0.918 0.035 1.157 0.082	0.938 0.063 1.023 0.073	0.939 0.064 1.022 0.072	0.936 0.062 1.082 0.076	0.820 0.374 16.100 1.136	0.743 0.478 23.436 1.653
60,6	**20 %**	**4σ**	**2**	0.823 0.623 56.501 1	0.922 0.058 1.376 0.024	0.922 0.050 1.328 0.023	0.902 0.069 1.096 0.019	0.907 0.069 1.076 0.019	0.918 0.076 1.164 0.020	0.876 0.662 46.609 0.825	0.477 0.388 35.210 0.623
40,2	**10 %**	**5σ**	**2**	0.940 0.301 5.036 1	0.954 0.039 1.205 0.239	0.955 0.036 1.205 0.239	0.967 0.054 0.934 0.185	0.965 0.052 0.990 0.196	0.965 0.052 0.990 0.196	0.944 0.264 5.766 1.145	0.898 0.383 10.776 2.140
60,6	**10 %**	**5σ**	**2**	0.921 0.469 16.536 1	0.970 0.060 1.396 0.084	0.970 0.050 1.315 0.079	0.978 0.047 0.883 0.053	0.979 0.047 0.883 0.053	0.976 0.055 0.978 0.059	0.927 0.565 21.252 1.285	0.697 0.344 25.046 1.515
40,2	**20 %**	**5σ**	**2**	0.874 0.489 14.615 1	0.960 0.036 1.147 0.078	0.959 0.035 1.157 0.079	0.964 0.060 1.003 0.069	0.965 0.062 0.994 0.069	0.962 0.062 1.084 0.074	0.883 0.389 16.382 1.121	0.786 0.497 24.023 1.644
60,6	**20 %**	**5σ**	**2**	0.860 0.631 57.000 1	0.964 0.056 1.328 0.023	0.966 0.049 1.307 0.023	0.960 0.072 1.114 0.019	0.964 0.064 1.037 0.018	0.965 0.077 1.169 0.021	0.896 0.667 47.263 0.829	0.510 0.409 36.904 0.647
Avr. prb				0.802 0.296 12.178 1	0.855 0.053 1.374 0.563	0.849 0.048 1.347 0.553	0.851 0.074 1.121 0.461	0.856 0.071 1.092 0.448	0.861 0.069 1.121 0.450	0.813 0.295 11.800 0.963	0.563 0.261 12.490 1.258

Table 3.4: Average Performance Measures of the Robust Regression Methods

	Scenario1				Scenario2				Scenario3			
Methods	PP	PO	IMSE	Ratio of IMSE	PP	PO	IMSE	Ratio of IMSE	PP	PO	IMSE	Ratio of IMSE
OLS	0.883	0.113	1.334	1	1	0.071	1.067	1	0.802	0.296	12.178	1
LTS	0.947	0.074	1.099	1.017	0.957	0.076	1.629	1.511	0.855	0.053	1.374	0.563
LMS	0.949	0.073	1.085	0.983	0.958	0.045	1.380	1.317	0.849	0.048	1.347	0.553
M	0.838	0.052	1.338	0.825	1	0.056	1.043	0.995	0.851	0.074	1.121	0.461
MM	0.908	0.047	1.298	0.814	1	0.058	0.981	0.936	0.856	0.071	1.092	0.448
LAD	0.956	0.064	1.058	0.796	1	0.054	0.993	0.959	0.861	0.069	1.121	0.450
LOWESS	0.799	0.119	1.128	0.841	0.948	0.157	1.419	1.235	0.813	0.295	11.800	0.963
MARS	0.852	0.125	1.318	0.986	0.635	0.083	1.095	1.001	0.563	0.261	12.490	1.258

The performances of the robust regression methods are tested using four performance measures which are detection capability (PP), false alarm rate (PO), improved mean square error (IMSE) and ratio of the IMSE. The average values of the robust regression methods' performance measures for the three simulation scenarios can be seen in Table 3.4.

As Table 3.4 demonstrates, the M-estimators which are M, MM and the LAD method seem to outperform the other methods according to many of the performance measures and scenarios. Multiple outlier detection scenarios are important for our study, so we should make our comments for each scenario respectively.

While comparing the seven robust regression methods, repeated ANOVA is used supported by SPSS 16 to test the null hypothesis that there is no statistical difference between the methods for each performance measure. Firstly, the assumptions of ANOVA for each measure are checked by residual plots. If the assumptions of ANOVA are not satisfied, logarithmic transformations of the performance measures are used. Different outlier locations are entered as subjects and robust methods are entered as treatments.

A repeated measures ANOVA is applied in this comparison study because there are 8 treatments (methods) and every treatment is to be used exactly once on each of the n objects (outlier locations). *Least Significant Difference (LSD)* adjustment with 0.05 confidence intervals is used to make pairwise comparisons. The Repeated ANOVA Tables are presented in Appendix A.

Comparison of the methods for Scenario 1: Randomly Scattered Regression Outliers in the Interior of X-space

The outliers have random levels of the independent variables with the same distribution as the clean observations but the dependent variable values are placed at $\boldsymbol{\delta_R}$ sigma away from the regression plane.

For each scenario twelf tables are constructed by SPPS. For each performance measure there are three tables. To test the null hypothesis that if there is no significant difference between the methods with respect to PP in Scenario-1, Tables A.1, A.2 and A.3 can be analyzed. As can be seen from Table A.1, the null hypothesis indicates that there is sphericity. So the sphericity assumption is assumed if the p-value is greater than 0.05. Since the sphericity is not satisfied for this case; the “Tests of Within-Subjects Effects” table, Greenhouse-Geisser p-value should be used from Table A.2. Since the p-value is smaller than 0.05, we can reject the null hypothesis which indicates the methods are equal. As a result, we can say that there is a significant difference between the methods with respect to PP.

To see which methods perform better, we should analyze the “Pairwise Comparisons” in Table A.3. To interpret the figure, the 95% confidence intervals should be checked. If the interval includes zero, then we can say that there is not a significant difference between these two methods with respect to the corresponding measure. To see which method’s performance is better, we should use the mean difference column. For the detection capability (PP) measure, higher values are preferred because this shows that this method’s outlier detection capacity is larger. For example, for the first row OLS (pp)-M (pp) < 0. This indicates that mean detection capability of OLS method is significantly smaller than the mean detection capability of M method. Therefore; we can say that M estimators outperform OLS estimators with respect to PP performance measure. The rest of the table can be interpreted in the same way. Table 3.5 demonstrates the competitive relationships according to Table A.3. While comparing the methods pairwisely, (<) means the method in the column outperforms the corresponding method in the row. The empty cells indicate that there is no significant difference between the method in the row and its corresponding method in the column.

From Tables A.4, A.5 and A.6, the null hypothesis that there is no significant difference between the methods with respect to PO can be tested for Scenario-1. As can be seen from Table A.4, the p-value is not greater than 0.05, i.e., sphericity is not satisfied so the "Tests of Within-Subjects Effects", Greenhouse-Geisser p-value should be used from Table A.5. Since the p-value is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal. As a result, we can say that there is a significant difference between the methods with respect to PO. To see which methods perform better, we should analyze the "Pairwise Comparisons" in Table A.6. These comparisons are given in Table 3.5.

Tables A.7, A.8 and A.9 are used to test if there is a significant difference among the methods with respect to IMSE for Scenario-1. As it can be seen from Table A.7 the p-value is not greater than 0.05, hence the sphericity assumption is not satisfied. Therefore the Greenhouse-Geisser p-value in Table A.8 should be used. Since it is smaller than 0.05, we can reject the null hypothesis that the methods are equal. Therefore it can be said that there is a significant difference between the methods with respect to IMSE. To see which methods perform better, we should analyze the "Pairwise Comparisons" in Table A.9. These comparisons are demonstrated in Table 3.5.

Tables A.27, A.28 and A.29 are used to test if there is a significant difference among the methods with respect to the ratio of IMSE for Scenario-1. As it can be seen from Table A.27 the p-value is not greater than 0.05, hence the sphericity assumption is not satisfied. Therefore the Greenhouse-Geisser p-value in Table A.28 should be used. Since it is smaller than 0.05, we can reject the null hypothesis that the methods are equal. Therefore it can be said that there is a significant difference between the methods with respect to IMSE. To see which methods perform better, we should analyze the "Pairwise Comparisons" in Table A.29. These comparisons are demonstrated in Table 3.5.

For the detection capability (PP); the LAD method outperforms the methods OLS, M, LTS, LMS, LOWESS and MARS. LOWESS and MARS nonparametric methods are outperformed by OLS, M, MM, LMS and LAD. As a result; LAD is a desirable method when the performance measure is detection capability.

For the false alarm rate (PO); OLS is outperformed by M, MM, LTS, LMS, LAD and MARS methods. LTS and LMS methods show stronger performance than M and MM. LTS is outperformed by LMS. LAD is superior to OLS and MM, but not as competitive as the methods LTS and LMS. LOWESS and MARS nonparametric methods are outperformed by M, MM, LTS, LMS and LAD. As a result; LTS and LMS are superior when the performance measure is false alarm rate.

For the improved mean squared error (IMSE); OLS is outperformed by M, MM, LAD and LOWESS. M and MM are superior to LTS, LMS and MARS. LAD outperforms LTS and LTS outperforms LMS. For the ratio of IMSE, the comments are same as the IMSE. As a result; LAD, M and MM are desirable methods when the performance measure is detection capability.

For scenario-1, we can say that LAD is competitive among the other methods.

Table 3.5: Pairwise Comparisons of the Robust Regression Methods for PP, PO, IMSE and ratio of IMSE for Scenario-1

Measure	**Methods**	**M**	**MM**	**LTS**	**LMS**	**LAD**	**LWS**	**MARS**
PP	OLS					< (.039)	> (.000)	> (.000)
	M					< (.044)	> (.000)	> (.000)
	MM						> (.000)	> (.000)
	LTS					< (.039)		
	LMS					< (.000)	> (.001)	> (.004)
	LAD						> (.000)	> (.000)
	LWS							
PO	OLS	< (.000)	< (.000)	< (.000)	< (.000)	< (.000)		< (.000)
	M			< (.013)	< (.003)		> (.000)	> (.000)
	MM			< (.008)	< (.001)	< (.046)	> (.000)	> (.000)
	LTS				< (.001)	> (.012)	> (.000)	> (.000)
	LMS					> (.001)	> (.000)	> (.000)
	LAD						> (.000)	> (.000)
	LWS							> (.031)

Table 3.5(cont'd): Pairwise Comparisons of the Robust Regression Methods for PP, PO, IMSE and ratio of IMSE for Scenario-1

Measure	**Methods**	**M**	**MM**	**LTS**	**LMS**	**LAD**	**LWS**	**MARS**
IMSE	OLS	< (.000)	< (.001)			< (.001)	< (.000)	
	M			> (.009)	> (.022)			> (.002)
	MM			> (.003)	> (.006)			> (.002)
	LTS				< (.002)	< (.000)		
	LMS					< (.000)		
	LAD							> (.001)
	LWS							> (.000)
Ratio of the IMSE	M			> (.007)	> (.014)			> (.001)
	MM			> (.003)	> (.006)			> (.002)
	LTS				< (.001)	< (.000)		
	LMS					< (.000)		
	LAD							> (.001)
	LWS							> (.000)

Scenario2: Regression Outliers in Multiple Point Clouds at the Centroid of X-space

This scenario compares the performance of the robust regression procedures when there are multiple Y-space outliers forming clouds at the centroid of X-space.

Table A.10 illustrates that the p-value is smaller than 0.05, so we cannot assume sphericity. Since the sphericity is not satisfied; Greenhouse-Geisser p-value is used from Table A.11 to determine if there exists significant difference between the methods with respect to PP for scenario2. Since it is smaller than 0.05, we can say that there is a significant difference between the methods. To see which methods outperform the others, Table A.12 is analyzed and Table 3.6 is demonstrated.

From Tables A.13 and A.14; the hypothesis that if there is a significant difference between the methods with respect to PO can be tested for Scenario-2. As can be seen from Table A.13, since the p-value is not greater than 0.05, sphericity is not satisfied for this case; the "Tests of Within-Subjects Effects", Greenhouse-Geisser p-value should be used from Table A.14. Since the p-value is not smaller than 0.05, we cannot reject the null hypothesis which indicates that the methods are equal with respect to PO. So part 2 of Table 3.6. cannot be used.

Tables A.15, A.16 and A.17 are used to test the hypothesis that if there is a significant difference between the methods with respect to IMSE for Scenario-2. As can be seen from figure A.15, since the p-value is not greater than 0.05, sphericity is not satisfied but the Greenhouse-Geisser p-value in Table A.16 is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal. Therefore; it can be said that there is a significant difference between the methods with respect to IMSE. To see which methods perform better, we should analyze the "Pairwise Comparisons" in Table A.17. These comparisons are demonstrated in Table 3.6.

Tables A.30, A.31 and A.32 are used to test the hypothesis that if there is a significant difference between the methods with respect to ratio of IMSE for Scenario-2. As can be seen from figure A.30, since the p-value is not greater than 0.05, sphericity is not satisfied but the Greenhouse-Geisser p-value in Table A.31 is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal. Therefore; it can be said that there is a significant difference between the methods with respect to the ratio of IMSE. To see which methods perform better, we should analyze the "Pairwise Comparisons" in Table A.32. These comparisons are demonstrated in Table 3.6.

Our results have shown that there is no significant difference between the methods for the PO measure but for the other three performance measures; we can analyze the table to see which methods outperform the others.

For the PP; OLS is superior to the methods LTS, LMS and MARS. LTS and LMS is outperformed by the methods M, MM and LAD. MARS is not as competitive as the other methods.

For the IMSE; OLS shows stronger performance than LTS and LMS but shows poorer performance than MM. LTS and LMS is outperformed by M, MM, LAD and MARS. The comments are the same for the ratio of the IMSE.

As a result; M, MM and LAD methods perform well under scenario-2.

Scenario3: Regression Outliers in Multiple Point Clouds: Independent Variables Randomly Scattered on the Interior of X-Space

In this case, the multiple outlier clouds are not placed at the centroid but at different locations in X-space for each replication.

