YANGON INSTITUTE OF ECONOMICS

Ph. D. PROGRAMME

ROBUST ESTIMATION TECHNIQUES APPLIED TO REGRESSION AND TIME SERIES ANALYSES

MAW MAW KHIN

NOVEMER, 2011

YANGON INSTITUTE OF ECONOMICS

Ph. D. PROGRAMME

ROBUST ESTIMATION TECHNIQUES APPLIED TO REGRESSION AND TIME SERIES ANALYSES

by

Maw Maw Khin

4 Ph. D. Res-Ah-2

NOVEMER, 2011

ROBUST ESTIMATION TECHNIQUES APPLIED TO REGRESSION AND TIME SERIES ANALYSES

Partial fulfillment of the requirement for the Degree of

Ph. D.

of the Department of Statistics,

Yangon Institute of Economics

Submitted by

Maw Maw Khin

November, 2011

CETIFICATION

I here by certify that the content of this dissertation is wholly my own work unless otherwise referenced or acknowledged. Information from sources is referenced with original comments and ideas from the writer him/herself.

Maw Maw Khin
4 Ph. D. Res- Ah- 2

ROBUST ESTIMATION TECHNIQUES APPLIED TO REGRESSION AND TIME SERIES ANALYSES

Board of Examiners

Professor Dr. Khin San Yee

(Chairman)

Rector

Yangon Institute of Economics

30/11/2011

Prof. Dr. Soe Win

(External Examiner)

Director General

Department of Higher Education

Upper Myanmar

24/11/2011

Prof. Dr. Lay Kyi

(Supervisor)

Professor and Head of Department

Department of Statistics

Yangon Institute of Economics

Prof. Dr. Thet Lwin
(Referee)
Professor / Head (Retd.)
Department of Statistics
Yangon Institute of Economics

U Kyaw Min Htun
(Member)
Pro-Rector (Retd.)
Yangon Institute of Economics

Prof. Daw Khin San Myint
(Member)
Professor / Head (Retd.)
Department of Statistics
Yangon Institute of Economics

Prof. Dr. Daw San Kyi
(Member)
Professor (Retd.)
Department of Statistics
Yangon Institute of Economics

Prof. Dr. Daw Khin May Than
(Internal Examiner)
Professor / Head
Department of Economics
Yangon University of Distance Education

Prof. Dr. Thaung Htay
(Internal Examiner)
Professor / Head
Department of Statistics
Monywa Institute of Economics

ABSTRACT

Regression analysis is an important statistical tool for studying the relationship between two or more variables. The frequently used approach to regression analysis is the method of ordinary least squares (OLS). But this method is vulnerable to outliers; even the presence of a single outlier can affect the precision of the regression coefficients. Therefore, many robust methods have been developed to handle data contaminated with outliers. This study attempts to investigate the effect of outliers on estimation of parameters in regression and time series analysis. The results about outlier robustness point out that the robust and classical methods both worked well data with no outliers indicating that their mean squares error (MSE) are quite close to each other. If there are outliers in the data, the robust methods perform better than the classical method. From the simulated results of distributional robustness, it is found that OLS estimation under a heavy-tailed distribution does not yield outlier robust estimates. Indeed, not only with the Gaussian distribution but also with the skewed distributions, OLS estimators collapse in the presence of small levels of outlier contamination. The Huber *M*-estimate and bisquare *M*-estimate estimate have shown to be more appropriate alternatives to the OLS in heavy-tailed distributions whereas the least median squares (LMS) estimates are better choices for skewed data. One best method could not be suggested in all situations; however the use of more than one method of exploratory data analysis is recommended in practice. In analyses with two real data, there observed that both data sets contain outlying observations. Thus, these results lead to the violation of the normality assumption. In such situations, the LMS estimation methods provide good results in estimating the true parameters. In time series analysis, the simulation results point out that ML estimator of the autoregressive parameter φ in the AR(1) model with innovation outlier (IO) is consistent, but inefficient. Under the AR(1) model with additive outlier (AO), the ML estimator is not even consistent with the desired property of consistency. The ML estimates are extremely sensitive to the presence of AO but not to the presence of IO. The RA estimates based on bisquare family have very good robustness properties for AR(1) with AO outlier and AR(1) model with IO outlier and they compare favorably with the *GM*-estimates. The results of the simulation and real data lead to the recommendation of the use of robust methods in both regression and time series analysis. Application of these robust methods provides outlier resistant estimates.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my gratitude to Professor Dr. Khin San Yee, Ractor, Yangon Institute of Economics, for giving me an opportunity to write this thesis.

Next, I am highly and tremendously grateful to my supervisor, Professor Dr. Lay Kyi, Head of Department of Statistics, Yangon Institute of Economics, for reading and supervising the entire work and providing many valuable suggestions for and constant criticism improvement.

My sincere thanks and gratitude go to Professor Dr. Soe Win, Director General, Department of Higher Education, Upper Myanmar, who willingly helped me in gathering paper and gave me, other relevant information for my thesis. Without his generous cooperation and interest, this research work has not been fully accomplished.

I am greatly indebted to Professor Dr. Thet Lwin, Head, Department of Statistics (Retd.), Yangon Institute of Economics, for his generous editorial assistance in my thesis. I am really indebted to Professor Daw Khin San Myint, Head, Department of Statistics (Retd.), Yangon Institute of Economics, Professor Dr. Daw San Kyi, Department of Statistics (Retd.), Yangon Institute of Economics, Professor Dr. Daw Khin May Than, Head, Department of Economics, Yangon University of Distance Education, for their constructive suggestions, comments, expert opinions and generous editorial assistance in my thesis.

I owe great gratitude to my mother, my brother who gave all round support to me. Without their kind encouragement, support, and understanding, writing of this thesis will not be possible.

TABLE OF CONTENTS

LIST OF TABLES

Appendix Tables

Table No.	Title	Page

LIST OF FIGURES

Figure No. **Title** **Page**

Appendix Figures

Figure No. **Title** **Page**

LIST OF ABBREVATIONS

ACM	Approximate Conditional Mean
AM	Approximate Maximum Likelihood
AO	Additive Outlier
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARMA	Autoregressive Moving Average
BIF	Bounded Influence Function
BP	Breakdown Point
C	Cauchy Distribution
c.d.f.	Cumulative Distribution Function
EIF	Empirical Influence Function
EXP	Exponential Distribution
GAM	Gamma Distribution
GM-estimate	Generalized Maximum Likelihood Type Estimate
GS-estimate	Generalized Studendized Location Estimate
HBP	High Breakdown Point
i.i.d.	Independent and Identically Distributed
IC	Influence Curve
IF	Influence Function
IMSL	International Mathematical and Statistical Libraries
IO	Innovation Outlier
IRLS	Iterative Reweighted Least Squares
LAV	Least Absolute Values
LMS	Least Median Squares
LOG	Logistic Distribution
LQD	Least Quartile Difference
LS	Level Shift
LTS	Least Trimmed Squares
M-estimate	Maximum Likelihood Type Estimate
MA	Moving Average
MAD	Median Absolute Deviation
MCD	Minimum Covariance Determinant

MD	Mahalanobis Distance
ME	Mean Errors
ML	Maximum Likelihood
MML	Modified Maximum Likelihood
MSE	Mean Squares Error
MVE	Minimum Volume Ellipsoid
N	Normal Distribution
OLS	Ordinary Least Squares
R-estimate	Rank Test Estimate
RA	Residual Autocovariances
RD	Robust Distance
RDL_1	Robust Distance Least Absolute Values
REF	Relative Efficiency
RLS	Reweighted Least Squares
sacf	Sample Autocorrelation Function
S-estimate	Studentized Location Estimate
SIF	Sample Influence Function
spacf	Sample Partial Autocorrelation Function
SPSS	Statistical Package for Social Science
STATA	Statistics Data Analysis
TC	Temporary Change
TRA	Truncated Residual Autocovariances
UNDP	United Nations Development Programme
WLS	Weighted Least Squares

CHAPTER I

INTRODUCTION

It is generally known that classical statistics depend on the assumptions of parametric models. Typically, assumptions are made on the structural and the stochastic parts of the model and optimal procedures are derived and used under these assumptions. Standard examples include such estimators as least squares estimators in linear models and their extensions, maximum likelihood estimators and corresponding likelihood-based tests, etc. Many classical statistical procedures are known for not being robust, because results violate stochastic assumptions and rely on a few sample observations. These procedures are optimal when the assumed model is exactly satisfied, but they are biased and/or inefficient when small deviations from the model are present. The results obtained by classical procedures can therefore mislead when it comes to applications. Consequently, it may cause issues when the classical methods are used for the underlying model. Hence, one would naturally like to employ an estimation method that is sufficiently resistant to outliers.

The term “robust” was coined in statistics by G.E.P. Box in 1953. Various definitions of greater or lesser mathematical rigidity exist, but in general, referring to a statistical estimator, it means “insensitive to small departures from the idealized assumptions for which the estimator is optimized.” The word “small” can have two different interpretations, both important: either fractionally small departures for all data points, or else fractionally large departures for a small number of data points. It is the second interpretation that leads to the notion of outlier points, the most significant points to statistical procedures. The aims of robust statistics are:

1. To describe the structure of best fitting the bulk of the data,
2. To identify outliers (for possible further treatment),
3. To give a warning about highly influential data points (leverage points),
4. To deal with deviations from the assumed correlation structures.

Five to ten percent of wrong values in the data appear to be the rule rather than the exception. Outliers may appear in data due to (i) gross errors, (ii) wrong classification of the data (outlying observations may not belong to the model followed

by the bulk of the data), (iii) grouping, and (iv) correlation in the data (Hampel et al., 1986).

Gross errors often show themselves as outliers, but not all outliers are gross errors. Gross errors or outliers are data severely deviating from the pattern set by the majority of the data. This type of error usually occurs due to mistakes in copying or computation. They can also be due to part of the data not fitting the same model, as in the case of data with multiple clusters. Gross errors are often the most dangerous type of errors. In fact, a single outlier can completely spoil the least squares estimate, causing it to break down. Consequently, the estimators may not be efficient estimators. Some outliers are genuine and may be the most important observations of the sample. Rounding and grouping errors result from the inherent inaccuracy in collecting and recording data which are usually rounded, grouped, or even roughly classified. The departure from an assumed model means that real data can deviate from the assumed distribution. The departure from the normal distribution can manifest itself in many ways, for instance, in the form of skewed (asymmetric) or longer-tailed distributions.

The theory of robust statistics deals with deviations from the assumptions on the model. Examples of deviations include the contamination of data by gross errors, rounding and grouping errors, and departure from an assumed distribution. Robust statistics is concerned with the construction of statistical procedures which are still reliable and reasonably efficient in a neighborhood of the model that is pointed by Huber (1981), Hampel, Ronchetti, Rousseeuw, and Stahel (1986), Maronna, and Martin, and Yohai (2006). Therefore it can be viewed as a statistical theory dealing with approximate parametric models and a bridge between the parametric approach and the nonparametric approach. It is a reasonable negotiation between the inflexibility of a strict parametric model and the potential difficulties of interpretation of a full nonparametric analysis.

Fundamental theories on robustness were proposed by Huber (1964), and Hampel (1971). Their assumptions could be considered as the foundations of modern robust statistics.

1.1 Rationale of the Study

The concept of robustness is not new in the field of statistics. Box (1953) first gave the word a statistical meaning. Since then, rapid development of the theory of robustness has risen to alternative approaches to robust statistical theory.

The earliest discussions and applications of robustness go back as far as the eighteenth-nineteenth century (Hampel et al., 1986 and Barnett and Lewis, 1978). For instance, as regards astronomical observations, there were 'unrepresentative', 'rogue' or 'outlying' observations, that is, outliers. In the early 1800s, scientific debate was concerned with the effects of outlying observations on estimates calculated by ordinary least squares (OLS).

Hampel (1971) introduced the concept of qualitative robustness. The main idea is to complement the notion of differentiability (influence function) with continuity conditions, with respect to the Prohorov distance. Hampel et al. (1986) considered the relationship of continuity with respect to qualitative robustness. Two other concepts for judging the robustness performance of an estimator are efficiency robustness and min-max robustness, which are also directly applicable to time series data (Martin and Yohai, 1985 and Huber, 1973). Quantitative robustness is built on the concept of a breakdown point, whereas infinitesimal robustness incorporates the influence function as the critical concept. In this study, outlier robustness as well as distributional robustness in regression analysis will be emphasized. In time series, only outlier robustness will be observed.

1.2 Related Literature Review

There have been many researches and papers on robust statistics. A pioneering work on robust statistics is due to Huber (1964) and Hampel (1971). Robust statistics is the generalization of the classical theory: it takes into account of model misspecification, and the inferences remain valid not only at the parametric model but also in the neighborhood model.

In practice, there are many types of robust estimators. In the regression context, least absolute values (LAV or L_1) regression is very resistant to observations with unusual values of the dependent variable *Y*. Estimates are found by minimizing the sum of the absolute values of the residuals. Ashar and Wallace (1963) studied the statistical properties of regression parameters estimated by minimization of L_1 norm. Huber (1964) explored the properties of L_1 regression in its robustness to wild fluctuations in the magnitude of residual elements.

Meyer and Glauber (1964) for the first time directly compared L_1 and L_2 regression estimators. They estimated their investment model by minimization of L_1 as well as L_2 norm and tested the regression equations obtained on post-sample data by using those equations to forecast the nine (in some cases eleven) observations subsequent to the period of fit. They found that with very few exceptions, the equations estimated by L_1 minimization outperformed the ones estimated by L_2 minimization even on criteria (such as sum of squared forecast errors) with respect to which, L_2 regression is ordinarily thought to be remarkably suitable or optimal.

The study by Oveson (1968) in his Doctoral research on the LAV estimator gave a new thrust to the investigation into the properties and applicability of the estimator. It was almost fully established that in the presence of errors generated by thick-tailed distribution, L_1 regression performed better than L_2 regression.

Mosteller and Tukey (1977) poined out that LAV was less affected than OLS by unusual values of the dependent variable *y*, it failed to account for leverage and thus had a breakdown point (BP) of 0.

Kim and Muller (2000) presented the asymptotic properties of two-stage quantile regression estimators. In their paper, they derived the asymptotic representation of the estimators and proved the asymptotic normality with quantile regression predictions. The asymptotic variance matrix and asymptotic bias were discussed. They also analyzed the asymptotic normality and the asymptotic covariance matrix with OLS predictions. The results obtained permitted valid inferences in structural models estimated by using quantile regressions, in which the possible endogenous of some explanatory variables was treated via ancillary predictive equations. Simulation results illustrated the usefulness of this approach.

Furno (2000) compared the performance of least absolute deviation (LAD) and OLS in the linear regression model with random coefficient autocorrelated (RCA) errors. The presence of thick-tailed error distribution led to the estimation of the RCA model by LAD estimator. It is known that when error follows a double exponential distribution, LAD coincides with maximum likelihood. In all other cases, the estimator is less affected by observations coming from tails, since it minimizes the absolute value and not the squared value of the residuals. In case of leptokurtic error distribution, the LAD estimator is particularly useful. Furno proved that the LAD estimator for randomly autocorrelated errors is asymptotically normal. The more general random coefficient ARMA models for the error term was also considered in

the study and the resulting heteroskedasticity was analyzed. Monte Carlo experiments revealed that LAD improved upon OLS in case of RCA errors, both in terms of bias reduction and efficiency gains. However, in the case of constant autocorrelation model, the results confirmed that LAD is not advantageous, especially in small samples, since its sampling distribution differs from the asymptotic one.

Huber (1964) proposed maximum likelihood type estimation (*M*-estimation) for regression which was a relatively straightforward extension of *M*-estimation for location. It represented a compromise between the efficiency of the OLS estimators and the resistance of the LAV estimators, both of which could be seen as special cases of *M*-estimation.

Hampel at al. (1986) showed that *M*-estimators have higher statistical efficiency but tolerate much lower percentages of outliers unless properly initialized. Rousseeuw and Leroy (1987) found out that *M*-estimates of location are highly robust, having a bounded influence function and a breakdown point of 0.5. In the paper, they showed that *M*-estimates for regression share these attributes for *y* but not the *x*s, resulting in a breakdown point of 0. Moreover, in some situation they performed no better than OLS.

Hampel et al. (1986) pointed out that the *M*-estimator had unbounded influence because it failed to account for leverage. In response to this problem, bounded influence generalized *M*-estimators (*GM*-estimators) had been proposed. The goal was to create weights that consider both vertical outliers and leverage points. Outliers are dealt with using a standard *M*-estimator, and leverage points are typically down-weighted according to their hat value.

Least median of squares estimation (LMS or LMedS) is based on Hampel's idea and was later proposed by Rousseeuw (1984). LMS replaces the summing of the squared residuals that characterizes OLS with the median of the squared residuals. The idea is that by replacing the sum with the more robust median, the resulting estimator will be resistant to outliers. Although this result was achieved at a breakdown point (BP) of 0.5, LMS estimator had deficiencies that limit its use.

Rousseeuw and Leroy (1987) found out that the LMS estimator additionally possesses the "short" property, that is, it provides a regression hyper plane around which 50% of the observations are most tightly packed in terms of absolute deviation of the residuals.

Rousseeuw and van Zomeren (1990) and Atkinson (1994) showed that the short property makes the LMS estimator a convenient tool for identifying outliers and potential nonhomogeneity.

Davies (1992) pointed out that there are however many estimators with the same BP and a number of studies in the statistical literature compare their relative merits. Among the various estimators the LMS estimator while robust to gross contamination was characterized by lack of local robustness that was related to its sensitivity to "inliers" due to the unboundedness of the influence function (IF) at small values.

Campbell and Galbraith (1993) found out that the short properties indicated that the LMS estimator could be a useful tool in applications to data that were likely to lead to a breakdown of other estimators. Closeness of results for standard estimators and the LMS could be used as evidence that contamination/outliers did not present a significant problem for the standard methodology.

Finite sample comparisons were made by Ferretti et al. (1999). They showed that as expected LMS did relatively well when there was substantial gross contamination (demonstrating its global robustness) but not as well against local contamination (lack of local robustness). Since many financial data series were very accurately recorded, but exhibit occasional extreme volatility LMS may provide valuable information.

Least trimmed squares estimation (LTS) principle was introduced by Rousseeuw (1984) to overcome the efficiency problems of the LMS estimation technique. The performance of this method was improved by Rousseeuw and Van Driessen (1999).

Hample (1975), Rousseeuw and Yohai (1984) proposed studentized location estimates (*S*-estimates). They showed that *S*-estimates were the solution that found the smallest possible dispersion of the residuals. With the same breakdown value, it had a higher statistical efficiency than LTS estimation.

The *MM*-estimators first proposed by Yohai (1987) had become increasingly popular and were perhaps now the most commonly employed robust regression techniques. They combined a high breakdown point (50%) with good efficiency (approximately 95% relative to OLS under the Gauss-Markov assumptions). The ''*MM*'' in the name refers to the fact that more than one *M*-estimation procedure is used to calculate the final estimates. Following from the *M*-estimation case, iteratively

reweighted least squares (IRLS) is employed to find estimates. It has both the high breakdown property and a higher statistical efficiency than *S*-estimation.

Hubert and Rousseeuw (1997) proposed a method, called robust distance least absolute values (RDL_1), that first downweights the leverage points in the space of the continuous regressors and then follows a weighted least absolute value fit for both continuous and categorical regressors. In their paper they observed that RDL_1 may suffer from the swamping effect due to its weights for the L_1 procedure being obtained by only considering the continuous design matrix.

Kosinski (1999) proposed a new method for the detection of outliers which was very resistant to high contamination of data. As classical covariance matrix was very sensitive to outliers, alternative covariance matrix had been proposed. The minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) were two of several multivariate location and scale estimators. These estimators had high finite-sample breakdown point. The use of estimators with high finite-sample breakdown point yields good performance according to masking effect.

The estimation of coefficients in a simple linear model is one of the oldest and most important problems and has received tremendous attention in the literature in statistics and econometrics. Lawrence and Arthur (1990) pointed out that most of the work reported was, however, based on the assumption of normality. It had been recognized that the underlying distribution in most situations was basically not normal, especially in Economics and Finance proposed by Huber (1981). In his study the solution was to develop efficient estimators of coefficients in multiple regression model when the underlying distribution was non-normal. Naturally, one would prefer closed form estimators which were fully efficient (or nearly so). Preferably, these estimators should also be robust to plausible deviations from an assumed model. The underlying distribution was assumed to be symmetric and to be student's *t* family. The method of modified maximum likelihood (MML) estimation was invoked.

Tiku et al. (1999) developed the MML estimators for simple linear regression with symmetric innovation and came up with the MML estimators for the first order autoregressive model with symmetric Innovation; Tiku et al. (2001) treated the MML estimator for the simple linear regression model with innovation from student's *t* family while Bian and Tiku (1997) adopted the Bayesian approach to study a standard multiple regression model with identical and independently distributed (i.i.d.) error term.

Tiku et al. (1999), Wong and Bian (2005) pointed out that the MML estimators had been extensively demonstrated by simulation study to be robust and really efficient and clearly superior to the traditional normal-theory estimators in all the models being studied, including the autoregressive model, simple linear regression model and simple linear regression model with autoregressive innovation. As the multiple linear regression model was a simple extension of the above models, the properties of the robustness and efficiency for its estimators would be similar to that of the simple linear regression. As such, the resulting estimators were explicit functions of sample observations and were asymptotically fully efficient. Since they were almost fully efficient for small sample sizes, they were noted as being remarkably robust.

Several authors have considered robust estimation procedures for parameters of time series in the presence of outliers. Denby and Martin (1979) showed that the *M*-estimator was robust to innovation outliers (IO), but not to additive outliers (AO). In fact, they showed that *M*-estimators could have asymptotic bias nearly as large as least squares estimators in the AO case. Martin and Yohai (1985) pointed out that the performance of *M*-estimators in the presence of outliers, especially AO, was not satisfactory. This was because in ARMA models, contaminated Y_t's would also be included in explanatory variables.

Denby and Martin (1979) first proposed a class of generalized maximum likelihood estimates (*GM*-estimates) for AR(1) model in the presence of outliers of other type. It was shown that those *GM*-estimates performed moderately well in the presence of AO and IO, the *M*-estimates performed much better in the presence of IO. Martin (1979) extended the *GM*-estimates to AR(*p*) model. He also discussed some theories and methodologies of robust estimation for time series with AO and IO as well as the problem of patch outliers.

Martin and Yohai (1985) pointed out that the performance of *M*-estimators in the presence of outliers, especially AO, was not satisfactory. In their paper, they showed that the *GM*-estimator could handle both AO and IO quite successfully in the AR(1) model. The problem with the *GM*-estimator was its performance when the order of AR structure increases. More specifically, since the *GM*-estimator was a low break down point estimator, when the proportion of outliers increased in the AR model, it will break down.

Martin and Yohai (1985) proposed a new class of estimators that were based on the combined use of robust filtering and *S*-estimates. For the case of an AR(1) model, they showed that these estimates dominated both the bounded influence (BIF) autoregressive estimator and the estimator suggested by Chang et al. (1988), while the BIF estimator dominated the estimator suggested by Chang et al(1988).

Busto and Yohai (1986) pointed out that the LS estimate was not very sensitive to the presence of innovation outliers. It was shown that if the innovations had finite variance, the asymptotic covariance matrix was independent of the distribution *F*.

It was pointed out by Bustos and Yohai (1986) that the *GM*-estimator has a complicated asymptotic covariance matrix. They proposed two new robust estimators based on residual autocovariances estimators (RA estimators) and truncated residual autocovariances estimators (TRA estimators). The proposed estimators were compared with LS estimator, *M* and *GM*-estimators for AR(1) and MA(1) models with AO and IO outliers. Based on Monte Carlo results, it was shown that RA estimators were not qualitatively robust when the model had the moving average part but much stable than LS and *M*-estimators in the presence of AO.

Based on re-weighted maximum likelihood estimator using Huber or redescending weights, Luceno (1998) proposed robust estimators in the presence of nonconsecutive multiple outliers in Autoregressive Moving Average (ARMA) (*p, q*) series. Another attempt in this direction was by Pena (1984) who discussed sample influence function for parameters in the presence of outliers in ARMA model.

In the area of robust estimations, high breakdown point (HBP) estimators have been widely suggested over the last decade. HBP estimators were concerned with the concept of the breakdown point of an estimator. Intuitively, the breakdown point measures the largest possible proportion of outliers in the data set an estimator can tolerate before it collapses to some nonsensical value. Maddala and Yong Yin (1997) pointed out that the performance of the *M*-estimator was not always satisfactory, HBP estimators have been introduced into time series analysis as well. Rousseeuw and Yohai (1984) introduced the class of *S*-estimators and Yohai (1987) introduced the *MM*-estimator in time series analysis.

Soe Win (2004) proposed adjustment diagnostics measure based on error variance (ADV) to identify the correct type of outlier and developed statistical time series diagnostic software (STDS) to diagnose the outliers in time series. The ADV

procedure can work well for isolated outliers but does not result satisfactorily in handling patch outliers.

Mya Thanda (2010) studied the detection and identification of outliers using the likelihood ratio and the adjustment diagnostics procedures. In this study, the simulation data as well as selected economic time series data were used. The simulated results suggested that the percentage of correct model selection declined as the value of outliers increased in the AO whereas it was not true in the IO of both AR(1) and MA(1) series. Based on the outliers detected, the most fitted model of each series was constructed for forecasting purpose.

1.3 Objectives of the Study

The objectives of the study are as follows:

(i) To analyze the robust estimates compared with the classical estimates and

(ii) To estimate the parameters of regression model and time series model by using simulation data as well as real data,

(iii) To explore the best estimation method among all available robust techniques used in regression and time series for simulation data and real data having heavy-tailed distributions.

1.4 Research Methodology

The study attempts to analyze the effect of outliers on estimation of parameters in regression and time series analysis. In attempting to achieve the objectives of the study, classical and robust methods are used, based on simulated as well as real data. Regarding regression analysis, two major problems such as outlier robustness and distributional robustness are investigated. The required data sets are generated by using multiple linear regression models with three explanatory variables. Then, these data sets are transformed into outlier contaminated data sets. After that, the performances are compared in terms of bias and MSE criteria and then the most suitable estimation method is chosen. For the time series analysis, the outlier robustness is considered. The simulated data sets are generated using AR(1) and MA(1) models. Then, the clean data gained are transformed into outlier contaminated data sets using AO and IO outliers. The results are compared in terms of ME and MSE criteria and then the most suitable estimation method is explored. With regard to

real data, secondary data are used in the study. These data are obtained from the Central Statistical Organization and UNDP.

1.5 Scope and Limitations of the Study

Simple linear regression and multiple linear regression methods are taken into account for the study. However, non-parametric regression analysis is excluded from the study. In this study, five well-known methods of linear regression like LAV, *M*, *MM*, LMS and LTS estimators are included but *L*, *R*, and *GS*-estimators are excluded. And for the time series analysis, only univariate case is considered. Four robust estimators (*M*, *GM*, ACM and RA) are considered but approximate maximum likelihood type (AM), truncated residual autocovariances (TRA) and filter smoother estimators are excluded in this study. Some types of outliers which can occur in a time series, only additive outlier (AO) and innovation outlier (IO) are used. Only non-seasonal Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) time series models are considered in this study. No comparison has been made between robust procedures and outlier detection procedures used in estimating parameters of contaminated models in this study.

1.6 Organization of the Study

This study is composed of six chapters. Chapter I is the introduction which includes rationale of the study, related literature review, objectives of the study, research methodology, scope and limitations of the study and organization of the study. Then, the concepts of robustness are shown in chapter II. In this chapter, influence function, gross-error sensitivity, local-shift sensitivity, rejection point, maximum-bias curve, breakdown point, min-max robustness and qualitative robustness are discussed. Then, robust estimators and their possible applications to regression and time series are mentioned in chapter III. After that, in chapter IV simulated data as well as real data are used to analyze the effect of the outliers in regression by using the OLS method and robust methods. In chapter V, certain robust methods for AR(1) and MA(1) models with AO and IO outliers are analyzed by using simulated data sets. In this chapter, the performances of classical and robust procedures in the presence of outliers in an ARMA model are analyzed by using real data set. Chapter VI is the conclusion.

CHAPTER II

CONCEPTS OF ROBUSTNESS

As mentioned in the earlier chapter, robustness means insensitivity to small departures from idealized assumptions for which the estimator is optimized. Robustness is usually used in the context of distributional robustness that is the actual noise distribution deviates from the nominal distribution. In most cases, the nominal noise distribution is independent and identically distributed (i.i.d.). The deviations, however, may also be due to model class selection errors, or there may be more than one statistical population present in the data set, and hence it is not possible to describe only with one set of parameters. Robust methods can be considered to be approximately parametric, that is, a parametric model can be used but some deviations from the strict model are also allowed.

To know more clearly about the robustness concept, mathematical setup has to be described first because it allows us to formalize the robustness thoughts. The notion of the sensitivity of an estimator T is put into theory considering a model characterized by a cumulative distribution function (c.d.f.) F and its neighborhood $F_{\varepsilon,G:}$ distributions $(1-\varepsilon)F+\varepsilon G$, where $\varepsilon \in (0,1/2)$ and G is an arbitrary probability distribution, which represents data contamination. Hence, not all data necessarily follow the pre-specified distribution, but the ε-part of data can come from a different distribution G. If $H \in F_{\varepsilon,G}$, the estimation method T is then judged by how sensitive or robust are the estimates $T(H)$ to the size of $F_{\varepsilon,G}$, or alternatively, to the distance from the assumed c.d.f. F. Two main concepts for robust measures analyze the sensitivity of an estimator to infinitesimal deviations, $\varepsilon \to o$, and to finite (large) deviations, $\varepsilon > 0$, respectively. Despite generality of the concept, easy interpretation and technical difficulties often limit our choice to point-mass distributions (Dirac measures) $G = \delta_x, x \in \mathbb{R}$, which simply represents an (erroneous) observation at point $x \in \mathbb{R}$.

The following are the basic concepts of robustness, statistical robustness and robustness properties.

2.1 Basic Concepts of Robustness in Statistics

Robustness, a fundamental concept in statistics, was first introduced by Box (1953). Since then, the theory of robustness has been developed as an important part in the field of statistics (Huber, 1981 and Hampel et al., 1986). The notion of robustness has different interpretations. One interpretation would be the distributional robustness, that is, robustness of the assumed model to minor departures from the model assumptions. Outlier-resistant or distributionally robust (so-called robust) statistics methods aim at constructing statistical procedures that are stable (robust) even when the underlying model is not perfectly satisfied by the available data set. An example of departure from the assumed model is the presence of outliers – observations that are very different from the rest of the data. Outliers are "bad" data in the sense that they deviate from the pattern set by the majority of the data. Hence, they tend to be doubtful in its common flow and may lack explanatory and predictive power regarding the common portion of the data. Robust models focus on the statistical properties of the bulk of the data without being distracted by outliers, while in classical models all data equally participate in the analysis.

The presence of outlying events, the so-called "low frequency / high severity" events, in the data creates a following inconsistency. On the one hand, the tail events correspond to the data that, despite their low frequency of occurrence, are often the most destructive for the institution as shown in Figure 2.1. In this sense, they cannot be ignored as they express important information regarding the process and may signal important flaws in the system. On the other hand, recent empirical findings suggest that classical methods will frequently fit neither the bulk of the data nor the outliers well, and the center and the tails of the data appear to conform to different laws.

Classical estimators that assign equal importance to all available data are highly sensitive to outliers and in the presence of just a few extreme values can produce arbitrarily large estimates of mean, variance, and other vital statistics. For example, a high mean and standard deviation values for data do not provide an indication as to whether this is generally due to large values of observations or just one outlier, and it may be difficult to give the right interpretation to such results.

On the contrary, robust methods take into account the underlying structure of the data and "separate" the bulk of the data from outlying events, a way of avoiding the upward bias in the vital statistics and forecasts. Robust methods do not aim at

throwing away extreme observations. They focus on the behavior of the bulk of the data that can be easily distorted by outliers. An important application of robust statistics uses them as a diagnostic technique to evaluate the sensitivity of the inference conducted under the classical model to the rare events and to reveal their possible economic role.

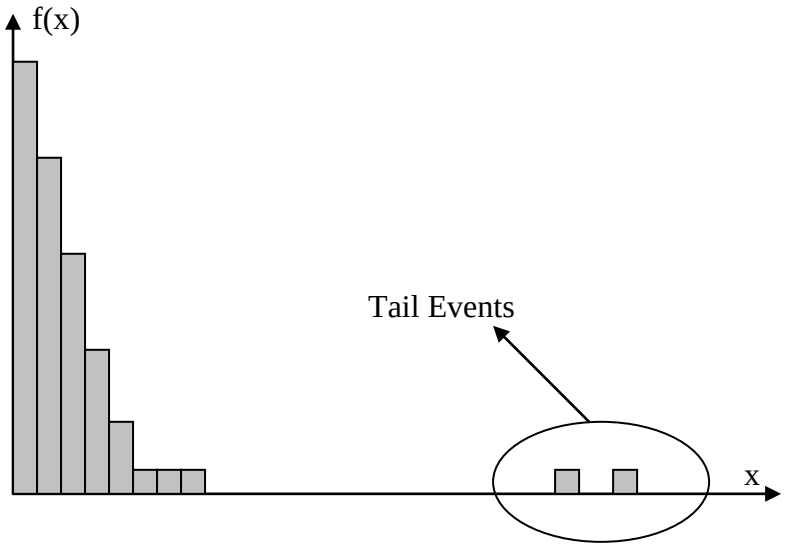


Figure 2.1 Histogram of Hypothetical Data

The classical model and the robust model are not competitive. The uses of both models are complementary to each other. Instead the use of the robust model should be encouraged. The results from both approaches are not expected to be the same, as they explain different phenomena dictated by the original data: the general tendency (the robust method) and the conservative view (the classical method).

Conducting robust or classical analysis of the data is a trade-off between safety and efficiency: although some information may be lost while discarding or diminishing the contribution of the outlying events, one can significantly improve forecasts and produce more reliable estimates by applying robust theory.

2.1.1 Statistical Robustness

Robustness, in general, refers to the ability of a procedure or an estimator to produce results that are insensitive to departures from ideal assumptions. This definition of robustness covers all scientific research. Lucas (1996) made the remark that robustness is a fascinating subject both from a theoretical and practical point of

view. With reference to general robustness theory, Hampel et al. (1986) defined statistical robustness as follows:

In a broad informal sense, robust statistics is a body of knowledge, partly formalized into ‘theories of robustness’, relating to deviations from idealized assumptions in statistics.

2.1.2 Outlier Robustness

Outliers can be thought of as observations in a data set that cause surprise in relation to the majority of the data. For example, surprising or extreme observations might be unusually large or unusually small values compared to the remaining data. Outliers are a common occurrence in data. They may be the result of an error in measurement of recording or transmission errors of exceptional phenomena such as earthquakes or strikes, or they may be due to the samples not being entirely from the same population. Apparent outliers may also be due to the values being the same, but nonnormal (in particular, heavy-tailed) distribution.

Outliers should be investigated carefully. Often they contain valuable information about the process under investigation or the data gathering and recording process. Before considering the possible elimination of these points from the data, one should try to understand why they appeared.

Outliers can be classified in statistics as outlying observations in linear regression, time series analysis, survey, directional and contingency table data (Barnett and Lewis, 1978). In the regression context, outliers are classified as *y*- and *x*-outliers. Outliers always entail both theoretical and practical problems. Usually, depending on our goal(s), we need one or more procedures that are robust, to protect against and detect outlying observations in the data. For instance, in the case of a forecasting model, it is of utmost importance to be able to detect, estimate the effects of, and interpret outliers. In some cases, outliers in a residual series may indicate omission of an explanatory variable from the model. Furthermore, the robust regression estimates are less biased than OLS and provide estimates of outliers that are more strikingly seen in residual series.

Many robust methods have been developed to handle data contaminated with outliers. Such methods are said to have outlier robustness. These robust methods can be used to detect outlying observations and to provide resistant results which are stable in the presence of outliers. For instance, if we are interested in estimating a

model parameter for a data contaminated with outliers from a random measurement error, it is of interest to use an estimator which is not sensitive to such outlying observations.

2.2 Main Concepts of Statistical Robustness

In this section, the main concepts of statistical robustness such as influence function, gross-error sensitivity, local-shift sensitivity, rejection point, maximum-bias curve, breakdown point, min-max robustness were described.