Table 3.6: Pairwise Comparisons of the Robust Regression Methods for PP, PO, IMSE and ratio of IMSE for Scenario2

Measure	**Methods**	**LTS**	**LMS**	**M**	**MM**	**LAD**	**LWS**	**MARS**
PP	OLS	> (.000)	> (.001)					> (.000)
	LTS			< (.000)	< (.000)	< (.000)	< (.047)	> (.015)
	LMS			< (.001)	< (.001)	< (.001)		> (.019)
	M							> (.000)
	MM							> (.000)
	LAD							> (.000)
	LWS							> (.000)
PO	OLS							
	LTS							
	LMS							
	M							
	MM							
	LAD							
	LWS							

Table 3.6 (cont'd): Pairwise Comparisons of the Robust Regression Methods for PP, PO, IMSE and ratio of IMSE for Scenario2

Measure	**Methods**	**LTS**	**LMS**	**M**	**MM**	**LAD**	**LOWESS**	**MARS**
IMSE	OLS	> (.002)	> (.002)		< (.014)			
	LTS			< (.002)	< (.000)	< (.001)	< (.043)	< (.009)
	LMS			< (.001)	< (.000)	< (.000)		< (.030)
	M							
	MM							
	LAD							
	LWS							
Ratio of the IMSE	LTS			< (.030)	< (.013)	< (.020)		< (.025)
	LMS			< (.006)	< (.000)	< (.001)		< (.016)
	M	> (.030)	> (.006)					
	MM							
	LAD							
	LWS							

To test the null hypothesis which indicates the methods are equal with respect to PP for Scenario3, Table A.18 can be analyzed. The sphericity assumption is not satisfied. We should look at the “Tests of Within-Subjects Effects” in Table A.19. Since the Greenhouse-Geisser p-value is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal with respect to PP. To see which methods perform better, we should analyze the “Pairwise Comparisons” in Table A.20. These results are demonstrated in Table 3.7.

From Tables A.21, A.22 and A.23 the null hypothesis that there is no significant difference between the methods with respect to PO can be tested for Scenario-3. As can be seen from Table A.21 the p-value is not greater than 0.05, sphericity is not satisfied for this case; the “Tests of Within-Subjects Effects”, Greenhouse-Geisser p-value should be used from Table A.22. Since the p-value is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal with respect to PO. As a result, we can say that there is a significant difference between the methods with respect to PO. To see which methods perform better, we should analyze the “Pairwise Comparisons” in Table A.23. These comparisons are given in Table 3.7.

Tables A.24, A.25 and A.26 are used to test the hypothesis that if there is a siginificant difference between the methods with respect to IMSE for Scenario-3. As can be seen from Table A.24, since the p-value is not greater than 0.05, sphericity is not satisfied but the Greenhouse-Geisser p-value in Table A.25 is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal. Therefore; it can be said that there is a significant difference between the methods with respest to IMSE. To see which methods perform better, we should analyze the “Pairwise Comparisons” in Table A.26. These comparisons are demonstrated in Table 3.7.

Tables A.33, A.34 and A.35 are used to test the hypothesis that if there is a siginificant difference between the methods with respect to IMSE for Scenario-

3. As can be seen from Table A.33, since the p-value is not greater than 0.05, sphericity is not satisfied but the Greenhouse-Geisser p-value in Table A.34 is smaller than 0.05, we can reject the null hypothesis which indicates that the methods are equal. Therefore; it can be said that there is a significant difference between the methods with respest to IMSE. To see which methods perform better, we should analyze the “Pairwise Comparisons” in Table A.35. These comparisons are demonstrated in Table 3.7.

Table 3.7 results have shown that; for the PP; OLS is outperformed by LTS, LMS, M, MM and LAD but is superior to MARS. LTS shows stronger performance than LMS and MM shows stronger performance than M. LOWESS and MARS are outperformed by OLS and other robust methods.

For the PO; OLS is outperformed by LTS, LMS, M, MM and LAD. LTS and LMS are superior to M, MM, LAD, LOWESS and MARS. LMS shows stronger performance than LTS. LOWESS and MARS are outperformed by OLS and other robust methods.

For the IMSE; OLS is outperformed by LTS, LMS, M, MM and LAD. LTS and LMS show weaker performance than M, MM and LAD. MM is superior to LAD. LOWESS and MARS are outperformed by OLS and other robust methods.The comments for the ratio of the IMSE are the same as IMSE’s.

As a result; LTS, MM and LAD methods perform well ınder scenario-3.

Table 3.7: Pairwise Comparisons of the Robust Regression Methods for PP, PO, IMSE and ratio of IMSE for Scenario3

Measure	**Methods**	**LTS**	**LMS**	**M**	**MM**	**LAD**	**LWS**	**MARS**
PP	OLS	< (.000)	< (.000)	< (.000)	< (.000)	< (.000)		> (.000)
	LTS		> (.010)				> (.004)	> (.000)
	LMS						> (.006)	> (.000)
	M				< (.011)		> (.002)	> (.000)
	MM						> (.001)	> (.000)
	LAD						> (.001)	> (.000)
	LOWESS							> (.000)
PO	OLS	< (.000)	< (.000)	< (.000)	< .000)	< (.000)		
	LTS		< (.000)	> (.002)	> (.001)	> (.000)	> (.000)	> (.000)
	LMS			> (.000)	> (.000)	> (.000)	> (.000)	> (.000)
	M						> (.000)	> (.000)
	MM						> (.000)	> (.000)
	LAD						> (.000)	> (.000)
	LWS							

Table 3.7(cont'd): Pairwise Comparisons of the Robust Regression Methods for PP, PO, IMSE and ratio of IMSE for Scenario3

Measure	**Methods**	**LTS**	**LMS**	**M**	**MM**	**LAD**	**LOWESS**	**MARS**
IMSE	OLS	< (.000)	< (.000)	< (.000)	< (.000)	< (.000)		
	LTS		< (.002)	< (.000)	< (.000)	< (.000)	> (.001)	> (.000)
	LMS			< (.000)	< (.000)	< (.000)	> (.001)	> (.000)
	M						>(.000)	> (.000)
	MM					>(.000)	>(.000)	> (.000)
	LAD						>(.000)	> (.000)
	LWS							> (.002)
Ratio of the IMSE	LTS	< (.000)	< (.017)	< (.014)	< (.004)	< (.002)	> (.006)	> (.000)
	LMS			< (.018)	< (.005)	< (.002)	> (.004)	> (.000)
	M						> (.000)	> (.000)
	MM					> (.000)	> (.000)	> (.000)
	LAD						> (.000)	> (.000)
	LWS							> (.002)

3.3. Discussion

For Scenario-1 if Tables 3.1 and 3.5 are analyzed, the LAD method's PP stands out among the other methods. Although OLS has good detection capability, it is unsatisfactory because it swamps clean observations indicated by the high false alarm rate. The M estimator does not perform well with respect to the PO in high density outlier cases as expected because its optimal distribution is close to normal. When the outlier density increases, the tails become heavier and the distribution becomes far from normal. The MM method gives exactly the same false alarm rates with the M method. Although the LTS and LMS methods' detection capability is low at 3σ outlier locations, they can be preferred for 4σ and over outlier locations because of their low false alarm rate and competitive detection capabilities. The order of the performances is as same as Wisnowski (1999). The nonparametric methods LOWESS and MARS cannot be preferred because of their low detection capabilities and high false alarm rates. The LAD estimators' IMSE and ratio of the IMSE are the lowest among other methods. At 3σ outlier locations, LOWESS is a better choice but at 4σ and beyond LAD is preferable as expected.

For Scenario-2 if Tables 3.2 and 3.6 are analyzed, again OLS, M, MM and LAD are superior in detection capability. As indicated by Wisnowski (1999), OLS's false alarm rate is high in the single cloud outlier locations because of the degradation in parameter estimates such that the clean data are no longer fit well. However, when there are two clouds, there is no swamping because there is an equal and opposite pull on the regression surface from each cloud that leaves the parameter estimates unchanged from those obtained with clean observations only. With respect to the improved mean square error and ratio of the IMSE, we can order the methods by increasing performance as MM, M, LAD> OLS > LMS, LTS> LOWESS, MARS. M and MM have nearly the same detection capability, false alarm rate, IMSE and ratio of IMSE.

Scenario-3 is the most challenging one among the three scenarios. OLS, LTS and LMS have the highest detection capabilities. The OLS does not fit the outlying clouds as evidenced by the high detection capability; however they do chase these observations enough to swamp some clean observations as stated by Wisnowski (1999). Moreover, improved mean square error of the OLS method is significantly high when compared with M, MM, LAD, LTS and LMS methods under “two clouds, 60 observations, 6 independent variables” outlier location scenarios. The IMSE and the ratio of the IMSE of the nonparametric methods are even larger than OLS and they also have high false alarm rates. Therefore, these methods are not preferable for this scenario. The M and MM estimators have higher detection capabilities and significantly better false alarm rates than OLS, especially for the high-dimension, high density and also two cloud scenarios. The LTS and LMS estimators are very outstanding methods with high detection capability, low false alarm rates and competitive improved mean square errors.

CHAPTER 4

COMPARISON OF OLS, M-REGRESSION, LAD AND LTS METHODS ON AN INDUSTRIAL DATA SET

In this chapter, OLS and three most promising methods, which are M, LAD and LTS, from the simulation study are compared with respect to the performance measures, coefficient of multiple determination and mean square error. These methods are performed on an industrial data set. Firstly, the data set is analyzed to see whether it is suitable for conducting robust methods. Residual plots are checked to see if there are outliers in the data and what type of outliers are they. Moreover, we check whether the data needs any transformation to validate the close to normality and constant variance assumption. Cross-validation approach is used to compare the methods performance and to see if there is a significant difference between the methods, Repeated ANOVA method is used. The results of the study are discussed at the end.

4.1. Description of the Data Set

Our data is taken from a real life manufacturing process which includes the sub-processes core, molding, melting, casting, fettling and painting, which was studied by Bakır (2007). The dependent variable is the percentage of defectives on a cylinder head. Without conducting any specific data analysis, the company records values of certain parameters hourly, daily or weekly to monitor the production processes. The requirement of the company is to determine the most influential parameters that cause defects on the last product.

Regression method is the most commonly used technique for determining the relationship between a dependent variable and a number of independent variables. Regression method is an easy to understand and interpret. In this study, values of independent variables are determined by sampling; therefore every individual item is associated with the average values of the batch that the item belongs to. There are some missing values in the data since the company did not record them. These values are eliminated by the proper methods and our comparison study is conducted without missing values. The basic data set includes 36 independent variables and 92 observations.

4.2. Cross-validation Approach, Applications and Performance Analysis

High dimension of the independent variables causes some shortcomings like the collinearity between variables and increase in calculation time; thus a model selection procedure is decided to be used. However, the variable selection procedures for robust regression are very limited and they are proposed by only specific robust regression methods. Hence, stepwise model selection procedure of multiple linear regression in Minitab 14 is used. For the implementation details of the selection procedure, Appendix B can be referred.

First the original data, with 92 observations and 36 independent variables, is analyzed by multiple linear regression to see whether there are outliers and the normality assumption is valid. Figure 4.1 demonstrates that the normality assumption hypothesis is rejected with the Anderson Darling p-value which is smaller than 0.005.

From Figure 4.2 the boxplot of the response value indicates that there are outliers in the data. Therefore, our data is suitable for conducting robust regression methods but we should also check that if we delete the outliers do the rest of the data distributed normally. Since in our study, we are interested in the data sets that are "close" to normal.

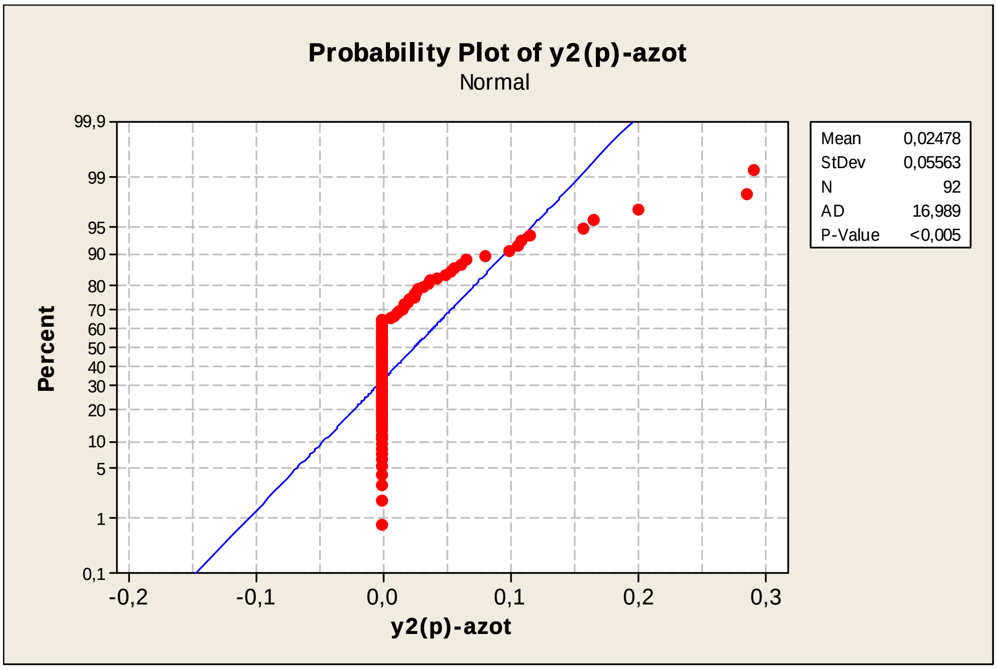

Figure 4.1: Anderson Darling Test and Normal Probability Plot for the Original Data with 92 Observations without Making any Transformation

The model selection was done by stepwise procedure using Minitab 14 with default values which are alpha to enter (0.15) and alpha to remove (0.15). The model selection output is presented in Appendix B.1. Eight significant variables are determined and observations 16, 17, 45, 49, 52, 71, 78, 88 are pointed as unusual observations which inavalidate the normality assumption. Moreover, if the residual plots in Figure B.1 are analyzed, both the normality and constant variance assumptions are not satisfied.

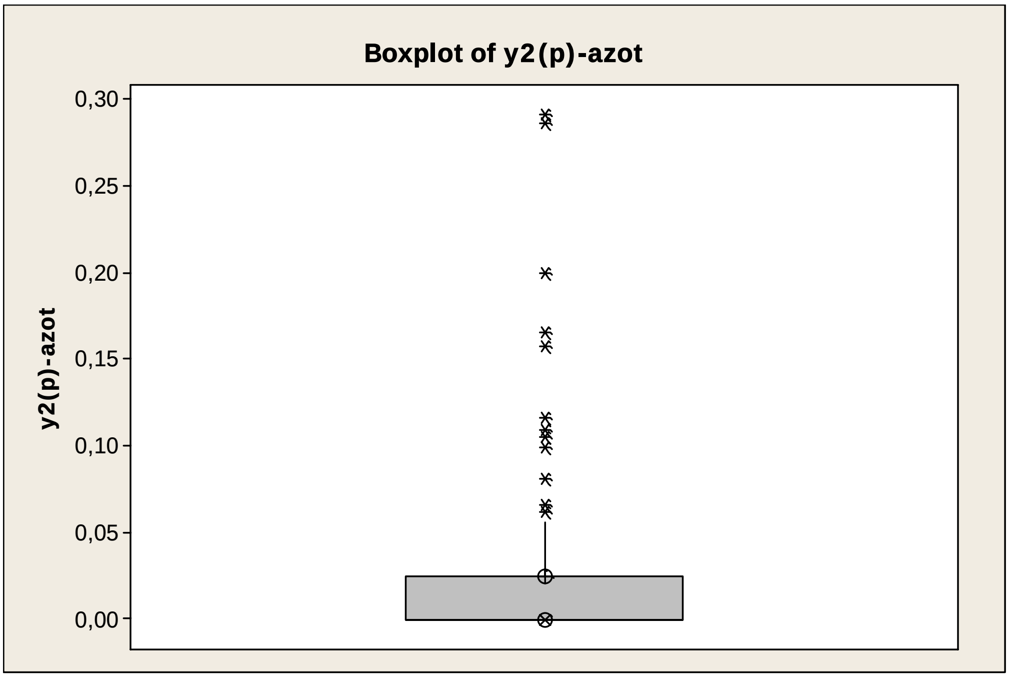

Figure 4.2: Boxplot of the Original Data with 92 Observations

Second, the original data with deleting the outliers without making any transformation is analyzed by multiple linear regression to see whether there are the normality assumption becomes valid. Figure 4.3 demonstrates that the normality assumption hypothesis is rejected with the Anderson Darling p-value which is smaller than 0.005. This means that although the unusual values are deleted, the remaining data of 84 observations still do not satisfy the normality assumption.

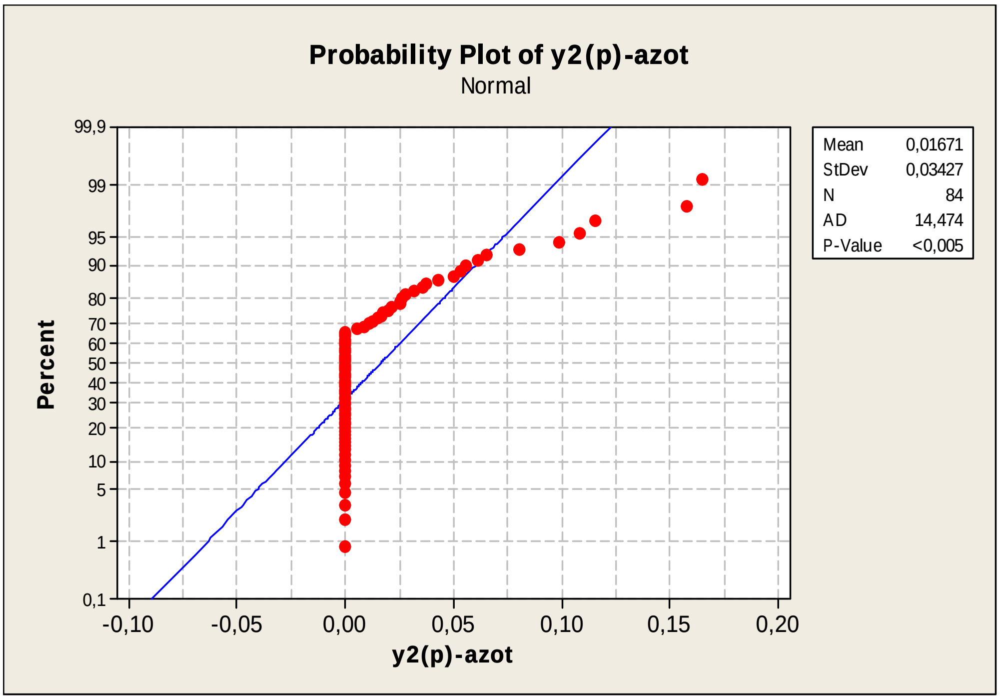

Figure 4.3: Anderson Darling Test and Normal Probability Plot for the Original Data by Deleting Possible Outliers

Also, we can see from Figure 4.4 there still seems to be unusual observations although we have deleted 10% of the original data.

Moreover Figure B.3 demonstrates that the constant variance assumption is not satisfied either. The variance of the data seems to be increasing as the response values increase. Since our data is a percentage data, logit transformation is suitable.

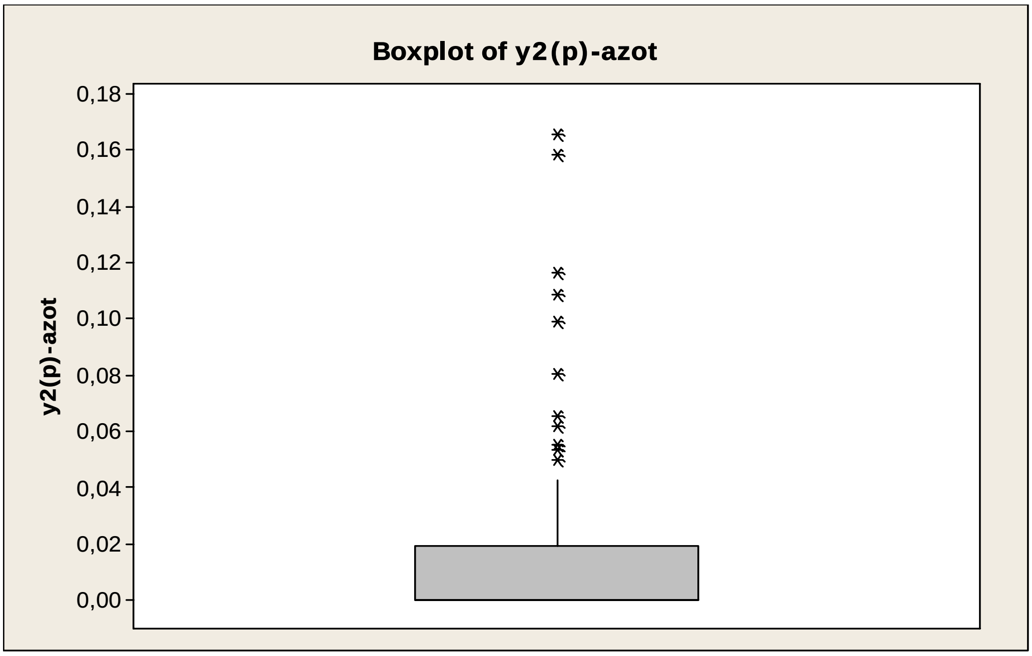

Figure 4.4: Boxplot of the Original Data by Deleting Possible Outliers

Logit transformation is applied to the original data with 92 observations and 36 independent variables. The transformed data is analyzed by multiple linear regression to see whether there are outliers and the normality assumption is valid. Figure 4.5 demonstrates that the normality assumption hypothesis is rejected with the Anderson Darling p-value which is smaller than 0.005. Boxplot of the transformed response value is demonstrated in Figure. 4.6.

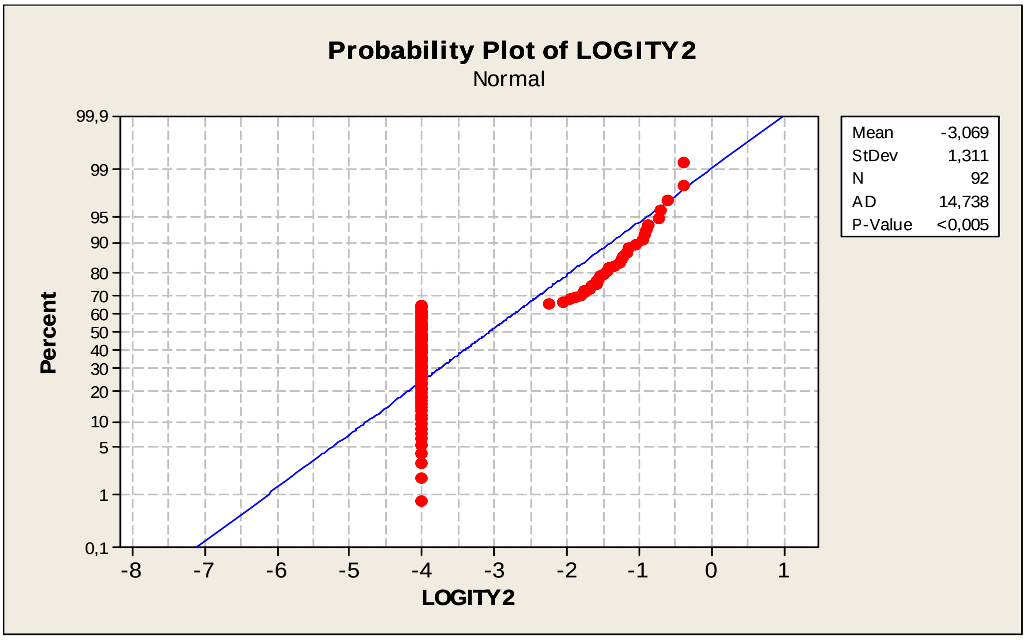

Figure 4.5: Anderson Darling Test and Normal Probability Plot for the Logit Transformed Data with 92 Observations

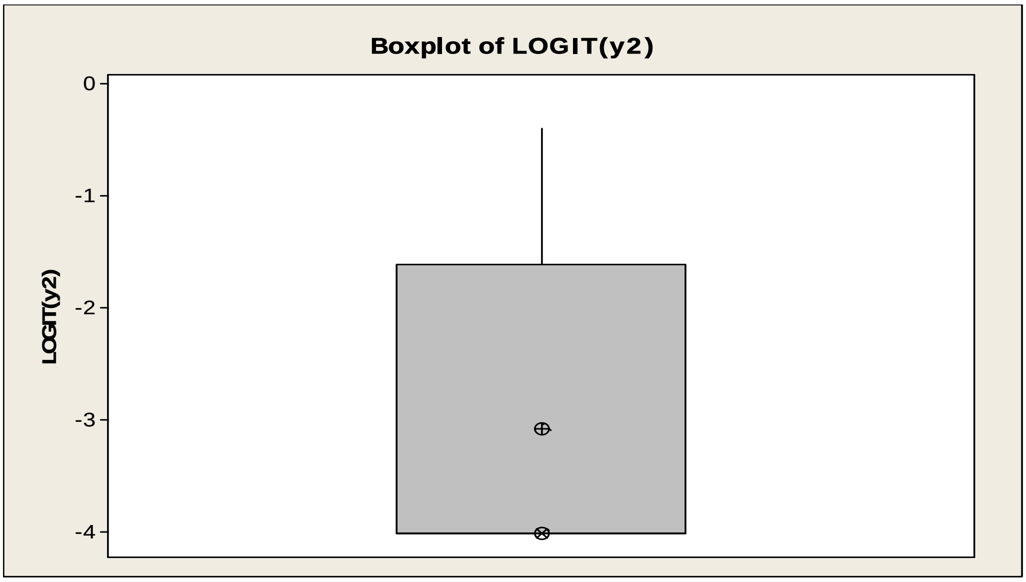

Figure 4.6: Boxplot of the Logit Transformed Data with 92 Observations

The model selection was done by stepwise regression procedure. The model selection output is presented in Appendix B.2. Four significant variables x9, x19, x22 and x28 are determined and observations 16, 21, 22, 70, 71, 72, 77, 78 are pointed as unusual observations which contaminate the normality assumption.

To adjust the nonconstant error variance logit transformation will be used because our dependent variable is a defective rate which is between 0 and 1. After the transformation; the residuals versus the fitted values plot in Figure B.2 the funnel type shape becomes less visible. Robust regression is suitable for this data because there are unusual observations and one of the basic advantages of robust regression is that it can be used with contaminated normal distributions.

A 3-fold and 3-replicate cross validation approach is used to compare the robust regression methods' performance. The model selection was conducted by using the train data sets. Then prediction performance measures which are coefficient of multiple determination and mean square error are calculated for each of the nine test data sets.

Coefficient of multiple determination R2 is defined as

$$R^2 = \frac{SS_R}{S_{yy}} = 1 - \frac{SS_E}{S_{yy}} \ .$$

It is a measure of the reduction in the variability of y obtained by using the independent variables in the model (Montgomery and Peck, 1991).

The residual mean square error is

$$MS_E = \frac{SS_E}{n-p} \ .$$

It is an unbiased estimator of σ^2. A model with a small residual mean square is usually preferred to a model with a small one (Montgomery and Peck, 1991).

Repeated ANOVA supported by SPSS 16 is used to see whether there is significant difference between the alternative methods. The Repeated ANOVA analysis is done with the average performance measures for each replication.

The average values of the OLS, M, LAD and LTS regression methods for the three replications can be seen in Table 4.1.

Table 4.1: Average Performance Measures of the Methods

ORDINARY LEAST SQUARES					
MSE	Rep1	1,653	**R^2**	Rep1	0,194
	Rep2	1,811		Rep2	0,219
	Rep3	1,637		Rep3	0,246
M					
MSE	Rep1	1,527	**R^2**	Rep1	0,212
	Rep2	1,506		Rep2	0,346
	Rep3	1,728		Rep3	0,218
LAD					
MSE	Rep1	1,576	**R^2**	Rep1	0,185
	Rep2	1,094		Rep2	0
	Rep3	1,528		Rep3	0,305
LTS					
MSE	Rep1	1,653	**R^2**	Rep1	0,141
	Rep2	1,890		Rep2	0,175
	Rep3	1.689		Rep3	0

As Table 4.1 is analyzed, there seems not much difference between the methods and M-regression seems to outperform the other methods.

To check the null hypothesis that there is no statistical difference between the methods with respect to MSE; Tables C.1 and C.2 are analyzed. As can be seen from Table C.1, the p-value is not greater than 0.05; thus sphericity is not satisfied for this case; the “Tests of Within-Subjects Effects”, Greenhouse-Geisser p-value should be used from Table C.2. Since the p-value is not smaller than 0.05, we can not reject the null hypothesis which indicates that the methods’ performances are equal with respect to MSE.

From Tables C.3 and C.4, the null hypothesis that there is no statistical difference between the methods with respect to R^2 can be analyzed. Table C.3 demonstrates that the p-value is not greater than 0.05, thus sphericity is not satisfied; the “Tests of Within-Subjects Effects”, Greenhouse-Geisser p-value should be used from Table C.4. Since the p-value is not smaller than 0.05, we can not reject the null hypothesis which indicates that the methods’ performances are equal with respect to R^2.

4.3. Discussion

Performance analysis study of OLS, M-regression, LAD and LTS indicates that the robust methods do not give better performance results than the classical OLS method. This demonstrates that our industrial data exhibit different behavior from the generated distributions and does not support the simulation results.

These results support the studies of Stigler (1977) and Hill and Dixon (1982) which are about robustness in real life. Stigler (1977) stated that most robustness studies have relied upon mathematical theory and computer simulated data. However, no matter how experienced the researcher in his choice of sampling distributions, there is no guarantee that the samples he generates are representative of real data. The real data can be correlated, biased, asymmetric or heterogeneous. He concluded that real data exhibit different behavior from the simulated data used in most robustness studies and this affects the recommendations for the choice of an estimator and relative performances of the estimators. His data sets have more extreme values than one would expect from normal samples. He found that light trimming provides some improvement. Hill and Dixon (1982) conduct a real life study on biomedical data by first transforming their data by a logarithmic function. The biomedical distributions can be asymmetric and have shorter tails and other anomalies. They recommend 15% trim is a 'safe' estimator to use when little is known about the underlying distribution. Both of these real life studies recommend using light trimming which means deleting the extreme values at the specified trimming ratio. In fact, this exactly corresponds to the classical way which is first deleting outliers and then fitting the OLS to the data without outliers.