2.2.1 Influence Function

The influence function (IF) or the influence curve (IC) is a local robustness measure. By definition, an IF measures the change in the value of an estimator when outliers are added to the sample. The IF is essentially the first derivative of an estimator, viewed as functional, at some distribution (in an infinite-dimensional space), and it is shown how it can be used to several local robustness properties which are defined and intuitively interpreted. The study of influence curves serves to extend our understanding of estimators (e.g., the three-part descending *M*-estimators, or the optimal robust estimators of scale, with the median deviation as limiting case). The influence curve can be drawn and looked at, and its various properties (qualitative shape, supremum, maximal slope, points and heights of jumps, points and intervals where it is zero, etc.) together with a bit of qualitative information about type or regularity of the estimator (how the influence curve behaves in a neighborhood, and how the limit which defines it is approached) tell us a lot about the detailed behavior of the estimator and about how the separate observations contribute to the estimated value.

The influence of infinitesimal contamination on an estimator is characterized by the IF, which measures the relative change in estimates caused by an infinitesimally small amount ε of contamination at x (Hampel et al., 1986). Let δ_x denote the probability measure which puts the unit mass at the point x, that is, the c.d.f. with a point mass at x. The IF of an estimator T at F is given by

$$IF(x;T,F) = \lim_{\varepsilon \to 0} \frac{T\{(1-\varepsilon)F + \varepsilon \delta_x\} - T(F)}{\varepsilon} \qquad (2.1)$$

at those x where the limit exists. In (2.1) T (F) denotes the value of the estimator of

the original distribution *F*. Similarly, $T\{(1-\varepsilon)F+\varepsilon\delta_x\}$ denotes the value of this estimator of the slightly contaminated distribution $(1-\varepsilon)F+\varepsilon\delta_x$. ε refers to the fraction of the perturbation, $0 < \varepsilon < 1$.

For each point x, the IF reveals the rate at which the estimator *T* changes if a wrong observation appears at x. In the case of sample mean $\bar{x} = T(F_n)$ for $\{x_i\}_{i=1}^n$ we obtain

$$IF(x;T,F_n)=\lim_{\varepsilon\to 0}\left\{(1-\varepsilon)\int udF_n(u)+\varepsilon\int ud\delta_x(u)-\int udF_n(u)\right\}/\varepsilon$$

$$=\lim_{\varepsilon\to 0}\left\{-\int udF_n(u)+\int ud\delta_x(u)\right\}=x-\bar{x}. \qquad (2.2)$$

The IF is the first derivative of a statistic *T* of an underlying distribution *F*. If the IF is bounded, then the effects of a small number of outliers are also bounded. As for its interpretation, one can say that IF measures the asymptotic (standardized) bias of the estimator *T* caused by contamination of *F*. There are also some studies in the literature on finite-sample versions of the IF, that is, the empirical influence function (EIF) and sample influence function (SIF). Not all estimators have an influence function, but all of them have a breakdown point. The IF can be further shown to be of the form

$$\text{IF} = (\text{constant})\ \psi(u).$$

Hence, the shape of IF depends only on the shape of the ψ-function, not the data distribution.

2.2.2 Gross-Error Sensitivity

The gross-error sensitivity expresses asymptotically the maximum effect a contaminated observation can have on the estimator. It is the maximum absolute value of the IF. The asymptotic bias of an estimator, defined as the maximum effect of the contamination of a given distribution with a proportion ε from an outlying distribution, is giving by ε (gross-error sensitivity). Unfortunately, it was reported that in general, poor gross-error sensitivity corresponds to higher Gaussian efficiency, and vice versa. The IF allows us to define various desirable properties of an estimation method. The largest influence of contamination on estimates can be formalized by the gross-error sensitivity,

$$\gamma(T,F)=\sup_{x\in\mathbf{R}} IF(x;T,F), \tag{2.3}$$

which under robustness consideration is finite and small. Even though such a measure can depend on *F* in general, the qualitative results (e.g., $\gamma(T,F)$being bounded) are typically independent of *F*. Hample (1974) calls (2.3) the gross-error sensitivity. It may be regarded as a measure of the maximum possible influence of any observation on the estimated coefficients. A disadvantage of (2.3) is that it is not location or scale invariant.

2.2.3 Local-Shift Sensitivity

Local-shift sensitivity measures the effect of the removal of a mass ε at *y* and its reintroduction at x. Therefore, it measures the effect of rounding and grouping errors on an estimator. For highest resistance, it is required that the local-shift sensitivity be bounded. For a continuous and differentiable IF, local-shift sensitivity is given by the maximum absolute value of the slope of IF at any point. The sensitivity to small changes in data, for example moving an observation from x to $y \in \mathbb{R}$, can be measured by the local-shift sensitivity

$$\lambda(T,F)=\sup_{x\neq y}\frac{\left\| IF(x;T,F)-IF(y;T,F)\right\|}{\left\|x-y\right\|}. \tag{2.4}$$

Also this quantity should be relatively small since we generally do not expect that small changes in data cause extreme changes in values or sensitivity of estimates. In general, a lower (hence better) local-shift sensitivity corresponds to higher Gaussian efficiency.

2.2.4 Rejection Point

The rejection point is defined as the point beyond which IF becomes zero. Except possibly through the auxiliary scale estimate, observations with residuals beyond the rejection point have zero influence. Hence they make no contribution to the final estimate. Estimators which have a finite rejection point are said to be redescending and are well protected against very large outliers. However, a finite rejection point usually results in the underestimation of scale. This is because when

the samples near the tails of a distribution are ignored, too little of samples may remain for the estimation process. This in turn adversely affects the efficiency of the estimator. An estimator is efficient if the variance of its estimate is as close as possible to the variance of the best estimator for a given distribution. For the Gaussian distribution the best estimator is the mean which also yields the minimum variance of the estimate. In general, it is best for a robust estimator to use as many of the good samples of a distribution as possible, in order to maintain a good efficiency. Another adverse effect of finite rejection is that if a large part of the sample is ignored, the objective function may have many local minima. As unlikely large or distant observations may represent data errors, their influence on estimates should become zero. Such a property is characterized by the rejection point,

$$\rho(T,F)=\inf_{r>0}\left\{r : IF(x;T,F)=0, \|x\|\geq r\right\}, \tag{2.5}$$

which indicates the non-influence of large observations. Alternatively, behavior of the estimator T can be studied for any finite amount ε of contamination.

2.2.5 Maximum-Bias Curve

The most common property viewed in this context is the estimator's bias $b(T;H)=E_H\{T_H\}-E_F\{T_F\}$ which measures a distance between the estimates for clean data, $T(F)$, and contaminated data, $T(H)$; $H\in F_{\varepsilon,\mathrm{G}}$ .The corresponding maximum-bias curve measures the maximum bias of T on $F_{\varepsilon,\mathrm{G}}$ at any ε:

$$B(\varepsilon,T)=\sup_{x\in\mathbf{R}} b\{T,(1-\varepsilon)F+\varepsilon\delta_x\}. \tag{2.6}$$

Although the computation of this curve is rather complex, Berrendero and Zamar (2001) provide general methodology for its computation in the context of linear regression. The maximum-bias curve is not only useful on its own, but allows to define further scalar measures of robustness.

2.2.6 Breakdown Point

The breakdown point (BP) introduced by Hampel (1971) is a measure of global stability for a statistical functional and as such is a typical robustness measure. However, the quest for high breakdown point estimators in the field of robust statistics has pushed the development, among other things, of general computational techniques and resampling algorithms which can be used in more general settings (Rousseeuw, 1984 and Rousseeuw and Leroy, 1987).

The BP is a global robustness measure (of reliability); it is often the first and most important number to be looked at before going into the details of local robustness properties. The BP takes values between 0 and 1. The BP is essentially the largest fraction of contamination which does not ruin an estimate. The definition of the BP contains no probability distributions. It is also often quite simple: for the arithmetic mean it is 0, for the median it is 1/2 (slightly less than 1/2 of the data can move to infinity while the median still stays in the range of the "good" data). Among scale estimators, standard deviation, mean deviation and range all have BP of 0, while the interquartile range (difference between 3^{rd} and 1^{st} quartile, perhaps with a factor) has BP = 1/4. But the counterpart of the median among scale estimators is the median (absolute) deviation or "MAD" (Hampel, 1974), which is the median of the absolute differences of the data from their median, and which has BP = 1/2. A zero value for the BP reflects extreme sensitivity of an estimator to outliers.

An estimator's robustness can be characterized in several ways. One concept that has received a lot of attention in recent years is an estimator's BP (Donoho and Huber, 1983 and Rousseeuw and Leroy, 1987); it is the smallest fraction of contamination that can produce an infinite bias in the estimator. For instance, in a univariate sample of size n, the average can be increased without limit if any single observation is made arbitrarily large; accordingly, the BP of the average is $1/n$, or zero asymptotically. On the other hand, if all the observations that exceed the sample median are increased arbitrarily, the median is unchanged; so its BP is essentially 50%. In other words, the BP accurately characterizes the average's well-known lack of robustness and the sturdiness of the median.

Contamination in excess of an estimator's BP is a sufficient condition for an indefinitely large bias, but it is not a necessary condition. As a practical matter, the extent of the bias obviously depends not only on the number of outliers but also on their magnitudes. Another important consideration is the fit between the

uncontaminated data and the model. In a linear regression where the valid observations have a very high value of coefficient of determination (R^2), the bias induced by several stray data points may be small even if the estimator itself is not highly robust. Conversely, if the valid observations have a low R^2, application of a high breakdown estimator may not avoid a large (but finite) bias due to the presence a few outliers. Accordingly, the BP is similar to estimation criteria like efficiency and consistency; its usefulness ultimately depends on its performance in actual applications (Zaman et al., 2001).

An important robustness theorem states that "50% is the highest possible value for the BP, since for larger amounts of contamination it becomes impossible to distinguish between the good and the bad parts of the sample". Actually, 50% is an asymptotic value for the maximum BP; in finite samples, the value is reduced by a degrees-of-freedom adjustment.

Like the univariate average, OLS regression is quite vulnerable to aberrant observations not only in the dependent variable ("regression outliers") but also among the regressors ("bad leverage points"). Either sort of data problem can produce a large bias, so OLS has a BP of $1/n$. On the other hand, the L_1 norm minimizes the sum of the absolute values of the residuals; as such, it estimates the conditional median of the dependent variable and might be expected to inherit the robustness of the univariate median. The L_1 norm is in fact highly resistant to regression outliers, but it performs no better than OLS when there are bad leverage points among the regressors; so its BP is also $1/n$ (Rousseeuw and Leroy, 1987).

The maximum BP is attained by LMS, LTS, *S*-estimators, and other procedures that behave like multivariate versions of the mode. These estimators are based on the 50% of observations that cluster most tightly around the regression plane, and they are unaffected by data lying outside that cluster. As a result, the high-breakdown methods are quite robust, but they are inefficient when the data set is uncontaminated. A researcher can apply a high-breakdown estimator initially and follow up with a more efficient estimator once any inconsistent observations have been identified, scrutinized, and either reinstated, down weighted, or removed (Rousseeuw and Leroy, 1987, Yohai, 1987, and Yohai and Zamar, 1988).

The most important is the BP (Hampel, 1971), which is defined as the smallest amount ε of contamination that can cause an infinite bias:

$$\varepsilon^*(T) = \inf_{\varepsilon \geq 0} \{\varepsilon : B(\varepsilon, T) = \infty\}. \quad (2.7)$$

The intuitive aim of this definition specifies the BP $\varepsilon^*(T)$ as the smallest amount of contamination that makes the estimator *T* useless. Note that in most cases $\varepsilon^*(T) \leq 0.5$. This definition and the upper bound however apply only in simple cases, such as location or linear regression estimation. The most general definition of BP formalizes the idea of "useless" estimates in the following way: an estimator is said to break down if, under contamination, it is not random anymore, or more precisely, it can achieve only a finite set of values. This definition is based on the fact that estimates are functions of observed random samples and are thus random quantities themselves unless they fail. Although the latter definition includes the first one, the latter one may generally depend on the underlying model *F*, for example in time series context.

Finite Sample Breakdown Point

The standard definition (Rousseeuw and Leroy, 1987) of a finite sample breakdown point in a regression estimator is as follows. Consider a regression estimator *T*(*X*); where *X* represents a sample of *n* observations, $X'(m,n)$ coincides with *X* for all but *m* of the observations that have been replaced with arbitrary unbounded values and define

$$bias(m, T, X) = \sup \| T(X'(m,n)) - T(X) \|, \quad (2.8)$$

where supremum is taken over all possible $X'(m,n)$. If the bias $= \infty$; it is said that the estimator "breaks down". The BP is then defined as the minimum amount of contamination to cause the breakdown of an estimator:

$$\varepsilon(T, X) = \min\left\{\frac{m}{n}; bias(m, T, X) = \infty\right\}. \quad (2.9)$$

Clearly the OLS estimator has a finite sample breakdown point 1/*n*; the LAV estimator by construction is flexible to high contamination in regression errors but not in the regressors and in the event of such contamination breaks down at 1/*n*. A number of regression estimators are resistant to contamination of the regressors as

well as errors. An example of such an estimator is LMS estimator introduced by Rousseeuw (1984) which has a finite sample breakdown point close to 1/2 ; specifically of Rousseeuw and Leroy (1987) demonstrates that generally the BP of LMS is $\frac{(\lfloor n/2 \rfloor - k + 2)}{n}$ where k is number of regressors.

2.2.7 Min-Max Robustness

The minimax approach borrowed from game theory was Huber (1964) elegant solution of the robustness problem, viewed as a game between nature which chooses a distribution of the data in a neighborhood of the model and the statistician who chooses an estimator in a given class. The payoff is the asymptotic variance of the estimator at a given distribution. Sometimes minimax solutions can be pessimistic, but it turned out that this was not the case here. The resulting estimator, Huber's estimator, became the basic building block of any robust procedure and is a basic tool beyond robust statistics. Many estimators have been developed for these problems such as *M*-estimates, LTS, LMS, *S*-estimates, LAV or L_1 estimates, rank test estimates (*R*-estimates) and *MM*-estimates.

2.3 Robustness Properties

Robustness properties can be formulated within two frameworks: ***qualitative*** and ***quantitative*** robustness. Qualitative robustness is concerned with the situation in which the shape of the underlying (true) data distribution deviates slightly from the assumed model. It focuses on questions like stability and performance loss over a family of such slightly deviating distributions. Quantitative robustness considers the situation in which the sensitivity of estimators to a proportion of unusual observations is studied.

Another important property of an estimator is equivariance. A linear regression estimator is equivariant if it transforms properly when a variable (either dependent or regressor) is recentered or rescaled. For example, if each observation on a particular continuous-valued regressor is multiplied by a positive constant *c*, the estimated regression coefficient should change by the factor 1/*c*. Some widely used equivariant estimators are OLS, the L_1 norm (least absolute deviations), and LTS (Rousseeuw and Leroy, 1987). On the other hand, orthogonal regression is a method

that lacks equivariance because the estimated coefficients are changed in a nonlinear way when a variable is rescaled.

Another concept (Tukey, 1976), closely related to robustness and which may be considered its data-oriented counterpart, is resistance. A resistant sequence of estimates should be (i) stable in the presence of a few atypical outliers (that is a small percentage of abnormal observations should not influence the estimates too much) and (ii) stable when all of the observations have small errors (e.g., round-off errors). Some robust estimators were discussed in the next chapter.

CHAPTER III

ROBUST ESTIMATORS

In this chapter different types of robust statistical estimators which were developed by several statisticians were discussed in the context of regression and time series analysis. From section 3.1 through section 3.14, some salient features as well as outstanding robustness features, together with respective drawbacks, of each of these robust estimators are presented.

A robust estimator is one whose performance remains unchanged, satisfactory and attractive even when the true distribution of the data deviates from the assumed distribution. Data sets, for which one often makes a Gaussian assumption, sometimes contain a small fraction of unusually large values or outliers. In this situation, typical statistical estimators might be affected by the cause of these unusually large values.

The goals of robust estimation are to find estimates (i) that are highly efficient under a central model and (ii) such that small changes in the distribution of sample produce small changes in the distribution of the estimates. Robust methods can be considered to be approximately parametric, that is, a parametric model is used but some deviations from the strict model are allowed.

Many statisticians in the field of robust inference have developed various sorts of robust statistical estimators. These estimators are different from each other in real situations. Some estimators are usually used in regression analysis, some are in time series and some are in both. In this chapter, some robust statistical estimators, which are usually applied for regression and time series analysis, are presented.

Robust techniques, including those sometimes labeled as resistant techniques, in an evolutionary manner, explaining how new methods evolved in response to limitations of existing ones. Several classes of estimators are discussed. They are: linear combinations of order statistics (*L*-estimators); least absolute values estimators (LAV); *M*-estimators (extending from *M*-estimates of location by considering the size of the residuals); generalized *M*-estimators (*GM*-estimators, which extend *M*-estimators by giving less weight to high influence points as well as to large residual points); least median squares estimators (LMS); least trimmed squares estimators (LTS); *R*-estimators (based on the ranks of the residuals); *S*-estimators (which

minimize a robust *M*-estimate of the residual scale); generalized *S*-estimators; *MM*-estimators (which are built on both *M*-estimation and *S*-estimation to achieve a high breakdown point with high asymptotic efficiency); minimum volume ellipsoid estimator (MVE); minimum covariance determinant estimators (MCD); residual autocovariances estimators (which replace with the robust estimate of the residuals autocovariances) (RA); and approximate conditional mean robust filter (ACM).

Some of these methods could be considered obsolete, but general descriptions are still provided in many research works in the literature of robustness because more recent developments in robust regression build on them. Among the above mentioned estimators, LAV, *M*, *GM*, LMS, LTS, *S*, and *MM*-estimators are usually used in regression as well as time series analysis. But, MVE and MCD estimators are usually used in regression analysis. However, RA and ACM estimators are usually used in time series analysis only. In this study, the robust estimators such as the LAV, *M*, LMS, LTS, and *MM*-estimators were chosen to analyze the effect of outliers in regression analysis. Moreover, the *M*, *GM*, ACM and RA estimators were selected to study the effect of outliers' contamination on the estimation of parameters of time series models.

3.1 Linear Combination Estimators

Linear combination estimators (*L*-estimators) are linear combinations of order statistics. They are of the form:

$$T_n(x_1,...,x_n)=\sum_{i=1}^{n} a_i x_{i:n}, \tag{3.1}$$

where $x_{1:n},...,x_{n:n}$ are the ordered samples of size n and the $a_i's$ are coefficients. One of the most widely used *L*-estimators for location estimation is α-trimmed mean, where αn samples from the both ends of the ordered set of samples do not contribute to the estimate.

As stated above, any estimator that is computed from a linear combination of order statistics is classified as an *L*-estimator. The first *L*-estimation procedure, which is somewhat more resistant than OLS, is least absolute values (LAV) estimator. It is also known as L_1 regression because it minimizes the L_1-norm (sum of absolute deviations); LAV is the simplest and earliest approach to bounded influence robust regression, predating OLS by about 50 years. Least squares regression also fits this

definition, and thus it is sometimes referred to as L_2, reflecting that the L_2-norm (the sum of squared deviations) is minimized. Other well-known *L*-estimators are the least median of squares (LMS) and the least trimmed squares (LTS) estimators.

3.2 Least Absolute Values Estimators

Due to the squaring of the residuals, OLS becomes extremely vulnerable to the presence of outliers. To handle with this, Edgeworth (1887) proposed a method consisting in minimizing the sum of the absolute values of the residuals rather than the sum of their squares. Least absolute values (LAV), sometimes called L_1 or least absolute deviation (LAD) regression, is also known as median regression. The LAV estimator is very resistant to observations with unusual *y* values. The LAV estimates are found by minimizing the sum of the absolute values of the residuals

$$\sum_{i=1}^{n} \left| y_i - \mathbf{x}_i' \boldsymbol{\beta} \right| = \sum_{i=1}^{n} \left| e_i \right| . \tag{3.2}$$

This estimator does protect against vertical outliers but not against bad leverage points. It has an efficiency of only 64% at a Gaussian error distribution (Huber, 1981).

The LAV can be seen as a case of the more general quantile regression. In this case, the objective function to be minimized can be written as

$$\sum_{i=1}^{n} \rho_{\alpha}(e_i), \tag{3.3}$$

where

$$\rho_{\alpha}(e_i) = \begin{cases} \alpha\, e_i & if\, e_i \geq 0 \\ (\alpha - 1) e_i & if\, e_i < 0 \end{cases} \tag{3.4}$$

and α is the quantile being estimated.

Although LAV is less affected than OLS by unusual *y* values, it fails to account for leverage (Mosteller and Tukey, 1977), and thus has a breakdown point BP of 0. Moreover, LAV estimates have relatively low efficiency. Following the case of the mean, under the assumption that y ~ $N(\mu, \sigma^2)$, the sampling variance of *y* for OLS is σ^2 / n; for LAV it is $\pi / 2 = 1.57$ times larger at $\pi\sigma^2 / 2n$ (in other words, about 64%

efficiency). The combination of the low breakdown point and low efficiency makes LAV less attractive than other robust regression methods still to be discussed.

3.3 Maximum Likelihood Type Estimators

Huber introduced the notion of *M*-estimators in (1964) (Hampel et al., 1986) which opened new gates in the theory of classical statistics. Afterwards several *M*-estimators were proposed from time to time and the theory of *M*-estimators got enriched by every day passed. A brief discussion over the concept of *M*-estimators is given in this section. In simple terms, the *M*-estimator minimizes some function of the residuals. The linear regression model is given by:

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \cdots + \beta_p X_{i,p} + \varepsilon_i \quad i = 1,2,...,n. \tag{3.5}$$

From this equation, the parameters $\beta_0, \beta_1, ..., \beta_p$ are estimated and then the fitted model is written as follows:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \cdots + \hat{\beta}_p x_{ip} + e_i = x_i' \hat{\beta} + e_i \tag{3.6}$$

where

$$e_i = y_i - x' \hat{\beta}_i . \tag{3.7}$$

A robust regression *M*-estimator minimizes the sum of a less rapidly increasing function of the residuals is given by

$$\min \sum_{i=1}^{n} \rho\left(y_i - x_i' \hat{\beta}\right) = \min \sum_{i=1}^{n} \rho\left(e_i\right) \tag{3.8}$$

where the function $\rho(.)$ gives the contribution of each residual to the objective function. A reasonable $\rho(.)$ should have the following properties:

- $\rho(e) \geq 0$
- $\rho(0) = 0$
- $\rho(e) = \rho(-e)$
- $\rho(e_i) \geq \rho\left(e_{i'}\right)$ for $\left|e_i\right| > \left|e_{i'}\right|$.

The solution obtained from Equation (3.8) is not scale equivariant, and thus the residuals must be standardized by a robust estimate of their scale $\hat{\sigma}$, which is estimated simultaneously.

$$\min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\sigma}\right) \tag{3.9}$$

As in the case of *M*-estimates of location, the median absolute deviation (MAD) is often used.

$$\hat{\sigma} = 1.4826 \times MAD \tag{3.10}$$

$$\text{MAD} = \text{median}\ (\ |\ e_{\text{i}} - \text{median}\ \{e_{\text{i}}\ \}|\ ) \tag{3.11}$$

Taking the derivative of Equation (3.9) and solving produces the score function

$$\sum_{i=1}^{n} \psi\left((y_i - x_i' \hat{\beta}) / \hat{\sigma}\right) x_{ik} = \sum_{i=1}^{n} \psi\left(e_i / \hat{\sigma}\right) x_i = 0 \tag{3.12}$$

with $\psi = \rho'$ which is called the IF. There is now a system of p + 1 equations, for which ψ is replaced by appropriate weights that decrease as the size of the residual increases. Define the weight function $w\ (e) = \psi(e)\ /e$, and let $w_{\text{i}} = w(e_{\text{i}})$. Then the Equations (3.12) becomes

$$\sum_{i=1}^{n} w_i\left(e_i / \hat{\sigma}\right) x_i = 0 \quad for \quad i = 1,2,...,n. \tag{3.13}$$

This is exactly the system of equations that can be solved by using the iterated reweighted least squares (IRLS) procedure.

The weights w_i are computed by using the residuals of each point to determine the influence of each residual to the fit. Estimators using weighting functions which reject completely observations farther than certain distance are called redescending. Among the most widely employed functions for weighting are Huber's, Andrew's, Hampel's and Tukey's. The shape of each weighting curve is depicted in Figure 3.1. The BP of *M*-estimators is shown to be ε = l / (p + l), where p is the number of parameters to be estimated (Hoaglin, Mosteller and Tukey, 1983).

For time series analysis, the *M*-estimators of parameter ϕ denoted by $\hat{\phi}_M$, proposed by Denby and Martin (1979), is defined by

$$\hat{\phi}_M = \min_{\phi'} \sum_{t=1}^{T-1} \rho\left(z_{t+1} - z_t \phi'\right), \tag{3.14}$$

where $\rho(.)$ is a symmetric robustifying loss function. Equivalently, $\hat{\phi}_M$ is the solution of the following equation:

$$\sum_{t=1}^{T-1} z_t \psi\left(z_{t+1} - z_t \hat{\phi}_M\right) = 0, \tag{3.15}$$

where $\psi(e) = \rho'(e)$ is called the IF. $\psi(.)$ is chosen to be a bounded function with $e\phi(e) \geq 0$, and usually $\psi'(0) = 1$. The most commonly used IFs are those from the Huber family and from bisquare family proposed by Beaton and Tukey (1974).

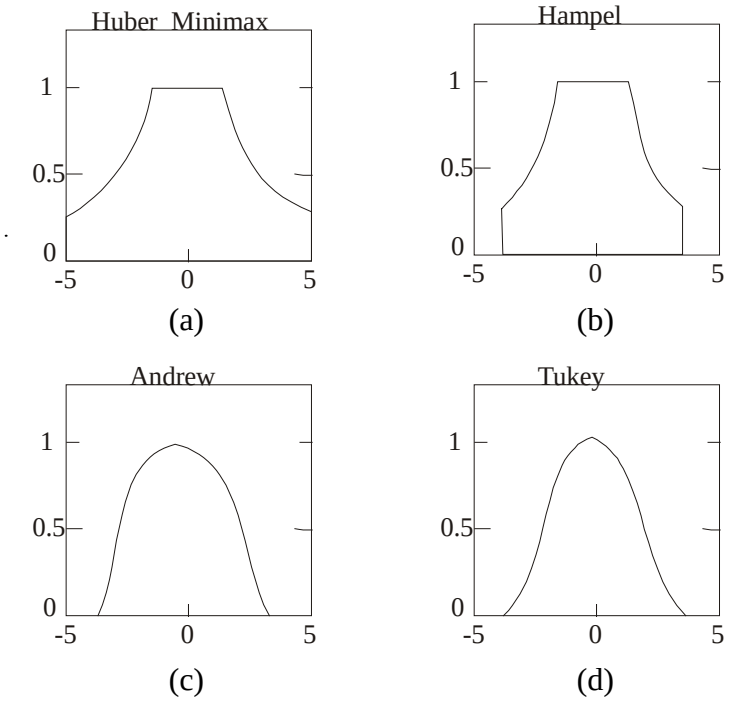


Figure 3.1 Shape of Widely Used Weighting Functions Based on: (a) Huber's; (b) Hampel's; (c) Andrew's sine; and (d) Tukey's biweight

There are different *M*-estimators according to the choice of ψ-function or weight function. They are Huber, Hampel, Andrews and Tukey *M*-estimators which are discussed in brief in the followings.

(a) Huber's *M*-estimators

Huber's uses the following ψ-function.

$$\psi(e)=\begin{cases} -a, & e<-a \\ e, & -a\le e\le a \\ a, & e>a \end{cases} \tag{3.16}$$

Huber's ψ-function is shown in Figure 3.2. Both OLS and Huber objective functions increase without bound as the residual departs from 0, but the OLS objective function increases more rapidly. OLS method assigns equal weight to each observation; the weights for the Huber estimator decline when $|e|>a$. The Huber's ψ-function takes into account the neighborhood of a normal model in a linear way. In Figure 3.2, it has a constant-linear-constant behavior, that is, it is constant beyond the specified bound $(-a\,to\,a)$ and is linear like mean within these bound. Like the OLS it assigns equal weights to all observations within its bound, which surely will result in its high efficiency but distant outliers still have a maximum influence (in the form of constant a), which lead to the efficiency losses of about 10-20 percent in typical cases with outliers. To cope with this problem redescending *M*-estimators were introduced.

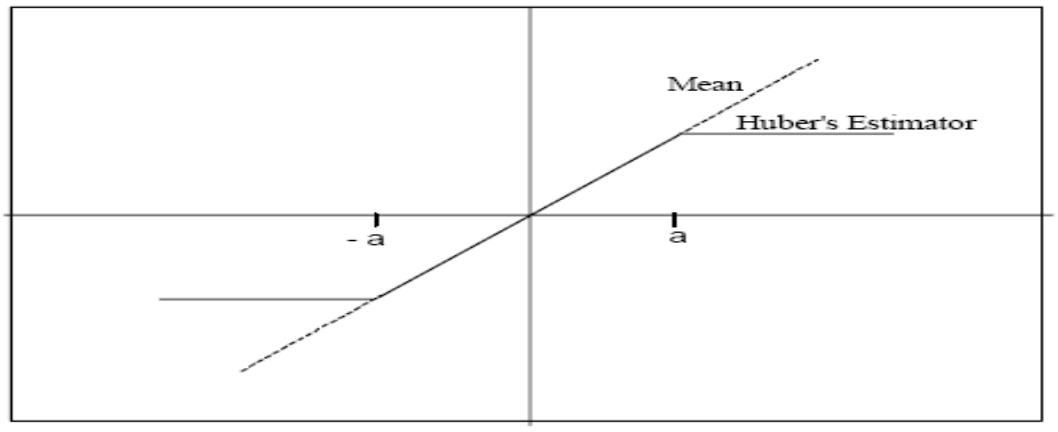


Figure 3.2 Shape of the ψ-Functions of Mean and Huber-Estimators

(b) Hampel's *M*-estimators or Redescending *M*-estimators

The redescending *M*-estimators were introduced by Hampel (Hampel et al., 1986), who used a three part-redescending estimator with ρ-function bounded and ψ-function becoming 0 for large $|e|$ can be shown in Figure 3.3. They reject distant outliers completely, but not suddenly, allowing a transitional zone of increasing doubt, and are therefore much more efficient than "hard" rejection rules; they are usually about as good to clearly better than Huber-estimators (Hampel et al., 1986). The logic of these estimators is that the very central observations (in the neighborhoods of 0) of the normal neighborhood receive maximum weight and as they depart from center their weights decline and as they reach the specified bounds their ψ-function becomes 0.

The Hampel's three-part redescending ψ-function is defined as follows:

$$\psi(e)=\begin{cases} e & |e|< a \\ a\,sign(e) & a \le |e| \le b \\ a(c-|e|)/(c-b) & b \le |e| \le c \\ 0 & |e| \ge c. \end{cases} \qquad (3.17)$$

From Figure 3.3, it can be concluded that the Hampel's three-part redescending estimator is still not a good one, as the abrupt changes in its slope are unappealing because of the abrupt changes in the way the data are used. So, the need of a ψ-function with a smoothly redescending nature was seriously felt. Several smoothly redescending *M*-estimators have been proposed from time to time.

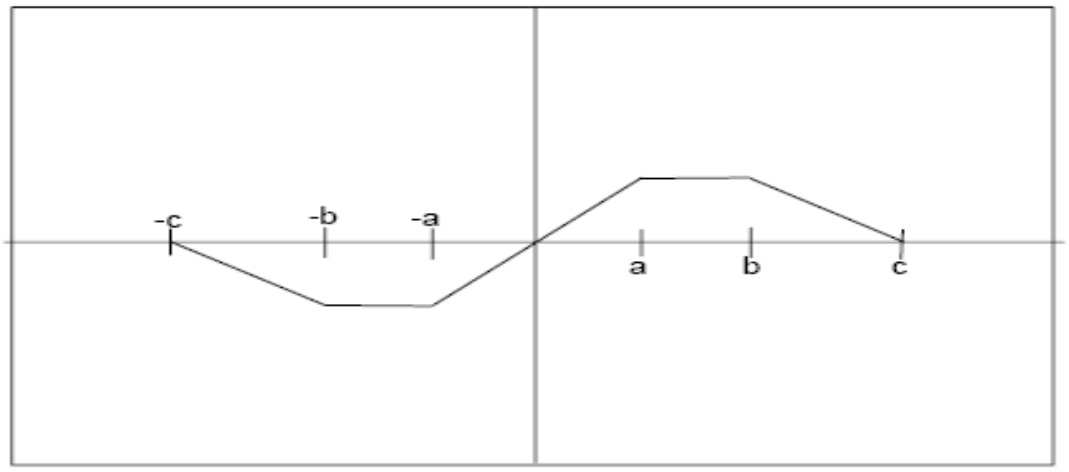


Figure 3.3 Hample's Three-Part ψ-Function

(c) Andrews *M*-estimators

A real improvement came from Andrews (1974), who used wave estimators (also called sine estimators), developed the following ψ-function. Andrews' wave estimators have smoothly redescending ψ-functions, shown in Figure 3.4. The Andrews ψ-function is given by,

$$\psi(e)=\begin{cases} a\sin\left(\dfrac{e}{a}\right) & |e|\le a \\ 0 & |e|>a. \end{cases} \tag{3.18}$$

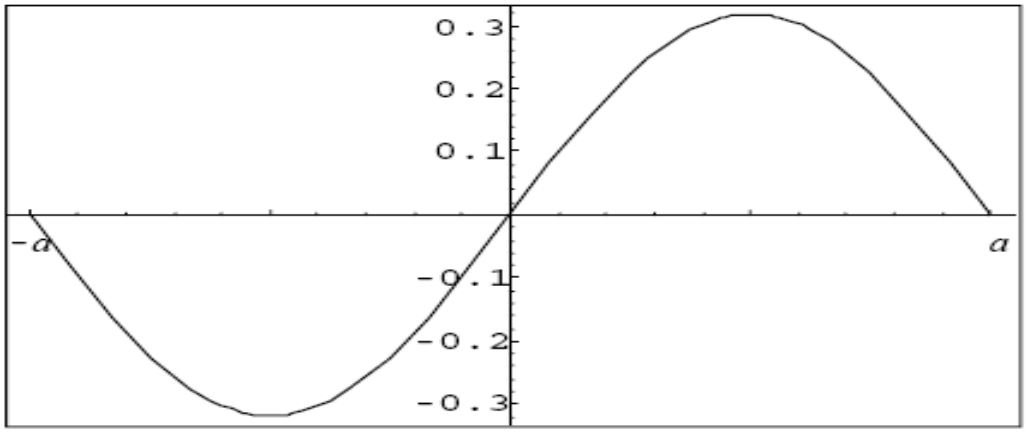


Figure 3.4 Andrews Wave ψ-Function

(d) Tukey's *M*-estimators

Tukey who used biweight estimators proposed another smoothly redescending ψ-function. The Tukey's redescending ψ-function is defined as follows:

$$\psi(e)=\begin{cases} e\left(1-\left(\dfrac{e}{a}\right)^2\right)^2 & |e|\le a \\ 0 & |e|>a. \end{cases} \tag{3.19}$$

The Tukey's redescending ψ-function is shown in Figure 3.5. It is clearly seen that, the Tukey's ψ-function declines as soon as e departs from 0, and are 0 for $|e| > a$. The Tukey's biweight estimators are also known as bisquare estimators (Mosteller and Tukey, 1977 and Hoaglin et al., 1983).

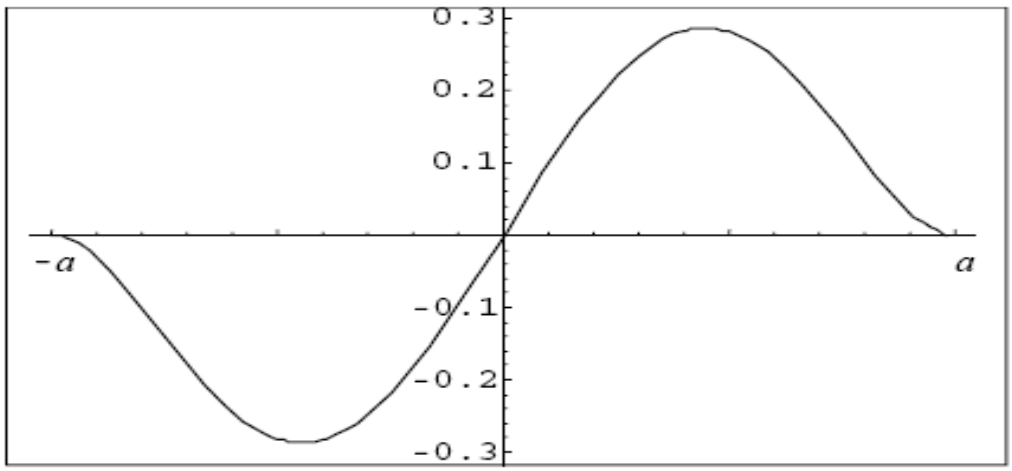


Figure 3.5 Tukey's Biweight ψ-Function

It is a well-known statistical truth that the arithmetic mean has the highest possible efficiency among all other estimators but unfortunately it is extremely sensitive to outliers and even a single outlier can have disastrous effects on it. Among *M*-estimators the most efficient estimator would be Huber's estimator with carefully chosen bounds. The ψ-function of Huber's estimator has a constant-linear-constant behavior (Hoaglin et al., 1983). Figure 3.2 shows that the central section of Huber's ψ-function is linear and within specified bounds it coincides with the ψ-function of the mean and in respect of efficiency it is its plus point. Within the specified bounds it is as efficient as mean can be.