For our study, one of the reasons of this result can be the complexity of the data. OLS and the three robust methods are not resistant to complexity of the data; that is they only assume linear relationships between the independent variables and the response variable. However, our data may require nonlinear or nonparametric relations. Therefore both the OLS and the robust regression methods cannot explain the data well. This can be proved by the low R^2 values in Table 4.1. The second reason can be the irrelevant variables in the data set. In fact we have reduced the number of the independent variables by using the stepwise regression procedure but the chosen variables still may not be relevant with the response variable. Maybe there are more relevant parameters that affect the process but are not noticed along the data collection.

Moreover, even if there is no significant difference between OLS and the three other robust methods; the usage of OLS is not appropriate in this data set because its basic assumption which is normality is not satisfied. That is we cannot say that the results are the same, so it is advisable to use OLS. The 'safer' way is to use the robust methods.

CHAPTER 5

CONCLUSIONS AND SUGGESTIONS FOR FUTURE STUDIES

This study brings three contributions to the field of robust regression.Firstly; the seven most promising robust regression methods which have not been compared together in literature are compared. In fact two of them, MARS and LOWESS, are classified in nonparametric regression methods in the literature but only their outlier detection capacities are in the scope of this study. A Monte Carlo simulation study is conducted using the three challenging scenarios for multiple outlier detection. As a result of the comparisons M, MM, LAD and LTS are the most promising robust methods among the others with respect to the performance measures PP, PO and IMSE.

Secondly, an improved performance measure, which is the mean square error of the robust regression lines without outliers, is developed. The idea comes from the relative efficiency of a robust regression estimator which is defined by the ratio of the MSE of the robust estimator over MSE of the OLS estimator. However in this ratio the MSE of the robust estimator is evaluated by using all residuals even the residuals of the outliers are very large in robust regression whereas the MSE of OLS is calculated by using only the normal data without outliers. In this idea, we behave as if we are penalizing the robust methods for giving large residuals to outliers. But as far as known, the fundamental aim of a robust method is fitting a regression line that is not so much pulled by outliers. Therefore, we fit the OLS and the robust methods' regression lines to the contaminated data but omit the residuals of the outliers while calculating the MSE. By this way, we have compared the methods at equal conditions.

Thirdly, since there is very limited number of real life applications on robust regression methods; we conducted a real life data analysis. An industrial data set is used to compare the robust methods LAD, M, LTS with OLS. As a result, for our real life data we see that there is no significant difference between the robust methods and the classical OLS method. We have explained this situation by complexity of the data and irrelevant variables. Moreover, even if the results of the OLS and the robust regression methods are the same; the model fitted by OLS is not valid because it is not applied with normality assumption satisfied. As a result, robust methods are more appropriate to deal with outliers even if their performances are the same with the classical methods since they do not have such strict assumptions.

As a future study alternative outlier placement scenarios, for example high magnitude and high contamination, can be performed based on the design parameters determined. In addition, we cannot compare some of the robust regression methods such as multiple stage general models because of unavailability of software. These methods can be coded, or their codes can be obtained and compared with the ones in our study. Furthermore, the variable selection procedures for robust regression are very limited and they are proposed by only specific robust regression methods. These procedures can be extended and proposed for all the robust methods we have mentioned.

REFERENCES

Anderson, C. (2001). A comparison of five regression methods with ordinary least squares: Relative effiency, bias and test of the null hypothesis. (Doctoral dissertation, University of North Texas, 2001).

Bakır, B. (2007). “Defect Cause Modeling with Decision Tree And Regression Analysis: A Case Study in Casting Industry” M.S Thesis, Middle East Technical University, Ankara, Turkey.

Barnett, V. and Lewis, T. (1994). *Outliers in Statistical Data*. 3rd edition, Wiley: Great Britain.

Barrodale, I. and Roberts, F.D.K. (1974). Algorithm 478: Solution of an overdetermined system of equations in the L1 norm. Communications of the Association for Computing Machinery, vol. 17, pp. 319-320.

Birkes, D. and Dodge, Y. (1993). *Alternative Methods of Regression.* New York, NY:John Wiley & Sons, Inc.

Box, G.E.P. and Tiao, G.C. (1962). A further look at robustness via bayes theorem. *Biometrika*, no. 49, pp. 419-432.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*. 2nd edition. New York, NY: Wiley.

Fox, J. (2002). *An R and S-PLUS Companion to Applied Regression.* Thousand Oaks: Sage Publications.

Friedman, J.H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics*, vol. 19, no. 1, pp. 1-67.

Hadi, A.S. and Simonoff, J.S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association,* vol. 88, no. 424, pp. 1264-1272.

Hastie, T., Tibshirani, R. and Friedman. J. (2001) *The Elements of Statistical Learning.* Canada: Springer-Verlang New York, Inc.

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

Hampel, F. R. (1971). A general qualitative definition of robustness. *Annals of Mathematical Statistics*, vol. 42, no. 6, pp. 1887-1896.

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, vol. 69, no. 69, pp. 383-393.

Hill, M.A. and Dixon, W.J. (1982). Robustness in real life: A study of clinical laboratory data. *International Biometric Society*, vol. 38, no. 2, pp. 377-396.

Hogg, R.V. (1974). Adaptive robust procedures: A partial review and some suggestions for future applications and theory. *Journal of the American Statistical Association*, vol. 69, no. 348, pp. 909-923.

Hogg, R.V. (1979). Statistical robustness: One view of its use in applications today. *The American Statistician*, vol. 33, no. 3, pp. 108-115.

Huber, P. J. (1964). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35, pp. 73-101.

Huber, P. J. (1972). *The 1972 Wald lecture robust statistics: A review*. The Annals of Mathematical Statistics, vol. 43, no. 4, pp. 1041-1067.

Huber, P. J. (1972). Robust Regression: Asymtotics, conjectures and monte carlo. *The Annals of Statistics*, vol. 1, no. 5, pp. 799-821.

Huber, P. J. (1981). *Robust Statistics*. New York, NY: John Wiley & Sons, Inc.

Lane, K. (2002). *What is robust regression and how do you do it?* Paper presented at the Annual Meeting of the Southwest Educational Research Association.

Montgomery, D.C. and Peck, E.A. (1991). Introduction to Linear Regression Analysis. New York, NY: John Wiley & Sons, Inc, 2nd edition.

Nevitt, J and Tam, H.P. (1997). *A comparison of robust and nonparametric estimators under the simple linear regression model*. Paper presented at the Annual Meeting of the American Educational Research Association.

Portney, S. and He, X. (2000). A robust journey in the new millenium. *Journal of the American Statistical Association*, vol. 95, no. 452, pp. 1331-1335.

Portney, S. and Welsh, A. (1992). Exactly what is being modelled by the systematic component in a heteroscedastic linear regression. *Statistical Probability of Letters*, 13, pp. 253-258.

Rocke, D. M. and Woodruff D. L. (1996). Identification of Outliers in Multivariate Data. Journal of the American Statistical Association, vol.91, no.435, pp. 1047-1061.

Rousseeuw, P. J., (1983). Multivariate estimation with high breakdown point. Paper presented at Pannonian Symposium on Mathematical Statistics and Probability, Bad Tatzmanndorf, Austria. Abstracts in IMS Bull. (1983), 12, p. 234, Appeared in (1985), *Mathematical Statistics and Applications*, vol. B, edited by W. Grossman, G. Pflug, I. Wincze, and W. Wertz, Reidel, Dordrecht, The Netherlands, pp. 283-297.

Rousseeuw, P. J., (1984). Least median of squares regression. *Journal of the American Statistical Association*. 79, pp. 871-880.

Rousseeuw, P. J. and Leroy, A. M. (1987). *Robust Regression and Outlier Detection.* New York, NY: John Wiley & Sons, Inc.

Rousseeuw, P. J. and Driessen, K.V. (2006). *Computing LTS regression for large data sets*. Data Mining and Knowledge Discovery, vol. 12, pp. 29-45.

Stigler, S.M. (1977). Do robust estimators work with real data? *The Annals of Statistics*, vol. 5, no. 6, pp. 1055-1098.

Takezawa, K. (2006). *Introduction to nonparametric regression.* New Jersey: John Wiley and Sons, Inc.

Tukey, J. (1960). A survey of sampling from contaminated distributions, *Contributions to Probability and Statistics*, ed 1. Olkin, Stanford, CA: Stanford University Press.

Venables, W. N. and Ripley, B. D. (1999). *Modern applied statistics with S-PLUS*, Third Edition. New York, NY: Springer-Verlag.

Wisnowski, J. W. (1999). Multiple outliers in linear regression: Advances in detection methods, robust estimation, and variable selection (Doctoral dissertation, Arizona State University, 1999).

Western, B. (1995). American Journal of Political Science. Concepts and suggestions for robust regression analysis, vol. 39, no. 3, pp. 786-817.

Wu, L.L. (1985). Robust m-estimaton of location and regression. Sociological Methodology, vol. 15, pp. 316-388.

Yohai, V., Stahel. W. A. and Zamar, R. H. (1991). A procedure for robust estimation and inference in linear regression. *Directions in Robust Statistics and Diagnostics*, Part 2, eds W.A. Stahel and S. W. Weisberg. New York: Springer- Verlag.

APPENDIX A

REPEATED ANOVA RESULTS FOR THE COMPARISON OF THE EIGHT REGRESSION METHODS

Table A.1: Test of Sphericity for Scenario-1 (PP)

Mauchly's Test of Sphericity[b]

Measure:PP

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	102,867	27	**,000**	,290	,356	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.2: Tests of Within-Subjects Effects of PP for Scenario-1

Tests of Within-Subjects Effects

Measure:PP

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	18,876	7	2,697	11,703	,000

Table A.2 (cont'd): Tests of Within-Subjects Effects of PP for Scenario-1

	Greenhouse-Geisser	18,876	2,029	9,306	11,703	**,000**
	Huynh-Feldt	18,876	2,490	7,580	11,703	,000
	Lower-bound	18,876	1,000	18,876	11,703	,006
Error(METHODS)	Sphericity Assumed	17,743	77	,230		
	Greenhouse-Geisser	17,743	22,314	,795		
	Huynh-Feldt	17,743	27,394	,648		
	Lower-bound	17,743	11,000	1,613		

Table A.3. Pairwise Comparisons of the Methods with respect to PP for Scenario-1

Pairwise Comparisons

Measure:PP

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
OLS	2	**,034**	,096	,732	**-,177**	**,244**
	3	**-,075**	,044	,121	**-,172**	**,023**
	4	**,514**	,333	,152	**-,220**	**1,247**
	5	**,216**	,137	,143	**-,085**	**,517**
	6	-,204*	,087	,039	-,396	-,012
	7	1,052*	,149	,000	,724	1,381
	8	,932*	,145	,000	,612	1,251

Table A.3(cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-1

M	1	**-,034**	,096	,732	**-,244**	**,177**
	3	**-,108**	,083	,217	**-,290**	**,074**
	4	**,480**	,356	,204	**-,303**	**1,263**
	5	**,182**	,176	,323	**-,205**	**,570**
	6	-,238*	,105	,044	-,469	-,007
	7	1,019*	,174	,000	,636	1,402
	8	,898*	,148	,000	,572	1,224
MM	1	**,075**	,044	,121	**-,023**	**,172**
	2	**,108**	,083	,217	**-,074**	**,290**
	4	**,588**	,329	,101	**-,136**	**1,312**
	5	**,291**	,133	,052	**-,003**	**,584**
	6	**-,130**	,073	,104	**-,291**	**,031**
	7	1,127*	,167	,000	,760	1,494
	8	1,006*	,160	,000	,655	1,358
LTS	1	**-,514**	,333	,152	**-1,247**	**,220**
	2	**-,480**	,356	,204	**-1,263**	**,303**
	3	**-,588**	,329	,101	**-1,312**	**,136**
	5	**-,297**	,281	,312	**-,916**	**,321**
	6	-,718*	,306	,039	-1,392	-,044
	7	**,539**	,259	,062	**-,032**	**1,110**
	8	**,418**	,281	,165	**-,200**	**1,037**

Table A.3(cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-1

LMS	1	**-,216**	,137	,143	**-,517**	**,085**
	2	**-,182**	,176	,323	**-,570**	**,205**
	3	**-,291**	,133	,052	**-,584**	**,003**
	4	**,297**	,281	,312	**-,321**	**,916**
	6	-,420*	,086	,000	-,610	-,230
	7	,836*	,193	,001	,412	1,261
	8	,716*	,198	,004	,279	1,152
LAD	1	,204*	,087	,039	,012	,396
	2	,238*	,105	,044	,007	,469
	3	**,130**	,073	,104	**-,031**	**,291**
	4	,718*	,306	,039	,044	1,392
	5	,420*	,086	,000	,230	,610
	7	1,257*	,182	,000	,856	1,658
	8	1,136*	,175	,000	,750	1,522
LOWESS	1	-1,052*	,149	,000	-1,381	-,724
	2	-1,019*	,174	,000	-1,402	-,636
	3	-1,127*	,167	,000	-1,494	-,760
	4	**-,539**	,259	,062	**-1,110**	**,032**
	5	-,836*	,193	,001	-1,261	-,412
	6	-1,257*	,182	,000	-1,658	-,856
	8	**-,121**	,064	,085	**-,261**	**,019**

Table A.3(cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-1

MARS	1	-,932*	,145	,000	-1,251	-,612
	2	-,898*	,148	,000	-1,224	-,572
	3	-1,006*	,160	,000	-1,358	-,655
	4	**-,418**	,281	,165	**-1,037**	**,200**
	5	-,716*	,198	,004	-1,152	-,279
	6	-1,136*	,175	,000	-1,522	-,750
	7	**,121**	,064	,085	**-,019**	**,261**

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

*. The mean difference is significant at the ,05 level.

Table A.4: Test of Sphericity for Scenario-1 (PO)

Mauchly's Test of Sphericity[b]

Measure:PO

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	178,949	27	**,000**	,201	,221	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.5: Tests of Within-Subjects Effects of PO for Scenario-1

Tests of Within-Subjects Effects

Measure:PO

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	2,479	7	,354	28,025	,000
	Greenhouse-Geisser	2,479	1,405	1,764	28,025	**,000**
	Huynh-Feldt	2,479	1,549	1,600	28,025	,000
	Lower-bound	2,479	1,000	2,479	28,025	,000
Error(METHODS)	Sphericity Assumed	,973	77	,013		
	Greenhouse-Geisser	,973	15,460	,063		
	Huynh-Feldt	,973	17,042	,057		
	Lower-bound	,973	11,000	,088		

Table A.6. Pairwise Comparisons of the Methods with respect to P0 for Scenario-1

Pairwise Comparisons

Measure:PO

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
OLS	2	,194*	,033	,000	,121	,267
	3	,195*	,038	,000	,112	,277
	4	,349*	,068	,000	,200	,497
	5	,397*	,068	,000	,247	,547
	6	,247*	,044	,000	,149	,345
	7	**-,026**	,013	,076	**-,056**	**,003**
	8	-,051*	,006	,000	-,063	-,039
M	1	-,194*	,033	,000	-,267	-,121
	3	**,000**	,007	,967	**-,015**	**,016**
	4	,154*	,052	,013	,040	,269
	5	,203*	,053	,003	,087	,319
	6	**,053**	,028	,083	**-,008**	**,114**
	7	-,221*	,038	,000	-,304	-,137
	8	-,245*	,036	,000	-,324	-,166
MM	1	-,195*	,038	,000	-,277	-,112
	2	**,000**	,007	,967	**-,016**	**,015**
	4	,154*	,048	,008	,048	,260
	5	,203*	,048	,001	,097	,309
	6	**,053***	,024	,046	**,001**	**,104**
	7	-,221*	,041	,000	-,312	-,130
	8	-,245*	,040	,000	-,332	-,158

Table A.6(cont'd): Pairwise Comparisons of the Methods with respect to P0 for Scenario-1

LTS	1	-,349*	,068	,000	-,497	-,200
	2	-,154*	,052	,013	-,269	-,040
	3	-,154*	,048	,008	-,260	-,048
	5	,049*	,012	,001	,023	,074
	6	-,101*	,034	,012	-,175	-,028
	7	-,375*	,075	,000	-,540	-,210
	8	-,399*	,070	,000	-,552	-,246
LMS	1	-,397*	,068	,000	-,547	-,247
	2	-,203*	,053	,003	-,319	-,087
	3	-,203*	,048	,001	-,309	-,097
	4	-,049*	,012	,001	-,074	-,023
	6	-,150*	,032	,001	-,221	-,079
	7	-,424*	,074	,000	-,587	-,260
	8	-,448*	,069	,000	-,601	-,295
LAD	1	-,247*	,044	,000	-,345	-,149
	2	**-,053**	,028	,083	**-,114**	**,008**
	3	**-,053***	,024	,046	**-,104**	**,000**
	4	,101*	,034	,012	,028	,175
	5	,150*	,032	,001	,079	,221
	7	-,274*	,048	,000	-,379	-,168
	8	-,298*	,045	,000	-,398	-,198
LOWESS	1	**,026**	,013	,076	**-,003**	**,056**
	2	,221*	,038	,000	,137	,304
	3	,221*	,041	,000	,130	,312
	4	,375*	,075	,000	,210	,540
	5	,424*	,074	,000	,260	,587
	6	,274*	,048	,000	,168	,379
	8	-,024*	,010	,031	-,046	-,003

Table A.6(cont'd): Pairwise Comparisons of the Methods with respect to P0 for Scenario-1

MARS	1	,051*	,006	,000	,039	,063
	2	,245*	,036	,000	,166	,324
	3	,245*	,040	,000	,158	,332
	4	,399*	,070	,000	,246	,552
	5	,448*	,069	,000	,295	,601
	6	,298*	,045	,000	,198	,398
	7	,024*	,010	,031	,003	,046

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.7: Test of Sphericity for Scenario-1 (MSE)

Mauchly's Test of Sphericity[b]

Measure:MSE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a] Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	159,762	27	**,000**	,198	,217	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.8: Tests of Within-Subjects Effects of MSE for Scenario-1

Tests of Within-Subjects Effects

Measure:MSE

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	,169	7	,024	7,106	,000
	Greenhouse-Geisser	,169	1,384	,122	7,106	**,012**
	Huynh-Feldt	,169	1,518	,111	7,106	,009
	Lower-bound	,169	1,000	,169	7,106	,022
Error(METHODS)	Sphericity Assumed	,262	77	,003		
	Greenhouse-Geisser	,262	15,219	,017		
	Huynh-Feldt	,262	16,703	,016		
	Lower-bound	,262	11,000	,024		

Table A.9: Pairwise Comparisons of the Methods with respect to MSE for Scenario-1

Pairwise Comparisons

Measure:MSE

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
OLS	2	,088*	,018	,000	,048	,127
	3	,095*	,021	,001	,049	,141
	4	**,008**	,036	,824	**-,070**	**,087**
	5	**,022**	,035	,538	**-,055**	**,100**
	6	,106*	,023	,001	,054	,157
	7	,076*	,007	,000	,059	,092
	8	**,007**	,008	,430	**-,011**	**,024**
M	1	-,088*	,018	,000	-,127	-,048
	3	**,007**	,004	,087	**-,001**	**,016**
	4	-,080*	,025	,009	-,135	-,024
	5	-,065*	,024	,022	-,119	-,012
	6	**,018**	,013	,202	**-,011**	**,047**
	7	**-,012**	,022	,596	**-,060**	**,036**
	8	-,081*	,020	,002	-,125	-,037
MM	1	-,095*	,021	,001	-,141	-,049
	2	**-,007**	,004	,087	**-,016**	**,001**
	4	-,087*	,022	,003	-,136	-,037
	5	-,072*	,022	,006	-,120	-,025
	6	**,011**	,011	,348	**-,013**	**,035**
	7	**-,019**	,024	,453	**-,073**	**,035**
	8	-,088*	,022	,002	-,138	-,039

Table A.9(cont'd): Pairwise Comparisons of the Methods with respect to MSE for Scenario-1

LTS	1	**-,008**	,036	,824	**-,087**	**,070**
	2	,080*	,025	,009	,024	,135
	3	,087*	,022	,003	,037	,136
	5	,014*	,004	,002	,006	,022
	6	,097*	,015	,000	,063	,132
	7	**,068**	,041	,124	**-,022**	**,157**
	8	**-,002**	,035	,964	**-,078**	**,075**
LMS	1	**-,022**	,035	,538	**-,100**	**,055**
	2	,065*	,024	,022	,012	,119
	3	,072*	,022	,006	,025	,120
	4	-,014*	,004	,002	-,022	-,006
	6	,083*	,014	,000	,052	,114
	7	**,053**	,040	,207	**-,034**	**,141**
	8	**-,016**	,034	,651	**-,091**	**,059**
LAD	1	-,106*	,023	,001	-,157	-,054
	2	**-,018**	,013	,202	**-,047**	**,011**
	3	**-,011**	,011	,348	**-,035**	**,013**
	4	-,097*	,015	,000	-,132	-,063
	5	-,083*	,014	,000	-,114	-,052
	7	**-,030**	,027	,290	**-,089**	**,029**
	8	-,099*	,023	,001	-,149	-,049
LOWESS	1	-,076*	,007	,000	-,092	-,059
	2	**,012**	,022	,596	**-,036**	**,060**
	3	**,019**	,024	,453	**-,035**	**,073**
	4	**-,068**	,041	,124	**-,157**	**,022**
	5	**-,053**	,040	,207	**-,141**	**,034**
	6	**,030**	,027	,290	**-,029**	**,089**
	8	-,069*	,013	,000	-,097	-,042

Table A.9(cont'd): Pairwise Comparisons of the Methods with respect to MSE for Scenario-1

MARS	1	**-,007**	,008	,430	**-,024**	**,011**
	2	,081*	,020	,002	,037	,125
	3	,088*	,022	,002	,039	,138
	4	**,002**	,035	,964	**-,075**	**,078**
	5	**,016**	,034	,651	**-,059**	**,091**
	6	,099*	,023	,001	,049	,149
	7	,069*	,013	,000	,042	,097

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.10: Test of Sphericity for Scenario-2 (PP)

Mauchly's Test of Sphericity[b]

Measure:PP

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHOD	,000	.	27	.	,311	,365	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHOD

.

Table A.11: Tests of Within-Subjects Effects of PP for Scenario-2

Tests of Within-Subjects Effects

Measure:PP

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	70,901	7	10,129	16,004	,000
	Greenhouse-Geisser	70,901	2,174	32,618	16,004	**,000**
	Huynh-Feldt	70,901	2,556	27,743	16,004	,000
	Lower-bound	70,901	1,000	70,901	16,004	,001
Error(METHODS)	Sphericity Assumed	66,454	105	,633		
	Greenhouse-Geisser	66,454	32,606	2,038		
	Huynh-Feldt	66,454	38,335	1,734		
	Lower-bound	66,454	15,000	4,430		

Table A.12: Pairwise Comparisons of the Methods with respect to PP for Scenario-2

Pairwise Comparisons

Measure:PP

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
OLS	2	1,014*	,221	,000	,542	1,485
	3	,980*	,228	,001	,494	1,466
	4	**,00**8	,022	,731	**-,040**	**,056**
	5	**,027**	,020	,200	**-,016**	**,069**
	6	**,038**	,026	,164	**-,017**	**,093**
	7	**,401**	,225	,095	**-,079**	**,881**
	8	2,250*	,463	,000	1,264	3,236
LTS	1	-1,014*	,221	,000	-1,485	-,542
	3	**-,034**	,065	,612	**-,173**	**,105**
	4	-1,006*	,215	,000	-1,464	-,548
	5	-,987*	,211	,000	-1,438	-,536
	6	-,976*	,210	,000	-1,424	-,528
	7	-,613*	,282	,047	-1,215	-,011
	8	1,236*	,447	,015	,283	2,190
LMS	1	-,980*	,228	,001	-1,466	-,494
	2	**,034**	,065	,612	**-,105**	**,173**
	4	-,972*	,223	,001	-1,448	-,496
	5	-,953*	,219	,001	-1,420	-,486
	6	-,942*	,218	,001	-1,406	-,478
	7	**-,579**	,294	,067	**-1,205**	**,047**
	8	1,270*	,484	,019	,238	2,302

Table A.12 (cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-2

.M	1	**-,008**	,022	,731	**-,056**	**,040**
	2	1,006*	,215	,000	,548	1,464
	3	,972*	,223	,001	,496	1,448
	5	**,019**	,019	,333	**-,021**	**,059**
	6	**,030**	,030	,333	**-,034**	**,094**
	7	**,393**	,236	,117	**-,111**	**,897**
	8	2,242*	,456	,000	1,269	3,215
MM	1	**-,027**	,020	,200	**-,069**	**,016**
	2	,987*	,211	,000	,536	1,438
	3	,953*	,219	,001	,486	1,420
	4	**-,019**	,019	,333	**-,059**	**,021**
	6	**,011**	,011	,333	**-,013**	**,035**
	7	**,374**	,222	,113	**-,100**	**,848**
	8	2,223*	,452	,000	1,261	3,186
LAD	1	**-,038**	,026	,164	**-,093**	**,017**
	2	,976*	,210	,000	,528	1,424
	3	,942*	,218	,001	,478	1,406
	4	**-,030**	,030	,333	**-,094**	**,034**
	5	**-,011**	,011	,333	**-,035**	**,013**
	7	**,363**	,215	,111	**-,094**	**,820**
	8	2,212*	,449	,000	1,255	3,169
LOWESS	1	**-,401**	,225	,095	**-,881**	**,079**
	2	,613*	,282	,047	,011	1,215
	3	**,579**	,294	,067	**-,047**	**1,205**
	4	**-,393**	,236	,117	**-,897**	**,111**
	5	**-,374**	,222	,113	**-,848**	**,100**
	6	**-,363**	,215	,111	**-,820**	**,094**
	8	1,849*	,443	,001	,905	2,792

Table A.12 (cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-2

MARS	1	-2,250*	,463	,000	-3,236	-1,264
	2	-1,236*	,447	,015	-2,190	-,283
	3	-1,270*	,484	,019	-2,302	-,238
	4	-2,242*	,456	,000	-3,215	-1,269
	5	-2,223*	,452	,000	-3,186	-1,261
	6	-2,212*	,449	,000	-3,169	-1,255
	7	-1,849*	,443	,001	-2,792	-,905

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.13: Test of Sphericity for Scenario-2 (PO)

Mauchly's Test of Sphericity[b]

Measure:PO

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	233,442	27	**,000**	,296	,344	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.14: Tests of Within-Subjects Effects of PO for Scenario-2

Tests of Within-Subjects Effects

Measure:PO

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	1,811	7	,259	2,394	,026
	Greenhouse-Geisser	1,811	2,070	,875	2,394	**,106**
	Huynh-Feldt	1,811	2,407	,752	2,394	,096
	Lower-bound	1,811	1,000	1,811	2,394	,143

Table A.14 (cont'd): Tests of Within-Subjects Effects of PO for Scenario-2

Error(METHODS)	Sphericity Assumed	11,348	105	,108		
	Greenhouse-Geisser	11,348	31,055	,365		
	Huynh-Feldt	11,348	36,109	,314		
	Lower-bound	11,348	15,000	,757		

Table A.15: Test of Sphericity for Scenario-2 (MSE)

Mauchly's Test of Sphericity[b]

Measure:MSE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	154,745	27	**,000**	,340	,408	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.16: Tests of Within-Subjects Effects of MSE for Scenario-2

Tests of Within-Subjects Effects

Measure:MSE

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	,485	7	,069	4,833	,000
	Greenhouse-Geisser	,485	2,377	,204	4,833	**,010**
	Huynh-Feldt	,485	2,854	,170	4,833	,006
	Lower-bound	,485	1,000	,485	4,833	,044
Error(METHODS)	Sphericity Assumed	1,506	105	,014		
	Greenhouse-Geisser	1,506	35,651	,042		
	Huynh-Feldt	1,506	42,811	,035		
	Lower-bound	1,506	15,000	,100		

Table A.17: Pairwise Comparisons of the Methods with respect to MSE for Scenario-2

Pairwise Comparisons

Measure:MSE

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
OLS	2	-,145*	,039	,002	-,228	-,062
	3	-,105*	,028	,002	-,165	-,045
	4	**,011**	,021	**,614**	**-,034**	**,055**
	5	,031*	,011	,014	,007	,055
	6	**,023**	,017	,199	**-,014**	**,060**
	7	**-,024**	,058	,682	**-,147**	**,099**
	8	**,021**	,040	,604	**-,064**	**,106**
LTS	1	,145*	,039	,002	,062	,228
	3	**,040**	,019	,055	**,000**	**,081**
	4	,156*	,042	,002	,066	,246
	5	,176*	,037	,000	,097	,255
	6	,168*	,041	,001	,081	,256
	7	,121*	,055	,043	,004	,238
	8	,166*	,055	,009	,048	,284
LMS	1	,105*	,028	,002	,045	,165
	2	**-,040**	,019	,055	**-,081**	**,001**
	4	,116*	,029	,001	,055	,177
	5	,136*	,023	,000	,088	,184
	6	,128*	,024	,000	,076	,180
	7	**,081**	,062	,214	**-,052**	**,214**
	8	,126*	,053	,030	,014	,238

Table A.17 (cont'd): Pairwise Comparisons of the Methods with respect to MSE for Scenario-2