The value a for the Huber and bisquare estimators is called a tuning constant; smaller values of a produce more resistance to outliers, but at the expense of lower efficiency when the errors are normally distributed. The tuning constant is generally picked to give reasonably high efficiency in the normal case; in particular, $a = 1.345\sigma$ for the Huber and $a = 4.685\sigma$ for the bisquare (where σ is the standard deviation of the errors) produce 95-percent efficiency when the errors are normal, and still offer protection against outliers. In an application, the standard deviation of the errors is estimated and to use these results in estimation process. Usually a robust measure of spread is employed in preference to the standard deviation of the residuals.

The main drawback of Huber's ψ-function is that distant outliers still have maximum (though bounded) influence and would lead certainly to certain percent efficiency losses. To avoid these losses, one can use smoothly redescending *M*-estimators such as Andrew's sine function, or Tukey's biweight function, with $\rho(.)$

being bounded and ψ-function continuously becoming zero for large absolute residual. They reject distant outliers completely, but not suddenly, allowing a transitional zone of increasing doubt, and are therefore much more efficient than "hard" rejection rules; they are usually about as good to clearly better than Huber-estimators. As in the case of *M*-estimation of location, the robustness of the estimator is determined by the choice of weight function.

An effective strategy for obtaining an *M*-estimate is as follows: the initial estimate is computed by using OLS estimation; then the *M*-estimate based on the Huber (*MH*-estimate) influence function is computed by the IRLS method will be described in subsection 4.4.4; the corresponding *MH*-estimate is used as a starting point for computing *M*-estimate based on the bisquare (*MB*-estimate) influence function, again using the IRLS method. The use of the Huber influence function ensures that a unique root of Equation (3.15) is obtained and the choice of the bisquare influence function leads to a much more robust estimator in the case of AO model (Denby and Martin, 1979). Since the robustness of the *M*-estimator is not satisfactory, a more robust estimator, called the generalized *M*-estimator (*GM*-estimator) is used.

3.4 Generalized Maximum Likelihood Type Estimators

The *M*-estimators have unbounded influence because it fails to account for leverage (Hampel et al., 1986). In response to this problem, bounded influence *GM*-estimators have been proposed by Hampel (1974). The goal was to create weights that consider both vertical outliers and leverage points. Outliers are dealt with using a standard *M*-estimator, and leverage points are typically down-weighted according to their hat value. The general *GM* class of estimators is defined by

$$\sum_{i=1}^{n} w_i(x_i)\psi\left(\frac{e_i}{v_i(x_i)\hat{\sigma}_e}\right)x_i = 0, \tag{3.20}$$

where ψ is the score function (as in the case of *M*-estimation, this is typically the Huber or biweight function), and the weights w_i and v_i initially depend on the model matrix $\boldsymbol{X}$ from an initial OLS fitted to the data but are updated iteratively.

The first *GM*-estimator proposed by Mallows includes only the w_i weights—that is, $v_i(x_i) = 1$ in Equation (3.20). The w_i are calculated from the hat values which values range from 0 to 1, a weight of $w_i = \sqrt{1-h_i}$ ensures that observations with high leverage receive less weight than observations with small leverage (that is, if $h_i > h_j$, $u_i < u_j$). Although this strategy seems sensible at first, it is problematic because even ''good'' leverage points that fall in line with the pattern in the bulk of the data are down-weighted, resulting in a loss of efficiency.

For time series, the basic idea of the *GM*-estimator is to modify the minimization problem so that the summands of the estimating Equation (3.14) are bounded and continuous functions of the data. Therefore, Martin and Yohai (1985) proposed to use *GM*-estimator. *GM*-estimates $\hat{\phi}, \hat{s}_e$ of autoregression parameters $\phi^T = (\phi_1, \phi_2, ..., \phi_p)$ and innovations scale s_e are obtained by solving the equations

$$\sum_{t=p}^{n-1} W(Z_t) Z_t w_t . \left(Z_{t+1} - Z_t^T \hat{\phi}\right) = 0 \tag{3.21}$$

$$\sum_{t=p}^{n-1} \chi . \left(\frac{Z_{t+1} - Z_t^T \hat{\phi}}{\hat{s}_e} \right) = 0 \tag{3.22}$$

where the observed time series is $Z_1, Z_2, ..., Z_n$, $Z_t^T = Z_t, Z_{t-1}, ..., Z_{t-p+1}$, χ is a bounded and continuous function, and $W(Z_t)$, w_t are nonnegative, data-dependent weight functions. Equation (3.21) provides a linear weighted least squares (WLS) estimate, linear in the case where the "big" weights $W(Z_t)$ and the "little" weights w_t are replaced by fixed weights; that is, weights independent of both the data Z_t and the estimate $\hat{\phi}$. Because the w_t (but not $W(Z_t)$) depend upon $\hat{\phi}$, the Equations in (3.21) are nonlinear.

The big weights $W(Z_t)$ are constructed so that $W(Z_t).Z_t$ is bounded and continuous, and the little weights w_t are constructed so that $w_t . \left(Z_{t+1} - Z_t^T \hat{\phi}\right)$ is bounded and continuous. This achieves the basic requirement for robustness that the summands of the estimating Equation (3.21) be bounded and continuous. Specifically, the weights w_t^j are obtained from a *psi*-function ψ_c, with tuning constant *c*, as follows:

$$w_t^j = \frac{\hat{s}_e^j \psi_c\left(\left(Z_{t+1} - Z_t^T \hat{\phi}^j\right)/\hat{s}_e^j\right)}{Z_{t+1} - Z_t^T \hat{\phi}^j}. \tag{3.23}$$

Two types of *psi*-functions are used, namely Huber's (Huber, 1964) favorite *psi*:

$$\psi_{H,chr}(e) = \begin{cases} e & |e| < c \\ c.\mathrm{sgn}(e) & |e| \geq c \end{cases} \tag{3.24}$$

and Tukey's bisquare functions (Mosteller and Tukey, 1977):

$$\psi_{B,cbr}(e) = \begin{cases} e\left(1 - e^2\right)^2 & |e| \leq c \\ 0 & |e| > c. \end{cases} \tag{3.25}$$

The separate tuning constants *chr* and *cbr* for the ψ-function applied to residuals are adjusted to obtain a compromise between high efficiency when the data are actually Gaussian, and robustness towards outliers. The "big" weights $W(Z_t)$ are also derived from a *psi*-function of either the Huber or Tukey type. When the residual $e_t = Z_{t+1} - Z_t^T \hat{\phi}$ is not too large, w_t will be close to one, whereas when $|e_t|$ is "very large"; for example, when Z_{t+1} is a gross outlier, w_t will be zero.

The only difficulty is that when w_t is based on the Tukey bisquare $\psi_{B,cbr}$, the equations in (3.21) have multiple solutions and starting the iteration (3.21) with OLS might lead to a poor solution. This difficulty is avoided when w_t is based on the Huber *psi*-function $\psi_{H,chr}$, since then (3.21) has an essentially unique solution. However, basing w_t on $\psi_{H,chr}$ does not result in as much robustness toward large outliers as does basing w_t on $\psi_{B,cbr}$. Thus, the strategy adopted is to iterate (3.21) a number of times ***iterh*** using w_t based on the Huber *psi*-function, followed by a number of iterations ***iterb*** using w_t based on the Tukey *psi*-function.

The *GM*-estimates are consistent under a perfectly observed autoregressive model, but they do not completely use the structure of time series when down weighting observations. Consequently, they have a complicated asymptotic covariance matrix and the calibration of the tuning constant when making robust the function that defines the estimates depends on the order of autoregressive operator.

3.5 Least Median Squares Estimators

First proposed by Rousseeuw (1984), least median of squares (LMS) replaces the summing of the squared residuals that characterizes OLS with the median of the squared residuals. The estimates are found by minimizing the median of squared residuals; that is,

$$\min\left(Median\left(y_i - \sum x_{ij}\,\beta_j\right)^2\right) = \min\left(Median\left(e_i^2\right)\right). \tag{3.26}$$

The idea is that, by replacing the sum with the more robust median, the resulting estimator will be resistant to outliers.

Rousseeuw and Croux (1993) pointed out that, it had at best a relative efficiency of 37%, and it did not have a well-defined influence function because of its convergence rate of $n^{-1/3}$ (Rousseeuw, 1984). Despite these limitations, LMS estimators can play an important role in the calculation of the much more efficient *MM*-estimators by providing initial estimates of the residuals. Another method developed by Rousseeuw (1984) is least trimmed squares (LTS) regression. Extending from the trimmed mean, LTS regression minimizes the sum of the trimmed squared residuals.

3.6 Least Trimmed Squares Estimators

The least trimmed squares (LTS) estimation principle was introduced by Rousseeuw (1984) to overcome the efficiency problems of the LMS estimation technique. This particular method can be chosen because of the good convergence rate, smoother objective function and more stable algorithm than the LMS method. The squared residuals are ordered in ascending, that is,

$$\left(e^2\right)_{(1)} \le \left(e^2\right)_{(2)} \le \ldots \le \left(e^2\right)_{(n)}. \tag{3.27}$$

To construct a high-breakdown method which is still equivariant, the OLS criterion is modified as follows:

$$\min_{(\hat{\beta}_0,\ldots,\hat{\beta}_p)} \sum_{i=1}^{h} \left(e^2\right)_{i:n}. \tag{3.28}$$

Since criterion (3.28) does not count the largest squared residuals, it allows the LTS fit to steer clear of outliers. The LTS method achieves the maximal breakdown point

ε = ([(n - p) / 2] + l) / n for h = [n / 2]+[(p + 1) / 2], where p is the number of parameters to be estimated and h is the number of observations included in the calculation of the estimator. When using LTS regression, σ can be estimated by

$$\hat{\sigma} = C_{h,n}\sqrt{\frac{1}{n}\sum_{i=1}^{h}\left(e^2\right)_{i:n}} \tag{3.29}$$

where e_i's are the residuals from the LTS fit, and $c_{h,n}$ makes $\hat{\sigma}$ consistent and unbiased at Gaussian error distributions (Rousseeuw and Leroy, 1987). The LTS scale estimator $\hat{\sigma}$ is itself highly robust. Therefore, regression outliers are identified by their standardized LTS residuals $e_i / \hat{\sigma}$.

This estimator basically finds a robust estimate by identifying the (n - h) points having the largest residuals as outliers, and discarding (trimming) them from the data set. The resulting estimates are essentially OLS estimates of the trimmed data set. It can be seen that h should be as close as possible to the number of good points in the data set, because the higher the number of good points used in the estimates, the more accurate the estimates are. In this case, LTS will yield the best possible estimate. One problem with LTS is that its objective function does not provide itself to mathematical optimization. Besides, the estimation of h itself is difficult in practice.

Atkinson and Cheng (1999) discussed the choice of h. They showed that one can get more stable results for the detection of outliers as well as highly efficient estimates when more data are fitted, provided h is small enough to exclude outlying cases. Zaman et al. (2001) suggested that [0.75n] is a reasonable value for h in most empirical studies. In contrast to OLS, the LTS estimator has a high BP which ensures it is robust enough to resist multiple outliers. The LTS fits only those h observations with the smallest residuals and trims the rest where $h \leq n$. The LTS depends only on the residuals; it generally will not trim the same number of observations from the upper and lower tails of the distribution.

The LTS objective function is based on hard rejection. That is, a given data point is either totally included in the estimation process or totally excluded from it. This is not a good strategy if there are points in the region of doubt. LTS suffers from a low efficiency, because it completely ignores part of the data.

The LTS by itself is not suited for inference because of its relatively low finite-sample efficiency. This can be resolved by carrying out a reweighted least squares (RLS) step. To each observation *i* one assigns a weight w_i based on its standardized LTS residual $e_i / \hat{\sigma}$, e.g. by putting $w_i = w(|e_i / \hat{\sigma}|)$ where *w* is a decreasing continuous function. A simpler way, but still effective, is to put $w_i = 1$ if $|e_i / \hat{\sigma}| \leq 2.5$ and $w_i = 0$ otherwise. Either way, the RLS fit $(\hat{\beta}_0, ..., \hat{\beta}_p)$ is then defined as follows:

$$\min_{(\hat{\beta}_0, ..., \hat{\beta}_p)} = \sum_{i=1}^{n} w_i e_i^2 \tag{3.30}$$

which can be computed quickly. The result inherits the breakdown value, but is more efficient and yields all the usual inferential output such as *t*-statistics, *F*-statistics, an R^2 statistic, and the corresponding *p*-values. These *p*-values assume that the data with $w_i = 1$ come from the regression model which is shown in Equation (3.5) whereas the data with $w_i = 0$ do not.

Although LTS is highly resistant, it suffers badly in terms of relative efficiency at about 8%. Its efficiency is so low that it is not desirable as a stand-alone estimator. Still, the LTS has merit in the role it plays in the calculation of other estimators. For example, the *GM*-estimators proposed by Coakley and Hettmansperger (1993) use LTS to obtain initial estimates of the residuals.

3.7 Rank Test Estimators

First proposed by Jaeckel (1972), rank test estimation (*R*-estimation) is a procedure based on the ranks. Consider replacing one factor in the OLS objective function $S(\beta) = \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta})^2$ by its rank. Here, $\underline{\mathbf{x}_i'}$ is $1 \times (p + 1)$ row vector and $\underline{\boldsymbol{\beta}}$ is $(p + 1) \times 1$ column vector of parameters. Thus if R_i is the rank of $y_i - \mathbf{x}_i'\boldsymbol{\beta}$, then we wish to

$$\min \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta}) R_i \,. \tag{3.31}$$

More generally, replacing the ranks (which are the integers $1,2,...,n$) by the score function $a(i) = 1,2,...,n,$ so that the objective function becomes

$$\min_{\beta} \sum_{i=1}^{n} (y_i - \mathbf{x}_i'\boldsymbol{\beta}) a(R_i) \tag{3.32}$$

where R_i is the rank of the i th residual and $a_n(i)$ is a monotone (nondecreasing) score function that satisfies $\sum_{i=1}^{n} a_n(i) = 0$.

Many possibilities have been proposed for the score function. The simplest, and perhaps most commonly employed, are the Wilcoxon scores, which directly find the rank of observations from the median, given by

$$a_n(i) = i - \left(\frac{n+1}{2} \right). \tag{3.33}$$

Median scores are a simple adjustment over the Wilcoxon scores,

$$a_n(i) = \sin\left(i - \left(\frac{n+1}{2} \right) \right). \tag{3.34}$$

Van der Waerden scores adjust the ranks according to the inverse of the normal probability density function Φ^{-1}:

$$a_n(i) = \Phi^{-1}\left(\frac{i}{n+1} \right). \tag{3.35}$$

Finally, bounded normal scores adjust the Van der Waerden scores by bounding them according to a constant, c:

$$a_n(i) = \min\left\{ c, \max\left(\Phi^{-1}\left(\frac{i}{n+1} \right), -c \right) \right\}. \tag{3.36}$$

An advantage of *R*-estimators over some others (such as *M*-estimators, and those extending from them) is that they are scale equivariant. They have some undesirable attributes, however. One problem is that the optimal choice for the score function is unclear. A second problem is that the objective function is invariant with respect to the intercept. If an intercept is not required, this is of no concern—it is simply not estimated. Even if one is needed, it can be calculated manually after fitting the model from the median of the residuals, so this limitation of *R*-estimators is surmountable.

More problematic is the fact that most *R*-estimators have a BP of 0. An exception is the bounded influence *R*-estimator of Naranjo and Hettmensperger (1994), which is also fairly efficient (90%-95%) when the Gauss-Markov assumptions are satisfied. Even for this estimator, however, the BP never reaches more than 0.20.

3.8 Studentized Location Estimators

OLS estimator depends on the minimization of the variance of the residuals. Hence, since the variance is highly sensitive to outliers, OLS is also largely influenced as well. For this reason, Rousseeuw and Yohai (1984) proposed to minimize a measure of dispersion of the residuals that is less sensitive to extreme values than the variance. They called this class of estimators the *S*-estimators. To increase robustness, the square function could be replaced by another loss function ρ (.) which awards less importance to large residuals. The estimation problem would now consist in finding the smallest robust scale of the residuals. This robust dispersion, that will be called $\hat{\sigma}^s$, satisfies the condition

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{e_i}{\hat{\sigma}^s}\right)=b, \tag{3.37}$$

where *b* is a constant defined as $b = E_{\Phi}\left[\rho(e)\right]$ and Φ represents the standard normal distribution. Differentiating Equation (3.37) and solving results in

$$\frac{1}{n}\sum_{i=1}^{n}\psi\left(\frac{e_i}{\hat{\sigma}^s}\right)=b, \tag{3.38}$$

where ψ is replaced with an appropriate weight function. As with most *M*-estimation procedures, either the Huber weight function or the biweight function is usually employed. The value of β that minimizes $\hat{\sigma}^s$ is then called an *S*-estimator. More formally, an *S*-estimator is defined as follows:

$$\hat{\beta}^s = \min_{\beta}\left\{\hat{\sigma}^s\left(e_1\left(\hat{\beta}\right),\ldots,e_n\left(\hat{\beta}\right)\right)\right\} \tag{3.39}$$

where $\hat{\sigma}^s$ is the robust estimator of scale as defined in (3.37).

Although *S*-estimators have a BP of 0.5, it comes at the cost of very low efficiency (approximately 30%) relative to OLS. *S*-estimators play an important role in calculating *MM*-estimates, which are far more efficient.

Among the estimators with high resistance to many outliers, it is found that the *S*-estimators are particularly attractive. Unlike the LTS and LMS estimators, the *S*-estimators smoothly down weigh outlying observations. This smooth handling of the outliers is not only natural from a practitioners' point of view, but also it tends to make the *S*-estimators more efficient relative to the LMS estimator. Further, as shown in Sakata and White (1995), the *S*-estimators are more efficient than the OLS estimator under various error distributions. In addition, they are computationally less demanding than the LTS. These are points of more attractiveness than LMS and LTS.

3.9 Generalized Studentized Location Estimators

Croux et al. (1994) proposed generalized studentized location estimators (*GS*-estimators) in an attempt to overcome the low efficiency of the original *S*-estimators. These estimators are computed by finding a *GM*-estimator of the scale of the residuals. A special case of the *GS*-estimator is the least quartile difference estimator (LQD), the parallel of which is using the interquartile range to estimate the scale of a variable. The LQD estimator is defined by

$$\min\left[Q_n\left(e_1,...,e_n\right)\right], \tag{3.40}$$

where

$$Q_n = \left\{ \left|e_i - e_j\right| \; ; \; i<j \right\}_{\binom{h_p}{2}:\binom{n}{2}} \tag{3.41}$$

and

$$h_p = \frac{n+p+1}{2} \tag{3.42}$$

and p is the number of parameters in the model. Putting more simply, this means that Q_{n} is the $\binom{h_p}{2}$th order statistic among the $\binom{n}{2}$ elements of the form $\left\{\left|e_i - e_j ; i<j\right|\right\}$. Although these estimators are more efficient than *S*-estimators, they have a ‘‘slightly increased worst-case bias’’ (Croux et al., 1994).

3.10 *MM*-Estimators

First proposed by Yohai (1987), *MM*-estimators have become increasingly popular and are perhaps now the most commonly employed robust regression technique. They combine a HBP (50%) with good efficiency (approximately 95% relative to OLS under the Gauss-Markov assumptions). The ‘‘*MM*’’ in the name refers to the fact that more than one *M*-estimation procedure is used to calculate the final estimates. These estimators are redescending *M*-estimators as defined in (3.9), but where the scale is fixed at $\hat{\sigma}^s$. The preliminary *S*-estimator is obtained from Equation (3.39). The objective function of *MM*-estimator is defined as

$$\min \sum_{i=1}^{n} \rho\left(\frac{e_i}{\hat{\sigma}^s}\right) \tag{3.43}$$

where $\hat{\sigma}^s$ is the robust estimator of scale as defined in (3.37). It has both the high breakdown property and a higher statistical efficiency than *S*-estimation. The *MM*-estimates are found that simultaneously have BP 0.5 and asymptotic efficiency for normal errors as close to one as desired. For both the initial *S*-estimate and the *MM*-estimate, two different weight functions can be used: Tukey's bisquare function and an optimal weight function introduced in Yohai and Zamar (1988). It is common to use a Tukey's bisquare $\rho(.)$ function for both the preliminary *S*-estimator and the final *MM*-estimator. Tukey's bisquare functions $\rho(.;c)$ and $\psi(.;c)$ are as follows:

$$\rho(e;c) = \begin{cases} \left(\frac{e}{c}\right)^6 - 3\left(\frac{e}{c}\right)^4 + 3\left(\frac{e}{c}\right)^2 & if \ \ |e| \le c \\ 1 & if \ \ |e| > c, \end{cases}$$

$$\Psi(e;c) = \begin{cases} \frac{6}{c}\left(\frac{e}{c}\right) - \frac{12}{c}\left(\frac{e}{c}\right)^3 + \frac{6}{c}\left(\frac{e}{c}\right)^5 & if \ \ |e| \le c \\ 0 & if |e| > c. \end{cases} \tag{3.44}$$

The Yohai and Zamar optimal functions $\rho(.;c)$ and $\psi(.;c)$ are as follows:

$$\rho(e;c)=\begin{cases} \dfrac{e^2}{2} & if \ \left|\dfrac{e}{c}\right|\le 2 \\ c^2\left[1.792+h_1\left(\dfrac{e}{c}\right)^2+h_2\left(\dfrac{e}{c}\right)^4+h_3\left(\dfrac{e}{c}\right)^6+h_4\left(\dfrac{e}{c}\right)^8\right] & if \ \ 2<\left|\dfrac{e}{c}\right|\le 3 \\ 3.25c^2 & if \ \left|\dfrac{e}{c}\right|>3, \end{cases}$$

$$\Psi(e;c)=\begin{cases} e & if \ \left|\dfrac{e}{c}\right|\le 2 \\ c\left[g_1\left(\dfrac{e}{c}\right)+g_2\left(\dfrac{e}{c}\right)^3+g_3\left(\dfrac{e}{c}\right)^5+g_4\left(\dfrac{e}{c}\right)^7\right] & if \ \ 2<\left|\dfrac{e}{c}\right|\le 3 \\ 0 & if \ \left|\dfrac{e}{c}\right|>3, \end{cases} \qquad (3.45)$$

where

$$g_1 = -1.944, \quad g_2 = 1.728, \quad g_3 = -0.312, \quad g_4 = 0.016$$
$$h_1 = g_1/2, \quad h_2 = g_2/4, \quad h_3 = g_3/6, \quad h_4 = g_4/8.$$

Yohai and Zamar (1988) showed that the $\rho(.)$ and ψ-functions given above are optimal in the following highly desirable sense: the *MM*-estimate has a BP of one-half, and minimizes the maximum bias under contamination distributions (locally for small fractions of contamination), subject to achieving a desired efficiency when the data is Gaussian. The Gaussian efficiency of the *MM*-estimate is controlled by the choice of the tuning constant *c*.

The preliminary *S*-estimator gives guarantee of a HBP, and the final *MM*-estimator allows a high Gaussian efficiency. The tuning constant *c* can be set to 1.547 for the *S*-estimator, to guarantee a 50% BP, and it can be set to 4.685 for the *MM*-estimator in Equation (3.43) to guarantee a 95% efficiency of the final estimator. If *c* = 2.697, the efficiency of the *MM*-estimator will be 70%. For computing the *MM*-estimator, the IRLS algorithm can be used, taking $\hat{\beta}_s$ as initial value.

3.11 Minimum Volume Ellipsoid Estimator

In the minimum volume ellipsoid estimator (MVE), proposed by Rousseeuw (1984), an ellipsoid of the smallest volume with a subset of h objects (no contaminated data) is constructed. In one of the proposed iterative algorithms, $n + 1$ object is selected iteratively at random in each of iterations and their mean and covariance are determined. Then, the ellipsoid containing exactly $h = ((n + p + 1)/2)$ data objects is found by deflating or expanding the data covariance. The steps of the algorithm are repeated until the subset of h objects yielding the smallest volume of the covariance ellipsoid is found.

The MVE estimator has highest possible BP value (50%). Geometrically, the estimator finds the minimum volume ellipsoid covering, or enclosing a given set of points. The MVE estimator is a generalization of the LMS estimator for high dimensional data sets, with the extra property of being equivariant to translation, scaling, orthogonal projection and affine transformations.

The objective function of MVE estimator is based on the h th quantile $d_{\mathrm{h:N}}$ of the Mahalanobis-type distances $d = (d_1,\ldots,d_{\mathrm{N}})$,

$$\mathrm{F_{MVE}} = \sqrt{\frac{1}{h}\sum_{i=1}^{h}\left(e^2\right)_{i:N}} \rightarrow \min \tag{3.46}$$

subject to $d_{\mathrm{h:N}} = \sqrt{\chi^2_{n,0.05}}$, where $\boldsymbol{C}$ is the scatter matrix estimate, and the Mahalanobis-type distances are computed as

$$d = diag\left(\sqrt{(X-T)^T C^{-1}(X-T)}\right) \tag{3.47}$$

where T is the center of minimal volume ellipsoid covering h points of $\boldsymbol{X}$. The covariance estimator of this is given by the ellipsoid. Because of the transform $x \rightarrow xA + b$ is an ellipsoid where A and b are the constants, MVE is an affine equivariant estimator such that any transformation on x does not affect the MVE.

3.12 Minimum Covariance Determinant Estimator

The minimum covariance determinant (MCD) method of Rousseeuw (1984) aims to find h observations out of n whose covariance matrix $\boldsymbol{C}$ has the lowest determinant. Since the covariance matrix is an $n \times n$ symmetric positive definite matrix, all p eigenvalues are positive. Furthermore, the determinant of a covariance

matrix equals the product of the p eigenvalues. Near linear dependencies among the set of p variables produce near-zero eigenvalues. Thus, a small value in the determinant reflects some linear patterns in the data. Consider all C_h^n subsets, and compute the determinant of the covariance matrix for each subset. The subset with the smallest determinant is used to calculate the usual $p\times1$ mean vector, $\boldsymbol{m}$, and corresponding $p\times p$ covariance matrix, $\boldsymbol{C}$. These estimators are called the MCD estimators.

In terms of asymptotic properties, the MCD is preferred over the MVE. Both have an $[(n - p - 1) / 2] / n$ BP, which is 50% asymptotically. Both are also affine equivariant. However, the MCD converges as $n^{-1/2}$ while MVE converges as $n^{-1/3}$. Thus, MCD has higher efficiency than does MVE.

3.13 Residual Autocovariances Estimators

Bustos and Yahoi (1986) proposed a new class of estimators that are based on residual autocovariances (RA estimators). The basic idea behind the RA estimates is to exhibit the OLS estimates in a form that involves (the usual nonrobust) covariance estimates of residuals, and then make OLS robust by making the covariance estimates robust in a natural way. Consider the OLS estimates by minimizing the following residuals sum of squares

$$\sum r_t^2(\lambda) \tag{3.48}$$

where $\boldsymbol{\lambda} = (\boldsymbol{\Phi},\boldsymbol{\theta}, \mu)$, the residuals $r_t(\lambda)$ are defined by

$$r_t(\lambda)=\theta(B)^{-1}\phi(B)(Z_t-\mu),$$

and it is assumed that $Z_t = \mu$ for $t \le 0$. The residuals can be computed recursively from

$$r_t=(Z_t-\mu)-\phi_1(Z_{t-1}-\mu)-\cdots-\phi_p(Z_{t-p}-\mu)+\theta_1 r_{t-1}+\cdots+\theta_q r_{t-q},$$

for t $\ge$ p +1, with initial conditions $r_p=\cdots=r_{p-q+1}=0$. Differentiating the Equation (3.48), the following system of p + q + 1 equations was obtained for the OLS estimates.

$$\left.\begin{aligned}\sum_{t=p+1}^{T} r_t(\partial r_t/\partial\phi_j)=0, \quad 1\le j\le p\\ \sum_{t=p+1}^{T} r_t(\partial r_t/\partial\theta_j)=0, \quad 1\le j\le q\\ \sum_{t=p+1}^{T} r_t(\partial r_t/\partial\mu)=0\ .\end{aligned}\right\} \tag{3.49}$$

It is easy to show from ARMA (p, q) processes

$$\phi(B)(Z_t - \mu) = \theta(B) r_t \quad \text{that} \tag{3.50}$$

$$\left.\begin{aligned} \partial r_t / \partial \phi_j &= -\theta^{-1}(B) B^j (Z_t - \mu) = -\phi^{-1}(B) r_{t-j} \\ \partial r_t / \partial \theta_j &= \theta^{-2}(B) \phi(B) B^j (Z_t - \mu) = \theta^{-1}(B) r_{t-j} \\ \partial r_t / \partial \mu &= -(1 - \phi_1 - \ldots - \phi_p)/(1 - \theta_1 - \ldots - \theta_q), \end{aligned}\right\} \tag{3.51}$$

where stationarity and invertibility of ARMA process ensures that $\phi^{-1}(B)$ and $\theta^{-1}(B)$ exit. By replacing parameter values with estimated parameter values in (3.51) and using the result in (3.49), the OLS estimates which will satisfy the following equations were gained.

$$\left.\begin{aligned} \sum_{t=p+1}^{T} r_t \phi^{-1}(B) r_{t-j} = 0, \quad 1 \le j \le p \\ \sum_{t=p+1}^{T} r_t \theta^{-1}(B) r_{t-j} = 0, \quad 1 \le j \le q \\ \sum_{t=p+1}^{T} r_t = 0. \end{aligned}\right\} \tag{3.52}$$

Call $s_i = s_i(\phi)$, $0 \le i \le \infty$, the series expansion coefficients of the operator $\phi^{-1}(B)$ and $t_i = t_i(\theta)$, $0 \le i \le \infty$, those of $\theta^{-1}(B)$; that is,

$$\phi^{-1}(B) = \sum_{i=o}^{\infty} s_i B^i \ , \ \theta^{-1}(B) = \sum_{i=o}^{\infty} t_i B^i \ .$$

Then, (3.52) can be written as

$$\begin{aligned} \sum_{t=p+1+j}^{T} r_t \sum_{h=0}^{\infty} r_{t-j-h} s_h = 0, \quad 1 \le j \le p \\ \sum_{t=p+1+j}^{T} r_t \sum_{h=0}^{\infty} r_{t-j-h} s_h = 0, \quad 1 \le j \le p \\ \sum_{t=p+1}^{T} r_t = 0. \end{aligned}$$

Using the initial conditions $r_t = 0$ for $t < p + 1$ and interchanging the order of summations, it can be seen as shown below

$$\left.\begin{aligned} \sum_{h=0}^{T-j-p-1} s_h\,\gamma_{h+j}(\lambda) = 0, \quad 1 \le j \le p \\ \sum_{h=0}^{T-j-p-1} t_h\,\gamma_{h+j}(\lambda) = 0, \quad 1 \le j \le q \\ \sum_{t=p+1}^{T} r_t = 0, \end{aligned}\right\} \tag{3.53}$$

where

$$\gamma_i(\lambda) = \sum_{t=p+1}^{T-i} r_{t+i}\, r_t . \tag{3.54}$$

The class of robust RA estimates can be defined by making the residual covariances γ_i robust. This may be done by replacing the γ_i**(λ)** in (3.53) by

$$\gamma_i(\lambda) = \sum_{t=p+1+i}^{T} \eta\left(r_t / \hat{\sigma}, r_{t-i} / \hat{\sigma}\right), \quad i = 1,2,\ldots . \tag{3.55}$$

In the last Equation of (3.53), r_t is replaced by $\psi\left(r_t / \hat{\sigma}\right)$, where $\eta : R^2 \to R$ *and* $\psi : R \to R$ are bounded and continuous functions and $\hat{\sigma}$ is a robust estimate of the scale of the a_t's. Then, the RA estimates are defined by the following $p + q + 1$ equations:

$$\left.\begin{aligned} \sum_{h=0}^{T-j-p-1} s_h\,\gamma_{h+j}(\lambda) = 0, \quad 1 \le j \le p \\ \sum_{h=0}^{T-j-p-1} t_h\,\gamma_{h+j}(\lambda) = 0, \quad 1 \le j \le q \\ \sum_{t=p+1}^{T} \psi\left(r_t / \hat{\sigma}\right) = 0. \end{aligned}\right\} \tag{3.56}$$

$\hat{\sigma}$ is computed simultaneously using,

$$\hat{\sigma} = Med\left(\left|r_{p+1}\right|,\ldots,\left|r_T\right|\right) / 0.6745. \tag{3.57}$$

It was assumed that η is odd in each variable and ψ is odd. Therefore, if the distribution of the innovations a_t is symmetric, the RA estimates will be Fisher consistent. This is because in this case we have

$$\lim_{t\to\infty} r_t(\lambda_0) = a_t ,$$

where λ_0 is the true parameter and

$$\left.\begin{aligned} E_F\left(\eta\left(a_t/\sigma, a_{t-i}/\sigma\right)\right) &= 0, \quad \forall\sigma, \quad i=1,2,... \\ E_F\left(\psi\left(a_t/\sigma\right)\right) &= 0, \quad \forall\sigma. \end{aligned}\right\} \tag{3.58}$$

Two ways of selecting η are

$$\left.\begin{aligned} \eta_M(u,v) &= \psi(u)\,\psi(v), \quad Mallows\,type \\ \eta_H(u,v) &= \psi(uv), \quad Hampel\,type \end{aligned}\right\} \tag{3.59}$$

where ψ is a continuous and odd function. They may be chosen, for example, to be in the Huber family given by

$$\psi_{H,c}(u) = \mathrm{sgn}(u)\min\left(|u|, c\right), \tag{3.60}$$

where *sgn* (*u*) is the sign function.

Another possibility is to take the ψ-function in a redesending family – for example, in the bisquare family proposed by Beaton and Tukey (1974). This family is defined by

$$\psi_{B,c}(u) = u\left(1-u^2/c^2\right)^2, \qquad 0 \le |u| \le c. \tag{3.61}$$

If η (*u*,*v*) = ψ(*u*)*v* , then the RA estimate given by (3.56) is asymptotically equivalent to the *M*-estimate minimizing

$$\sum_{t=p+1}^{\infty} \rho\left(r_t/\sigma\right), \tag{3.62}$$

where $\rho' = \psi$.

Therefore the class of RA estimates contains the class of *M*-estimates. In particular the OLS estimate is obtained putting η (*u*,*v*) = *uv* and ψ(*u*) = *u*. Since the s_h and t_h decay exponentially, for numerical computations it is not necessary to consider all of the terms in the first *p* + *q* equations of (3.56).