M	1	**-,011**	,021	,614	**-,055**	**,034**
	2	-,156*	,042	,002	-,246	-,066
	3	-,116*	,029	,001	-,177	-,055
	5	**,020**	,020	,329	**-,023**	**,063**
	6	**,013**	,020	,541	**-,030**	**,055**
	7	**-,035**	,059	,564	**-,161**	**,091**
	8	**,010**	,047	,828	**-,089**	**,110**
MM	1	-,031*	,011	,014	-,055	-,007
	2	-,176*	,037	,000	-,255	-,097
	3	-,136*	,023	,000	-,184	-,088
	4	**-,020**	,020	,329	**-,063**	**,023**
	6	**-,008**	,008	,347	**-,025**	**,009**
	7	**-,055**	,064	,405	**-,192**	**,082**
	8	**-,010**	,045	,828	**-,107**	**,087**
LAD	1	**-,023**	,017	,199	**-,060**	**,014**
	2	-,168*	,041	,001	-,256	-,081
	3	-,128*	,024	,000	-,180	-,076
	4	**-,013**	,020	,541	**-,055**	**,030**
	5	**,008**	,008	,347	**-,009**	**,025**
	7	**-,047**	,069	,503	**-,194**	**,100**
	8	**-,002**	,050	,966	**-,108**	**,104**
LOWESS	1	**,024**	,058	,682	**-,099**	**,147**
	2	-,121*	,055	,043	-,238	-,004
	3	**-,081**	,062	,214	**-,214**	**,052**
	4	**,035**	,059	,564	**-,091**	**,161**
	5	**,055**	,064	,405	**-,082**	**,192**
	6	**,047**	,069	,503	**-,100**	**,194**
	8	**,045**	,048	,361	**-,057**	**,147**

Table A.17 (cont'd): Pairwise Comparisons of the Methods with respect to MSE for Scenario-2

MARS	1	**-,021**	,040	,604	**-,106**	**,064**
	2	-,166*	,055	,009	-,284	-,048
	3	-,126*	,053	,030	-,238	-,014
	4	**-,010**	,047	,828	**-,110**	**,089**
	5	**,010**	,045	,828	**-,087**	**,107**
	6	**,002**	,050	,966	**-,104**	**,108**
	7	**-,045**	,048	,361	**-,147**	**,057**

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.18: Test of Sphericity for Scenario-3 (PP)

Mauchly's Test of Sphericity[b]

Measure:PP

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	333,927	27	**,000**	,311	,344	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept

Within Subjects Design: METHODS

Table A.19: Tests of Within-Subjects Effects of PP for Scenario-3

Tests of Within-Subjects Effects

Measure:PP

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	15,730	7	2,247	44,132	,000
	Greenhouse-Geisser	15,730	2,175	7,231	44,132	**,000**
	Huynh-Feldt	15,730	2,411	6,525	44,132	,000
	Lower-bound	15,730	1,000	15,730	44,132	,000
Error(METHODS)	Sphericity Assumed	8,198	161	,051		
	Greenhouse-Geisser	8,198	50,030	,164		
	Huynh-Feldt	8,198	55,449	,148		
	Lower-bound	8,198	23,000	,356		

Table A.20: Pairwise Comparisons of the Methods with respect to PP for Scenario-3

Pairwise Comparisons

Measure:PP

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a] Lower Bound	Upper Bound
OLS	2	-,265*	,052	,000	-,373	-,156
	3	-,242*	,054	,000	-,355	-,130
	4	-,294*	,049	,000	-,395	-,193
	5	-,315*	,048	,000	-,415	-,215
	6	-,312*	,042	,000	-,399	-,225
	7	**-,008**	,057	,886	**-,127**	**,110**
	8	,579*	,068	,000	,439	,719
LTS	1	,265*	,052	,000	,156	,373
	3	,022*	,008	,010	,006	,039
	4	**-,030**	,056	,600	**-,145**	**,085**
	5	**-,050**	,050	,326	**-,154**	**,053**
	6	**-,047**	,035	,193	**-,121**	**,026**
	7	,256*	,079	,004	,093	,420
	8	,844*	,105	,000	,626	1,061
LMS	1	,242*	,054	,000	,130	,355
	2	-,022*	,008	,010	-,039	-,006
	4	**-,052**	,056	,361	**-,167**	**,063**
	5	**-,073**	,051	,165	**-,178**	**,032**
	6	**-,070**	,036	,067	**-,144**	**,005**
	7	,234*	,078	,006	,073	,395
	8	,821*	,105	,000	,604	1,039

Table A.20 (cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-3

M	1	,294*	,049	,000	,193	,395
	2	**,030**	,056	,600	**-,085**	**,145**
	3	**,052**	,056	,361	**-,063**	**,167**
	5	-,021*	,007	,011	-,036	-,005
	6	**-,018**	,021	,415	**-,062**	**,027**
	7	,286*	,084	,002	,113	,459
	8	,873*	,088	,000	,692	1,055
MM	1	,315*	,048	,000	,215	,415
	2	**,050**	,050	,326	**-,053**	**,154**
	3	**,073**	,051	,165	**-,032**	**,178**
	4	,021*	,007	,011	,005	,036
	6	**,003**	,016	,849	**-,030**	**,036**
	7	,307*	,085	,001	,131	,482
	8	,894*	,091	,000	,706	1,083
LAD	1	,312*	,042	,000	,225	,399
	2	**,047**	,035	,193	**-,026**	**,121**
	3	**,070**	,036	,067	**-,005**	**,144**
	4	**,018**	,021	,415	**-,027**	**,062**
	5	**-,003**	,016	,849	**-,036**	**,030**
	7	,304*	,080	,001	,139	,468
	8	,891*	,091	,000	,702	1,080
LOWESS	1	**,008**	,057	,886	**-,110**	**,127**
	2	-,256*	,079	,004	-,420	-,093
	3	-,234*	,078	,006	-,395	-,073
	4	-,286*	,084	,002	-,459	-,113
	5	-,307*	,085	,001	-,482	-,131
	6	-,304*	,080	,001	-,468	-,139
	8	,587*	,069	,000	,444	,730

Table A.20 (cont'd): Pairwise Comparisons of the Methods with respect to PP for Scenario-3

MARS	1	-,579*	,068	,000	-,719	-,439
	2	-,844*	,105	,000	-1,061	-,626
	3	-,821*	,105	,000	-1,039	-,604
	4	-,873*	,088	,000	-1,055	-,692
	5	-,894*	,091	,000	-1,083	-,706
	6	-,891*	,091	,000	-1,080	-,702
	7	-,587*	,069	,000	-,730	-,444

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.21: Test of Sphericity for Scenario-3 (PO)

Mauchly's Test of Sphericity[b]

Measure:PO

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	423,356	27	,000	,185	,191	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.22: Tests of Within-Subjects Effects of PO for Scenario-3

Tests of Within-Subjects Effects

Measure:PO

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	23,746	7	3,392	51,638	,000
	Greenhouse-Geisser	23,746	1,294	18,352	51,638	**,000**
	Huynh-Feldt	23,746	1,339	17,740	51,638	,000
	Lower-bound	23,746	1,000	23,746	51,638	,000
Error(METHODS)	Sphericity Assumed	10,577	161	,066		
	Greenhouse-Geisser	10,577	29,760	,355		
	Huynh-Feldt	10,577	30,786	,344		
	Lower-bound	10,577	23,000	,460		

Table A.23: Pairwise Comparisons of the Methods with respect to PO for Scenario-3

Pairwise Comparisons

Measure:PO

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
OLS	2	,801*	,099	,000	,596	1,006
	3	,844*	,099	,000	,639	1,050
	4	,664*	,106	,000	,446	,883
	5	,673*	,105	,000	,456	,891
	6	,678*	,098	,000	,476	,881
	7	**,643**	,021	,998	**-,043**	**,043**
	8	**,051**	,038	,199	**-,029**	**,130**
LTS	1	-,801*	,099	,000	-1,006	-,596
	3	,044*	,006	,000	,032	,055
	4	-,136*	,038	,002	-,215	-,058
	5	-,127*	,034	,001	-,197	-,058
	6	-,123*	,024	,000	-,172	-,073
	7	-,801*	,099	,000	-1,006	-,595
	8	-,750*	,086	,000	-,928	-,573
LMS	1	-,844*	,099	,000	-1,050	-,639
	2	-,044*	,006	,000	-,055	-,032
	4	-,180*	,034	,000	-,251	-,109
	5	-,171*	,030	,000	-,232	-,109
	6	-,166*	,021	,000	-,209	-,124
	7	-,844*	,100	,000	-1,052	-,637
	8	-,794*	,085	,000	-,969	-,618

Table A.23(cont'd): Pairwise Comparisons of the Methods with respect to PO for Scenario-3

M	1	-,664*	,106	,000	-,883	-,446
	2	,136*	,038	,002	,058	,215
	3	,180*	,034	,000	,109	,251
	5	**,009**	,007	,226	**-,006**	**,024**
	6	**,014**	,017	,422	**-,021**	**,048**
	7	-,664*	,110	,000	-,891	-,438
	8	-,614*	,090	,000	-,800	-,428
MM	1	-,673*	,105	,000	-,891	-,456
	2	,127*	,034	,001	,058	,197
	3	,171*	,030	,000	,109	,232
	4	**-,009**	,007	,226	**-,024**	**,006**
	6	**,005**	,012	,709	**-,021**	**,030**
	7	-,673*	,109	,000	-,899	-,448
	8	-,623*	,088	,000	-,805	-,441
LAD	1	-,678*	,098	,000	-,881	-,476
	2	,123*	,024	,000	,073	,172
	3	,166*	,021	,000	,124	,209
	4	**-,014**	,017	,422	**-,048**	**,021**
	5	**-,005**	,012	,709	**-,030**	**,021**
	7	-,678*	,101	,000	-,887	-,470
	8	-,628*	,082	,000	-,797	-,458
LOWESS	1	**-,643**	,021	,998	**-,043**	**,043**
	2	,801*	,099	,000	,595	1,006
	3	,844*	,100	,000	,637	1,052
	4	,664*	,110	,000	,438	,891
	5	,673*	,109	,000	,448	,899
	6	,678*	,101	,000	,470	,887
	8	**,051**	,051	,335	**-,056**	**,157**

Table A.23(cont'd): Pairwise Comparisons of the Methods with respect to PO for Scenario-3

MARS	1	**-,051**	,038	,199	**-,130**	**,029**
	2	,750*	,086	,000	,573	,928
	3	,794*	,085	,000	,618	,969
	4	,614*	,090	,000	,428	,800
	5	,623*	,088	,000	,441	,805
	6	,628*	,082	,000	,458	,797
	7	**-,051**	,051	,335	**-,157**	**,056**

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.24: Test of Sphericity for Scenario-3 (MSE)

Mauchly's Test of Sphericity[b]

Measure:MSE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	615,444	27	**,000**	,152	,153	,143

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.25: Tests of Within-Subjects Effects of MSE for Scenario-3

Tests of Within-Subjects Effects

Measure:MSE

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	17,916	7	2,559	22,754	,000
	Greenhouse-Geisser	17,916	1,062	16,871	22,754	**,000**
	Huynh-Feldt	17,916	1,071	16,735	22,754	,000
	Lower-bound	17,916	1,000	17,916	22,754	,000
Error(METHODS)	Sphericity Assumed	18,110	161	,112		
	Greenhouse-Geisser	18,110	24,424	,741		
	Huynh-Feldt	18,110	24,623	,735		
	Lower-bound	18,110	23,000	,787		

Table A.26: Pairwise Comparisons of the Methods with respect to MSE for Scenario-3

Pairwise Comparisons

Measure:MSE

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
OLS	2	,552*	,126	,000	,292	,812
	3	,560*	,126	,000	,299	,821
	4	,646*	,129	,000	,379	,914
	5	,654*	,129	,000	,388	,920
	6	,639*	,125	,000	,380	,897
	7	**,024**	,017	,163	**-,010**	**,059**
	8	**-,068**	,036	,073	**-,142**	**,007**
LTS	1	-,552*	,126	,000	-,812	-,292
	3	,008*	,002	,002	,003	,013
	4	,094*	,018	,000	,056	,133
	5	,102*	,016	,000	,070	,135
	6	,087*	,012	,000	,062	,112
	7	-,528*	,134	,001	-,806	-,250
	8	-,619*	,129	,000	-,887	-,352
LMS	1	-,560*	,126	,000	-,821	-,299
	2	-,008*	,002	,002	-,013	-,003
	4	,086*	,017	,000	,050	,122
	5	,094*	,014	,000	,065	,124
	6	,079*	,011	,000	,057	,101
	7	-,536*	,135	,001	-,815	-,257
	8	-,628*	,129	,000	-,895	-,360

Table A.26 (cont'd): Pairwise Comparisons of the Methods with respect to MSE for Scenario-3

M	1	-,646*	,129	,000	-,914	-,379
	2	-,094*	,018	,000	-,133	-,056
	3	-,086*	,017	,000	-,122	-,050
	5	**,008**	**,004**	**,057**	**,000**	**,016**
	6	**-,007**	,009	,406	**-,026**	**,011**
	7	-,622*	,139	,000	-,909	-,336
	8	-,714*	,134	,000	-,992	-,436
MM	1	-,654*	,129	,000	-,920	-,388
	2	-,102*	,016	,000	-,135	-,070
	3	-,094*	,014	,000	-,124	-,065
	4	**-,008**	,004	,057	**-,016**	**,000**
	6	-,016*	,006	,017	-,028	-,003
	7	-,630*	,138	,000	-,915	-,346
	8	-,722*	,133	,000	-,997	-,447
LAD	1	-,639*	,125	,000	-,897	-,380
	2	-,087*	,012	,000	-,112	-,062
	3	-,079*	,011	,000	-,101	-,057
	4	**,007**	,009	,406	**-,011**	**,026**
	5	,016*	,006	,017	,003	,028
	7	-,615*	,134	,000	-,892	-,338
	8	-,706*	,129	,000	-,974	-,439
LOWESS	1	**-,024**	,017	,163	**-,059**	**,010**
	2	,528*	,134	,001	,250	,806
	3	,536*	,135	,001	,257	,815
	4	,622*	,139	,000	,336	,909
	5	,630*	,138	,000	,346	,915
	6	,615*	,134	,000	,338	,892
	8	-,092*	,026	,002	-,146	-,038

Table A.26 (cont'd): Pairwise Comparisons of the Methods with respect to MSE for Scenario-3

MARS	1	**,068**	,036	,073	**-,007**	**,142**
	2	,619*	,129	,000	,352	,887
	3	,628*	,129	,000	,360	,895
	4	,714*	,134	,000	,436	,992
	5	,722*	,133	,000	,447	,997
	6	,706*	,129	,000	,439	,974
	7	,092*	,026	,002	,038	,146

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

Table A.27: Test of Sphericity for Scenario-1 (Ratio of IMSE)

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	132,537	20	**,000**	,227	,249	,167

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.28: Tests of Within-Subjects Effects of the ratio of IMSE for Scenario-1