3.14 Approximate Conditional Mean Robust Filter

Martin (1979) proposed the approximate conditional mean robust filter estimator. These robust filters are nonlinear functions of the data which are designed to give good estimates of Z_t in the presence of outliers. A class of robust filter cleaners which are used to obtain the one-step-ahead predictions $\hat{Y}_t^{t-1} = \hat{Z}_t^{t-1}$, and thereby compute the prediction residuals $u_t = Y_t - \hat{Y}_t^{t-1}$ appearing in the loss function. These filter cleaners are sometimes called approximate conditional mean type (ACM) filter cleaners because of an approximate optimality result are obtained. The term filter refers to an estimate $\hat{Z}_t$ of Z_t which is based on the present and past data $Y^t = (Y_1,...,Y_t)'$. The filter cleaner computes robust estimates $\hat{Z}_t$ of the vector Z_t according to the following recursion:

$$\hat{Z}_t = \Phi\, \hat{Z}_{t-1} + \rho_t\, s_t\, \psi\left(\frac{Y_t - \hat{Y}_t^{t-1}}{s_t}\right) \tag{3.63}$$

where $\rho_t = m_t / s_t^2$, with m_t being the first column of the *k*×*k* matrix M_t, which is computed recursively as

$$M_{t+1} = \Phi P_t \Phi' + Q, \tag{3.64}$$

$$P_t = M_t - w\left(\frac{y_t - \hat{y}_t^{t-1}}{s_t}\right)\frac{m_t m_t'}{s_t^2}. \tag{3.65}$$

The ψ is a robustifying *psi*-function, $Q = \sigma^2 rr'$, r is a *k*×1 column vector defined as $(1, -\theta_1, -\theta_2, ..., -\theta_{k-1})'$ with θ_i = 0 for *i* > *q* in case *p* > *q*. w is a weight function which is given as follows:

$$w(u) = \frac{\psi(u)}{u} \tag{3.66}$$

The time-varying scale s_t is defined by

$$s_t^2 = m_{11,t}, \tag{3.67}$$

where $m_{11,t}$ is the 1-1 element of M_t, the robust one-step-ahead predictors of Y_t and Z_t are

$$\hat{Y}_t^{t-1} = \hat{Z}_t^{t-1} = \left(\Phi\, \hat{Z}_{t-1}\right)_1, \tag{3.68}$$

and the cleaned data at time *t* is given by

$$\hat{Z}_t = \left(\hat{Z}_t\right)_1. \tag{3.69}$$

With the scaling (3.67), we will have $\hat{Z}_t = Y_t$ a large fraction of the time when there are rather few outliers in the series. This is we use the term *filter cleaner*. The weight function w should have the same qualitative properties as a good robustifying ψ-function, namely: boundedness, continuity and perhaps compact support. A common compact support for ψ and w results in the following desirable behavior of the filter cleaner: if an observation Y_t deviates from its prediction $\hat{Y}_t^{t-1}$ by a sufficiently large amount, then $\hat{Z}_t$ will be the pure prediction $\hat{Z}_t = \Phi\, \hat{Z}_{t-1}$, and the filtering-error covariance is set equal to the one-step prediction-error covariance $P_t = M_t$. The later idea has often been implemented as a so-called hard-rejection rule: set $\hat{Z}_t = \Phi\, \hat{Z}_{t-1}$ and $P_t = M_t$ if $|u_t| > c\, s_t$, replacing (3.67) by $s_t^2 = m_{11,t} + \sigma_0^2$ in general noise component. Typically, c = 3 has been used according to a time-honored habit, and the procedure accordingly is termed a ***3- sigma- edit rule***. This corresponds to the choices

$$\psi_{HR}(u) = \begin{cases} u, & |u| < c \\ 0, & |u| \geq c, \end{cases} \tag{3.70}$$

$$w_{HR}(u) = \begin{cases} 1, & |u| < c \\ 0, & |u| \geq c. \end{cases} \tag{3.71}$$

The two-part redescending ψ-function is as follows:

$$\psi_{HA}(u) = \begin{cases} u & |u| \leq a, \\ \dfrac{a}{b-a}(b-u) & a < u \leq b, \\ \dfrac{-a}{b-a}(b+u) & -b \leq u < -a \\ 0 & |u| > b. \end{cases} \tag{3.72}$$

CHAPTER IV

ROBUSTNESS IN REGRESSION MODELS

Regression analysis is an important tool for any quantitative research. For the regression analysis, OLS method can produce bad estimates when the error distribution is not normal, particularly when the errors are heavy-tailed. It explores the relationship between dependent and explanatory variables. Many hypotheses claimed by economic theories can be tested by applying a regression model on real world data. The OLS method is a mostly applied method in regression technique. The application of this specific method requires several assumptions. A researcher should be aware of the fact that the OLS method performs poorly if these assumptions are not fulfilled.

In the last two centuries, various strategies were introduced to test whether the model assumptions are fulfilled or not. Besides, various more general regression techniques are available based on less stringent conditions. Until the mid-20^{th} century, violations of the model assumptions were treated independently from any common error source. But in particular, outlying observations within the data can cause violations of model assumptions and thereby it can have a huge impact on regression results.

Robust regression analyses have been developed as an improvement to OLS estimation in the presence of outliers and provide information about what a valid observation is and whether this should be thrown out. The primary purpose of robust regression analysis is to fit a model which represents the information in the majority of the data. In this context, robust regression is to employ a fitting criterion that is not as vulnerable as OLS to unusual data. One remedy is to remove influential observations before using the OLS fit.

Robust regression analysis provides an alternative to an OLS regression model when fundamental assumptions are unfulfilled by the nature of the data. When the estimates of the parameters of statistical regression models and tests assumptions, it is frequently found that the assumptions are substantially violated. Sometimes, the variables can be transformed to confirm to those assumptions. Often, however, a transformation will not eliminate or satisfy the leverage of influential outliers that bias the prediction and distort the significance of parameter estimates. Under these

circumstances, to the best of the present researcher's knowledge, robust regression that is resistant to the influence of outliers may be the only reasonable remedy. This chapter focuses on the (robustness-) performance of estimators if outliers occur within the data set.

In this chapter, section 4.1 introduces the linear regression models. The OLS estimates for regression model are discussed in section 4.2. Section 4.3 describes the types of outliers that can be found in regression analysis and discusses the methods of outlier detection. Section 4.4 is mainly concerned with the some robust estimates (LAV, *M*, *MM*, LMS and LTS) that are applied to the simulated as well as real data in regression analysis. In this section, the computation algorithms of these estimators are described. Robustness in dummy variable is presented in section 4.5. In this section, the RDL_1 estimates and its computation algorithm is discussed. In section 4.6, these robust estimates are applied to the simulated data as well as real data to illustrate how the robust methods outperform the OLS when data contain the outlying observations.

4.1 Linear Regression Model

Regression analysis is a statistical method that utilizes the relation between two or more quantitative variables so that one variable can be predicted from others. The linear regression model is written as follows:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \varepsilon_i \tag{4.1}$$

for $i = 1,2,...,n$ where y_i stands for the response variable and x_{i1} to x_{ip} are the regressors (explanatory variables). The constant term is denoted by β_0. Classical theory, be assumed as the ε_i have a Gaussian distribution with mean 0 and variance σ^2. The parameters $\beta_0, \beta_1, ..., \beta_p$ and σ are estimated from *n* observations of the form $(x_{i1}, ..., x_{ip}, y_i)$. Applying a regression estimator to the data yields *p* + 1 regression coefficients $\hat{\beta}_0, ..., \hat{\beta}_p$. The residual e_i of case *i* is defined as

$$e_i(\hat{\beta}_0, ..., \hat{\beta}_p) = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}). \tag{4.2}$$

Although coefficients can be estimated in several ways, the underlying idea is usually applied to get as close as possible to the true value by reducing the magnitude of the residuals, as measured by an aggregate prediction error. There are various

methods for estimation of regression coefficients in the above linear regression model (4.1). The most commonly used is the method of least squares. In the case of the well-known OLS, this aggregate prediction error is defined as the sum of squared residuals. This method was discussed in the following section.

4.2 Ordinary Least Squares Estimates for Regression Model

A linear regression analysis, one of the most important statistical tools, is very useful for many fields. Almost regression analysis relies on the method of least squares for estimation of the parameters in the model. The OLS regression method of Gauss and Legendre computes the parameters $(\hat{\beta}_0,...,\hat{\beta}_p)$ that minimize the sum of squares of $e_i\left(\hat{\beta}_0,...,\hat{\beta}_p\right)$. Formally, it can be written as

$$Q = \min_{(\hat{\beta}_0,...,\hat{\beta}_p)} \sum_{i=1}^{n} e_i^2 . \tag{4.3}$$

Q is the sum of the squared vertical deviations from the hyperplane $H = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p$. Taking the derivative of (4.3) with respect to $\boldsymbol{\beta} = \left(\beta_0, \beta_1,..., \beta_p\right)$, obtain the normal equations

$$\boldsymbol{X}^T \boldsymbol{X}\ \hat{\boldsymbol{\beta}} = \boldsymbol{X}^T \boldsymbol{Y} \tag{4.4}$$

and solving these equations give the least squares estimator of $\boldsymbol{\beta}$

$$\hat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{Y} . \tag{4.5}$$

The vector of predicted or fitted values is $\hat{\boldsymbol{Y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}} = \boldsymbol{HY}$ where $\boldsymbol{H} = \boldsymbol{X}\left(\boldsymbol{X}^T \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T$ is called the hat matrix. The *i* th entry of $\hat{Y}$ is the *i* th fitted value (or predicted value) $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i,1} + \cdots + \hat{\beta}_p X_{i,p} = x_i^T \hat{\beta}$ for observation Y_i while the *i* th residual is $e_i = Y_i - \hat{Y}_i$. The vector of residuals is $\boldsymbol{e}$ = (**I -H)Y**. The OLS criterion allows computing the coefficients explicitly from the data and is optimal if the errors e_i follow a Gaussian distribution. Recently, people have realized that actual data do not often satisfy Gauss' assumptions, and it can have dramatic effects on the OLS results. One of the basic assumptions of regression analysis is equality of the error variance along the predicted line: a condition called homoskedasticity which provides a degree of uniformity to the confidence intervals. If the residual distribution is normally distributed, the analyst can determine where the level of significance or

rejection regions begin. Even if the sample size is large, the influence of the outlier can increase the local and possibly even the global error variance. This inflation of error variance decreases the efficiency of estimation.

Another form of violation resides in the lack of independence of observations. When this happens, the R^2, F and t values become inflated. Failures of these assumptions can influence output toward false statistical significance. However, failure of basic classical regression model assumptions can be detected using the proper tests. Another assumption is a normality of the residuals. When there are violations of normality assumption of the residuals in OLS regression analysis, the estimation of significance becomes impaired.

Several assumptions need to be judged whether the OLS regression model is valid or not. When the regression model does not meet the fundamental assumptions, the prediction and estimation of the model may become biased. Residuals, differences between the values predicted by the model and the real data, which are very large, can seriously distort the prediction. When these residuals are extremely large, they are called outliers. A problem encountered in the application of regression is the presence of an outlier or outliers in the data. Small sample taken from a different population, which has outliers, can have an impact on statistical inference. Even one outlying observation can destroy OLS estimation, resulting in parameter estimates that do not provide useful information for the majority of the data. The outliers will make the error variance inflate. The confidence interval becomes stretched. The estimation cannot become asymptotically consistent. When outliers inflate the error variance, they undermine the model of power to detect the outliers.

OLS regression possesses the lowest possible breakdown point of $1/n$, where n denotes the number of observations. This indicates that OLS cannot handle a single outlier because one outlier can be sufficient to move the coefficient estimates arbitrarily far away from the actual underlying values. Thus, outliers cause unreliable coefficient estimates if OLS is applied. This weakness of OLS estimation in outlying observations has been demonstrated in various studies (Hampel et al., 1986, Huber, 1973, and Rousseeuw and Leroy, 1987). Reliable results are provided by OLS if and only if outlier diagnostic and treatment tools such as robust regression methods or robust regression diagnostics are applied as well. A least squares analysis assigns weights to each observation equally in achieving the parameter estimates. Therefore, it has to be sensitive to outlying cases. The robust methods, however, enable the

observations to be weighted unequally, so that observations with large residuals are down-weighted. The application of these methods ensures the non-inclusion or the appropriate down-weighting of outliers in the analysis.

A drawback of OLS is that, by considering squared residuals, it tends to award an excessive importance to observations with very large residuals and, consequently, distort parameters' estimation in case of existence of outliers. That is why the outliers in regression are needed to discuss as shown below before discussing robust estimation procedures.

4.3 Outliers in Regression

Outliers play an important role in regression. Outliers in the response variable represent model failure. Such observations are called outliers. Outliers with respect to the predictors are called leverage points. They can affect the regression model, too. Their response variables need not be outliers. Observation whose inclusion or exclusion results in substantial changes in the fitted model (coefficients, fitted values) is said to be influential. For this, about the types of outliers that can be found in regression analysis, their effects on regression coefficients and outliers detection were discussed in following subsection 4.3.1 and 4.3.2.

4.3.1 Types of Outliers in Regression

According to Rousseeuw and Van Zomeren (1990), there are several kinds of outliers. They proposed vertical outlier, good leverage point and bad leverage point. A point (x_i , y_i) which does not follow the linear pattern of the majority of the data but whose x_i is not outlying is called a vertical outlier. A point (x_i , y_i) whose x_i is outlying is called a good leverage point which follows the pattern of the majority, and a bad leverage point otherwise. To summarize, a data set can contain four types of points: regular observations, vertical outliers, good leverage points, and bad leverage points. Of course, most data sets do not have all four types. These types of outliers are shown in diagrammatic form.

Figure 4.1 shows these four types in simple regression. Point *A* clearly deviates from the typical linear relationship between the dependent (*Y*) and the independent (*X*) variable. Such ‘vertical’ outlier is characterized by an unusual observation in the dependent variable. The impact of vertical outliers on the

estimation of regression coefficients is usually small and mainly affects the regression intercept. If unusual observations occur in the set of independent variables, these outliers are called leverage points. If such leverage point deviates from the linear relationship described by the majority of observations it is called ‘bad leverage point’ such as Point *B* in Figure 4.1. Due to the exposed position of the outlier it has a leverage effect on the coefficient estimation. In contrast, a leverage point is called ‘good leverage point’ if it does not deviate from the typical relationship. Good leverage points are no outliers and even improve the regression inference as these points reduce standard errors of coefficient estimates.

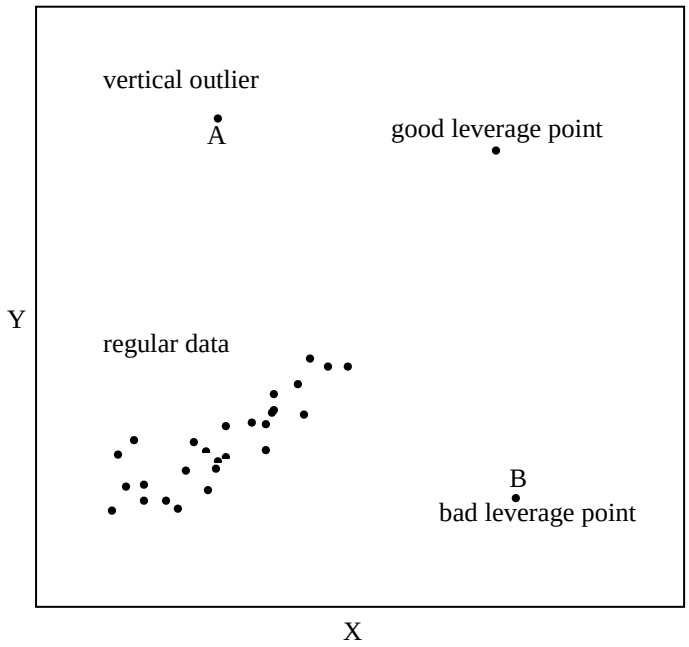


Figure 4.1 Simple Regression Data with Points of All Four Types

Rousseeuw and Van Zomeren (1990) pointed out that high leverages can affect the estimated slope of the regression line in OLS, thus they may cause more serious problems than other outliers which might only affect the estimated intercept term. Moreover, their occurrence in regression models may move to some low leverage as well as high leverage and it can turn in vice versa. These two concepts are called masking and swamping in linear regression (Rousseeuw and Leory, 1987). Furthermore, the range of explanatory variables increases when they exist in regression analysis. Thus, the multiple coefficient determination statistics (R^2) which is a well known and popular measure of goodness-of-fit in the regression models will increase even by any changes of a single *x* variable (Ryan, 1997). In addition, high

leverages may be the prime source of collinearity-influential observations whose presence can make collinearity and can destroy the existing collinearity pattern among the *x* variables (Hadi, 1992). In this respect, the identification of high leverage points to prevent their effect on linear regression becomes necessary.

4.3.2 Detection of Outliers in Regression Analysis

Outlier detection is one of the most important tasks in data analysis. There are various methods for detecting outliers in the literature of regression analysis. A simple outlier diagnostic tool is the scatter plot that enables the detection of outliers in simple regression cases. However, this is impossible if the dimension of the problem exceeds the simple regression case and the number of observations is very large.

Outlier diagnostics based on residual plots might suffer from outliers (Rousseeuw and Leroy, 1987), in particular for bad leverage points. Outliers can tilt the (original) regression line and have small regression residuals. Thus, outliers might not be discovered in residual plots. Other diagnostic tools are required to identify outlying or influential observations. However, they may involve additional problems. Studentized and jackknifed residuals, Cooks distances and other diagnostics based on hat matrix elements, for instance, are vulnerable to the so called masking effect. If more than one outlier occurs, these outlier diagnostics might not be able to detect a single one because one outlier can be masked by the presence of others.

One way to identify possible multivariate outliers is to calculate a distance from each point to a center of the data. Mahalanobis distance (MD) is one of these well-known multivariate methods for detecting high leverage points as well. MD gives a one-dimensional measure of how a point is far from a location with respect to a shape. Utilizing MD, the points, which are unusually far away from a location are found and call those points outlying.

Although it is a reliable diagnostic tool for detecting high leverage points, it suffers from masking problem. Most of the classical diagnostic methods fail to identify the multiple high leverage points due to their masking effects. Data sets with multiple outliers are subject to problems of masking and swamping. Although it is still quite easy to detect a single outlier by means of the MD, this approach no longer suffices for multiple outliers because of the masking effect, by which multiple outliers do not necessarily have large MD.

Masking occurs when a group of outlying points skews the mean and covariance estimates toward these points and the resulting distance of the outlying point from the mean is small. While, swamping occurs when a group of outlying points skews the mean and covariance estimates toward these points and away from other inlying points and the resulting distance from the inlying points to the mean is large. Problems of masking can be resolved by using robust estimates of shape and location, which by definition are less affected by outliers. Multiple-case diagnostics or high-breakdown diagnostics have to be employed instead. Therefore, robust regression and outlier identification based on robust regression residuals for identification were applied and adequate treatment of outliers. Two robust estimators of multivariate location and scatter, MVE and MCD, have already been discussed in sections 3.11 and 3.12.

Regression diagnostics aim to detect observations of one or more of these above stated types. There are many numerical and graphic diagnostics for detecting outliers and leverage points. Among them three plots such as the Mahalanobis distance, the standardized residuals and the diagnostic plot were used for this study. These three diagnostics plots were discussed in the followings.

(a) Mahalanobis Distance

Outlying points are less likely to enter into the calculation of the robust procedures, so they will not be able to influence the parameters used in the MD. The inlying points, which all come from the underlying distribution, will completely determine the estimate of the location and shape of the data. For a *p*-dimensional multivariate sample $x_i\,(i=1,2,...,n)$, the MD is defined as:

$$MD(x_i)=\sqrt{(X-T(X))'C(X)^{-1}(X-T(X))}\quad for\, i=1,...,n \tag{4.6}$$

where:

$T(X)$ = The estimated multivariate location which is usually the multivariate arithmetic mean

$C(X)$ = The estimated covariance matrix which is usually the sample covariance Matrix.

If there are only a few outliers, large values of MD, indicate that the point x_i is an outlier (Barnett and Lewis, 1978). Any value of which the MD exceeds the cutoff $\sqrt{\chi^2_{p,0.975}}$ is considered as outliers where p is the number of explanatory variables.

(b) Standardized Residuals

Standardized residuals are defined as

$$e_i / \hat{\sigma} \tag{4.7}$$

where $\hat{\sigma}$ denotes a robust scale estimate based on the LTS residuals which is estimated from the Equation (3.29). Standardized residuals help us to distinguish between well-fitting and non-fitting observations by comparing their absolute values to some yardstick,

$$\text{compare } |e_i| / \hat{\sigma} \text{ to 2.5.}$$

The yardstick 2.5 is used since it would determine a (roughly) 99% tolerance interval for e_i if they had a standard Gaussian distribution. Since the standardized residuals approximate the e_i, an observation can be considered as non-fitting if its standardized residual lies (far) outside this tolerance region.

(c) The Diagnostic Plot

The diagnostic plot makes the complete classification into the four types.

Rousseeuw and Leroy (1987) introduced the robust distance given by

$$RD(x_i) = \sqrt{(X - T_R(X))' C_R(X)^{-1} (X - T_R(X))} \quad for\ i = 1, ..., n \tag{4.8}$$

where, $T_R(X)$ and $C_R(X)$ are robust location and shape estimate such as MCD or MVE. By using a robust location and shape estimate in the RD, outlying points will not skew the estimates and can be identified as outliers by large values of the RD. Unfortunately, using robust estimates gives RD with unknown distributional properties. The use of $\sqrt{\chi^2_{p,0.975}}$ quantile as cutoff point for RD will be prone to declare some good, low leverage and high leverage points and often leads to identifying too many points as outliers (Rousseeuw and Van Zomeren, 1990). In general, each observation (x_i) is given by weight

$$w_i = 1 \quad if \;\; RD(x_i) \le \sqrt{\chi^2_{p,0.975}} \;\; and$$
$$w_i = 0 \quad otherwise. \tag{4.9}$$

The resulting one-step reweighted mean and covariance matrix

$$T_1(X) = \left(\sum_{i=1}^{n} w_i x_i\right) / \left(\sum_{i=1}^{n} w_i\right)$$

$$C_1(X) = \left(\sum_{i=1}^{n} w_i \,(x_i - T_1(x))(x_i - T_1(x))'\right) / \left(\sum_{i=1}^{n} w_i - 1\right) \tag{4.10}$$

have a better finite-sample efficiency than the MCD. The updated robust distances RD are then obtained by inserting $T_1(x)$ and $C_1(x)$ into (4.8).

4.4 Robust Estimates for Regression Model

Robust regression is an important tool for analyzing data that are contaminated with outliers. It can be used to detect outliers and to provide resistant (stable) results in the presence of outliers. The main idea of robust regression is to give little weight to outlying observations in order to isolate the true underlying relationship.

In this section, the five robust estimators such as LAV, LMS, LTS, *M*, and *MM*-estimators were chosen to describe the calculation procedures. In the previous chapter, these estimators were discussed in detail. In this section, the computation algorithms of these estimators were explained vividly.

4.4.1 Algorithm for Computing LAV Estimates

The LAV algorithm is an iterative algorithm and it converges in a finite number of iterations. It can be described as follows:

1. Compute the initial estimates from a weighted least squares of absolute residuals.
2. Estimate the quantile of the dependent variable, the median, by taking the raw sum of absolute deviations around the unconditional median.
3. Find the regression coefficients that minimize the objective function.

In our studies, it was found that the algorithm generally converges after a few less than 10 iterations.

4.4.2 Algorithm for Computing LMS Estimates

The LMS algorithm is a resampling algorithm and it performs four steps. The key idea is to start with sub samples of size *p* are drawn from *n* observations. A detailed procedure is shown below:

1. Drawing sub samples of size *p* from *n* observations.
2. A trial estimate β_J is computed from each sub sample.
3. For each β_J, the corresponding LMS objective function with respect to the whole data set can be determined.
4. Repeat steps 1 until 3, and keep the estimate with lowest objective function value that is the fit with the lowest median of squared residuals is retained.

Once the above algorithm has been carried out, it remains possible to assign weights to the observations based on their LMS residuals, and then to perform a reweighed least squares analysis.

4.4.3 Algorithm for Computing LTS Estimates

The LTS estimation algorithm is a resampling algorithm and starts from randomly drawn *p*-subsets. For each *p*-subsets the coefficients of the hyperplane through the points in the subset are calculated. Then the algorithm obtains the estimate with a lowest value for the objective function. The algorithm formally described in the following steps:

1. Draw a random subset of *p* observations. For small data sets it is possible to consider all *p*-subsets, whereas for larger data sets many *p*-subsets are drawn at random.
2. Compute hyperplane through these *p* observations and obtain an initial estimate of the slope and the intercept.
3. If regression with intercept, adjustment is made.
4. Evaluate the objective function at this estimate.
5. Repeat steps 1 until 4, and keep the estimate with lowest objective function value.

4.4.4 Algorithm for Computing *M*-Estimates

An iterative procedure is necessary to find *M*-estimates for regression. A single step is impossible because the residuals can not be found until the model is

fitted, and the estimates can not be found without the residuals. As a result, iteratively reweighted least squares (IRLS) is employed:

1. Setting the iteration counter at $I = 0$, an OLS regression is fitted to the data, finding initial estimates of the regression coefficients $\hat{\beta}^{(0)}$.
2. The residuals are extracted from the preliminary OLS regression, $e_i^{(0)}$, and used to calculate initial estimates for the weights.
3. A weight function is then chosen and applied to the initial OLS residuals to create preliminary weights, $w(e_i^{(0)})$.
4. The first iteration, $I = 1$, uses weighted least squares (WLS) to minimize $\sum w_i^{(1)} e_i^2$ and thus obtain $\hat{\beta}^{(1)}$. In matrix form, with W representing the $n \times n$ diagonal matrix of individual weights, the solution is $\hat{\beta}^{(1)} = \left(X^T W X\right)^{-1} X^T W y$.
5. The process continues by using the residuals from the initial WLS to calculate new weights, $w_i^{(2)}$.
6. The new weights $w_i^{(2)}$ are used in the next iteration, $I = 2$, of WLS to estimate $\hat{\beta}^{(2)}$.
7. Steps 4–6 are repeated until the estimate of $\hat{\beta}$ stabilizes from the previous iteration.

Specifically telling, at each of the q iterations, the solution is $\hat{\beta}^{(I)} = \left(X^T W_q X\right)^{-1} X^T W_q y$, where $\underset{(n \times n)}{W_q} = diag\left\{w_i^{(I-1)}\right\}$. The iteration process continues until $\hat{\beta}^{(I)} - \hat{\beta}^{(I-1)} \cong 0$. Typically, the solution is considered to have converged when the change in estimates is no more than 0.01% from the previous iteration. The asymptotic covariance matrix of $\hat{\beta}$ is

$$V\left(\hat{\beta}\right) = \frac{E\left(\psi^2\right)}{\left[E\left(\psi'\right)\right]^2} (X'X)^{-1}$$

using $\sum \left[\psi\left(e_i\right)\right]^2$ to estimate E(ψ^2), and $\sum \left[\psi'\left(e_i\right)/n\right]^2$ to estimate $\left[E\left(\psi'\right)\right]^2$ produce the estimated asymptotic covariance matrix, $\hat{V}\left(\hat{\beta}\right)$ (which is not reliable in small samples).

4.4.5 Algorithm for Computing *MM*-Estimates

The *MM*-estimation algorithm consists of three steps. It can be described as follows:

1. To compute the *S*-estimators, it is necessary to select N subsets of p observation (defined as p-subset). For each p-subset, the equation of the hyperplane that fits all points perfectly is obtained yielding a trial solution of (3.43) and residuals are computed.
2. Based on these residuals obtained in step 1, a scale estimate is computed for each p-subset. An approximation for the final scale estimate is given by the trial value that leads to the smallest scale over all p-subsets.
3. At this stage, it is required to compute the final *MM*-estimates.

4.5 Robustness in Dummy Variables

In the previous sections, it was found that both response and regressors are continuous, but in practice, data can be mixed with both continuous and categorical regressor variables. Thus, dummy variables in regression needs to be considered. The classical linear regression model shown in (4.1) can alternatively be described as

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i, \tag{4.11}$$

where $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1,2,...,n,$ the explanatory variables x_{ij} are often quantitative. In real situation, qualitative variables can be included in regression model as explanatory variables. This situation often occurs in social sciences, where the explanatory variables may include gender, ethnic background, professional occupation, marital status and so on.

If there are qualitative explanatory variables, then it is conventional in practice to encode such regressors by binary dummy variables. Extending model (4.11) to a model with continuous and discrete regressors, the model is expressed in the following form:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{k=1}^{q} \delta_k I_{ik} + \varepsilon_i, \; i = 1,...,n \tag{4.12}$$

where I_k denotes the dummy variable with elements either zero or one. The OLS method fits the model (4.12) in a nonrobust way. For instance, it is possible to apply

the standard calculations by processing the dummy variables in the same manner as the continuous ones, as described by Draper and Smith (1981) and Chatterjee and Price (1977).

Unfortunately, the OLS method is very sensitive to outliers. Even a small fraction of contamination can influence the regression estimates. In the above stated model (4.12), both outliers in the *y*-direction and in the *x*-direction can occur. The $p + q$ coefficients $\beta_1,...,\beta_p$ and $\delta_1,...,\delta_q$ in (4.12) are estimated in a robust way. Armstrong and Frome (1977) developed a linear programming algorithm to obtain the least absolute values (L_1) estimate. This approach is already more robust against outliers in the *y*-direction, but not against outliers in the *x*-direction which can still tilt the estimated hyperplanes.

One of the mostly applied methods of robust regression is *M*-estimation (also OLS and L_1 belong to this class). This approach can also be applied to the model (4.12), as done by Birch and Myers (1982) for the case of one categorical variable. One then has to solve a system of $p + q + 1$ implicit equations, using an IRLS algorithm. But *M*-estimators are still vulnerable to leverage points.

Therefore, it seems natural to extend regression methods that are able to withstand a positive percentage of contamination, including leverage points. Typical examples are the LMS estimator and the LTS estimator (Rousseeuw, 1984), and the class of *S*-estimators (Rousseeuw and Yohai, 1984). However, Hubert and Rousseeuw (1996) pointed out that, these estimators in (4.12) can not be simply calculated by treating the dummy variables in the same way as the continuous regressors, since this would lead to a problem of singular matrices. The typical algorithm for LMS regression in the model (4.11) starts by drawing a subset of $p + 1$ observations. Then the hyperplane through these $p + 1$ points is obtained, and the corresponding objective function computed. This procedure is repeated often, and the best fit is kept. But in the case of $p + q$ regressors of which q are binary variables, a large majority of the ($p + q + 1$)-subsets will be of less than full rank, hence the hyperplanes cannot be computed.

Therefore, RDL_1 proposed by Hubert and Rousseeuw (1997), which is particularly used in the case when the regression model consists of binary regressors. They proposed the RDL_1 estimator which can withstand contaminations in the data for model (4.12). RD stands for robust distance, and L_1 indicates the least absolute

values method in the estimates of the regression coefficients. The computation algorithm of RDL_1 estimates is described in the followings.

4.5.1 Algorithm for Computing RDL_1 Estimates

The RDL_1 consists of three steps: identifying leverage points, downweighting the leverage points when estimating the parameters, and estimating the residual scale.

1. In the first step, the MVE estimator introduced by Rousseeuw (1984) is applied to compute the robust *RD* for the continuous predictors. The *RD* have already mentioned in Equation (4.8) of subsection (4.3.2). These distances (4.8) are used to identify the leverage points for the space of continuous regressors and to be the weights for estimating the regression coefficients by a weighted L_1 procedure at the second step.
2. In the second step, the parameters ($\boldsymbol{\beta}$, $\boldsymbol{\delta}$) of model (4.12) are estimated by a weighted L_1 procedure

$$\min_{\beta,\delta} \sum_{i=1}^{n} w_i |e_i|, \tag{4.13}$$

where $e_i = y_i - \left(\beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \sum_{k=1}^{q} \delta_k I_{ik} \right)$, and

$$w_i = \min\left\{1, \frac{p}{(RD(x_i))^2}\right\}, \; for\; i = 1,...,n \tag{4.14}$$

applied to the observations $(y_i, 1, x_{i1}, ..., x_{ip}, I_{i1}, ..., I_{iq})$.

3. In the final step, the scale of the residuals $\hat{\sigma}$ is estimated from the Equation (3.10).

The robust estimate $(\hat{\beta}, \hat{\delta}, \hat{\sigma})$ can be used to detect regression outliers, by flagging the observations whose absolute standardized residual $|e_i / \hat{\sigma}|$ exceeds 2.5. The entire three-stage procedure is called the RDL_1 estimation. The applications of RDL_1 estimator to the simulated data as well as a real data are shown in section 4.6.

4.6 Simulation and Real Data Study

In this section, simulated data as well as real data were used to analyze the effect of the outlier in regression by using the OLS method and robust methods. The statistical software packages namely S-PLUS 2000, STATA 10, and SPSS 13.0 were used to obtain the desire estimates throughout the analyses. In the following subsections, the performances of OLS and robust estimators were analyzed by simulations. Then, simulated data were used to show that the robust methods outperform the classical method in presence of outliers. The real data study was followed then.

4.6.1 Simulation Study

(a) Analysis of Outlier Robustness in Multiple Linear Regression

To study the outlier robustness, a multiple linear regression model with three explanatory variables was used. The data sets were generated from the following model:

$$y_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + e_i,\ i = 1,2,...,n \tag{4.15}$$

where all regression coefficients are fixed $\beta_0 = 5$ and $\beta_j = 1$, for each $i = 1,2,\ldots,n$ and $j = 1,2,\ldots,p$. The explanatory variables were randomly generated from a normal distribution with mean 0 and unit variance. The errors were assumed to be i.i.d. with $N(0,0.5)$.The data sets were generated under three regressors ($p = 3$) and the sample sizes were ($n = 30$ and $n = 40$) respectively. The true y's were calculated from the Equation (4.15). In this simulation study, two types of outlier namely vertical outlier and bad leverage point were studied because they give different effects in the estimation of parameters of the regression model. After generating the data sets, two scenarios were considered in the following manners. They were seen as follows

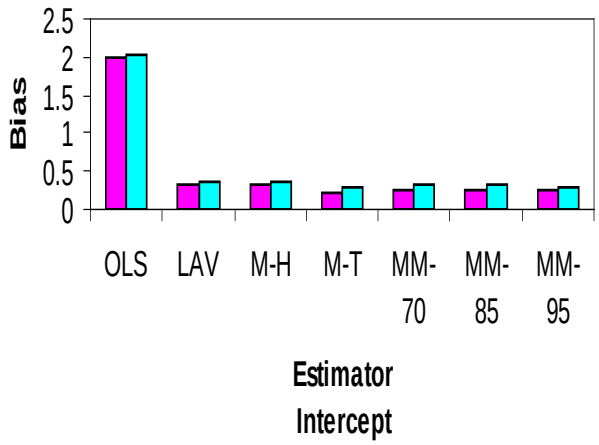
(i) outliers in the independent variable: 10% of the y observations set to be vertical outliers by multiplying constant number 5 and keeping the others.

(ii) outliers in both y and x: 10% of both y and x observations were modified to be vertical outlier and bad leverage points and the remaining were unchanged. The vertical outlier was obtained by multiplying 5 to its y value and the bad leverage point was obtained by adding 10 to its x value.

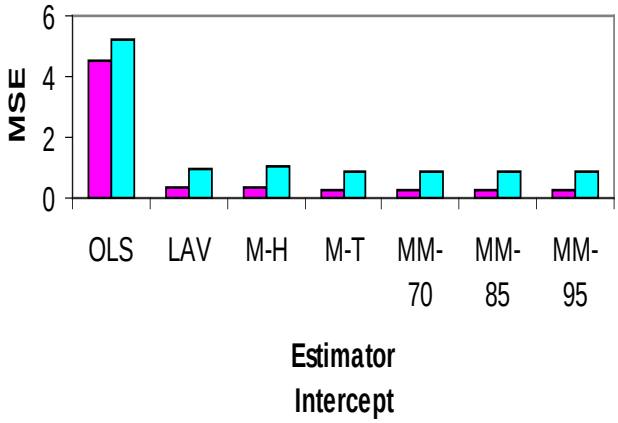
All simulations were done with 100 replications. To measure the robustness, the bias (that is the average of the estimated parameters minus the true value) and the mean squared errors (that is the variance of the estimated parameters plus the square of the bias) were used. For the first scenario, among the robust methods the LAV, *M* and *MM*-estimators were applied to this simulated data set because these estimators are robust subject to the vertical outliers. Then, this procedure was repeated 100 times and each time the parameters of OLS, LAV, *M*- (using Huber and Turkey) and *MM*-estimators (with a 70%, an 85% and a 95% efficiency) were estimated. On the basis of all the estimated parameters, the bias and the MSE were computed and the results were presented in Appendix A, Table (A.4.1). Figure 4.2 summarizes the results of simulations where $n = 30$ and $n = 40$ observations and three predictors. Bars represent bias and MSE for each estimator.

It is seen in Appendix Table (A.4.1) and Figure 4.2 that in the presence of vertical outliers, both the bias and MSE obtained from the *MM*-estimators (with a 70%, an 85% and a 95% efficiency), Huber and Turkey-*M*, and LAV are much close to each other but inferior to the OLS estimator. Their patterns shown in Figure 4.2(a) to (d) are intermingled and so no methods have a preferable bias and MSE in this case.

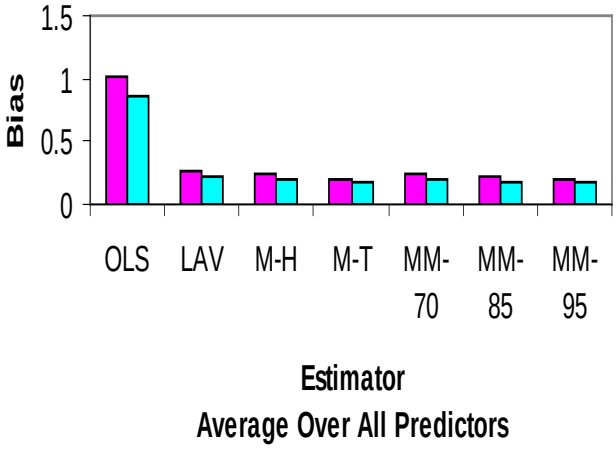
In the case of second scenario, the LAV, *M*, *MM*, LTS and LMS estimators were applied to this simulated data set. The results are shown in Appendix A, Table (A.4.2) and Figure 4.3. According to Figure 4.3, the bias and MSE obtained from the Huber and Turkey-*M* are the smallest, followed by the *MM*-estimators (with a 70%, an 85% and a 95% efficiency) and LTS estimator in presence of vertical outlier and bad leverage points. In this case, the LMS behaves differently but just slightly, and have a bias and an MSE comparable to that of Huber and Turkey-*M* and *MM*-estimators. The OLS method also indicated in Figure 4.3(a) to (d) performs much worst in these situations. Therefore, the low bias and MSE values of the Huber and Turkey-*M* and *MM*-estimators are in line with the asymptotic robustness properties. As expected, OLS is a relatively less efficient estimator whatever the type of outliers occurred in the data.



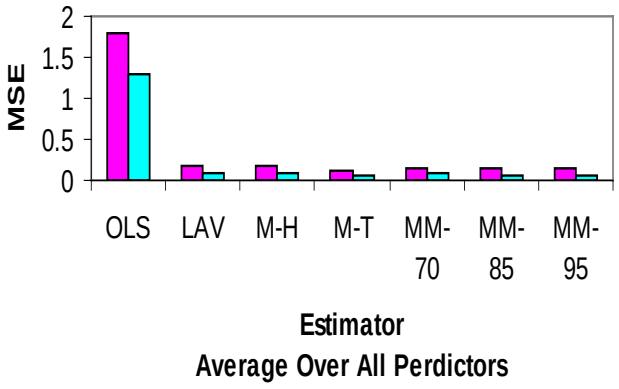
(a)



(b)



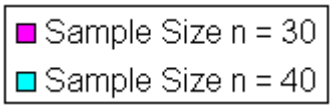
(c)

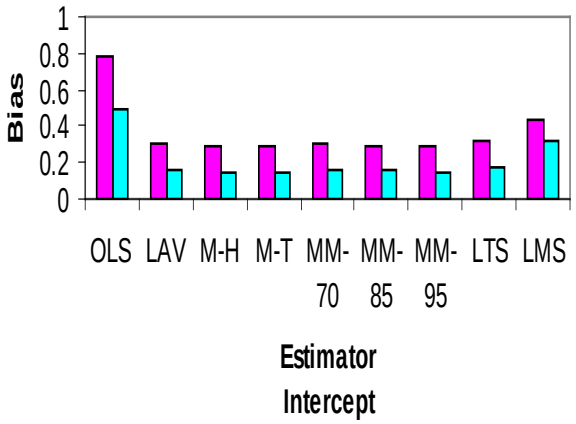


(d)

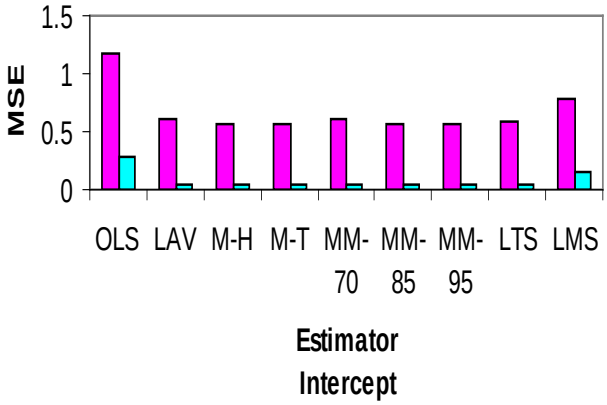
Figure 4.2 Bias and MSE of Simulated Data with Vertical Outliers

Source: Appendix Table (A.4.1)

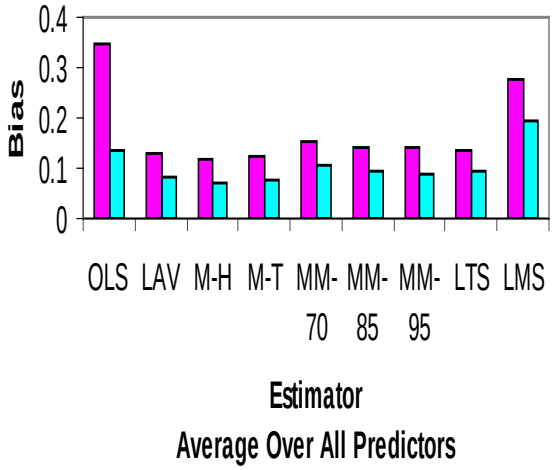




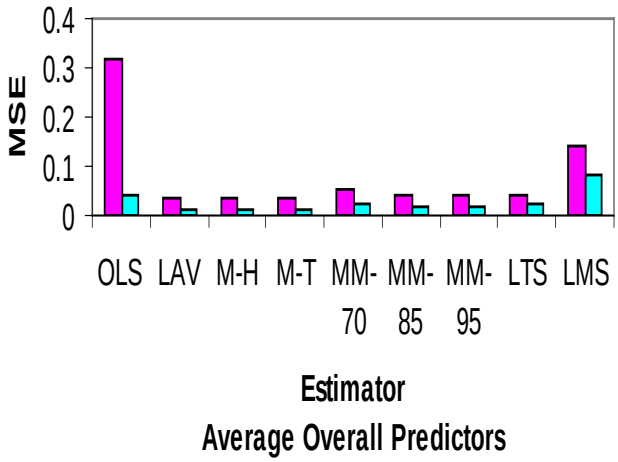
(a)



(b)



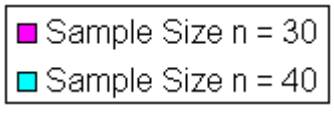
(c)



(d)

Figure 4.3 **Bias and MSE of Simulated Data with Vertical Outlier and Bad Leverage Points**

Source: Appendix Table (A.4.2)



(b) Analysis of Distributional Robustness in Multiple Linear Regression

Another simulation study was carried out to analyze the effects of outliers on parameter estimation. When performing such a simulation study, different error structures were taken into account. In this study, the multiple linear regression model with three explanatory variables was used. In this model, the intercept and slopes were equal to one. These explanatory variables were generated from standard normal distribution. In this simulation study, the errors which contain outliers were generated using heavy-tailed distribution (compare to standard normal distribution) such as logistic, Cauchy and skewed independent data sets like gamma and exponential distribution. Thus, the errors were simulated from the following densities: *N*(0,1), *LOG*(0,1), *EXP*(1), *C*(0,1), and *GAM*(1,0.5). Table 4.1 shows the notation and parameters of distributions, which are used in the simulation process.

Table (4.1) Notation and Parameters of Distribution

Distribution	Notation and Parameters	p.d.f. f(x)
Normal	$x \sim N(\mu,\sigma^2)$, $-\infty < \mu < +\infty,\ \sigma > 0$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-[(x-\mu)/\sigma]^2/2}, -\infty < x < +\infty$
Logistic	$x \sim LOG(\theta,\eta),\ \theta > 0$	$\frac{1}{\theta}\frac{\exp[(x-\eta)/\theta]}{\{1+\exp[(x-\eta)/\theta]\}^2}, -\infty < x < +\infty$
Exponential	$x \sim EXP(\theta), \theta > 0$	$\frac{1}{\theta}\exp\left(\frac{-x}{\theta}\right),\ x > 0$
Cauchy	$x \sim C(\mu,\sigma)$, $-\infty < \mu < +\infty, \sigma > 0$	$\frac{1}{\pi\sigma\left[1+\left(\frac{x-\mu}{\sigma}\right)^2\right]},\ -\infty < x < +\infty$
Gamma	$x \sim GAM(\theta,k), \theta > 0,\ k > 0$	$\frac{1}{\theta^k \Gamma(k)} x^{k-1} e^{-x/\theta},\ x > 0$

In each case, 10 replications were simulated and regression coefficients of OLS, LAV, Huber and Turkey *M*-estimates, LMS and LTS were calculated. To compare the properties of the estimation procedures, the mean squared errors (MSE) and bias of the estimated coefficients were computed using the following formulas

$$MSE = \frac{1}{10}\sum_{i=1}^{10}\left(\hat{\beta}_i - \beta\right)^2$$

$$Absolute\ Bias = \frac{1}{10}\sum_{i=1}^{10}\left|\hat{\beta}_i - \beta_i\right|$$

Overall results of the methods under study and corresponding MSE and bias of 10 simulations for each estimation method were shown in Appendix A, Tables (A.4.3 to A.4.7) and Appendix B, Figures (B.4.1 to B.4.5). Those figures illustrate the results of MSE and bias for the coefficients of multiple linear regression model with three explanatory variables (p = 3).

Based on the results of normal distributions, the bias of OLS is the smallest as expected, followed by the bias of Turkey and Huber-*M*, respectively. Moreover, in this case, the MSE of OLS is the smallest followed by the values of MSE of Huber and Turkey-*M*, respectively. It is found that, the OLS method is more efficient than the robust methods under normal error distribution. Thus, the low bias and MSE values of the OLS method are in line with the asymptotic robustness properties. In this normal distribution, the bias and MSE of LMS are much greater which followed by the biases and MSEs of LTS and LAV methods. The LMS method performs much worst in this case. [See in Appendix A, Table (A.4.3)]

Yet, as for logistic distributions, the bias of OLS, Turkey-*M* and Huber-*M* are close to each other and perform better than LAV, LTS and LMS methods. In this case, the MSE of OLS, Turkey and Huber-*M*, LAV and LTS methods are much close to each other but this value for LMS is significantly larger. Furthermore, although biases and MSEs of OLS, Turkey and Huber-*M* are significantly smaller than the bias and MSE of LMS, their patterns as shown in Appendix B, Figure B.4.2(a) to (h) are intermingled and so no methods have a preferable bias and MSE in this situation. The bias and MSE of LMS are much greater than the others. [See in Appendix A, Table (A.4.4)]

In exponential distributions, the LAV, LTS and Turkey-*M* are close to each other, but inferior to the Huber-*M* in terms of intercept. The bias of LMS is the smallest in this case. The OLS method as shown in part (a) of Appendix B, Figure B.4.3 performs much worst in these situations. The MSE of OLS, for this situation is much greater which followed by the MSE values of Huber-*M*, Turkey-*M*, LTS and LAV respectively. The MSE of LMS is the smallest in this case. As indicated in Appendix B, Figure B.4.3(c) to (h), the general pattern of the bias and MSE values for

all methods are intermingled so that no preferred method could be chosen for the study of slope coefficients. [See in Appendix A, Table (A.4.5)]

In Cauchy distribution, regarding intercept, the biases of the robust methods are so close to each other, but the OLS method as shown in Appendix B, Figure B.4.4 (a) performs much worst in this situation. Furthermore, the MSE of OLS is significantly larger than the MSE of robust methods. The robust methods are so close to each other and their pattern as shown in Appendix Figure B.4.4(b) is intermingled. So, no methods have a preferable MSE in this case. The similar results are found in the study of slope coefficients. The OLS method performs much worst based on bias criterion in this study. The biases of LAV, LTS, Turkey and Huber-*M* and LMS are so close to each other. In this case, the MSE of OLS is significantly larger than the MSE of robust methods. From this study, it is found that, the general patterns of the bias and MSE values for all robust methods are intermingled so that no preferred method can be selected for this case. [See in Appendix A, Table (A.4.6)]

In the case of gamma distribution, the intercept of LAV, LTS and Turkey-*M* are close to each other, but inferior to the Huber-*M* depending on bias criterion. In this situation, the bias of LMS is significantly smaller than the bias of other methods. The OLS method as described in Appendix B, Figure B.4.5(a) performs much worst. In addition, in this case, the MSE of LAV, LTS and Turkey-*M* are close to each other. The MSE of LMS is the smallest and performs better than the other methods. The OLS method as described in Appendix Figure B.4.5(b) performs much worst. Moreover, the bias and MSE of Turkey-*M* is the smallest in terms of the slope. It is closely followed by the bias and MSE values of Huber-*M*, LAV, LTS and OLS. [See in Appendix A, Table (A.4.7)]

(c) Analysis of Effect of Outliers in Multiple Linear Regression Model with Categorical Variable

Next, to show the fact that the robust procedure outperforms the classical procedure in dummy variable regression model, simulation with data set contaminated by different types of outliers was carried out. Observations can be classified into good data, good leverage points, vertical outliers, and bad leverage points in regression analysis. Taking this idea, a set of observations consisting of these four types of model (4.12) with two continuous variables (p = 2), one categorical variable (q = 1), and n = 30 was used. The data were generated by the following model

$$y_i = 9 + x_{i1} + x_{i2} + I_{i1} + \varepsilon_i \text{ , } i = 1,\ldots,30, \qquad (4.16)$$

where both x_{i1} and x_{i2} follow a standard normal distribution, I_{i1} is a binomial distribution with a success rate 0.5, and ε_i is a normal distribution with mean zero and standard deviation 0.5. Once these 30 observations have been generated, cases 25 and 26 are then set to be vertical outliers by doubling their y values and keeping the others. Cases 27 and 28 are bad leverage points by adding 9 to their x_1 values and keeping the others as well. Case 29 and 30 are good leverage points by adding 9 to both x_1 and x_2 values and reproducing the corresponding y values as model (4.16). The resulting simulated data are presented in Appendix A, Table (A.4.8). The RDL_1, LTS and OLS estimates are applied to these simulated data and results are summarized in Table (4.2).

First, the RDL_1 estimate is applied to these simulated data. Part (a) of Appendix Figure B.4.6 shows the standardized residuals. Case 25, 26, 27, 28, 29, and 30 are revealed as outliers. This is due to the fact that the weight (4.14) is calculated by the continuous design matrix without considering the model fitting. The resulting weights are shown in Appendix B, Figure B.4.6(b). Therefore, case 27, 28, 29, and 30 are outlying observations from $\boldsymbol{X}$ space and will be given relatively small weights as shown in part (b) of Appendix Figures B.4.6. This makes case 29 and 30 become bad leverage points in the diagnostic plot of Appendix Figure B.4.6(c). The cutoff values are indicated ± 2.5 and $\sqrt{\chi^2_{2,0.975}}$ by horizontal and vertical lines. These results point

out that the RDL_1 method results in the swamping effect due to its weights for the L_1 procedure obtained from the application of continuous design matrix.

Thus, the LTS method is applied using the same data. Parts (a), and (b) of Appendix B, Figure B.4.7 show the results of robust standardized residuals and the diagnostic plot, respectively. From part (a), the LTS procedure identifies cases 25, 26, 27, and 28 to be outliers. It gives weight 1 for both cases 29 and 30. The corresponding diagnostic plot also divides all points into the right categories as the original configuration of these data being generated. The fit from the LTS method ignores outlying observations, which gives the MSE of 0.3042.

Next the OLS method is applied using the simulated data. Both the visual sketch [Figure 4.4(a)] and the large MSE of OLS for the simulated data set argue that the data are highly influenced by outliers. The OLS regression estimators often break down in the presence of vertical outliers or bad leverage points. It is evident from the graphical sketch of data as the OLS line is pulled towards the middle of the two groups of the data points rendering it an unrepresentative line. A Gaussian Q-Q plot Figure 4.4(b) confirms that the residuals are roughly normally distributed. Only a small number of outliers cause the distribution to be heavier-tailed.

According to the result of LTS analysis, the observations (25, 26, 27, and 28) gained from the simulated data are excluded and the remaining data are rerun using the OLS method. Table (4.3) presents the regression results for the two techniques. New regression represents the regression results after eliminating the outlier found through LTS analysis. The intercept and slope coefficients changed and all are statistically significant at 1% level. The fact that the F and R^2 values increased indicates that the new regression is well matched with those remaining data. The OLS line fits the simulated (non-contaminated) data well with a reasonable MSE of 0.1230. Figure 4.5(a) shows the OLS residuals without considering the cases 25, 26, 27, and 28. Cases 29 and 30 are obviously located near the regression surface. Figure 4.5(b) suggests that the residuals are approximately normally distributed.

This simulation also shows that the RDL_1 regression clearly outperforms OLS regression as well as LTS regression. The OLS regression is the best when data are free from outliers.

Table (4.2) OLS, LTS and RDL_1 Regression Models Fitted to the Simulated Data

Method	Coefficients				MSE
	Constant	x_1	x_2	I_{i1}	
OLS	8.84***	0.43**	1.34***	2.47**	6.1504
LTS	9.22***	1.16**	0.72***	1.37***	0.3042
RDL_1	8.94***	0.82***	0.77***	1.40***	0.1849

Note: (1) Significant at *** 1%, **5%, * 10%

Source: Appendix Table A.4.8

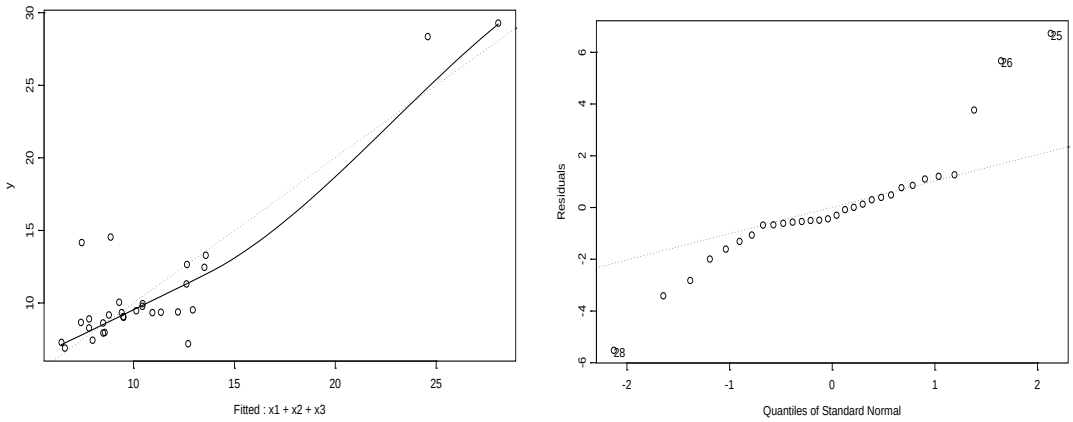
Table (4.3) Regression Results

OLS Regression				New OLS Regression		
Variable	Coefficient	Standard Errors of Coefficients	*t* Statistics	Coefficient	Standard Errors of Coefficients	*t* Statistics
Constant	8.8410 ***	0.5926	14.9203	9.1344***	0.0872	104.7172
x_1	0.4276 **	0.1737	2.4611	1.0139***	0.0412	24.6214
x_2	1.3386 ***	0.2355	5.6847	0.9480***	0.0504	18.8205
I_{i1}	2.4729 **	0.9456	2.6153	0.9954***	0.1484	6.7095
MSE = 6.1504, R^2 = 0.7995, F = 34.55***				MSE = 0.1230, R^2 = 0.996, F =2046.865***		

Note: (1) Absolute value of *t* statistics in parentheses

(2) Significant at *** 1%, **5%, * 10%

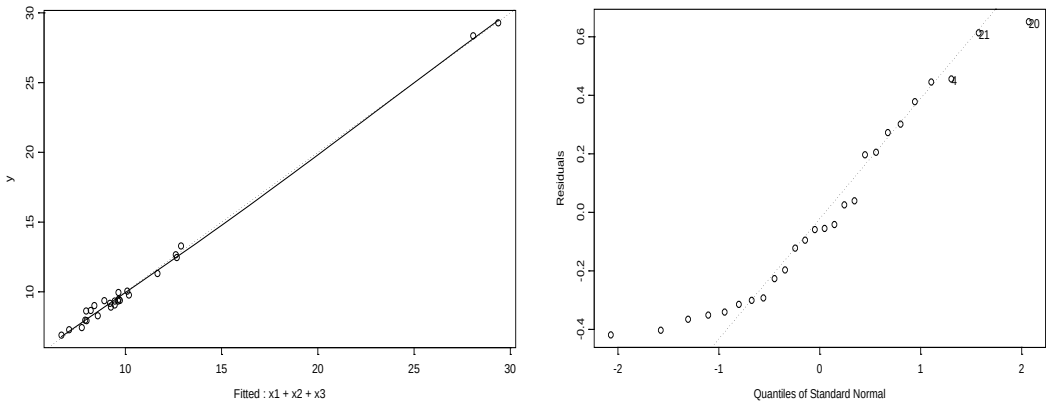
Source: Appendix Table A.4.8



(a) (b)

Figure 4.4 Simulated Contaminated Data Set Using the OLS: (a) scatter plot with OLS line; and (b) quantiles standard normal plot

Source: Appendix Table A.4.8



(a) (b)

Figure 4.5 Simulated Non-contaminated Data Set Using the OLS: (a) scatter plot with OLS line; and (b) quantiles standard normal plot

Source: Appendix Table A.4.8

4.6.2 Simple Linear Regression Using Real Data Study

In order to study the effect of regression outliers, it is required to start with simple linear regression. An attempt was made to explain the export of maize as a function of the production of maize. The required data set was taken from the Statistical Year Books. In this study, the dependent variable was the export of maize (Thousand Metric Tons) and the independent variable was the production of maize (Thousand Tons). The other factors such as sown acres and price etc. were not considered in this study. Only the export of maize related with the production was considered for the study. The scatter plot of export versus production was shown in the part (a) of Appendix B, Figure B.4.9. It shows that there exists a linear relationship between them and some outliers appear. The linear regression model of export of maize on the production of maize can be written as follows:

$$EXP_t = \beta_0 + \beta_1 PROD_t + \varepsilon_t \qquad (4.17)$$

where,

EXP = export of maize

PROD = production of maize

β_0 = intercept

β_1 = slope

ε_t = error.

All of the computed results from OLS method are summarized in Appendix A, Table (A.4.10). It is clear that from this table that the intercept is statistically insignificant but the slope is highly significant at 1% level. As expected, the production of maize coefficient is positive. The *F* value is 37.943 which is highly significant at 1% level and the coefficient of determination (R^2) is 0.558. This result shows that the production of Maize explains about 56 percent of the variation in export of maize. This value may signal that additional explanatory variables should be considered in the model or curvilinear regression. Moreover, Durbin-Watson (D.W.) statistic is 1.770 which lies between d_u = 1.502 and 4 - d_u = 2.498. This result suggests that there are no serially correlated residual values in the estimated model (4.17). Furthermore, residual analysis can be used to evaluate the appropriateness of the regression model that has already been fitted to the data.

First, the residuals of the fitted model are checked whether they satisfy the normality assumption by drawing the normal probability plot which is shown in part

(a) of Appendix B, Figure B.4.8. It shows that most of the points differ from the straight line. As a result of this, the residuals of the fitted model are not normally distributed. After that, it needs to confirm whether the variance of the residuals is constant or not. Then, the standardized predicted values are drawn on the horizontal axis against the corresponding standardized residual values on the vertical axis as shown in part (b) of Appendix Figure B.4.8. From this figure, a funnel shape is clearly seen. Thus, it can be concluded that the assumption of equal variance is violated. These results suggest that the transformation is needed. The model (4.17) can be transformed (by taking natural logarithms of both the dependent and independent variables) into the following model and then the regression was rerun using the transformed variables in place of the original variables. Thus, a new regression model can be described as follows:

$$\ln EXP_t = \beta_0' + \beta_1' \ln PROD_t + e_t \tag{4.18}$$

where

$\beta_0' = \ln \beta_0$

$\beta_1' = \ln \beta_1$

$e_t = \ln \varepsilon_t$.

In this new regression model, the log of export maize is the dependent variable and the log of production of maize is the independent variable. From this model, the parameters are estimated and the results are shown in Appendix A, Table (A.4.10). This Table suggests that the regression coefficients are statistically significant at 1% level. Furthermore, the *F* value is highly significant at 1% level. Moreover, the value of D.W. is 1.28 < d_l = 1.373. The low D.W. statistic of 1.28 strongly suggests the presence of positive first-order serial correlation. In order to correct for the presence of positive first-order serial correlation, the Cochrane-Orcutt iterative procedure is used to estimate ρ (correlation coefficient) from the estimated residuals of the new model. Then, both variables (dependent and independent variables) of the new model are transformed by using $\hat{\rho}$. Next, another regression model which can be denoted as transformed model is developed and proceeded to the usual OLS estimation. The transformed model can be shown as follows:

$$EXP_t^* = \beta_0^* + \beta_1^* PROD_t^* + u_t \tag{4.19}$$

where

$$\beta_0^* = \beta_0(1-\hat{\rho})$$

$$EXP_t^* = (\ln EXP_t - \hat{\rho}\ \ln EXP_{t-1})$$

$$PROD_t^* = (\ln PROD_t - \hat{\rho}\ \ln PROD_{t-1})$$

$$u_t = (\varepsilon_t - \hat{\rho}\varepsilon_{t-1}).$$

The results of transformed model are also shown in Appendix A, Table (A.4.10). From this Table, it is observed that the intercept is significant at 10% level and the slope is statistically significant at 1% level. Moreover, the value of *F* of the transformed model decreases to 13.604. According to the part (c) of Appendix Figure B.4.8, the residuals of transformed model do not satisfy the normality assumption. According to Appendix Figure B.4.8(d), the pattern in it is not so clear that it inclines to suspect heteroscedasticity. Based on this result, it can be concluded that the transformed model does not fit the data.

Appendix B, Figure B.4.9(a) depicts a situation in which there is a significant simple linear relationship between production and export of maize. However, it would be possible that a curvilinear model between the two variables might be more appropriate. Thus, it can be deduced from Appendix Figure B.4.9(a) that the curvilinear model may be a better fit and should be evaluated in place of the simple linear model. The curvilinear relationship between the production of maize and the export of maize can be expressed as follows:

$$EXP_t = \beta_0 + \beta_1 PROD_t + \beta_{11} PROD_t^2 + \varepsilon_t \qquad (4.20)$$

where,

β_0 = intercept

β_1 = linear effect on Export

β_{11} = curvilinear effect on Export

ε_t = error.

An alternative approach to the curvilinear regression model in Equation (4.20) is to center the data by subtracting the mean of the explanatory variable from each value in the model. This centered regression model can be described as

$$EXP_t = \beta_0 + \beta_1 (PROD_t - mean) + \beta_{11} (PROD_t - mean)^2 + \varepsilon_t. \qquad (4.21)$$

The values of the three regression coefficients (β_0, β_1, and β_{11}) in model (4.21) are estimated and the results are shown in Appendix A, Table (A.4.11). In this table, the regression coefficient (β_{11}) is statistically significant at 5% level. Based on this result, it can be noted that the curvilinear model is significantly better than the linear model in representing the relationship between export and production of maize. Moreover, the regression coefficient (β_1) is statistically significant at 1% level. This result implies that the curvilinear model that includes the linear effect is significantly better than the model that includes only the curvilinear effect. The value of R^2 is 0.644 which can be interpreted to mean that 64.4% of the variation in export can be explained by the curvilinear relationship between export and production. Based on the result of *F* value, it can be remarked that there is a significant curvilinear relationship between production and export of maize.

After estimating of regression, the aptness of the fitted regression model is evaluated by using the residual analysis. From Appendix Figure B.4.9(b), the residuals of the fitted curvilinear model do not satisfy the normality assumption. Appendix Figure B.4.9(c) shows more variation on the high side than on the low side. It can be said that the variance of the residuals increase as the predicted values increases. Thus, the fitted model does not satisfy the basic assumptions. These results suggest that the data can contain some outlying observations.

Both variables have to be transformed to meet those assumptions. However, the transformation does not eliminate or attenuate the leverage of influential outliers that bias the prediction and distort the significance of parameter estimates. Even though the curvilinear effect is included in the model, the reliable results are not obtained from the OLS fit. Therefore, alternative methods that can detect and resist outliers are needed so that reliable results can be obtained in the presence of outliers as well.

Before calculating the robust estimates, the unusual observations are detected by a robust diagnostic plot. The regression outliers are detected with standardized LTS residuals, and leverage points are diagnosed by robust distances RD (x_i) based on the MCD. Figure 4.6(c) shows the diagnostic display for the fitted LTS regression. It automatically makes the observations classify into four types.

The part (c) of Figure 4.6 shows that two observations (23 and 25) have residuals with robust distance above the horizontal value and thus are detected as

vertical / regression outliers. The observations 27 and 29 exceed the cutoff line $\sqrt{\chi^2_{2,0.975}}$ = 2.72, and detected as bad leverage points. Although observations 21, 24, 26, 28, 30, 31, and 32 exceed the cutoff point, they are detected as good leverage points because they fall within the two horizontal lines. As a result of the robust diagnostics plot, the production and export of maize data set contains two vertical outliers, two bad leverage points and seven good leverage points. Therefore, the different robust methods such as LAV, LMS, LTS, *MH*, *MT*, and *MM* are applied to the production and export of maize data and results are shown in Table (4.4).

Table (4.4) gives the estimates from several robust regressions fitted to the same data. Comparisons are made the OLS to the robust methods. The OLS method is used to build a model that can be predicted export of maize based on production of maize. The diagnostics plots and preliminary analyses suggest that the performance of the model is not quite satisfactory. From Table (4.4), it is clear that the fit from OLS gives the intercept term quite different to that of robust fits. Thus, there is a serious risk that the OLS estimator is strongly attracted by these vertical outliers and bad leverage points. The good leverage points are beneficial and they can improve the precision of the regression coefficient. The impact of vertical outliers on the estimation of regression coefficients is usually small and it mainly affects the regression intercept. Bad leverage points can be harmful to the results of OLS method because they can change the OLS fit drastically. Therefore, the results of the OLS clearly show that the intercept of this method is larger than the other robust methods and the slope coefficients are markedly different. It is observed that the OLS fit is highly influenced by outliers as it has the largest MSE of 1823.8878, thus this fit represents neither good nor bad data points very well.

The standardized residuals plot shown in Figure 4.6(a) shows that there are no outliers at all because the horizontal band of the occurrence between the standardized residuals of -2.5 and 2.5. It can also be concluded that the data set contains no outliers at all because all the standardized residuals fall nicely within the band. However, it is observed that robust fits suggest a different idea about the nature of the data. The LTS method shows a severe robustness by detecting four observations as outlying observations and deletes these four observations (two vertical outliers and two bad leverage points). Moreover, this method also gives the second smallest MSE among all other robust methods.

Due to the two bad leverage points, many differences between estimates are found out to a greater degree. But, the LAV, *MM*, LMS and LTS estimators reveal the same story regarding the effect of curvilinear. Moreover, intercepts and slope coefficients of linear effect of these robust estimates are markedly different. The MSE of *M*-estimates (using Huber and Turkey weights) are quite different. These estimators are not robust owing to bad leverage points. The LMS estimates detect the observations 23, 25 27, 29 and 30 as outliers and discard them. After that, the MSE of 557.9275 is obtained. The LMS and LTS estimators are resistant to the vertical outliers as well as the bad leverage points. According to the results, it can also be seen that the LMS estimator is the most suitable one which has the smallest MSE. The fitted LMS regression is given below:

$$\hat{EXP}_t = 66.7144 + 0.2314\,(\hat{PROD}_t - mean) - 0.0001(\hat{PROD}_t - mean)^2.$$

Table (4.4) OLS and Robust Regression Models Fitted to the Production and Export of Maize Data

Estimation Methods	β_0	β_1	β_{11}	MSE
OLS	88.6880*** (8.43)	0.3008*** (5.95)	-0.0003** (2.64)	1823.8878
LAV	76.2001*** (6.47)	0.2252*** (4.08)	-0.0001 (0.88)	960.4243
MH	75.7844*** (8.78)	0.2527*** (6.09)	-0.0002** (2.05)	657.9225
MT	66.3655*** (7.83)	0.2917*** (7.16)	0.0003*** (3.40)	852.0561
MM	59.9677*** (3.08)	0.1859** (2.06)	-0.0001 (0.40)	836.3664
LMS	66.7144***	0.2314***	-0.0001**	557.9275
LTS	60.5182***	0.1901***	-0.0001**	617.0256

Note: (1) MH *M*-estimate with $\psi(.)$ of the Huber type;

(2) MT *M*-estimate with $\psi(.)$ of the Turkey type.

(3) Absolute values of *t* statistics in parentheses

(4) Significant at *** 1%, **5%, * 10%

Source: Appendix Table (A. 4.9)

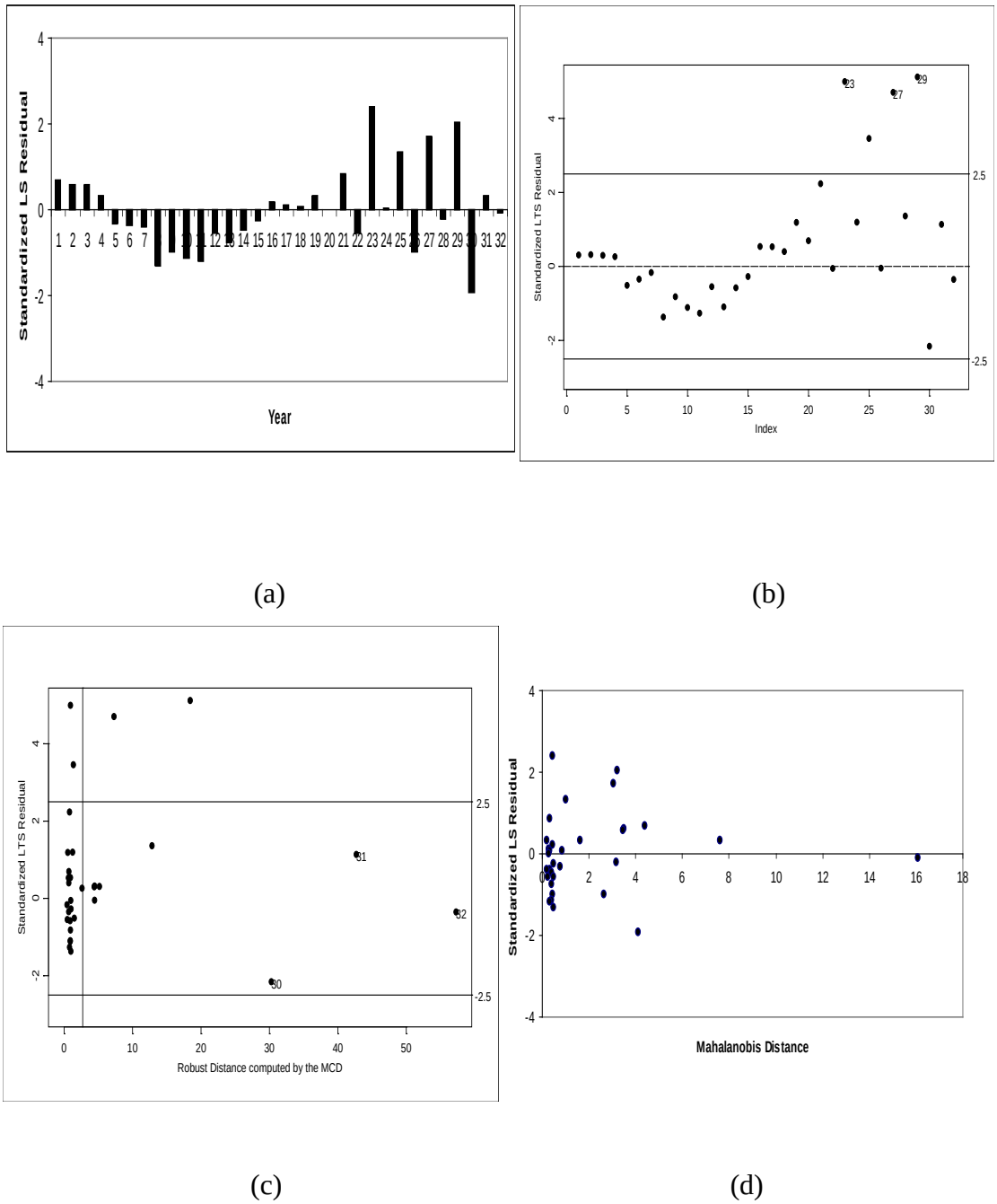


Figure 4.6 **(a) standardized residuals plot from OLS; (b) standardized residuals plot from LTS; (c) robust diagnostic plot; and (d) analogous plot based on classical estimates**

Source: Appendix Table (A. 4.9)

4.6.3 Multiple Linear Regression Using Real Data Study

The maternal mortality ratio (per 100,000 live births) was studied as a function of the contraceptive prevalence rate (% of married women aged 15-49), adult literacy rate (% aged 15 and older), births attended by skilled health personnel (%), total fertility rate (birth per women), life expectancy at birth (years), physicians (per 100,000 people), and public expenditure on health (% of GDP). The data contain 34 countries which were obtained from the Human Development Report (2007/2008) published by the UNDP. Concerning these, a model was constructed that explains the maternal mortality ratio for sample of 34 countries based on these seven explanatory variables. So, the distribution of the individual variable was examined first. It was found that the dependent variable was not linear. To make linearity, this dependent variable was transformed by taking the logarithm. Next, whether or not these explanatory variables shown above were individually related to the dependent variable were checked. Then, the correlation coefficient for each of the explanatory variables with the dependent variable was computed. It was found that these explanatory variables were highly correlated with the dependent and thus they were used in model building process. Thus, a model to predict the maternal mortality ratio of countries was built based upon the contraceptive prevalence rate, adult literacy rate, births attended by skilled health personnel, total fertility rate, life expectancy, physicians and public expenditure on health. In this study, it can be expected that the maternal mortality ratio might be inversely related to the contraceptive prevalence rate, adult literacy rate, births attended by skilled health personnel, life expectancy, physicians and public expenditure on health, but it can directly be related to the total fertility rate. The multiple linear regression model is seen as follows:

$$\begin{aligned}\ln MMR_i = {} & \beta_0 + \beta_1 CPR_i + \beta_2 ALR_i + \beta_3 BABSHP_i + \\ & \beta_4 TFR_i + \beta_5 LE_i + \beta_6 PHY_i + \beta_7 PEOH_i + \varepsilon_i\end{aligned} \tag{4.22}$$

where,

MMR	= maternal mortality ratio
CPR	= contraceptive prevalence rate
ALR	= adult literacy rate
BABSHP	= births attended by skilled health personnel
TFR	= total fertility rate

LE = life expectancy

PHY = physicians,

PEOH = public expenditure on health.