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	,663	6	,110	7,378	,000
	Greenhouse-Geisser	,663	1,365	,486	7,378	**,011**
	Huynh-Feldt	,663	1,492	,444	7,378	,008
	Lower-bound	,663	1,000	,663	7,378	,020
Error(METHODS)	Sphericity Assumed	,988	66	,015		
	Greenhouse-Geisser	,988	15,014	,066		
	Huynh-Feldt	,988	16,416	,060		
	Lower-bound	,988	11,000	,090		

Table A.29: Pairwise Comparisons of the Methods with respect to the ratio of IMSE for Scenario-1

Pairwise Comparisons

Measure:MEASURE_1

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	**,011**	,006	,083	**-,002**	**,024**
	3	-,192*	,058	,007	-,319	-,065
	4	-,158*	,054	,014	-,277	-,039
	5	**,029**	,023	,235	**-,022**	**,079**
	6	**-,016**	,040	,692	**-,105**	**,072**
	7	-,162*	,038	,001	-,247	-,078
2	1	**-,011**	,006	,083	**-,024**	**,002**
	3	-,203*	,054	,003	-,321	-,085
	4	-,169*	,050	,006	-,278	-,060
	5	**,018**	,020	,385	**-,025**	**,061**
	6	**-,027**	,044	,547	**-,125**	**,070**
	7	-,173*	,042	,002	-,265	-,082
3	1	,192*	,058	,007	,065	,319
	2	,203*	,054	,003	,085	,321
	4	,034*	,008	,001	,017	,050
	5	,221*	,043	,000	,126	,315
	6	**,176**	,088	,071	**-,017**	**,369**
	7	**,030**	,074	,697	**-,134**	**,194**

Table A.29 (cont'd): Pairwise Comparisons of the Methods with respect to the ratio of IMSE for Scenario-1

4	1	,158*	,054	,014	,039	,277
	2	,169*	,050	,006	,060	,278
	3	-,034*	,008	,001	-,050	-,017
	5	,187*	,038	,000	,103	,271
	6	,142	,084	,118	-,042	,326
	7	-,004	,071	,956	-,159	,151
5	1	-,029	,023	,235	-,079	,022
	2	-,018	,020	,385	-,061	,025
	3	-,221*	,043	,000	-,315	-,126
	4	-,187*	,038	,000	-,271	-,103
	6	-,045	,047	,358	-,149	,058
	7	-,191*	,040	,001	-,278	-,104
6	1	,016	,040	,692	-,072	,105
	2	,027	,044	,547	-,070	,125
	3	-,176	,088	,071	-,369	,017
	4	-,142	,084	,118	-,326	,042
	5	,045	,047	,358	-,058	,149
	7	-,146*	,027	,000	-,206	-,086
7	1	,162*	,038	,001	,078	,247
	2	,173*	,042	,002	,082	,265
	3	-,030	,074	,697	-,194	,134
	4	,004	,071	,956	-,151	,159
	5	,191*	,040	,001	,104	,278
	6	,146*	,027	,000	,086	,206

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

*. The mean difference is significant at the ,05 level.

Table A.30: Test of Sphericity for Scenario-2 (ratio of IMSE)

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	144,937	20	,000	,372	,440	,167

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.31: Tests of Within-Subjects Effects of the ratio of IMSE for Scenario-2

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	4,692	6	,782	3,880	,002
	Greenhouse-Geisser	4,692	2,231	2,103	3,880	,027
	Huynh-Feldt	4,692	2,639	1,777	3,880	,020
	Lower-bound	4,692	1,000	4,692	3,880	,068

Table A.31 (cont'd): Tests of Within-Subjects Effects of the ratio of IMSE for Scenario-2

Error(METHODS)	Sphericity Assumed	18,138	90	,202		
	Greenhouse-Geisser	18,138	33,471	,542		
	Huynh-Feldt	18,138	39,592	,458		
	Lower-bound	18,138	15,000	1,209		

Table A.32: Pairwise Comparisons of the Methods with respect to the ratio of MSE for Scenario-2

Pairwise Comparisons

Measure:MEASURE_1

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	**,195**	,129	,151	**-,080**	**,469**
	3	,517*	,216	,030	,057	,976
	4	,576*	,205	,013	,139	1,012
	5	,553*	,212	,020	,100	1,005
	6	**,276**	,220	,229	**-,193**	**,746**
	7	,511*	,206	,025	,072	,950
2	1	**-,195**	,129	,151	**-,469**	**,080**
	3	,322*	,102	,006	,105	,539
	4	,381*	,084	,000	,201	,561
	5	,358*	,088	,001	,170	,546
	6	**,082**	,213	,707	**-,373**	**,536**
	7	,316*	,117	,016	,068	,565

Table A.32 (cont'd): Pairwise Comparisons of the Methods with respect to the ratio of MSE for Scenario-2

3	1	-,517*	,216	,030	-,976	-,057
	2	-,322*	,102	,006	-,539	-,105
	4	**,059**	,059	,331	**-,066**	**,184**
	5	**,036**	,058	,544	**-,088**	**,160**
	6	**-,240**	,205	,258	**-,677**	**,196**
	7	**-,006**	,089	,951	**-,195**	**,184**
4	1	-,576*	,205	,013	-1,012	-,139
	2	-,381*	,084	,000	-,561	-,201
	3	**-,059**	,059	,331	**-,184**	**,066**
	5	**-,023**	,016	,182	**-,058**	**,012**
	6	**-,299**	,220	,194	**-,769**	**,170**
	7	**-,065**	,078	,421	**-,231**	**,102**
5	1	-,553*	,212	,020	-1,005	-,100
	2	-,358*	,088	,001	-,546	-,170
	3	**-,036**	,058	,544	**-,160**	**,088**
	4	**,023**	,016	,182	**-,012**	**,058**
	6	**-,276**	,229	,247	**-,765**	**,213**
	7	**-,042**	,088	,644	**-,230**	**,147**
6	1	**-,276**	,220	,229	**-,746**	**,193**
	2	**-,082**	,213	,707	**-,536**	**,373**
	3	**,240**	,205	,258	**-,196**	**,677**
	4	**,299**	,220	,194	**-,170**	**,769**
	5	**,276**	,229	,247	**-,213**	**,765**
	7	**,235**	,178	,207	**-,145**	**,614**

Table A.32 (cont'd): Pairwise Comparisons of the Methods with respect to the ratio of MSE for Scenario-2

7	1	-,511*	,206	,025	-,950	-,072
	2	-,316*	,117	,016	-,565	-,068
	3	**,006**	,089	,951	**-,184**	**,195**
	4	**,065**	,078	,421	**-,102**	**,231**
	5	**,042**	,088	,644	**-,147**	**,230**
	6	**-,235**	,178	,207	**-,614**	**,145**

Based on estimated marginal means

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

*. The mean difference is significant at the ,05 level.

Table A.33: Test of Sphericity for Scenario-3 (ratio of the IMSE)

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	412,578	20	,000	,208	,215	,167

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table A.34: Tests of Within-Subjects Effects of the ratio of the MSE for Scenario-3

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	14,355	6	2,392	21,389	,000
	Greenhouse-Geisser	14,355	1,250	11,479	21,389	,000
	Huynh-Feldt	14,355	1,288	11,146	21,389	,000
	Lower-bound	14,355	1,000	14,355	21,389	,000
Error(METHODS)	Sphericity Assumed	15,436	138	,112		
	Greenhouse-Geisser	15,436	28,761	,537		
	Huynh-Feldt	15,436	29,622	,521		
	Lower-bound	15,436	23,000	,671		

Table A.35: Pairwise Comparisons of the Methods with respect to the ratio of MSE for Scenario-3

Pairwise Comparisons

Measure:MEASURE_1

(I) METHODS	(J) METHODS	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	,011*	,004	,017	,002	,019
	3	,103*	,039	,014	,023	,182
	4	,115*	,036	,004	,041	,189
	5	,113*	,032	,002	,046	,180
	6	-,399*	,132	,006	-,673	-,126
	7	-,695*	,160	,000	-1,026	-,364
2	1	-,011*	,004	,017	-,019	-,002
	3	,092*	,036	,018	,017	,167
	4	,105*	,033	,005	,035	,174
	5	,102*	,030	,002	,041	,164
	6	-,410*	,130	,004	-,679	-,141
	7	-,706*	,158	,000	-1,033	-,378
3	1	-,103*	,039	,014	-,182	-,023
	2	-,092*	,036	,018	-,167	-,017
	4	**,013**	**,006**	,061	**,000**	**,026**
	5	**,010**	,013	,435	**-,017**	**,037**
	6	-,502*	,111	,000	-,732	-,272
	7	-,798*	,146	,000	-1,100	-,495

Table A.35 (cont'd): Pairwise Comparisons of the Methods with respect to the ratio of MSE for Scenario-3

4	1	-,115*	,036	,004	-,189	-,041
	2	-,105*	,033	,005	-,174	-,035
	3	**-,013**	,006	,061	**-,026**	**,001**
	5	**-,002**	,008	,788	**-,019**	**,014**
	6	-,515*	,109	,000	-,739	-,290
	7	-,810*	,144	,000	-1,107	-,513
5	1	-,113*	,032	,002	-,180	-,046
	2	-,102*	,030	,002	-,164	-,041
	3	**-,010**	,013	,435	**-,037**	**,017**
	4	**,002**	,008	,788	**-,014**	**,019**
	6	-,512*	,108	,000	-,736	-,289
	7	-,808*	,142	,000	-1,102	-,514
6	1	,399*	,132	,006	,126	,673
	2	,410*	,130	,004	,141	,679
	3	,502*	,111	,000	,272	,732
	4	,515*	,109	,000	,290	,739
	5	,512*	,108	,000	,289	,736
	7	-,295*	,072	,000	-,444	-,147
7	1	,695*	,160	,000	,364	1,026
	2	,706*	,158	,000	,378	1,033
	3	,798*	,146	,000	,495	1,100
	4	,810*	,144	,000	,513	1,107
	5	,808*	,142	,000	,514	1,102
	6	,295*	,072	,000	,147	,444

Based on estimated marginal means

*. The mean difference is significant at the ,05 level.

a. Adjustment for multiple comparisons: Least Significant Difference (equivalent to no adjustments).

APPENDIX B

MODEL SELECTION AND RESIDUAL ANALYSIS FOR THE ORIGINAL DATA

B.1. Model Selection Output for the Original Data without Making any Transformation

Stepwise Regression: y2(p)-azot versus x2; x3; ...

Alpha-to-Enter: 0,15 Alpha-to-Remove: 0,15

Response is y2(p)-azot on 35 predictors, with N = 92

Step	1	2	3	4	5	6
Constant	-0,04380	0,51583	1,05162	0,81391	1,56444	1,63715
x22	0,00398	0,00370	0,00365	0,00323	0,00331	0,00362
T-Value	4,44	4,21	4,23	3,79	3,93	4,21
P-Value	0,000	0,000	0,000	0,000	0,000	0,000
x14		-0,113	-0,113	-0,125	-0,117	-0,116
T-Value		-2,51	-2,57	-2,92	-2,75	-2,74
P-Value		0,014	0,012	0,004	0,007	0,008
x9			-0,167	-0,253	-0,259	-0,273
T-Value			-2,16	-3,07	-3,18	-3,36
P-Value			0,034	0,003	0,002	0,001
x8				0,0146	0,0156	0,0177
T-Value				2,55	2,75	3,05
P-Value				0,012	0,007	0,003
x29					-0,25	-0,22
T-Value					-1,78	-1,59
P-Value					0,079	0,114
x12						-0,00064
T-Value						-1,55
P-Value						0,126
S	0,0507	0,0492	0,0483	0,0468	0,0462	0,0459
R-Sq	17,96	23,39	27,24	32,30	34,70	36,48
R-Sq(adj)	17,05	21,67	24,75	29,19	30,90	32,00
Mallows C-p	13,2	8,5	5,7	1,5	0,5	0,3

Step	7	8	9
Constant	0,3583	0,3453	0,2595
x22	0,00333	0,00369	0,00363
T-Value	3,83	4,12	4,08
P-Value	0,000	0,000	0,000
x14	-0,118	-0,111	-0,105
T-Value	-2,81	-2,66	-2,52
P-Value	0,006	0,009	0,014
x9	-0,288	-0,300	-0,315
T-Value	-3,55	-3,71	-3,89
P-Value	0,001	0,000	0,000
x8	0,0171	0,0173	0,0196
T-Value	2,99	3,04	3,34
P-Value	0,004	0,003	0,001
x29	-0,32	-0,35	-0,38
T-Value	-2,12	-2,35	-2,52
P-Value	0,037	0,021	0,014
x12	-0,00071	-0,00076	-0,00077
T-Value	-1,73	-1,85	-1,90
P-Value	0,087	0,067	0,061
x26	0,00120	0,00130	0,00138
T-Value	1,63	1,77	1,88
P-Value	0,108	0,081	0,063
x36		-1,21	-1,25
T-Value		-1,50	-1,56
P-Value		0,136	0,124
x34			-0,49
T-Value			-1,47
P-Value			0,147
S	0,0454	0,0451	0,0448
R-Sq	38,42	40,05	41,58
R-Sq(adj)	33,29	34,28	35,17
Mallows C-p	-0,1	-0,1	0,0

Regression Analysis: y2(p)-azot versus x8; x9; ...

```
The regression equation is
y2(p)-azot = 0,345 + 0,0173 x8 - 0,300 x9 - 0,000756 x12 - 0,111 x14
             + 0,00369 x22 + 0,00130 x26 - 0,352 x29 - 1,21 x36
```

Predictor	Coef	SE Coef	T	P
Constant	0,3453	0,9414	0,37	0,715
x8	0,017339	0,005697	3,04	0,003
x9	-0,29974	0,08084	-3,71	0,000
x12	-0,0007558	0,0004079	-1,85	0,067
x14	-0,11147	0,04185	-2,66	0,009
x22	0,0036923	0,0008965	4,12	0,000

x26	0,0012979	0,0007353	1,77	0,081
x29	-0,3520	0,1498	-2,35	0,021
x36	-1,2142	0,8074	-1,50	0,136

S = 0,0451005   R-Sq = 40,1%   R-Sq(adj) = 34,3%

Analysis of Variance

Source	DF	SS	MS	F	P
Regression	8	0,112805	0,014101	6,93	0,000
Residual Error	83	0,168826	0,002034		
Total	91	0,281632			

Source	DF	Seq SS
x8	1	0,007593
x9	1	0,024676
x12	1	0,001304
x14	1	0,027176
x22	1	0,036651
x26	1	0,001512
x29	1	0,009294
x36	1	0,004601

Unusual Observations

Obs	x8	y2(p)-azot	Fit	SE Fit	Residual	St Resid
16	40,8	0,28570	0,11454	0,01541	0,17116	4,04R
17	39,8	0,20000	0,09243	0,01116	0,10757	2,46R
45	39,0	0,00000	0,03226	0,02710	-0,03226	-0,89 X
49	38,2	0,00000	-0,02771	0,03478	0,02771	0,97 X
52	35,0	0,00000	0,00667	0,02970	-0,00667	-0,20 X
71	39,6	0,00000	0,09361	0,01499	-0,09361	-2,20R
78	40,5	0,29060	0,11242	0,01381	0,17818	4,15R
88	39,0	0,10560	0,01395	0,01506	0,09165	2,16R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