The OLS method is applied using this data and the parameters $\beta_0,...,\beta_7$ from the model (4.22) are estimated and the results are shown in Appendix A, Table (A.4.13). From this table, it is obtained that most of the variance inflationary factor (VIF) values for each explanatory variable are relatively small, ranging from a high of 6.219 for BABSHP to a low of 2.672 for CPR. Due to these results, neither of the slope coefficients (CPR, ALR, TFR and PHY) is individually statistically significant. Moreover, the CPR and ALR variables are not only statistically insignificant but also wrong sign. In addition, it is found that the *F* value is highly significant at 1% level which is shown in Appendix A, Table (A.4.16). Hence, it can be remarked that there is multicollinearity among the set of explanatory variables.

Thus, a widely used procedure, "search" called stepwise regression was applied for exploring the "best" regression model. Appendix A, Table (A.4.14) presents the results obtained from the SPSS STEPWISE procedure. According to these results, the four variables (CPR, ALR, TFR, and PHY) were dropped out from the model (4.22) and the new model under consideration thus contains three explanatory variables: LE, PEOH and BABSHP. The new model can be seen as follows:

$$\ln MMR_i = \beta_0 + \beta_1 LE_i + \beta_2 PEOH_i + \beta_3 BABSHP_i + \varepsilon_i. \qquad (4.23)$$

In Appendix Table (A.4.14) the stepwise method shows a statistically significant negative effect of each of the explanatory variables (LE, PEOH and BABSHP) and suggests that when the CPR, ALR, TFR, and PHY variables are removed, a slight change in the values of coefficients of the remaining variables is found out. As expected, the coefficients of LE, PEOH and BABSHP are negative. The intercept value slightly increases from 5.802 to 5.863. In addition, the slope coefficients of the LE, PEOH and BABSHP change from -0.046, -0.084 and -0.010 to -0.044, -0.110 and -0.008 respectively. Moreover, the significance level of these coefficients changes from 5% to 1%. The value of *F* increases from 36.196 to 76.645 which is also shown in Appendix A, Table (A.4.15).

After that, the explanatory variables to be included in the model have been selected; now a residual analysis is used to evaluate the aptness of the fitted model. Thus, the diagnostic plots which are shown in the part (a) and (b) of Appendix B, Figure B.4.10 are used for the study. The part (a) of Appendix Figure B.4.10 suggests that the residuals of the fitted model (4.23) do not follow the normality assumption because some of the points do not fall in a straight line. It can be observed from the part (b) of Appendix Figure B.4.10 that there is no apparent pattern between the standardized residual and predicted value. The residuals appeared to be evenly spread above and below the mean value for the predicted value. Based on this result, it can be remarked that the fitted model appeared to be appropriate. According to the part (a) of Appendix Figure B.4.10, the new model violates the normality assumption. This result implies that, the data contain some outliers. Thus, the alternative procedures are used to achieve the robustness properties.

Before application of robust methods, the types of unusual observations are investigated using a robust diagnostic plot. This plot is shown in Figure 4.7(b). It reveals that observation 13 is vertical outlier and observations 21, 23, 24, 25, and 34 are good leverage points. Therefore, it can be concluded that according to the robust diagnostic plot, data contain one vertical outlier and five good leverage points. The four different robust estimates: *M*, *MM*, LMS and LTS are applied to the same data set and the results of both OLS and robust methods are shown in Table (4.5).

Table (4.5) shows the estimates from several robust regression fitted to the maternal mortality data. Although there are small differences between them, the *MH*, *MT* and *MM*-estimators tell a similar story regarding the effects of LE, PEOH and BABSHP. Their intercepts of *M*-estimates and *MM*-estimate are not too much different. The slope coefficients of the LE, PEOH and BABSHP are nearly identical regardless of which regression method is employed. Due to the unusual observations, it is observed that the MSE of OLS is the largest value among the others but magnitudes are not very different. Besides, the LTS regression detects the country Kuwait as outlier and discards them, gives the MSE of 0.2453. Moreover, the LMS gives a different idea about the nature of the data. It detects Kuwait, El Salvador and Indonesia as outliers and discards them. Consequently, it gives the smallest MSE of 0.2399. In this study, it is found that the best result is obtained using the LMS regression. The fitted LMS regression model is given below:

$$\ln \hat{MMR} = 6.0566 - 0.0092\,\hat{LE} - 0.0485\,\hat{PEOH} - 0.0455\,\hat{BABSHP}.$$

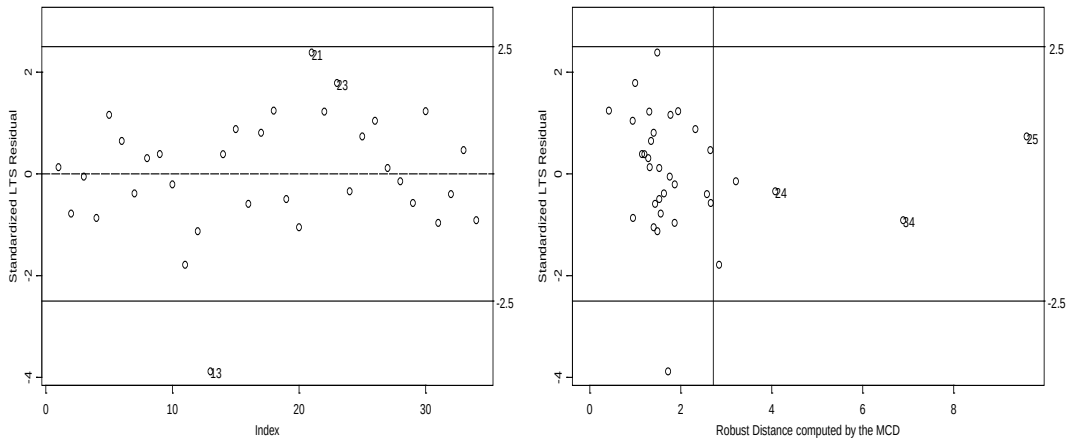
Table (4.5) OLS and Robust Regression Models Fitted to the Maternal Mortality Data

Estimation Methods	β_0	β_1	β_2	β_3	MSE
OLS	5.8628*** (11.4370)	-0.0080*** (3.1180)	-0.0436*** (4.9270)	-0.1102*** (4.2630)	0.2904
MH	5.7020*** (12.3681)	-0.0078*** (3.3627)	-0.0411*** (5.1639)	-0.1189*** (5.1164)	0.2632
MT	5.6681*** (11.7596)	-0.0075*** (3.1178)	-0.0406*** (4.8787)	-0.1229*** (5.0571)	0.2713
MM	5.6092*** (8.7760)	-0.0072*** (2.2810)	-0.0396*** (3.6030)	-0.1291*** (3.9200)	0.2869
LMS	6.0566***	-0.0092***	-0.0485***	-0.0455***	0.2399
LTS	5.4550***	-0.0069***	-0.0379***	-0.1366***	0.2453

Note: (1) Absolute values of *t* statistics in parentheses

(2) Significant at *** 1%, **5%, * 10%

Source: Appendix Table(A.4.12)



(a) (b)

Figure 4.7 Maternal Mortality Data Set Using the LTS Robust Procedure: (a) plot of the standardized residuals; and (b) diagnostic plot

Source: Appendix Table(A.4.12)

CHAPTER V

ROBUSTNESS IN TIME SERIES MODELS

In the previous chapter, about the types of outliers and their effects on regression have already been discussed. The underlying ideas of regression and time series analysis is almost the same. In regression analysis, the explanatory variable is related to the dependent variable whereas in autoregressive modeling for time series analysis, it was found that the explanatory variable is its lag value of the dependent variable.

The concept of robustness in statistics is defined as the lack of sensitivity of a particular inferential procedure to departures from the model assumptions. A theory of robustness (Huber, 1981, Hampel, Ronchetti, Rousseeuw, and Stahel, 1986) has been developed in order to be used for departures arising from contaminated observations called outliers which are present in a "clean" data set generated from a known model. In time series analysis, the essential concept is resistance, which is related to the notion of robustness from a data-oriented point of view. This concept has the important advantage that it can be applied without special assumptions about the model and, on the other hand, observations can be dependent or non-identically distributed (Martin and Yohai, 1985).

The basic theory of robustness has been developed in the i.i.d. context, but the study of outliers in time series is a more complicated task mainly due to the structure of the adjacent correlated observations. In addition, in time series, the influences of outliers can be shown depending on their type, relative position, number and magnitude and – as mentioned before – on the model structure underlying the time series.

Typically maximum likelihood and least squares estimators have poor robustness properties and the performance of these approaches is sensitive to influential cases and the departure of the error distribution. Moreover, these procedures are affected adversely by the presence of outliers. To overcome these shortcomings, several alternative robust estimators like *M*-estimators (Huber, 1981) and L_1-based estimators have been proposed. These methods were developed to be applied in time series models recently.

Robust estimation in the time series context is a difficult task because different types of outliers may occur in any data set. For instance, outliers can replace with or be added to some observations of the stochastic process. They can also be found in the innovation driving process. Furthermore, the configuration of time points in which the contaminations like isolated and patchy outliers occur gives different effects on estimation of parameters of time series models. Outliers should be investigated carefully. Often they contain valuable information about the process under investigation or the data gathering and recording process. Before considering the possible elimination of these points from the data, one should try to understand why they appeared.

Some of the robust estimators for time series parameters proposed in the literature are *M*-estimators, *GM*-estimators (Denby and Martin, 1979), RA estimators (Bustos and Yohai, 1986) and ACM estimator (Martin, 1979). These estimators were applied to simulated as well as real data to estimate the parameters of ARMA and ARIMA models and their performances were compared with the ML estimator in the presence of AO or IO. In this chapter, these robust methods were used to detect outlying observations and to provide resistant results which are stable in the presence of outliers. For instance, if we are interested in estimating a model parameter for a data contaminated with outliers from a random measurement error, it is of interest to use an estimator which is not sensitive to such outlying observations.

This chapter is organized as follows. Time series models such as AR, MA, ARMA and ARIMA models are mentioned in section 5.1, followed by maximum likelihood estimates for ARMA Model in section 5.2. The outliers in time series are discussed in section 5.3. Section 5.4 is mainly concerned with some robust estimates (*M*, *GM*, ACM and RA) that are used for this study. In this section, the computation algorithms of these estimators are described. In section 5.5, these robust estimates are applied to simulated data as well as real data to illustrate how the robust methods outperform the ML method when data contain the AO and IO outliers.

5.1 Time Series Models

Based on a finite number of available observations, a finite order parametric model was constructed to describe a time series model. In this section, autoregressive model (AR), moving average model (MA), autoregressive moving average model (ARMA) and autoregressive integrated moving average model (ARIMA) were described.

5.1.1 Autoregressive Model

Autoregressive models are widely used by many statisticians and they are really helping tools in the estimation of parameters, especially, in carrying out time series analysis using regression methods. An autoregressive model of order *p*, AR(*p*), can be defined as follows:

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \phi_2 \tilde{Z}_{t-2} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t , \tag{5.1}$$

where $\tilde{Z}_t = Z_t - \mu$, $\phi_1, \phi_2, ..., \phi_p$ are constants and a_t are independent and identically distributed $N(0, \sigma_a^2)$ random variables with $\sigma_a^2 < \infty$.The process defined by (5.1) is called an autoregressive process of order *p*, or more concisely, an AR(*p*) process. This Equation (5.1) can be written in the equivalent form as follows:

$$(1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)\tilde{Z}_t = a_t$$

or

$$\phi(B)\tilde{Z}_t = a_t , \tag{5.2}$$

where $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p)$. ***B*** is a backshift operator (that is, $B(\tilde{Z}_t) = \tilde{Z}_{t-1}$). Since $\sum_{j=1}^{p} |\phi_j| < \infty$, the process is always invertible. To be stationary, the roots of $\phi(B) = 0$ must lie outside of the unit circle. The AR processes are useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock.

5.1.2 Moving Average Model

The moving average model of order q can be written as:

$$\tilde{Z}_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}. \tag{5.3}$$

Equation (5.3) can be written in the equivalent form

$$\tilde{Z}_t = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q) a_t$$

or

$$\tilde{Z}_t = \theta(B) a_t, \tag{5.4}$$

where $\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q)$.

A finite moving average model is always stationary because $1 + \theta_1^2 + \cdots + \theta_q^2 < \infty$. This moving average model is invertible if the roots of $\theta(B) = 0$ lie outside of the unit circle. Moving average models are useful in describing phenomena in which events produce an immediate effect which only lasts for a short period of time.

5.1.3 Autoregressive Moving Average Model

A stationary autoregressive moving average model is obtained by combining the equations for an AR model given by (5.1) and an MA model given by (5.3). An ARMA(p, q) model can be written in the form

$$\tilde{Z}_t = \phi_1 \tilde{Z}_{t-1} + \cdots + \phi_p \tilde{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q} \tag{5.5}$$

where a_t is a white noise process; that is, the a_t's are uncorrelated, and have zero mean and variance σ^2. The process a_t is sometimes called the innovations process. The parameters $\phi_1, ..., \phi_p$ are the autoregressive coefficients, and the parameters $\theta_1, ..., \theta_q$ are the moving average coefficients. If the innovations a_t are Gaussian (the process $\tilde{Z}_t$ is Gaussian) and are uncorrelated, then they are also independent. This is a frequently used assumption. The ARMA model of (5.5) is often written in the form

$$\phi(B)\ \tilde{Z}_t = \theta(B)\, a_t, \tag{5.6}$$

where

$$\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$$

and

$$\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q .$$

For the process to be invertible, it requires that the roots of $\theta_q(B) = 0$ lie outside the unit circle. To be stationary, it is required that the roots of $\phi_p(B) = 0$ lie outside the unit circle. Assuming that $\phi_p(B) = 0$ and $\theta_q(B)$ = 0 share no common roots this process refers to an ARMA(*p*, *q*) process or model, in which *p* and *q* are used to indicate the orders of the associated autoregressive and moving average polynomials, respectively. Equation (5.6) can be written as follows:

$$\tilde{Z}_t = \phi^{-1}(B)\theta(B)a_t \tag{5.7}$$

$$\tilde{Z}_t = \frac{\theta(B)}{\phi(B)}a_t = \frac{1 - \theta_1 B - \cdots - \theta_q B^q}{1 - \phi_1 B - \cdots - \phi_p B^p}a_t$$

the mixed autoregressive moving average model can be thought of as the output $\tilde{Z}_t$ from a linear filter, whose transfer function is the ratio of two polynomials $\theta(B)$ and $\phi(B)$, when the input is white noise a_t.

5.1.4 Autoregressive Integrated Moving Average Model

In the previous subsection, the stationary processes have already been discussed. However, many applied time series, particularly those arising from economic and business areas are nonstationary. Regarding covariance stationary processes, nonstationary time series can occur in many different ways. They could have nonconstant means μ_t, time varying second moments such as nonconstant variance σ_t^2, or have both of these properties. In this subsection, the construction of a very useful class of homogeneous nonstationary time series model the autoregressive integrated moving average (ARIMA) model was introduced. The nanstationary autoregressive integrated moving average model can be written as follows:

$$\phi_p(B)(1 - B)^d Z_t = \theta_0 + \theta_q(B)a_t \tag{5.8}$$

where the stationary AR operator $\phi_p(B) = (1 - \phi_1 B - \cdots - \phi_p B^p)$ and the invertible MA operator $\theta_q(B) = (1 - \theta_1 B - \cdots - \theta_q B^q)$ share no common factors. The parameter θ_0 plays very different roles for $d = 0$ and $d > 0$. When $d = 0$, the original process is

stationary and θ_0 is related to the mean of the process, that is, $\theta_0 = \mu(1 - \phi_1 - \cdots - \phi_p)$. However, when $d \geq 1$, θ_0 is called the deterministic trend term and, it is often left out from the model if it is unnecessary to be considered.

The resulting homogeneous nonstationary model in (5.8) refers to the autoregressive integrated moving average model of order (*p, d, q*) and is denoted as the ARIMA(*p, d, q*) model. When *p* = 0, the ARIMA(*p, d, q*) model is also called the integrated moving average model of order (*d, q*).

5.2 Outliers in Time Series

Outliers are aberrant observations that are away from the rest of the data. They can be caused by recurrent events such as recording errors or non-recurrent events such as changes in economic policies, wars, disasters and so on. They tend to occur if errors have fat-tailed distributions which might lead to large disturbances. Sometimes, outliers appear through misspecification of estimated relationships (linear instead of nonlinear relationships, omitted variables and so on).

There is no issue that outliers can cause problems with inference using the traditional methods. The only problem is that how outliers should be tackled, that is, whether they are excluded or included. According to Legendre (1805), he suggested to throw these observations out. So did Edgeworth (1887). However, if outliers are caused by misspecification of the relationships estimated, a proper course is to change the specification. If outliers are caused by fat-tailed error distributions, a proper course is to use robust methods (Maddala and Yong Yin, 1997). Thus, there are three courses of action one can take:

(i) throwing the rascals out.

(ii) leaving them in but under control (robust methods) or

(iii) changing the model.

For (i), it is required to identify the outliers. For the linear regression model, there is a plethora of diagnostics which were discussed in previous chapter. However, many of them are not applicable for time series data.

In time series analysis, outliers can cause biases in parameter estimation as well as misspecification, resulting in misleading conclusion. For this reason, several outlier detection and robust estimation procedures have been proposed in the literature

for time series analysis. The methods of outliers' detection are excluded from this study.

In time series problem, because successive observations are correlated, outliers can cause more problems for detection. Fox (1972) first addressed outlier problems in time series by classifying outliers as additive outliers (AO) and innovation outliers (IO). These two types of outliers and other robustness problems in time series were discussed extensively in the time series literature (Denby and Martin, 1979, and Hampel et al., 1986). Another types of disturbances were introduced by Chen and Tiao (1990). They are the level shift (LS) and temporary change in level (TC). In the following subsection, these types of outliers in time series analysis were described and their plots of AO, IO, LS and TC were shown in Figure 5.1.

5.2.1 Additive Outlier

Additive outlier (AO) represents a disturbance which is committed to a particular observation. Mathematically, the observed time series is seen as

$$Y_t = Z_t + w_a I_t^{(d)} \tag{5.9}$$

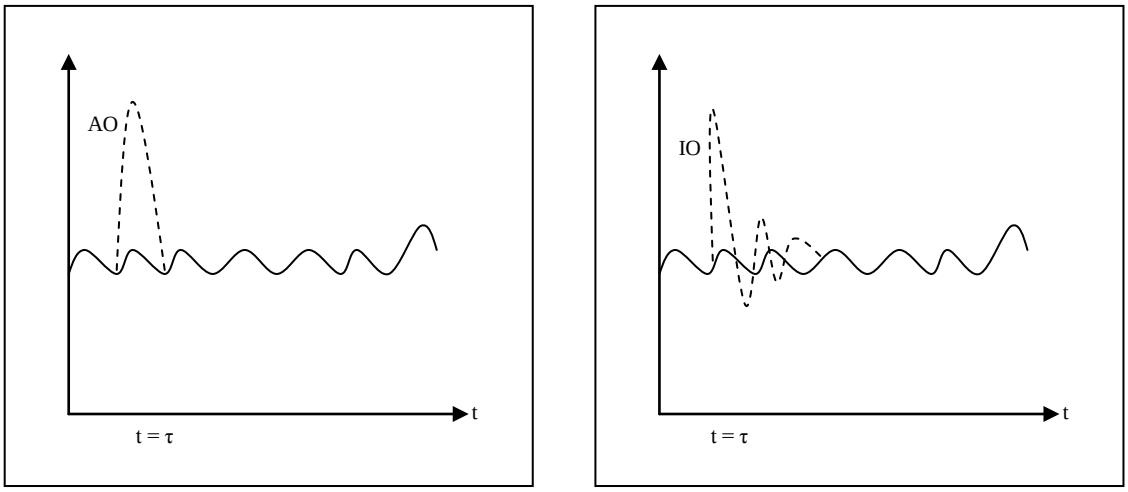
where Y_t is a contaminated time series, Z_t is an outlier-free time series, w_a denotes the magnitude of the disturbance and $I_t^{(d)}$ is an indicator variable defined by

$$I_t^{(d)} = \begin{cases} 1 & if \quad t = d \\ 0 & if \quad t \neq d. \end{cases}$$

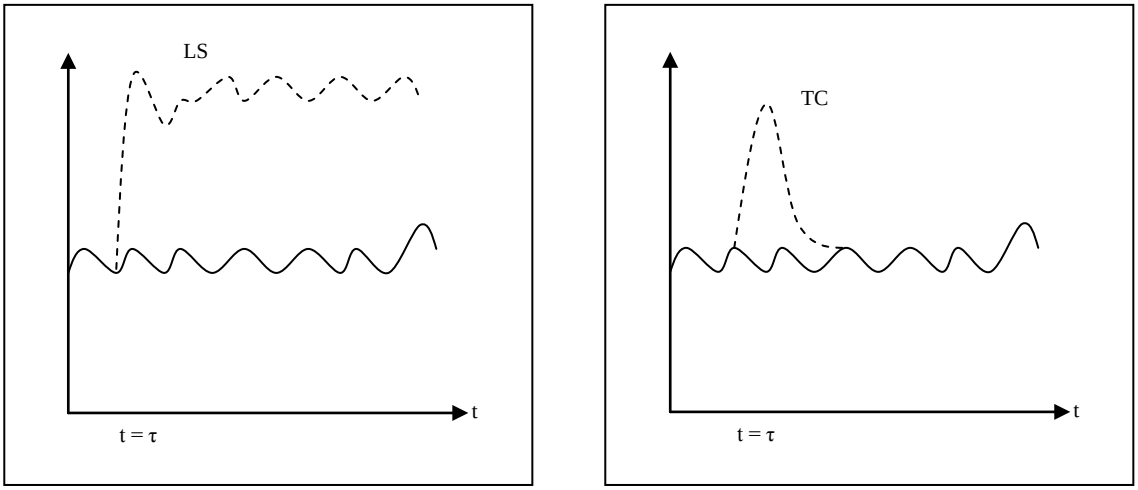
In other words, for an AO model

$$Y_t = Z_t \text{ if } t \neq d \text{ and } Y_d = Z_d + w_a \text{ otherwise.}$$

The typical reason for an AO is a recording or measurement error. Outbreaks of wars, strikes, an abrupt change in the market structure of some group of commodities, a technical change or new equipment in a communication system, or simply unexpected geophysical phenomena (e.g., earthquakes) are all possible causes of AOs.



(a) (b)



(c) (d)

Figure 5.1 Types of Outliers (a) the plot of additive outlier; (b) the plot of innovation outlier; (c) the plot of level shift; and (d) the plot of temporary change

5.2.2 Innovation Outlier

Another type of outlier is called an innovation outlier (IO), which is a disturbance in the innovational series $\{a_t\}$ and may affect every subsequent observation of the series. Mathematically, an IO model is

$$Y_t = \frac{\theta(B)}{\phi(B)}\left(a_t + w_v\, I_t^{(d)}\right) \tag{5.10}$$

where $I_t^{(d)}$ is defined as shown above and w_v denotes the magnitude of the disturbance. Rewriting the model as

$$Y_t = Z_t + \frac{\theta(B)}{\phi(B)} w_v I_t^{(d)}.$$

It can be seen that an IO effects the series through its own dynamic $\frac{\theta(B)}{\phi(B)}$ and it becomes part of the system thereafter. In practice, an IO often indicates an onset of certain chances in the system.

Of course, many other types of disturbance can occur in time series. The AO and IO models only two many possibilities.

5.2.3 Level Shift

Mathematically, a level shift (LS) can be described by

$$Y_t = Z_t + \frac{w_s}{(1-B)} I_t^{(d)} \tag{5.11}$$

where w_s is the amount of shift in the level of Z_t. It can be written as

$$\frac{1}{(1-B)} = 1 + B + B^2 + \cdots$$

The above model (5.11) can be seen as follows:

$$Y_t = \begin{cases} Z_t & for \quad t < d \\ Z_t + w_s & for \quad t \geq d. \end{cases}$$

Thus, the fixed constant w_s is added to every observation one or after d. Such a level shift is permanent.

5.2.4 Temporary Change

In some cases, the effect of a level shift is only temporary. In a mathematical model, such a shift is described as:

$$Y_t = Z_t + \frac{w_c}{(1-\delta B)} I_t^{(d)}, \; 0 < \delta < 1. \tag{5.12}$$

Since

$$\frac{1}{(1-\delta B)}=1+\delta B+\delta^2 B^2+\cdots$$

the magnitudes of level shift at times $d, d+1, d+2, \ldots$ are $w_c, \delta w_c, \delta^2 w_c, \ldots$. Thus, the initial shift is w_c and the subsequent shifts are discounted at the rate δ. With $0 < \delta < 1$, the shift decays exponentially to zero. Such a temporary level shift refers to a transient change model. In practice, the values of δ are a predetermined constant. Its value may be 0.8 or 0.7.

The difference between AO and IO is that in fact an AO is interpreted as an outlying observation added after the realization to affect a single observation and an IO as an outlying observation added during the realization with influence on all succeeding observations.

In autoregressive models, AOs are a cause of much greater concern than IOs because leverage points (outliers in the *x*-direction) create bigger problems than outliers in the *y*-direction. For example, in the case of an AR(1) model, one IO yields one outlier in the response variable and a number of "good" leverage points ("good" refers to the fact that the leverage points lie close to the fitted line determined by the majority of the data), which actually improve the accuracy of the parameter estimate. Therefore, one IO only affects one residual.

On the other hand, one AO results in one outlier in the vertical direction and one "bad" leverage point ("bad" refers to the fact that the leverage point does not lie close to the fitted line determined by the majority of the data). Thus, AO also affects the next residual inflating two consecutive residuals.

Although OLS has been used extensively in statistics, particularly, in regression analysis, it has shortcomings. One of the weakest points of the method is its high sensitivity to outliers: one sufficiently large outlier can ruin the estimate. One explanation for this high sensitivity of OLS to outliers is that squaring the residuals magnifies the effects of these extreme data points.

In the time series setting, OLS estimation methods also exhibit a lack of robustness to outliers. The OLS estimator of the autoregressive parameter ϕ in the

AR(1) with innovation outlier is consistent, but inefficient. For the AR(1) with AO, the OLS estimator is not even consistent (Denby and Martin, 1979).

The important point with IO is that the ARMA(*p*, *q*) model is still the exact model for the observations. However, if an outlier occurs at t_0, then a_{t_0} will affect not only Z_{t_0}, but many future observations. Meanwhile, the effect disappears. Bustos and Yohai (1986) give several results showing that IO does not affect too seriously the OLS estimators of autoregressive and moving average parameters of an ARMA model.

To improve this sensitivity to outliers, statisticians began to develop robust estimation methods starting around 1960 (Hampel, 1971). In the regression setting, robust regression methods are definitely designed to minimize the effect of outliers while retaining much of the sensitivity and precision of OLS in the absence of outliers.

5.3 Maximum Likelihood Estimates for ARMA Model

There are some parameters estimation methods for time series model. They are ordinary least squares (OLS), maximum likelihood (ML) and moment methods. Among them, the maximum likelihood method was chosen for this study. The maximization of likelihood function can be achieved by using three methods: leading to exact maximum likelihood, conditional maximum likelihood and unconditional maximum likelihood functions. The most commonly used method is a conditional maximum likelihood method which can be seen as below.

Conditional Maximum likelihood Estimation

This method was developed by Box and Jenkin (1970). For the general stationary ARMA(*p*, *q*) model which was described in (5.5), the joint probability density of $a = (a_1, a_2, ..., a_n)'$ is given by

$$P\left(a\middle|\phi,\mu,\theta,\sigma_a^2\right)=\left(2\pi\sigma_a^2\right)^{-n/2}\exp\left(-\frac{1}{2\sigma_a^2}\sum_{t=1}^{n}a_t^2\right). \tag{5.13}$$

Rewrite the Equation (5.5) as

$$a_t=\theta_1 a_{t-1}+\cdots+\theta_q a_{t-q}+\tilde{Z}_t-\phi_1\tilde{Z}_{t-1}-\cdots-\phi_p\tilde{Z}_{t-p}. \tag{5.14}$$

The likelihood function of parameters $(\phi, \mu, \theta, \sigma_a^2)$ is seen as follows. Let $Z = (Z_1, Z_2, ..., Z_n)'$ and assume the initial conditions $Z_* = (Z_{1-p}, ..., Z_{-1}, Z_0)'$ and $a_* = (a_{1-q}, ..., a_{-1}, a_0)'$. The conditional log-likelihood function

$$\ln L_*(\phi, \mu, \theta, \sigma_a^2) = -\frac{n}{2} \ln 2\pi\sigma_a^2 - \frac{S_*(\phi, \mu, \theta)}{2\sigma_a^2} \tag{5.15}$$

where

$$S_*(\phi, \mu, \theta) = \sum_{t=1}^{n} a_t^2(\phi, \mu, \theta | Z_*, a_*, Z) \tag{5.16}$$

is the conditional sum of squares function. The quantities $\hat{\phi}, \hat{\mu}, \text{and } \hat{\theta}$, which maximize Equation (5.15) are called the conditional maximum likelihood estimators. Since $\ln L_*(\phi, \mu, \theta, \sigma_a^2)$ involves the data only through $S_*(\phi, \mu, \theta)$, these estimators are the same as the conditional least squares estimators obtained from minimizing the conditional sum of squares function $S_*(\phi, \mu, \theta)$, which does not contain the parameter σ_a^2.

There are a few alternatives for specifying the initial conditions Z_* and a_*. Based on the assumptions that $\{Z_t\}$ is stationary and $\{a_t\}$ is a series of i.i.d. $N(0, \sigma_a^2)$ random variables, the unknown Z_t is replaced by the sample mean $\bar{Z}$ and the unknown a_t by its expected value of 0. For the model in (5.5), it is assumed that $a_p = a_{p-1} = \cdots = a_{p+1-q} = 0$ and calculates a_t for $t \geq (p + 1)$ using the afore mentioned model (5.5) thus it becomes

$$S_*(\phi, \mu, \theta) = \sum_{t=p+1}^{n} a_t^2(\phi, \mu, \theta | Z), \tag{5.17}$$

which is mostly used in computer programs.

After obtaining the parameter estimates $\hat{\phi}, \hat{\mu}, \text{and } \hat{\theta}$, the estimate $\hat{\sigma}_a^2$ of σ_a^2 is calculated from

$$\hat{\sigma}_a^2 = \frac{S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})}{d.f.}, \tag{5.18}$$

where the number of degrees of freedom (d.f.) equals the number of terms used in the sum of $S_*(\hat{\phi}, \hat{\mu}, \hat{\theta})$ minus the number of parameters estimated. If (5.17) is used to calculate the sum of squares, d.f. = $(n - p) - (p + q + 1) = n - (2p + q + 1)$. For other models, the d.f. should be adjusted accordingly.

5.4 Robust Estimates for ARMA Model

In real data application, OLS and ML estimates are very sensitive to the various types of outliers. They are not efficient and inconsistent when data contain outlying observations. Therefore, robust estimation methods are necessary because these estimates are not much affected by outliers. In this section, the computation algorithms of four robust estimators were discussed. The *M*-estimator and *GM*-estimator were described in subsection 5.4.1 and 5.4.2. Subsection 5.4.3 and 5.4.4 mention ACM estimators as well as RA estimators.

5.4.1 Algorithm for Computing of *M*-Estimates

The *M*-estimators algorithm is an iterative procedure and it contains three steps. They are described as follows:

1. Obtain an initial estimate of ϕ, say $\hat{\phi}_O$, from OLS.

2. Based on $\hat{\phi}_O$ in step 1, the initial weights are computed

$$w\left(e_{t+1}^0\right) = \frac{\psi\left(e_{t+1}^0\right)}{e_{t+1}^0}$$

where $e_{t+1}^0 = Z_{t+1} - Z_t\hat{\phi}_0$, for $t = 1,...,T\text{-}1$.

3. For $j = 0$ to convergence do

$$\hat{\phi}_{j+1} = \left(\sum_{t=1}^{T-1} w\left(e_{t+1}^j\right) Z_t^2\right)^{-1} \left(\sum_{t=1}^{T-1} w\left(e_{t+1}^j\right) Z_t Z_{t+1}\right)$$

$$e_{t+1}^j = Z_{t+1} - Z_t\hat{\phi}_j .$$

Convergence can be defined in several ways: relative change in the estimates, relative change in the scaled residuals, relative change in weights, and preselected number of steps.

5.4.2 Algorithm for Computing of *GM*-Estimates

An iterative procedure is necessary to find *GM*-estimates for time series. They are solved by an IRLS method. The *GM*-estimation algorithm can be described as follows:

1. To obtain an initial estimate of ϕ, say $\hat{\phi}_O$, it is necessary to use OLS.
2. Given $\hat{\phi}_O$, compute the initial weights as

$$\sum_{t=p}^{n-1} W(Z_t) Z_t w_t^j . \left(Z_{t+1} - Z_t^T \hat{\phi}^{j+1}\right) = 0 \quad , j = 0,1,...,iter$$

   where ***iter*** is the desired number of iterations.
3. $\hat{s}_e^j$ at iteration j = 0,1,2,… is obtained.
4. Repeat steps 1 until 3, and keep convergence have been reached.

5.4.3 Algorithm for Computing of ACM Estimates

The ACM estimation algorithm consists of five steps and computes the *GM*-estimates as an initial estimate. A detailed description of the algorithm is as follows:

1. Compute initial estimates of ϕ and s_e using *GM*-estimates.
2. Compute the Ψ-function is given by (3.70).
3. Obtain the scale estimate s_t and m_t.
4. Use a robust filter cleaner to obtained the cleaned observations $\hat{Z}_t$.
5. Compute new estimates using the cleaned observations $\hat{Z}_t$.

5.4.4 Algorithm for Computing of RA Estimates

The iteration algorithm used for RA estimates for a stationary and invertible ARMA(p, q) is as follows. Suppose that in the i th iteration we have estimates $\hat{\phi}^{(i)}$ and $\hat{\sigma}^{(i)}$ respectively. The estimate for the (i + 1) th iteration is,

1. Compute the residuals $e_t\left(\hat{\phi}^{(i)}\right)$, for $p + 1 \leq i \leq T$.
2. Modify the residuals by applying the Ψ-function,

$$e_t^* = \hat{\sigma}^{(i)} \psi\left(\frac{e_t\left(\hat{\phi}^{(i)}\right)}{\hat{\sigma}^{(i)}}\right)$$

   where Ψ are given by (3.60) and (3.61).

3. Calculate a new "pseudo-observations" process Z_t^*, using $\hat{\phi}^{(i)}$ and the modified residuals e_t^*, by putting

$$Z_t^* = \hat{\mu} + \left(\hat{\phi}^{(i)}(B)\right)^{-1} \hat{\theta}^{(i)}(B) e_t^* .$$

4. Compute $\hat{\phi}^{(i+1)}$ as the OLS of ϕ for Z_t^*.

5. At this stage, compute the scale estimate $\hat{\sigma}^{(i+1)}$.

In this computation, the median of the absolute values of the residuals divided by 0.6745 was considered as a robust scale estimate of residuals in step (5).

5.5 Simulation and Real Data Study

The effect of the outlier in time series model was analyzed using the ML method and robust methods in this section. Thus, a simulation study was conducted to compare the ML method with the robust methods for AR(1) and MA(1) models with AO and IO. The statistical software package S-PLUS 2000 was available and it was used to obtain the desire robust estimates throughout the analyses. In the following subsections, the performances of the ML estimator and robust estimators by simulations of AR(1) and MA(1) models with AO and IO outliers were analyzed first. Then, the real data study was followed up with it.

5.5.1 Simulation Study

In order to study the effect of outliers in estimation of parameters in time series model, each of 30 outlier free series of AR(1) with $\phi = 0.50$, MA(1) with $\theta = 0.50$ for n = 50, μ = 0 were generated using S-PLUS software. In this simulation study, let Y_t is an observed series and Z_t is an outlier free series. We consider the observed series $\{Y_t\}$ were contaminated by two scenarios, given by

(i) $Y_t = Z_t + X_t v_t$ for AO

(ii) $Y_t = Z_t + v_t$ for IO

where $\{Z_t\}$, $\{X_t\}$ and $\{v_t\}$ are independent and identically distributed (i.i.d.) series. For AR(1) model, the $\{Z_t\}$ assumed to be a stationary,

$$Z_t = \phi Z_{t-1} + a_t \qquad (5.19)$$

where ϕ is the autoregressive parameter and the a_t is the error term with mean 0 and variance 1. For MA(1) model, the $\{Z_t\}$ assumed to be an invertible,

$$Z_t = a_t - \theta a_{t-1} \tag{5.20}$$

where θ is the moving average parameter.

In first scenario, the Bernoulli process $\{X_t\}$ satisfies $P(x_t = 1) = \gamma$ and $P(x_t = 0) = 1 - \gamma$ so that the fraction of contamination is γ. This contamination fraction was set at 0%, 1%, 5 % and 10%. The $\{v_t\}$ has the form

$$v_t = \begin{cases} 10, & for\ x_t = 1,\ for\ \ t = 10,20,30,40,50 \\ 0, & otherwise. \end{cases}$$

For 1% contaminated, 10 is added to Z_{10}, for 5% contaminated, 10 is added to Z_{10} and Z_{20} and for 10% contaminated, 10 is added to $Z_{10}, Z_{20}, Z_{30}, Z_{40}$ and Z_{50}. For the second scenario, the $\{v_t\}$ has the following mixture density

$$F = (1-\gamma)N(0,\sigma^2) + \gamma\, N(0,\tau^2)$$

with γ = 0.01, 0.05 and 0.10 and σ = 1 and τ = 3.

For each estimator and for each type of simulated series, the mean error (ME) and the mean squared error (MSE) were computed using the following formulas

$$ME = \left\| ave_j \hat{\phi}_j - \phi \right\| \quad , \quad MSE = ave_j \left\| \hat{\phi}_j - \phi \right\|^2,$$

where ave_j denotes the average across the 30 replications and $\|.\|$ is the Euclidian norm operator. The words “efficient” and “robust” for the same concept of performance, that is an efficient (or robust) estimator is called so when its sample MSE is low.

For each of these models, five different estimates: the ML estimate, the *GM*-estimate, the ACM type robust filter, the RA estimate based on the Huber family $\psi_{H,chr}$ given by (3.60) (RAH-estimate) and the RA estimate based on the bisquare family $\psi_{B,cbr}$ given by (3.61) (RAB-estimate) were computed. In this study, the tuning constant *c* in the Huber family is 1.645 and in the bisquare family is 5.58.

The estimators included in the simulations as well as real data applications are as follows:

ML: The maximum likelihood estimates;

MH: M-estimates with $\Psi(.)$ of the Huber type, c = 1.345;

MB: M-estimates with $\Psi(.)$ of the bisquare type, c = 4.685;

GM: GM-estimates;

ACM: Approximate conditional mean etsimates;

RAH: RA-estimates with $\Psi(.)$ of the Huber type, c = 1.645; and

RAB: RA-estimates with $\Psi(.)$ of the bisquare type, c = 5.58.

To study the effect of outlier contamination on the bias and efficiency of the maximum likelihood (ML) and the robust estimators, model (5.19) is used. Tables (5.1) to (5.3) show the results of the simulation study with 30 replications of sample of size 50 for purely AR(1) model and for an AR(1) model with additive outliers (ARAO) and for an AR(1) model with innovation outliers (ARIO). The tables contain the average of the 30 estimators obtained for ϕ = 0.50, the corresponding MSEs as well as relative efficiencies (REF), defined as the ratio of the MSE of the ML estimator to the MSE of the robust estimator under consideration. The results shown in Table (5.1) suggest that there are no outliers, the MSE of ML and robust estimates are quite close to each other.

According to Table (5.2), it is clear that the resulting large MEs show just how sensitive the ML estimates are when the observations contain a small fraction of AO and IO. It is found that, the MEs and MSEs of ML estimates of AR(1) are not able to cope with contaminated situations, neither AO nor IO. It is clear that even a small fractions of outliers (γ = 1%) has a very large influence on the ML estimates. These estimators show a less resistance in terms of the mean value, it also has smaller efficiency than the robust estimators.

Moreover, the *GM*-estimator can handle both AO and IO quite successfully in the AR (1) model in line with the Monte Carlo results of Martin and Yohai (1985). Besides, the RA estimator based on the bisquare family compare favorably with the ML, *GM* and ACM estimators. From Table (5.3), it is clear that when the percentage of contamination increased to 5%, the RA estimator based on the Huber family yield a mean that is close to the true parameter for ARIO(1) model.

Tables (5.4) and (5.5) give the results of simulation study using 30 replications corresponding to (i) an MA(1) model with additive outliers (MAAO) model, and (ii) an MA(1) model with innovation outliers (MAIO) model. The tables contain the average of the 30 estimators obtained for $\theta = 0.50$. It is found out that the ML estimates are not very sensitive to the presence of innovation outliers. These results are in line with the Monte Carlo simulation results of Busto and Yohai (1986).

For MAAO model with $\gamma = 0.01$ and 0.05, the RAB performs much better than the ML estimates. However, under the MAIO model with $\gamma = 0.01$, the RAB is more robust than the ML estimates. When moving average terms are present, the *GM*-estimates are neither resistant nor robust. The simulation results indicate that the RA provides a good approximation to the true parameter of AR and MA models with AO and IO in line with the Monte Carlo simulation results of Busto and Yohai (1986).

Table (5.1) Simulation Results for the AR(1) Model with ϕ = 0.5

Estimates	γ = 0 %		
	ME	MSE	REF
ML	0.2681	0.0922	1.0000
GM	0.2583	0.0853	1.0809
ACM	0.2608	0.0926	0.9957
RAH	0.1244	0.0847	1.0885
RAB	0.2522	0.0865	1.0659

Source: Calculation Based on Simulation Data Sets

Table (5.2) Simulation Results for the ARAO(1) Model with ϕ = 0.5

Estimates	γ = 1%			γ = 5%			γ = 10%		
	ME	MSE	REF	ME	MSE	REF	ME	MSE	REF
ML	0.3672	0.1514	1.0000	0.4336	0.2000	1.0000	0.5240	0.2826	1.0000
GM	0.2545	0.0910	1.6637	0.2663	0.0882	2.2673	0.2661	0.0948	2.9810
ACM	0.2592	0.0950	1.5937	0.2554	0.0876	2.2829	0.2732	0.1066	2.6510
RAH	0.1295	0.1138	1.3304	0.1496	0.0899	2.2245	0.4417	0.0532	5.3120
RAB	0.2380	0.0837	1.8088	0.2278	0.0711	2.8127	0.1003	0.0304	9.2961

Source: Calculation Based on Simulation Data Sets

Table (5.3) Simulation Results for the ARIO(1) Model with ϕ = 0.5

Estimates	γ = 1%			γ = 5%			γ = 10%		
	ME	MSE	REF	ME	MSE	REF	ME	MSE	REF
ML	0.3493	0.1380	1.0000	0.4053	0.1935	1.0000	0.3577	0.1665	1.0000
GM	0.2749	0.0934	1.4775	0.2669	0.0985	1.9645	0.2812	0.1081	1.5402
ACM	0.2742	0.0946	1.4588	0.2794	0.1087	1.7801	0.3004	0.1220	1.3648
RAH	0.3340	0.1068	1.2921	0.1414	0.0658	2.9407	0.2202	0.1332	1.2500
RAB	0.2602	0.0881	1.5664	0.2759	0.0961	2.0135	0.2815	0.1088	1.5303

Source: Calculation Based on Simulation Data Sets

Table (5.4) Simulation Results for the MAAO(1) Model with θ = 0.5

Estimates	γ = 1%			γ = 5%			γ = 10%		
	ME	MSE	REF	ME	MSE	REF	ME	MSE	REF
ML	0.3409	0.1298	1.0000	0.3287	0.1275	1.0000	0.2852	0.0987	1.0000
GM	0.7888	0.6377	0.2035	0.7952	0.6501	0.1961	0.7936	0.6513	0.1515
ACM	0.7898	0.6373	0.2037	0.7851	0.6293	0.2026	0.7741	0.6163	0.1601
RAH	0.3156	0.1501	0.8648	0.3994	0.1863	0.6844	0.4855	0.2555	0.3863
RAB	0.2499	0.0781	1.6620	0.2882	0.1047	1.2178	0.5097	0.2824	0.3495

Source: Calculation Based on Simulation Data Sets

Table (5.5) Simulation Results for the MAIO(1) Model with θ = 0.5

Estimates	γ = 1%			γ = 5%			γ = 10%		
	ME	MSE	REF	ME	MSE	REF	ME	MSE	REF
ML	0.1989	0.0640	1.0000	0.2001	0.0581	1.0000	0.2119	0.0682	1.0000
GM	0.7769	0.6219	0.1029	0.7752	0.6150	0.0945	0.7474	0.5653	0.1206
ACM	0.7738	0.6172	0.1037	0.7529	0.5893	0.0986	0.7478	0.5658	0.1205
RAH	0.2743	0.1343	0.4765	0.2710	0.1230	0.4724	0.3227	0.1387	0.4917
RAB	0.1923	0.0599	1.0684	0.2260	0.0810	0.7173	0.2246	0.0742	0.9191

Source: Calculation Based on Simulation Data Sets

5.5.2 Real Data Study

The performance of robust procedures in the presence of outlier in an ARMA model was evaluated in this section. So, the outlier contaminated series had to be selected. Besides, the performances of robust estimates and ML estimates were compared based on MSE criterion.

Two series such as a daily average number of defects per truck and a yearly export of maize in Myanmar were chosen to analyze the effect of outliers in estimation of parameters of time series model in this section. The truck data set is a very popular data set and several statisticians studied it from time to time. To analyze

the effect of outliers in parameter estimation, we consider the truck data shown in Appendix B, Figure B.5.1 analyzed earlier by Wei (1990). According to the studied of Wei (1990), it was found out that this series contained two AO outliers and two IO outliers based on the results of outlier detection procedure. He pointed out that an AO occurs at time point t = 30 in first iteration, an IO occurs at time point t = 9 in second iteration, an AO occurs at time point t = 7 in third iteration and an IO occurs at time point t = 4 in fourth iteration.

(a) Analysis of Daily Average Number of Defects per Truck

Manufacturing defects of Truck series is a daily average number of defects per truck found in the final inspection at the end of the assembly line of a truck manufacturing plant. The data consist of 45 observations of consecutive business day during November 4 to January 10, as reported in Wei (1990). The Appendix Figure B.5.1 suggests a stationary process with constant mean and variance. For the identification of the order of p and q, the sample autocorrelation function (*sacf*) and sample partial autocorrelation function (*spacf*) are computed and plotted as shown in Appendix A, Table (A.5.2) and Figure 5.2. It shows that the *sacf* decays exponentially and the *spacf* has a single spike at lag 1 indicates that the series is likely to be an AR(1) or ARMA(1, 0) model. Hence, the tentative model is

$$(1-\phi B)(Z_t-\mu)=a_t \tag{5.21}$$

and the parameters of model (5.21) are estimated using the ML method. The following result is obtained:

$$\underset{(0.136)}{(1-0.43B)}\underset{(0.076)}{(Z_t-1.79)}=a_t \tag{5.22}$$

where the values in the parentheses under the estimates refer to the standard errors of those estimates. They are all highly significant at 1% level. Moreover, the *sacf* and *spacf* for the fitted model (5.22) as shown in Appendix A, Table (A.5.3) and Appendix B, Figure B.5.2 which indicate that the estimated residuals are a white noise process. From this result, it can be concluded that there is no outlier at all.

Although the AR(1) model is an adequate model for data, it contains four outliers according to the studied of Wei (1990). Therefore, the robust procedures are

required for this study. The parameters of Equation (5.21) are estimated using the *M*-estimator (Huber and bisquare), the *GM*-estimator, ACM type robust filter method, and RA estimators (Huber and bisquare) and the results are presented in Table (5.6). This table suggests that there is an evident that the data contain some outliers because the classical estimate and robust estimates differ each other. Besides, the MSE of RA estimate based on bisquare family (RAB) is the smallest, followed by the MSE of RA estimate based on Huber family (RAH), ACM, *M*-estimate based on bisquare, *GM* and *M*-estimate based on Huber respectively. In this study, RAB-estimate is more robust for a contaminated ARMA(1, 0) model. Thus, the estimated model become

$$\underset{(0.1351)}{(1-0.44B)}\underset{(0.076)}{(Z_t-1.79)}=a_t \tag{5.23}$$

where the values in the parentheses under the estimates refer to the standard errors of those estimates.

To check model adequacy, the residual *acf* and *pacf* of the fitted model as shown in Appendix A, Table (A.5.4) are all small and exhibit no patterns. Moreover, the values of Box-Ljung statistic are not significant at 5% level for all lags. Thus, the fitted ARMA(1, 0) model in (5.23) is adequate for the data using RA procedure based on bisquare family. In this study, it is found that the RAB-estimate is more robust for AR(1) model.

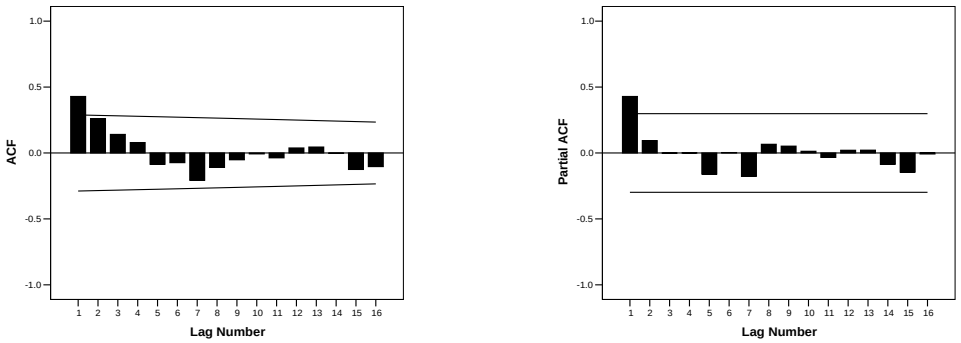


Figure 5.2 The *sacf* and *spacf* of the Truck Series

Source: Appendix Table (A.5.2)

Table (5.6) Summary Results of the Truck Series

Estimates	$\hat{\phi}$	Se($\hat{\phi}$)	MSE
ML	0.4289	0.1362	0.2103
MH	0.4465	0.1240	0.1249
MB	0.4503	0.1133	0.1185
RAH	0.4350	0.1357	0.1059
RAB	0.4436	0.1351	0.1047
GM	0.4662	0.4470	0.1200
ACM	0.5172	0.4063	0.1173

Source: Appendix Table (A.5.1)

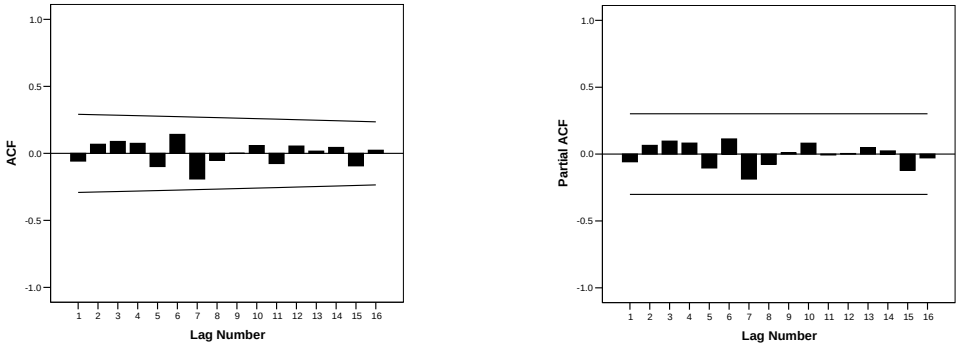


Figure 5.3 The *sacf* and *spacf* for the Residual of the RAB-AR(1) Model
Source: Appendix Table (A.5.4)

(b) Analysis of Yearly Export of Maize

In this study, the data were also obtained from the Statistical Year Books which are published by the Central Statistical Organization (CSO) over the period of 1976 to 2008. A plot of this series is shown in Appendix B, Figure B.5.3. This plot indicates that the series is nonstationary both in the mean and the variance. First, this series was transformed by taking a logarithmic transformation. Then, the transformed data are plotted in Appendix B, Figure B.5.4 and it shows a trend with a constant variance. This trend has to be removed by differencing before a model can be identified. The very slowly decaying *sacf* was shown in Appendix A, Table (A.5.7) and Figure 5.4 further support to need for differencing. Hence, the sample *sacf* and *spacf* for the differenced series are calculated in Appendix A, Table (A.5.8) with their plots in Appendix B, Figure B.5.5. The *sacf* tails off and the *spacf* cuts off after lag 1. It suggests that an ARIMA(1, 1, 0) or ARI(1, 1) is a possible model. Whether the

deterministic trend term θ_0 is needed or not, the *t*-ratio is examined $t = \overline{W} / S_{\overline{w}}$ = 0.09195 / 0.08942 =1.0283, which is not significant. Hence, the following ARIMA(1, 1, 0) model is considered as our proposed model:

$$(1-\phi B)(1-B)(ln\, Z_t - \mu) = a_t \tag{5.24}$$

and the estimated model is

$$\underset{(0.1486)}{(1+0.56B)}(1-B)(\ln Z_t - 3.69) = a_t \tag{5.25}$$

where the value in the parenthesis under the estimate refers to the standard error of this estimate.

The residual *acf* of this fitted model as shown in Appendix A, Table (A.5.9) are small and exhibit no patterns. In other words, the residuals from this modified model are white noise. Thus, the fitted ARIMA(1, 1, 0) model is adequate. Although the results all indicate a white noise phenomenon, a white noise series is itself an outlier contaminated series[1]. Appendix Figure B.5.4 shows a single outlier (AO) occurs at time *t* =1988. So, the ML estimate as shown in (5.25) is not robust in the presence of AO. Therefore, the robust procedures are used to estimate the parameters of the ARIMA(1, 1, 0) model and results are shown in Table (5.7). Due to the presence of a single AO outlier, the ML and robust estimates differ substantially. The MSE of ML and robust estimates are greatly different. Besides, the robust estimates are not markedly affected by outlier, whereas the ML procedure is. The ML estimate has the largest MSE among the robust methods in the presence of a single AO outlier. The MSE of RA estimate based on bisquare family (RAB) is the smallest, followed by the MSE of RA estimate based on Huber family (RAH), ACM, *GM* and *M*-estimate based on Huber and bisquare respectively. In this study, RAB-estimate is more robust for ARIMA(1, 1, 0) model with a single AO outlier. Therefore, the estimated model becomes

$$\underset{(0.1468)}{(1+0.58B)}(1-B)(\ln Z_t - 3.69) = a_t \tag{5.26}$$

[1] Wei (1990), Time Series Analysis Univariate and Multivariate Methods, U.S.A: Addison-Wesley Publishing Co., Chapter 9, section 9.5, pp. 203

where the value in the parenthesis under the estimate refers to the standard error of this estimate.

To check model adequacy, the residual *acf* and *pacf* and of the fitted model as shown in Appendix Table (A.5.10) are small and exhibit no patterns. Moreover, the values of Box-Ljung statistic are not significant at 5% level for all lags. So, the fitted ARIMA (1, 1, 0) model in (5.26) is adequate for the data using RA procedure based on bisquare family. In this study, it is observed that the RAB-estimate is more robust for ARIMA(1, 1, 0) model.

Table (5.7) Summary Results of the Maize Export Series

Estimates	$\hat{\phi}$	Se($\hat{\phi}$)	MSE
ML	-0.5597	0.1488	0.6025
MH	-0.6887	0.0984	0.2104
MB	-0.7648	0.0875	0.2283
RAH	-0.5651	0.1494	0.1268
RAB	-0.5758	0.1468	0.1249
GM	-0.7068	0.6283	0.2065
ACM	-0.6794	0.5610	0.1727

Source: Appendix Table (A.5.6)

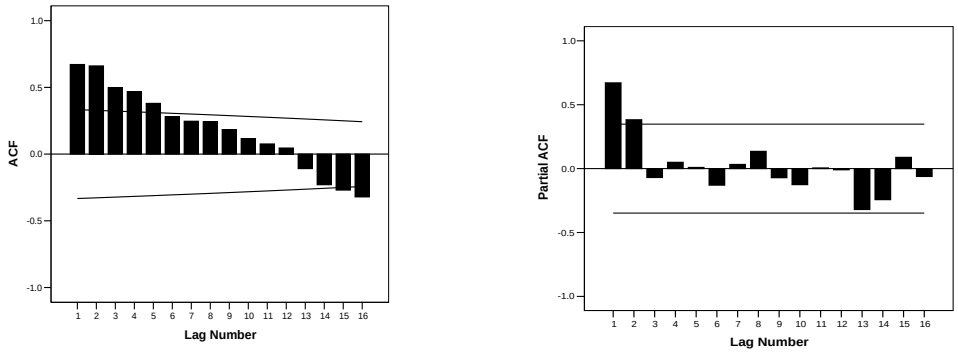


Figure 5.4 The *sacf* and *spacf* for Natural Logarithms of the Export Maize

Source: Appendix Table (A.5.7)

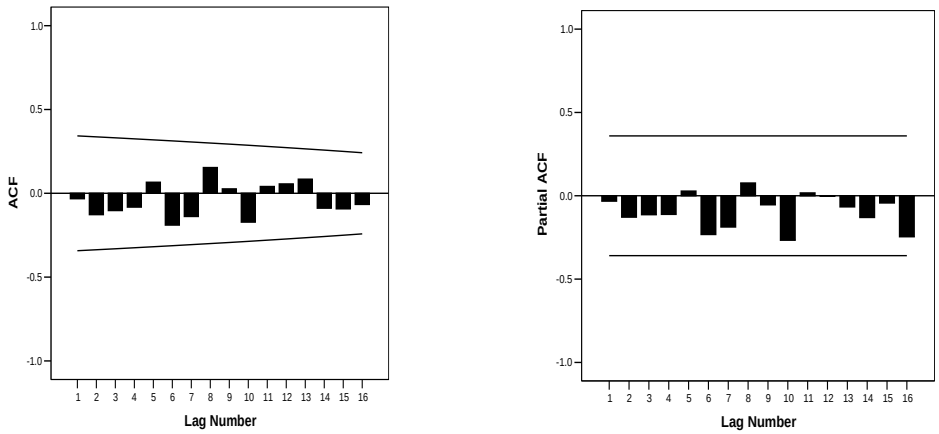


Figure 5.5 The *sacf* and *spacf* for the Residual of the RAB-ARIMA(1, 1, 0) Model

Source: Appendix Table (A.5.10)

CHAPTER VI

CONCLUSION

Many classical statistical procedures are known for not being robust, because results violate stochastic assumptions and rely on a few sample observations. These procedures are optimal when the assumed model is exactly satisfied, but they are biased and/or inefficient when small deviations from the model are present. The results obtained by classical procedures can therefore mislead when it comes to applications. Consequently, it may cause issues when the classical methods are used for the underlying model.

Outliers may appear in data due to gross errors, wrong classification of the data, grouping, and correlation in the data. Gross errors or outliers are data severely deviating from the pattern set by the majority of the data. This type of error usually occurs due to mistakes in copying or computation. They can also be due to part of the data not fitting the same model, as in the case of data with multiple clusters. Gross errors are often the most dangerous type of errors. In fact, a single outlier can completely spoil the least squares estimate, causing it to break down. Consequently, the estimators may not be efficient estimators. Some outliers are genuine and may be the most important observations of the sample.

Rounding and grouping errors result from the inherent inaccuracy in collecting and recording data which are usually rounded, grouped, or even roughly classified. The departure from an assumed model means that real data can deviate from the assumed distribution. The departure from the normal distribution can manifest itself in many ways, for instance, in the form of skewed or longer-tailed distributions. Hence, one would naturally like to employ an estimation method that is sufficiently resistant to outliers.

Robust statistics is concerned with the construction of statistical procedures which are still reliable and reasonably efficient in a neighborhood of the model. Robust statistics is the generalization of the classical theory: it takes into account of model misspecification, and the inferences remain valid not only at the parametric model but also in the neighborhood model.

Two major problems such as data robustness and distributional robustness usually arise in regression analysis. In time series analysis, outliers can cause biases in

parameter estimation as well as misspecification, resulting in misleading conclusion. Robust procedures in linear regression and time series, that can overcome the above two problems, have been investigated by many researchers. The present dissertation tries to analyze these problems and explore suitable methods for contaminated data sets in regression and time series analysis. In order to analyze the effect of outliers on the estimation of parameters in regression and times series model, the classical and the robust estimation techniques are used. The performances of these estimators are analyzed by simulations as well as real data. Major findings can be summarized as follows:

From the study of outlier robustness, the multiple linear regression with three explanatory variables is used to generate the data sets. These clean data sets are transformed into outlier contaminated data sets. In this simulation study, two scenarios are analyzed. According to the findings of the first scenario, it is shown that the *MM*-estimates (with a 70%, an 85% and a 95% efficiency), Huber and Turkey *M*-estimates, and LAV estimates are more resistant and efficient in the presence of vertical outliers. The OLS estimates provide poor estimates of true parameters of the regression model. Similarly, the Huber and Turkey *M*-estimates and *MM*-estimates are in line with the asymptotic robustness properties in the presence of both vertical and bad leverage points. As expected, OLS is a less efficient estimator whatever the type of outliers present in the data.

For the study of distribution robustness in regression, the performances of six regression methods for two important classes of distributions namely symmetric and skewed are investigated. Different error structures like normal, logistic, exponential, Cauchy and gamma distributions are used. It is found that, the OLS method is more efficient than the robust methods under normal error distribution. In this case, the LMS method performs much worst. In logistic distribution, the Turkey-*M* estimator is more robust than other estimation methods. Although no preferred robust method can be chosen in exponential and Cauchy distributions, the robust methods clearly outperform the OLS method. Moreover, it is shown that the OLS method performs much worst in the study of gamma distribution. The LMS method is more resistant in this distribution.

The simulation results indicate that when outliers exist, other alternatives of the OLS are more appropriate. Selecting a more efficient alternative to the OLS method is closely related to the type of data and so it is advisable to use several

alternative methods in data analysis. In cases of skewed distributions, the performance of OLS is inferior as compared to other methods. Based on bias and MSE criteria, the LMS is more suitable for the exponential and gamma distributions.

In symmetric distributions investigated here, the MSEs are very close to one another for the sample sizes larger than 50 and so none of the estimation methods is superior in such circumstances. However, this is not true for the cases of the skewed distributions where the OLS method has shown to be far inferior from the other methods of estimation. Compared to MSE criterion, the bias criterion fluctuated more and this fluctuation persists even for larger sample sizes. This instability of biases created some difficulties and confusion in finding the optimum estimation in some situations.

When series are outlier contaminated (γ = 0.01, 0.05 and 0.10), an overall result is that outliers adversely affect the bias as well as MSE of OLS estimators. It is found that OLS estimation under a heavy-tailed distribution does not yield outlier robust estimates. Indeed, not only with the Gaussian distribution but also with the skewed distributions, OLS estimators collapse in the presence of small levels of outlier contamination.

Moreover, it is illustrated that the RDL_1 method clearly outperforms the OLS and LTS methods in the regression model involved with dummy variables. In this study, the RDL_1 and LTS methods detect several outliers whereas the least squares residuals do not reveal any outlier. Based on the mean squared error (MSE) criterion, the RDL_1 estimator is more resistant but it suffers from swamping effect. The OLS method is much worst in this case. According to the result of LTS analysis, the observations (25, 26, 27, and 28) gained from the simulated data are excluded and the remaining data are rerun using the OLS method. It is found that there is no outliers in the data OLS estimators are more robust.

In analysis with real data, a curvilinear regression model is fitted to the production and export of maize data. It is found that the estimated residuals of fitted model do not follow the normality assumption. In this study, both variables have to be transformed to meet basic assumptions. However, the transformation does not eliminate or attenuate the leverage of influential outliers that bias the prediction and distort the significance of parameter estimates. Even though the curvilinear effect is included in the model, the reliable results are not obtained from the OLS fit. In such a

situation, the LMS estimation methods provide good results in estimating the true parameters. Similarly, a multiple regression model with three explanatory variables is fitted to the maternal mortality ratio data. It is found that the residuals of the fitted model do not satisfy the normality assumption. The LMS estimation method is more robust in this study.

Outliers in an economic time series include deviations that occur because of unusual events such as policy changes, environmental regulations, economic changes, advertising promotions, supply interruptions, natural disasters, wars, strikes and similar events. In time series, outliers can take several forms. Among than additive outlier (AO) and innovation outlier (IO) are focused in this study. The difference between AO and IO is that in fact an AO is interpreted as an outlying observation added after the realization to affect a single observation and an IO as an outlying observation added during the realization with influence on all succeeding observations.

In the case of an AR(1) model, one IO yields one outlier in the response variable and a number of "good" leverage points ("good" refers to the fact that the leverage points lie close to the fitted line determined by the majority of the data), which actually improve the accuracy of the parameter estimate. Therefore, one IO only affects one residual. On the other hand, one AO results in one outlier in the vertical direction and one "bad" leverage point ("bad" refers to the fact that the leverage point does not lie close to the fitted line determined by the majority of the data). Thus, AO also affects the next residual inflating two consecutive residuals. The presence of such outliers in a time series can also have substantial effects on parameters estimation.

The effect of outliers on estimation of parameter in time series models is analyzed by using simulation. In this simulation study, the time series models such as AR(1) with AO outlier, AR(1) with IO outlier, MA(1) with AO outlier and MA(1) with IO outlier are investigated and the parameters of each model is estimated by using ML method and robust methods. The results from simulation comparisons indicate that the RA estimates based on bisquare family have very good robustness properties for AR(1) with AO outlier and AR(1) model with IO outlier and they compare favorably with the *GM*-estimates. When the percentage of contamination increased to 5%, the RA estimator based on the Huber family yielded a mean that is close to the true parameter for AR(1) with IO outlier model.

Moreover, the *GM*-estimator can handle both AO and IO quite successfully in the AR(1) model. The ML estimates of AR(1) are not able to cope with contaminated situations, neither AO nor IO. It is clear that even a small fraction of outliers (γ = 1%) has a very large influence on the ML estimates. The RA estimates also behave robustly in terms of efficiency for the MA(1) model with AO when γ = 0.01 and 0.05. When moving average terms are present, the *GM*-estimates are neither resistant nor robust. Under the MA(1) model with IO outlier, it is preferable to use the ML estimates. On the other hand, the ML estimate is extremely sensitive to the presence of AOs. Since small fraction of AOs may cause a large bias in the ML estimates, they are more dangerous than IOs.

In real data application, two series namely a daily average number of defects per truck and a yearly export of maize in Myanmar are chosen to analyze the effect of outliers in estimating parameters in the time series model. The truck data set is a very popular data set and several statisticians studied it from time to time. It is noted that the truck data contained both the AO as well as the IO. The result of truck data points out that the RA estimate based on bisquare family is more efficient under the AR(1) with AO and IO. Under the AR(1) model with AO and IO, the MSEs of *MH*, *MB*, RAH, RAB, *GM* and ACM estimators are quite close to each other. The RA estimates perform better than the ML estimates and other robust methods. The estimate based on bisquare performs better than those based on Huber.

A similar result is obtained from the analysis of data about export of maize. This data set contains an AO outlier and the ARIMA(1,1,0) model is fitted to it. Due to the presence of a single AO outlier, it is found that the ML and robust estimates differed substantially. The ML estimate has the largest MSE among the robust methods in the presence of a single AO outlier. The MSE of RA estimate based on bisquare family (RAB) is found to be the smallest; it is followed by the MSE of RA estimate based on Huber family (RAH), ACM, *GM* and *M*-estimate based on Huber and bisquare respectively. In this study, the RA estimate based on bisquare family is found to be more robust under the ARIMA(1,1,0) model with a single AO outlier.

To sum up, when there are no outliers in the data, the robust and classical methods both worked well, indicating that the values of MSE are quite close to each other. When there are outliers in the data, the robust methods perform better than the classical method.

REFERENCES

1. Andrews, D. F. (1974), A Robust Method for Multiple Linear Regression, *Technometrics*, vol.16, 523-531.

2. Armstrong, R.D., and E.L. Frome (1977), A Special Purpose Linear Programming Algorithm for Obtaining Least Absolute Value Estimators in a Linear Model with Dummy Variables, *Communications in Statistics*: *Simulation and Computation*, vol. B6, 383-398.

3. Ashar, V. C., and T. D. Wallace (1963), A Sampling Study of Minimum Absolute Deviation Estimators, *Operations Research*, vol. 11, 747-758.

4. Atkinson, A. C. (1994), Fast Very Robust Methods for the Detection of Multiple Outliers, *Journal of American Statistical Association,* vol. 89, 1329-1339.

5. Atkinson, A. C., and T. C. Cheng (1999), Computing Least Trimmed Squares Regression with the Forward Search, *Statistics and Computing*, vol.9, 251-263.

6. Atkinson, A. C., and M. Riani (2000), *Robust Diagnostic Regression Analysis*, New York : Springer-Verlag.

7. Barnett, V., and T. Lewis (1978), *Outliers in Statistical Data*, New York: John Wiley and Sons.

8. Beaton, A.E., and J.W. Turkey (1974), The Fitting of Power Series, Meaning Polynomials, Illustrated on Band Spectroscopic Data, vol. 16, 147-185.

9. Berrendero, J. R., and R. H. Zamar (2001), Maximum Bias Curves for Robust Regression with Non-Elliptical Regressors, *The Annals of Statistics*, vol. 29, 224-251.

10. Boente, G., R. Fraiman, and V.J. Yohai (1987), Qualitative Robustness for Stochastic Processes, *The Annals of Statistics*, vol. 15, 1293-1312.

11. Box, G.E.P. (1953), Non-Normality and Tests of Variances, *Biometrika*, vol. 40, 318- 335.

12. Box, G.E.P., and G.M. Jenkins (1970), *Time Series Analysis*: *Forecasting and Control*, Revised Edition, San Francisco: Holen Day Inc.

13. Birch, J. B., and Myers (1982), Robust Analysis of Covariance, *Biometrics*, vol. 38, 699-713.

14. Busto, O. H., and V.J. Yohai (1986), Robust Estimates for ARMA Models, *Journal of American Statistical Association*, vol. 81, 155-168.

15. Campbell, B., and J. W. Galbraith (1993), Inference in Expectations Models of the Term Structure: a Non-Parametric Approach, *Empirical Economics*, vol. 18, 623-638.

16. Chang, I., G. C. Tiao, and C. Chen (1988), Estimation of Time Series Parameters in the Presence of Outliers, *Technometrics*, vol. 30, 2, 193-204.

17. Chatterjee, S., and B. Price (1977), *Regression Analysis by Example*, New York: John Wiley.

18. Coakley, C.W., and T.P. Hettmansperger (1993), A Bounded Influence, High Breakdown, Efficient Regression Estimator, *Journal of American Statistical Association,* vol. 88, 872-880.

19. Croux, C., P. J. Rousseeuw, and O. Hossjer (1994), Generalized S-Estimators, *Journal of American Statistical Association,* vol. 89, 1271-1281.

20. Davies, L. (1992), The Asymptotic of Rousseeuw's Minimum Volume Ellipsoid Estimator, *The Annals of Statistics*, vol. 20, 1828-1843.