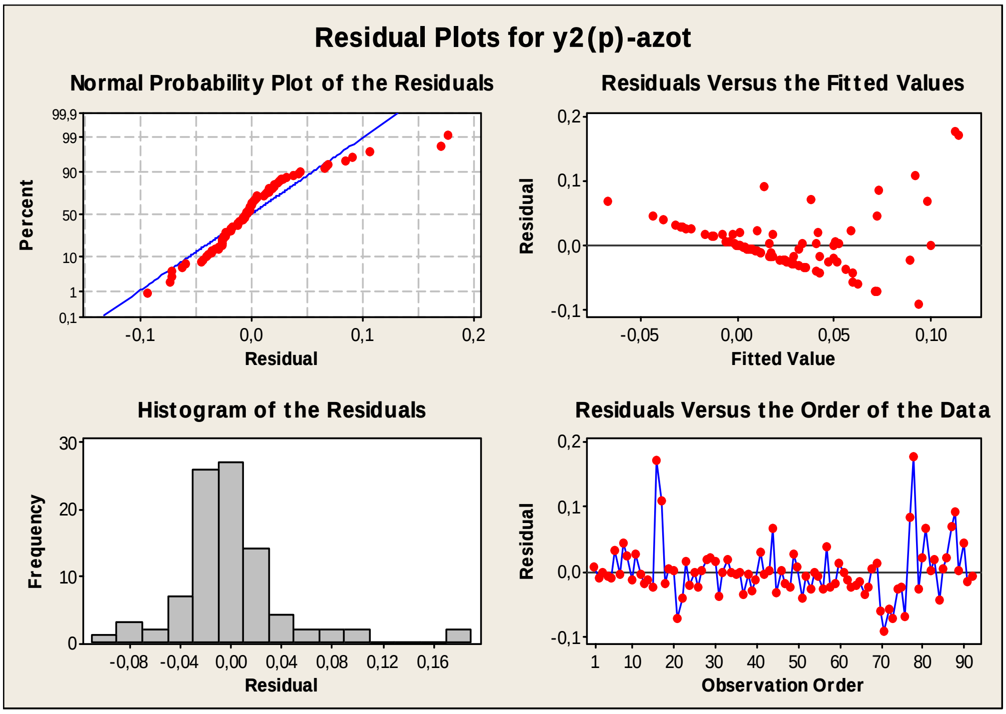

Figure B.1: Residual Plots for the Original Data with 92 observations without Making Transformation

B.2. Model Selection Output for the Original Data with Making Logit Transformation

Stepwise Regression: LOGITY2 versus x2; x3; ...

```
  Alpha-to-Enter: 0,15  Alpha-to-Remove: 0,15


Response is LOGITY2 on 35 predictors, with N = 92


Step               1        2        3        4
Constant     -5,5917  -7,3578   3,1234   0,7232

x22            0,146    0,130    0,129    0,119
T-Value         8,37     7,33     7,40     6,76
P-Value        0,000    0,000    0,000    0,000

x19                     0,115    0,118    0,114
T-Value                  2,93     3,07     3,01
P-Value                 0,004    0,003    0,003

x9                                -3,3     -3,3
T-Value                          -2,20    -2,22
P-Value                          0,030    0,029

x28                                       0,175
T-Value                                    1,99
P-Value                                   0,049

S              0,989    0,950    0,930    0,914
R-Sq           43,75    48,69    51,38    53,50
R-Sq(adj)      43,12    47,54    49,72    51,36
Mallows C-p     13,0      6,2      3,3      1,5
```

Regression Analysis: LOGITY2 versus x9; x19; x22; x28

```
The regression equation is
LOGITY2 = 0,72 - 3,26 x9 + 0,114 x19 + 0,119 x22 + 0,175 x28


Predictor      Coef  SE Coef      T      P
Constant      0,723    4,875   0,15  0,882
x9           -3,258    1,470  -2,22  0,029
x19         0,11433  0,03800   3,01  0,003
x22         0,11946  0,01768   6,76  0,000
x28         0,17540  0,08796   1,99  0,049


S = 0,914286   R-Sq = 53,5%   R-Sq(adj) = 51,4%


Analysis of Variance
```

Source	DF	SS	MS	F	P
Regression	4	83,675	20,919	25,02	0,000
Residual Error	87	72,725	0,836		
Total	91	156,400			

Source	DF	Seq SS
x9	1	4,686
x19	1	28,350
x22	1	47,315
x28	1	3,324

Unusual Observations

Obs	x9	LOGITY2	Fit	SE Fit	Residual	St Resid
16	3,21	-0,3980	-2,7830	0,2138	2,3850	2,68R
21	3,23	-4,0000	-2,0690	0,2153	-1,9309	-2,17R
22	3,18	-4,0000	-1,7981	0,1754	-2,2019	-2,45R
70	3,15	-4,0000	-1,6953	0,3277	-2,3047	-2,70R
71	3,19	-4,0000	-1,7263	0,3840	-2,2736	-2,74RX
72	3,26	-4,0000	-1,4408	0,3663	-2,5592	-3,06R
77	3,16	-0,7270	-2,5885	0,1202	1,8615	2,05R
78	3,17	-0,3876	-2,2078	0,1913	1,8202	2,04R

R denotes an observation with a large standardized residual.
X denotes an observation whose X value gives it large influence.

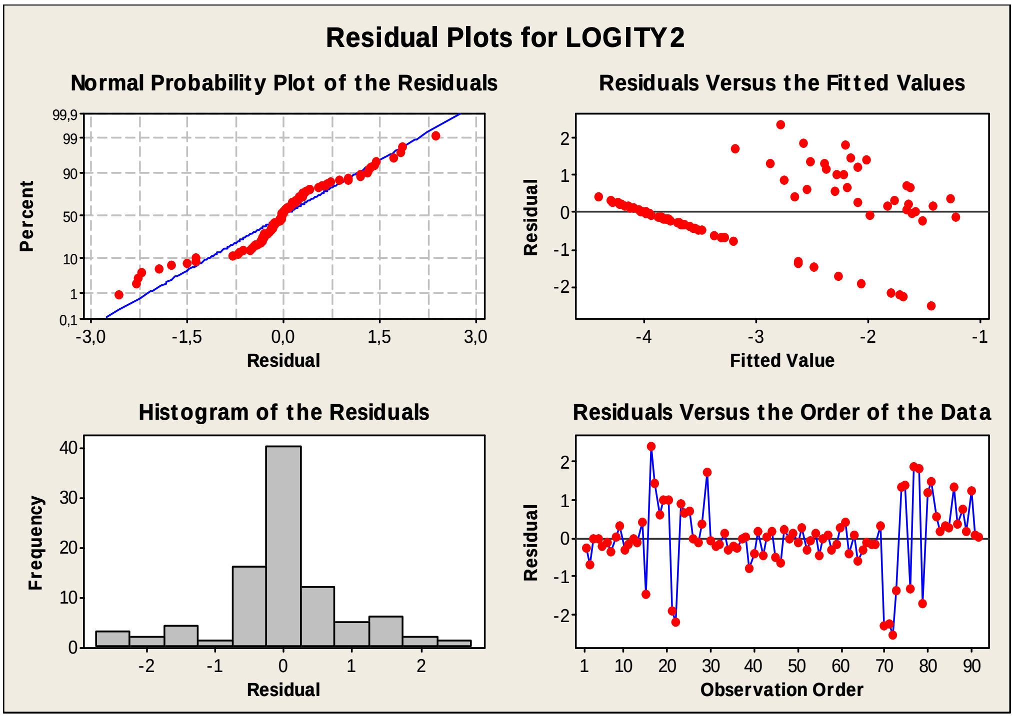

Figure B.2: Residual Plots for the Original Data with 92 observations with Making Logit Transformation

B.3. Model Selection Output for the Original Data with Deleting Outliers without Making any Transformation

Stepwise Regression: y2(p)-azot versus x2; x3; ...

```
  Alpha-to-Enter: 0,15  Alpha-to-Remove: 0,15


Response is y2(p)-azot on 35 predictors, with N = 84


Step                1         2         3         4         5
Constant     -0,03151   0,45003   0,71027   0,95893   0,99699

x22           0,00284   0,00271   0,00236   0,00215   0,00229
T-Value          4,84      4,81      4,09      3,67      3,89
P-Value         0,000     0,000     0,000     0,000     0,000

x9                       -0,150    -0,138    -0,147    -0,145
T-Value                   -2,88     -2,69     -2,88     -2,87
P-Value                   0,005     0,009     0,005     0,005

x20                                -0,0071   -0,0068   -0,0060
T-Value                              -2,10     -2,02     -1,80
P-Value                              0,039     0,046     0,075

x11                                          -0,0110   -0,0114
T-Value                                        -1,67     -1,75
P-Value                                        0,099     0,083

x3                                                     -0,0019
T-Value                                                  -1,47
P-Value                                                  0,145

S              0,0304    0,0291    0,0285    0,0282    0,0280
R-Sq            22,25     29,47     33,15     35,43     37,18
R-Sq(adj)       21,30     27,73     30,64     32,16     33,16
Mallows C-p       6,0       0,0      -2,0      -2,6      -2,5
```

Regression Analysis: y2(p)-azot versus x3; x9; x11; x20; x22

```
The regression equation is
y2(p)-azot = 0,997 - 0,00190 x3 - 0,145 x9 - 0,0114 x11 - 0,00605 x20
             + 0,00229 x22

Predictor        Coef    SE Coef      T      P
Constant       0,9970     0,2515   3,96  0,000
x3          -0,001896   0,001286  -1,47  0,145
x9           -0,14507    0,05062  -2,87  0,005
x11         -0,011429   0,006514  -1,75  0,083
x20         -0,006047   0,003354  -1,80  0,075
x22         0,0022914  0,0005888   3,89  0,000


S = 0,0280148   R-Sq = 37,2%   R-Sq(adj) = 33,2%
```

```
Analysis of Variance

Source          DF         SS         MS     F      P
Regression       5  0,0362341  0,0072468  9,23  0,000
Residual Error  78  0,0612166  0,0007848
Total           83  0,0974507


Source  DF     Seq SS
x3       1  0,0008638
x9       1  0,0088952
x11      1  0,0072507
x20      1  0,0073377
x22      1  0,0118867


Unusual Observations

Obs    x3  y2(p)-azot      Fit   SE Fit  Residual  St Resid
 23  33,7     0,09900  0,04308  0,00589   0,05592     2,04R
 66  33,7     0,00010  0,05596  0,01098  -0,05586    -2,17R
 71  32,0     0,15790  0,04923  0,00710   0,10867     4,01R
 74  33,7     0,16540  0,05738  0,00740   0,10802     4,00R
 82  34,0     0,11610  0,04792  0,00596   0,06818     2,49R

R denotes an observation with a large standardized residual.
```

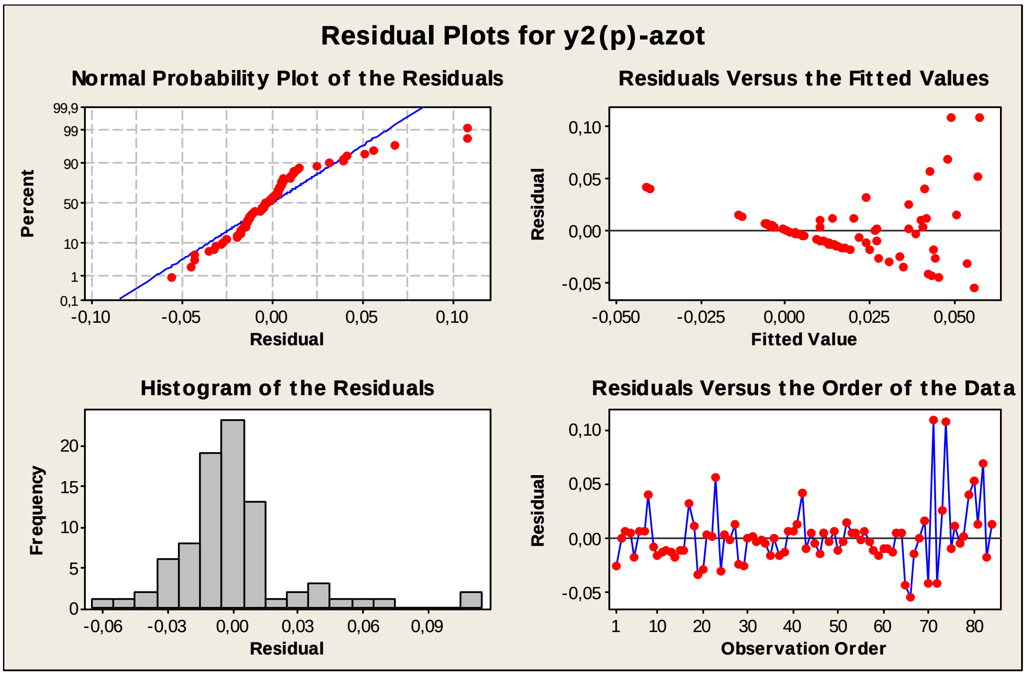

Figure B.3: Residual Plots for the Original Data by Deleting outliers without Making any Transformation

APPENDIX C

REPEATED ANOVA RESULTS FOR THE COMPARISON OF THE OLS, HUBERM, LAV AND LTS REGRESSION METHODS

Table C.1: Test of Sphericity for the Industrial Data with respect to MSE

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	.	5	.	,395	,637	,333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table C.2: Tests of Within-Subjects Effects with respect to MSE for the Industrial Data

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	,213	3	,071	2,008	,214
	Greenhouse-Geisser	,213	1,186	,180	2,008	**,283**
	Huynh-Feldt	,213	1,911	,111	2,008	,252
	Lower-bound	,213	1,000	,213	2,008	,292
Error(METHODS)	Sphericity Assumed	,212	6	,035		
	Greenhouse-Geisser	,212	2,371	,089		
	Huynh-Feldt	,212	3,822	,055		
	Lower-bound	,212	2,000	,106		

Table C.3: Test of Sphericity for the Industrial Data with respect to R^2

Mauchly's Test of Sphericity[b]

Measure:MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
METHODS	,000	.	5	.	,606	1,000	,333

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept
Within Subjects Design: METHODS

Table C.4: Tests of Within-Subjects Effects with respect to R^2 for the Industrial Data

Tests of Within-Subjects Effects

Measure:MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
METHODS	Sphericity Assumed	,049	3	,016	4,723	,051
	Greenhouse-Geisser	,049	1,819	,027	4,723	**,098**
	Huynh-Feldt	,049	3,000	,016	4,723	,051
	Lower-bound	,049	1,000	,049	4,723	,162
Error(METHODS)	Sphericity Assumed	,021	6	,003		
	Greenhouse-Geisser	,021	3,637	,006		
	Huynh-Feldt	,021	6,000	,003		
	Lower-bound	,021	2,000	,010		