21. Davies, L. (1993), Aspects of Robust Linear Regression, *The Annals of Statistics*, vol. 21, 1843-1899.

22. Denby, L. and D. Martin (1979), Robust Estimation of the First-Order Autoregressive Parameter, *Journal of American Statistical Association,* vol. 74, 140-145.

23. Donoho, D. L., and P.J. Huber (1983), The Notion of Breakdown-Point, in P. J. Bickel, K. A. Doksum and J. L. Hodges, eds., *A Festschrift for Erich Lehmann*, WadsWorth, Belmont, CA, 157-184.

24. Draper, N. R., and H. Smith (1981), *Applied Regression Analysis*, New York: John Wiley.

25. Edgeworth, F. Y. (1887), On Observations Relating to Several Quantities, *Hermathena*, vol. 6, 279-285.

26. Ferretti, N., D. Kelmansky, V. J. Yohai, and R.H. Zamar (1999), A Class of Locally and Globally Robust Regression Estimates, *Journal of American Statistical Association,* vol. 94, 174-188.

27. Fox, A.J. (1972), Outliers in Time Series, *Journal of the Royal Statistical Society, Series B*, vol. 34, 350-363.

28. Furno, M (2000), LAD Estimation with Random Coefficient Autocorrelated Errors, www.eco.unicas.it/docente/furno/Text-rca.pdf

29. Hadi, A. S. (1992), Identifying Multiple Outliers in Multivariate Data, *Journal of the Royal Statistical society, Series B,* vol. 54, 761-771.

30. Hampel, F. R. (1971), A General Qualitative Definition of Robustness, *The Annals of Mathematical Statistics*, vol. 42, 1887-1896.

31. Hampel, F. R. (1974), The Influence Curve and Its Role in Robust Estimation, *Journal of American Statistical Association*, vol. 69, 383-393.

32. Hampel, F. R. (1975), Beyond Location Parameters: Robust Concepts and Methods, *Proceedings of the 40th Session of the ISI*, vol. 46, 375-391.

33. Hampel, F. R., E. M. Ronchetti, P. J. Rousseeuw, and W.A. Stahel (1986), *Robust Statistics, The Approach Based on Influence Functions,* New York: John Wiley and Sons.

34. Hoaglin, D.C., F. Mosteller, and J. W. Turkey (1983), *Understanding Robust and Exploretory data Analysis,* New York: John Wiley.

35. Huber, P. J. (1964), Robust Estimation of a Location Parameter, *Annals of Mathematical Statistics*, vol. 35(1): 73-101.

36. Huber, P. J. (1973), Robust Regression: Asymptotics, Conjectures and Monte Carlo, *The Annals of Statistics*, vol. 1, 799-821.

37. Huber, P. J. (1981), *Robust Statistics*, New York: John Wiley and Sons.

38. Hubert, M., and P. J. Rousseeuw (1996), Robust Regression with a Categorical Covariable, *Robust statistics, Data analysis, and Computer Intensive Methods*, New York : Springer.

39. Hubert, M., and P.J. Rousseeuw (1997), Robust Regression with Both Continuous and Binary Regressors, *Journal of Statistical Planning and Inference*, vol. 57, 153-163.

40. Jaeckel, L.A. (1972), Estimating Regression Coefficients by Minimizing the Dispersion of the Residuals, *Annals of Mathematical Statistics*, vol. 43: 1449-1458.

41. Konsinski, A. S. (1999), A Procedure for the Detection of Multivariate Outliers, *Computational Statistics & Data Analysis*, vol. 29, 145-161.

42. Lucas, A. (1995), An Outlier Robust Unit Root Test with an Application to the Extended Nelson-Plosser Data, *Journal of Econometrics*, vol. 66, 153-173.

43. Luceno, A. (1998), Detecting Possibly Non-Consecutive Outliers in Industrial Time Series, *Journal of Royal Statistical Society, Series B*, vol. 60, 295-310.

44. Maddala, G.S., and Y. Yong (1997), Outliers, Unit Roots and Robust Estimation of Nonstationary Time Series, *Handbook of Statistics*, vol. 15, 237-266.

45. Maronna, R. A., O. H. Bustos, and V.J. Yohai (1979), Bias- and Efficiency-Robustness of General M-estimators for Regression with Random Carriers, in

T. Gasser and M. Rosenblatt, eds., *Smothing Techniques for Curve Estimation*, Lecture Notes in Mathematics 757, Springer-Verlag, Berlin, 91-116.

46. Maronna, R. A., D. Martin, and V. J. Yohai (2006), *Robust Statistics*, New York: John Wiley and Sons.

47. Martin, R. D. (1979), *Robust Estimation for Time Series Autoregressions*, New York: Academic Press, Inc.

48. Martin, R.D., and V. J. Yohai (1985), Robustness in Time Series and Estimating ARMA Models, *Handbook of Statistics*, vol. 3, 85-118.

49. Meyer, J. R., and R. R. Glauber (1964), *Investment Decisions: Economic Forecasting and Public Policy*, Cambridge, Massachusetts: Harvard Business School Press.

50. Mili, L., and C. W. Coakley (1993), Robust Estimation in Structured Linear Regression, *Annals of Statistics*.

51. Mosteller, F., and J. W. Tukey (1977), *Data Analysis and Regression*, Philippines: Addison-Wesley Publishing Company, Inc.

52. Mya Thanda (2010), ARIMA Modeling for Selected Economic Time Series with Outliers, Unpublished Ph. D Thesis.

53. Oveson, R. M. (1968), Regression Parameter Estimation by Minimizing the Sum of Absolute Errors, Unpublished Ph. D. Thesis, Harvard University, Cambridge.

54. Pena, D. (1984), Influential Observations, *MRC Tech*. *Rep*. 2718.

55. Rousseeuw, P. J. (1984), Least Median of Squares Regression, *Journal of American Statistical Association*, vol. 79, 871-880.

56. Rousseeuw, P. J., and C. Croux (1993), Alternatives to the Median Absolute Deviation, *Journal of American Statistical Association*, vol. 88, 1273-1283.

57. Rousseeuw, P. J., and K. Van Driessen (1999), A Fast Algorithm for the Minimum Covariance Determinant Estimator, *Technometric*s, vol. 41, 212-223.

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

59. Rousseeuw, P. J., and V. J. Yohai (1984), Robust Regression by Means of S-Estimators. Robust and Nonlinear Time series Analysis, *Lecture Notes in Statistics*, vol. 26, 256-272.

60. Rousseeuw, P. J., and B. C. Van Zomeren (1990), Unmasking Multivariate Outliers and Leverage Points, *Journal of American Statistical Association,* vol. 85, 633-639.

61. Ryan, T. P. (1997), *Modern Regression Methods*, New York: John Wiley and Sons.

62. Sakata, S., and H. White (1995), An Alternative Definition of Finite Sample Breakdown.

63. Soe Win (2004), Adjustment Diagnostic for Outliers in Time Series, Unpublished Ph. D Thesis, University of Hyderabad, India.

64. Tukey, J. W. (1960), A Survey of Sampling from Contaminated Distributions. In *Contributions to Probability and Statistics*, Stanford: Stanford University Press.

65. Tukey, J. W. (1976), *Exploratory Data Analysis*, Addison-Wesley, Reading. MA.

66. Wei, W. W. S. (1990), *Time Series Analysis Univariate and Multivariate Methods*, U.S.A : Addison-Wesley Publishing Co.

67. Yohai, V. J. (1987), High Breakdown-point and High Efficiency Robust Estimates for Regression, *The Annals of Statistics*, vol. 15, 642-656.

68. Yohai, V.J., and R. Maronna (1979), Asymptotic Behavior of M-estimates for the Linear Model, *The Annals of Statistics* , vol. 7, 258-268.

69. Yohai, V. J., and R. H. Zamar (1988), High Breakdown Estimates of Regression by Means of Minimization of an Efficient Scale, *Journal of American Statistical Association*, vol. 83, 406-413.

70. Zaman, A., P. J. Rousseeuw, and M. Orhan (2001), Econometric Applications of High-breakdown Robust Regression Techniques, *Economics Letters*, vol. 71, 1-8.

APPENDICES

APPENDIX A

Table (A.4.1) Bias and MSE for OLS and Robust Methods of Simulated Data with Vertical Outliers

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n = 30	OLS	Bias	1.9748	0.9043	1.0060	1.1331
		MSE	4.4833	1.7279	1.6046	2.0718
	LAV	Bias	0.3237	0.3173	0.2570	0.2460
		MSE	0.3378	0.3391	0.1148	0.1133
	M-H	Bias	0.3151	0.2832	0.2277	0.2381
		MSE	0.3284	0.3018	0.0916	0.0958
	M-T	Bias	0.2354	0.2419	0.1813	0.1787
		MSE	0.2809	0.2519	0.0625	0.0574
	MM-(0.70)	Bias	0.2481	0.2794	0.2143	0.2081
		MSE	0.2905	0.2940	0.0934	0.0762
	MM(0.85)	Bias	0.2403	0.2513	0.1914	0.1892
		MSE	0.2800	0.2744	0.0690	0.0647
	MM-(0.95)	Bias	0.2387	0.2506	0.1932	0.1813
		MSE	0.2827	0.2744	0.0729	0.0593
n = 40	OLS	Bias	2.0424	0.9613	0.7830	0.8337
		MSE	5.2361	1.4872	1.0443	1.3226
	LAV	Bias	0.3744	0.1748	0.2479	0.2070
		MSE	0.9587	0.0454	0.1463	0.1095
	M-H	Bias	0.3800	0.1810	0.2233	0.1918
		MSE	1.0227	0.0544	0.1283	0.1095
	M-T	Bias	0.2994	0.1446	0.2092	0.1579
		MSE	0.8796	0.0346	0.1100	0.0643
	MM-(0.70)	Bias	0.3171	0.1702	0.2202	0.1726
		MSE	0.8723	0.0455	0.1151	0.0696
	MM(0.85)	Bias	0.3099	0.1543	0.2111	0.1599
		MSE	0.8772	0.0407	0.1081	0.0602
	MM-(0.95)	Bias	0.2999	0.1511	0.2153	0.1556
		MSE	0.8763	0.0380	0.1107	0.0585

Simulation setup: simulations = 100, contamination = 10%

Source: Calculations based on simulation data

Table (A.4.2) Bias and MSE for OLS and Robust Methods of Simulated Data with Vertical Outlier and Bad Leverage Points

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n = 30	OLS	Bias	0.7798	0.1761	0.1618	0.6959
		MSE	1.1732	0.0558	0.0464	0.8442
	LAV	Bias	0.3016	0.0832	0.0983	0.2025
		MSE	0.6038	0.0135	0.0166	0.0829
	M-H	Bias	0.2894	0.0796	0.0855	0.1934
		MSE	0.5636	0.0124	0.0120	0.0743
	M-T	Bias	0.2896	0.0926	0.0932	0.1814
		MSE	0.5688	0.0201	0.0180	0.0624
	MM-(0.70)	Bias	0.3057	0.1286	0.1261	0.2095
		MSE	0.6098	0.0397	0.0387	0.0852
	MM-(0.85)	Bias	0.2966	0.1175	0.1165	0.1872
		MSE	0.5638	0.0309	0.0315	0.0665
	MM-(0.95)	Bias	0.2905	0.1218	0.1112	0.1841
		MSE	0.5686	0.0387	0.0262	0.0674
	LMS	Bias	0.3150	0.0992	0.0999	0.2046
		MSE	0.5869	0.0233	0.0215	0.0762
	LTS	Bias	0.4382	0.2175	0.2680	0.3365
		MSE	0.7850	0.1027	0.1484	0.1802
n = 40	OLS	Bias	0.4899	0.1193	0.1449	0.1400
		MSE	0.2817	0.0240	0.0556	0.0357
	LAV	Bias	0.1614	0.0775	0.0855	0.0762
		MSE	0.0462	0.0097	0.0170	0.0094
	M-H	Bias	0.1408	0.0637	0.0714	0.0708
		MSE	0.0352	0.0067	0.0152	0.0082
	M-T	Bias	0.1432	0.0786	0.0804	0.0769
		MSE	0.0362	0.0163	0.0177	0.0099
	MM-(0.70)	Bias	0.1559	0.1138	0.1013	0.0944
		MSE	0.0444	0.0296	0.0242	0.0172
	MM(0.85)	Bias	0.1546	0.1016	0.0886	0.0905
		MSE	0.0407	0.0253	0.0182	0.0144
	MM-(0.95)	Bias	0.1462	0.0962	0.0887	0.0800
		MSE	0.0366	0.0252	0.0203	0.0107
	LMS	Bias	0.1701	0.1078	0.0953	0.0855
		MSE	0.0515	0.0290	0.0228	0.0129
	LTS	Bias	0.3212	0.2157	0.1847	0.1872
		MSE	0.1621	0.0978	0.0715	0.0750

Simulation setup: simulations = 100, contamination = 10%

Source: Calculations based on simulation data

Table (A.4.3) Performances of OLS and Robust Methods of Normal Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =10	OLS	Bias	0.3086	0.2048	0.2738	0.2051
		MSE	0.1757	0.0644	0.1107	0.1176
	LAV	Bias	0.4002	0.2181	0.3676	0.2434
		MSE	0.2544	0.0695	0.2465	0.0907
	M-H	Bias	0.2974	0.2060	0.2266	0.2026
		MSE	0.1453	0.0685	0.0938	0.1038
	M-T	Bias	0.3015	0.3553	0.3434	0.3000
		MSE	0.1488	0.3310	0.2575	0.1934
	LTS	Bias	0.3432	0.2608	0.3072	0.2854
		MSE	0.1584	0.1668	0.1986	0.1045
	LMS	Bias	0.5234	0.5989	0.5645	0.4626
		MSE	0.4554	0.5986	0.5141	0.8200
n =20	OLS	Bias	0.1912	0.1865	0.2020	0.1403
		MSE	0.0602	0.0617	0.0744	0.0276
	LAV	Bias	0.2116	0.2230	0.2501	0.2515
		MSE	0.0602	0.0777	0.1012	0.1050
	M-H	Bias	0.2020	0.2012	0.2022	0.1145
		MSE	0.0701	0.0647	0.0720	0.0226
	M-T	Bias	0.1963	0.1941	0.2146	0.1186
		MSE	0.0641	0.0634	0.0719	0.0235
	LTS	Bias	0.2531	0.2828	0.2812	0.2082
		MSE	0.1131	0.1352	0.1375	0.0776
	LMS	Bias	0.5271	0.4909	0.4668	0.3154
		MSE	0.4082	0.3397	0.3225	0.1779
n =30	OLS	Bias	0.1642	0.1703	0.1450	0.1526
		MSE	0.0376	0.0494	0.0304	0.0334
	LAV	Bias	0.2056	0.2868	0.1536	0.1546
		MSE	0.0603	0.1480	0.0319	0.0306
	M-H	Bias	0.1940	0.1567	0.1554	0.1365
		MSE	0.0473	0.0446	0.0286	0.0249
	M-T	Bias	0.1860	0.1578	0.1491	0.1342
		MSE	0.0443	0.0441	0.0273	0.0255
	LTS	Bias	0.2227	0.1474	0.1598	0.0857
		MSE	0.0677	0.0400	0.0319	0.0132
	LMS	Bias	0.2395	0.1996	0.3296	0.3494
		MSE	0.0948	0.0549	0.1444	0.2118
n=50	OLS	Bias	0.1609	0.2201	0.1345	0.0908
		MSE	0.0328	0.0588	0.0265	0.0104
	LAV	Bias	0.1434	0.2307	0.1200	0.0979
		MSE	0.0284	0.0807	0.0200	0.0132
	M-H	Bias	0.1602	0.2016	0.1367	0.0905
		MSE	0.0315	0.0586	0.0254	0.0097
	M-T	Bias	0.1643	0.1964	0.1331	0.0857
		MSE	0.0331	0.0550	0.0242	0.0090
	LTS	Bias	0.1679	0.2641	0.1527	0.0923
		MSE	0.0388	0.0878	0.0316	0.0106
	LMS	Bias	0.2383	0.3356	0.1687	0.2540
		MSE	0.0977	0.1592	0.0445	0.1204

Table (A.4.3) (Contd.)

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =80	OLS	Bias	0.1396	0.1424	0.0917	0.0710
		MSE	0.0232	0.0279	0.0110	0.0108
	LAV	Bias	0.1446	0.1501	0.1034	0.0653
		MSE	0.0266	0.0358	0.0161	0.0075
	M-H	Bias	0.1358	0.1374	0.0877	0.0641
		MSE	0.0231	0.0294	0.0106	0.0098
	M-T	Bias	0.1413	0.1304	0.0893	0.0611
		MSE	0.0252	0.0273	0.0108	0.0010
	LTS	Bias	0.1184	0.1761	0.1093	0.1218
		MSE	0.0207	0.0478	0.0168	0.0268
	LMS	Bias	0.2239	0.1566	0.1329	0.2327
		MSE	0.0748	0.0343	0.0397	0.0876
n =100	OLS	Bias	0.1253	0.1307	0.0826	0.0697
		MSE	0.0229	0.0217	0.0089	0.0092
	LAV	Bias	0.1559	0.1067	0.0858	0.0728
		MSE	0.0395	0.0184	0.0108	0.0080
	M-H	Bias	0.1308	0.1307	0.0742	0.0696
		MSE	0.0255	0.0230	0.0080	0.0090
	M-T	Bias	0.1330	0.1255	0.0777	0.0705
		MSE	0.0266	0.0217	0.0086	0.0091
	LTS	Bias	0.1249	0.1651	0.0681	0.0894
		MSE	0.0290	0.0333	0.0070	0.0102
	LMS	Bias	0.2392	0.2121	0.2529	0.1743
		MSE	0.0761	0.0825	0.0919	0.0394

Source: Calculations based on simulation data

Table (A.4.4) Performances of OLS and Robust Methods of Logistic Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n = 10	OLS	Bias	0.6008	0.5992	0.6534	0.6685
		MSE	0.5172	0.6685	0.6845	0.7979
	LAV	Bias	0.6979	0.7139	0.8887	0.8093
		MSE	0.8873	0.9214	1.1780	1.0937
	M-H	Bias	0.6727	0.5797	0.7264	0.7172
		MSE	0.8082	0.7134	0.8194	0.9156
	M-T	Bias	0.7320	0.6707	0.7471	0.8297
		MSE	0.8665	1.0874	0.8327	1.3582
	LTS	Bias	0.8595	0.5997	0.9042	0.9234
		MSE	1.1525	0.7990	1.1726	1.2751
	LMS	Bias	1.3391	0.7972	1.1474	1.3129
		MSE	2.3115	1.1169	1.4996	2.7096
n = 20	OLS	Bias	0.2985	0.4272	0.4600	0.3331
		MSE	0.1225	0.2685	0.2970	0.1586
	LAV	Bias	0.3014	0.4991	0.4726	0.4228
		MSE	0.1230	0.4735	0.4369	0.2700
	M-H	Bias	0.2786	0.4063	0.4641	0.3586
		MSE	0.1150	0.2573	0.3216	0.1796
	M-T	Bias	0.2655	0.4036	0.4596	0.3478
		MSE	0.1071	0.2581	0.3210	0.1744
	LTS	Bias	0.3326	0.5109	0.6293	0.4290
		MSE	0.1207	0.5581	0.5929	0.2601
	LMS	Bias	0.4507	0.6996	1.0963	0.9802
		MSE	0.2784	0.8511	1.6086	1.6227

Table (A.4.4) (Contd.)

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =30	OLS	Bias	0.2664	0.2581	0.3643	0.1759
		MSE	0.1078	0.2136	0.1852	0.0522
	LAV	Bias	0.2543	0.3313	0.3976	0.2949
		MSE	0.0914	0.2682	0.1793	0.1186
	M-H	Bias	0.2501	0.2630	0.3493	0.1980
		MSE	0.0871	0.2155	0.1685	0.0577
	M-T	Bias	0.2459	0.2625	0.3455	0.1945
		MSE	0.0787	0.2153	0.1732	0.0555
	LTS	Bias	0.2799	0.3272	0.4033	0.4543
		MSE	0.1065	0.1966	0.1945	0.2414
	LMS	Bias	0.5637	0.9437	0.5797	0.3814
		MSE	0.3630	1.1606	0.4564	0.2530
n =50	OLS	Bias	0.1597	0.2760	0.2470	0.3024
		MSE	0.0493	0.1430	0.0891	0.1107
	LAV	Bias	0.2593	0.3333	0.2142	0.2767
		MSE	0.0869	0.1754	0.0626	0.1248
	M-H	Bias	0.1625	0.2748	0.2582	0.2643
		MSE	0.0448	0.1357	0.1042	0.0886
	M-T	Bias	0.1554	0.2684	0.2747	0.2420
		MSE	0.0403	0.1312	0.1154	0.0813
	LTS	Bias	0.2061	0.2707	0.2730	0.2834
		MSE	0.0819	0.1287	0.1338	0.1440
	LMS	Bias	0.5312	0.5433	0.6406	0.4673
		MSE	0.3731	0.5662	0.8952	0.3356
n =80	OLS	Bias	0.1427	0.1660	0.2487	0.1713
		MSE	0.0305	0.0381	0.0849	0.0392
	LAV	Bias	0.1856	0.1278	0.2423	0.1935
		MSE	0.0646	0.0230	0.1177	0.0472
	M-H	Bias	0.1496	0.1269	0.2279	0.1816
		MSE	0.0329	0.0248	0.0833	0.0421
	M-T	Bias	0.1524	0.1208	0.2276	0.1701
		MSE	0.0328	0.0219	0.0862	0.0384
	LTS	Bias	0.1775	0.1307	0.2478	0.1787
		MSE	0.0520	0.0251	0.1040	0.0510
	LMS	Bias	0.3531	0.2450	0.2971	0.3855
		MSE	0.1638	0.0851	0.1457	0.1698
n=100	OLS	Bias	0.1248	0.1285	0.1792	0.1459
		MSE	0.0185	0.0285	0.0525	0.0304
	LAV	Bias	0.1408	0.0968	0.1893	0.2056
		MSE	0.0289	0.0160	0.0593	0.0569
	M-H	Bias	0.1349	0.0985	0.1723	0.1672
		MSE	0.0228	0.0192	0.0559	0.0364
	M-T	Bias	0.1412	0.0949	0.1700	0.1670
		MSE	0.0249	0.0163	0.0568	0.0373
	LTS	Bias	0.1737	0.1063	0.2237	0.2346
		MSE	0.0422	0.0196	0.0996	0.0734
	LMS	Bias	0.3532	0.2339	0.2436	0.3622
		MSE	0.1769	0.0805	0.1291	0.1676

Source: Calculations based on simulation data

Table (A.4.5) Performances of OLS and Robust Methods of Exponential Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =10	OLS	Bias	0.8869	0.4223	0.1918	0.2208
		MSE	0.8400	0.3801	0.0503	0.0698
	LAV	Bias	0.8068	0.4030	0.1430	0.1966
		MSE	0.7369	0.3578	0.0402	0.0632
	M-H	Bias	0.7993	0.4430	0.1904	0.2118
		MSE	0.6795	0.4550	0.0454	0.0675
	M-T	Bias	0.7601	0.4563	0.2263	0.2151
		MSE	0.6432	0.4974	0.0705	0.0734
	LTS	Bias	0.6763	0.5282	0.2992	0.2817
		MSE	0.5482	0.7216	0.1790	0.2005
	LMS	Bias	0.7246	0.7137	0.3599	0.4490
		MSE	0.6232	0.8143	0.1961	0.3464
n =20	OLS	Bias	0.9875	0.1521	0.2115	0.1170
		MSE	1.0145	0.0323	0.0713	0.0157
	LAV	Bias	0.7316	0.1719	0.1408	0.1462
		MSE	0.6027	0.0533	0.0362	0.0288
	M-H	Bias	0.8483	0.1322	0.1651	0.1312
		MSE	0.7599	0.0231	0.0450	0.0207
	M-T	Bias	0.8008	0.1512	0.1453	0.1400
		MSE	0.6883	0.0425	0.0403	0.0273
	LTS	Bias	0.7500	0.2065	0.1719	0.1410
		MSE	0.6225	0.0678	0.0560	0.0298
	LMS	Bias	0.4050	0.1926	0.2943	0.2452
		MSE	0.1790	0.0509	0.1541	0.1608
n =30	OLS	Bias	0.9582	0.1562	0.1160	0.1363
		MSE	0.9524	0.0279	0.0216	0.0311
	LAV	Bias	0.7201	0.1580	0.1097	0.1080
		MSE	0.5627	0.0354	0.0169	0.0189
	M-H	Bias	0.8055	0.1222	0.1107	0.1184
		MSE	0.7929	0.0246	0.0208	0.0207
	M-T	Bias	0.7354	0.1253	0.1247	0.0962
		MSE	0.5590	0.0254	0.0232	0.0136
	LTS	Bias	0.7095	0.1063	0.0962	0.1040
		MSE	0.5175	0.0238	0.0126	0.0150
	LMS	Bias	0.4909	0.1890	0.1376	0.1206
		MSE	0.2991	0.0686	0.0308	0.0281
n=50	OLS	Bias	0.9627	0.1069	0.1066	0.1008
		MSE	0.9419	0.0137	0.0165	0.0132
	LAV	Bias	0.6673	0.0875	0.1037	0.0739
		MSE	0.4551	0.0127	0.0142	0.0090
	M-H	Bias	0.7972	0.0706	0.0975	0.0903
		MSE	0.6482	0.0073	0.0123	0.0099
	M-T	Bias	0.6945	0.0716	0.1055	0.0732
		MSE	0.4922	0.0081	0.0143	0.0077
	LTS	Bias	0.6937	0.0684	0.1143	0.1075
		MSE	0.4889	0.0068	0.0170	0.0161
	LMS	Bias	0.4415	0.0908	0.0847	0.1069
		MSE	0.2040	0.0117	0.0140	0.0188

Table (A.4.5) (Contd.)

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =80	OLS	Bias	0.9860	0.0661	0.0961	0.0642
		MSE	0.9849	0.0060	0.0132	0.0046
	LAV	Bias	0.6673	0.0854	0.0918	0.0806
		MSE	0.4537	0.0134	0.0118	0.0091
	M-H	Bias	0.8195	0.0667	0.0769	0.0513
		MSE	0.6854	0.0063	0.0097	0.0030
	M-T	Bias	0.6996	0.0743	0.0823	0.0474
		MSE	0.4996	0.0083	0.0093	0.0037
	LTS	Bias	0.7229	0.0657	0.0987	0.0494
		MSE	0.5319	0.0064	0.0175	0.0035
	LMS	Bias	0.4021	0.0549	0.0550	0.0742
		MSE	0.1784	0.0052	0.0055	0.0101
n =100	OLS	Bias	0.9979	0.0574	0.0942	0.0614
		MSE	1.0088	0.0052	0.0132	0.0061
	LAV	Bias	0.6812	0.0877	0.1129	0.0625
		MSE	0.4737	0.0142	0.0162	0.0055
	M-H	Bias	0.8355	0.0667	0.0767	0.0415
		MSE	0.7119	0.0058	0.0097	0.0022
	M-T	Bias	0.7175	0.0704	0.0853	0.0385
		MSE	0.5241	0.0080	0.0102	0.0025
	LTS	Bias	0.7280	0.0683	0.0822	0.0420
		MSE	0.5378	0.0066	0.0118	0.0026
	LMS	Bias	0.4288	0.0730	0.1042	0.0431
		MSE	0.1981	0.0077	0.0156	0.0032

Source: Calculations based on simulation data

Table (A.4.6) Performances of OLS and Robust Methods of Cauchy Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n = 10	OLS	Bias	2.6781	1.7431	2.8656	2.4600
		MSE	22.0394	8.3795	19.2347	17.3125
	LAV	Bias	0.9250	1.1730	1.2433	1.2245
		MSE	1.3160	2.7178	3.2583	2.4527
	M-H	Bias	1.1229	1.0392	1.7524	1.2515
		MSE	2.0275	1.6833	9.6928	2.2957
	M-T	Bias	1.0268	0.7001	1.6003	1.0101
		MSE	1.7355	0.8209	10.472	1.5833
	LTS	Bias	0.9644	0.7875	1.2490	1.2296
		MSE	1.4443	1.0583	3.2550	2.6758
	LMS	Bias	0.7397	0.8135	0.9280	1.2161
		MSE	1.0858	1.2536	1.5602	3.3787
n = 20	OLS	Bias	2.2200	1.7870	1.6756	1.2306
		MSE	8.3993	6.6070	3.9963	3.4722
	LAV	Bias	0.3561	0.2274	0.5412	0.3553
		MSE	0.2330	0.0982	0.4833	0.1745
	M-H	Bias	0.3121	0.2656	0.5713	0.3249
		MSE	0.2045	0.1198	0.6117	0.1842
	M-T	Bias	0.3683	0.3388	0.5146	0.3765
		MSE	0.1670	0.1502	0.4428	0.1782
	LTS	Bias	0.3011	0.3056	0.5078	0.2908
		MSE	0.1566	0.1157	0.4263	0.1383
	LMS	Bias	0.3995	0.2609	0.4618	0.4490
		MSE	0.2390	0.1140	0.3158	0.2748

Table (A.4.6) (Contd.)

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =30	OLS	Bias	4.3323	5.8791	1.5400	1.5029
		MSE	97.9254	235.9916	5.4072	9.3752
	LAV	Bias	0.2233	0.2265	0.4099	0.1742
		MSE	0.0657	0.0824	0.2798	0.0398
	M-H	Bias	0.2551	0.1866	0.5672	0.2752
		MSE	0.0884	0.0459	0.4956	0.1044
	M-T	Bias	0.1766	0.1929	0.5003	0.3177
		MSE	0.0427	0.0554	0.3877	0.1151
	LTS	Bias	0.2158	0.2688	0.6248	0.3137
		MSE	0.0806	0.1131	0.6523	0.1682
	LMS	Bias	0.3576	0.3978	0.5022	0.4450
		MSE	0.2412	0.2607	0.3258	0.3863
n =50	OLS	Bias	2.4380	3.3922	1.2825	2.1126
		MSE	23.9028	67.9288	4.1653	20.6503
	LAV	Bias	0.1430	0.1777	0.2044	0.2255
		MSE	0.0331	0.0488	0.0637	0.0804
	M-H	Bias	0.1170	0.1531	0.2824	0.2241
		MSE	0.0227	0.0377	0.1051	0.0967
	M-T	Bias	0.1387	0.1916	0.2494	0.2393
		MSE	0.0314	0.0454	0.0846	0.0848
	LTS	Bias	0.1595	0.2127	0.3489	0.2284
		MSE	0.0418	0.0522	0.1596	0.0937
	LMS	Bias	0.2338	0.2931	0.2854	0.2233
		MSE	0.0738	0.1162	0.1304	0.1053
n =80	OLS	Bias	1.5801	2.3042	1.0224	1.4368
		MSE	8.4181	31.4442	3.7895	7.2829
	LAV	Bias	0.1200	0.1266	0.0915	0.1574
		MSE	0.0197	0.0213	0.0193	0.0388
	M-H	Bias	0.1139	0.1350	0.1586	0.1872
		MSE	0.0231	0.0249	0.0342	0.0622
	M-T	Bias	0.1237	0.2073	0.1480	0.2121
		MSE	0.0361	0.0481	0.0401	0.0670
	LTS	Bias	0.1428	0.2518	0.1406	0.2555
		MSE	0.0561	0.0658	0.0374	0.0884
	LMS	Bias	0.3140	0.3346	0.1615	0.1461
		MSE	0.1215	0.1532	0.0465	0.0392
n=100	OLS	Bias	1.5374	2.0641	1.0080	1.0555
		MSE	6.4297	22.3868	2.9185	3.0349
	LAV	Bias	0.1039	0.1206	0.0469	0.1557
		MSE	0.0154	0.0180	0.0073	0.0350
	M-H	Bias	0.0824	0.1186	0.1001	0.1559
		MSE	0.0112	0.0188	0.0162	0.0421
	M-T	Bias	0.1182	0.1855	0.1105	0.1698
		MSE	0.0206	0.0421	0.0205	0.0457
	LTS	Bias	0.1355	0.2186	0.1286	0.1704
		MSE	0.0280	0.0627	0.0240	0.0406
	LMS	Bias	0.2323	0.1966	0.1657	0.2029
		MSE	0.0838	0.0522	0.0391	0.0536

Source: Calculations based on simulation data

Table (A.4.7) Performances of OLS and Robust Methods of Gamma Distribution

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =10	OLS	Bias	2.3935	0.8216	0.6492	0.5599
		MSE	7.3810	1.2113	0.7848	0.4866
	LAV	Bias	2.3062	1.1021	0.6106	0.7862
		MSE	7.3690	2.8982	0.5796	1.0408
	M-H	Bias	2.2569	0.8284	0.6746	0.5358
		MSE	6.8373	1.2261	0.8309	0.4439
	M-T	Bias	2.0283	0.9482	0.6500	0.5055
		MSE	5.6560	2.1358	0.7479	0.3862
	LTS	Bias	2.1743	1.0324	0.9422	0.6483
		MSE	6.5913	2.3942	2.3276	0.5631
	LMS	Bias	1.5070	0.8757	0.6243	0.4788
		MSE	2.7178	1.8015	0.4890	0.3060
n =20	OLS	Bias	2.1462	0.3684	0.3527	0.5616
		MSE	4.9348	0.2569	0.2105	0.4424
	LAV	Bias	1.5007	0.4235	0.3415	0.5842
		MSE	2.4972	0.2949	0.2167	0.4840
	M-H	Bias	1.9125	0.3363	0.3164	0.4549
		MSE	4.0267	0.2019	0.1985	0.2724
	M-T	Bias	1.7407	0.3463	0.3121	0.4594
		MSE	3.5293	0.2229	0.1919	0.2590
	LTS	Bias	1.8206	0.4639	0.4945	0.5835
		MSE	3.8613	0.3158	0.6900	0.4384
	LMS	Bias	1.0030	0.3097	0.4122	0.5455
		MSE	1.3840	0.1284	0.3720	0.5856
n =30	OLS	Bias	2.0303	0.3283	0.1996	0.4099
		MSE	4.2804	0.1445	0.0567	0.3006
	LAV	Bias	1.4922	0.2176	0.1931	0.3677
		MSE	2.4788	0.0725	0.0701	0.2350
	M-H	Bias	1.7707	0.2070	0.1685	0.3525
		MSE	3.2870	0.0587	0.0403	0.1965
	M-T	Bias	1.5254	0.1952	0.1656	0.2685
		MSE	2.5688	0.0571	0.0462	0.1467
	LTS	Bias	1.5158	0.2370	0.2220	0.2505
		MSE	2.3865	0.0768	0.0645	0.1009
	LMS	Bias	1.1215	0.3789	0.3694	0.5825
		MSE	1.7866	0.3366	0.2039	0.5754
n=50	OLS	Bias	1.9751	0.2472	0.1459	0.2949
		MSE	4.0195	0.0927	0.0295	0.1209
	LAV	Bias	1.4359	0.1300	0.1245	0.3035
		MSE	2.2758	0.0249	0.0304	0.1276
	M-H	Bias	1.6864	0.1899	0.0695	0.2437
		MSE	2.9755	0.0484	0.0103	0.0785
	M-T	Bias	1.4592	0.1921	0.0731	0.2034
		MSE	2.3139	0.0459	0.0086	0.0483
	LTS	Bias	1.4650	0.1921	0.0969	0.2202
		MSE	2.2362	0.0463	0.0173	0.0793
	LMS	Bias	0.9384	0.1761	0.2219	0.3415
		MSE	1.0656	0.0455	0.0657	0.3257

Table (A.4.7) (Contd.)

Sample Size	Estimation Method		β_0	β_1	β_2	β_3
n =80	OLS	Bias	1.9789	0.1801	0.1203	0.1844
		MSE	3.9942	0.0428	0.0223	0.0474
	LAV	Bias	1.3392	0.1363	0.1095	0.2038
		MSE	1.8845	0.0277	0.0202	0.0717
	M-H	Bias	1.6679	0.1108	0.0727	0.1702
		MSE	2.8548	0.0245	0.0174	0.0448
	M-T	Bias	1.4060	0.1706	0.0934	0.1567
		MSE	2.0780	0.0372	0.0183	0.0435
	LTS	Bias	1.4632	0.1718	0.1490	0.1665
		MSE	2.1798	0.0401	0.0340	0.0429
	LMS	Bias	0.8983	0.1404	0.2250	0.1698
		MSE	0.8930	0.0288	0.0967	0.0487
n =100	OLS	Bias	1.9481	0.1523	0.1187	0.1160
		MSE	3.8355	0.0294	0.0225	0.0207
	LAV	Bias	1.2564	0.0945	0.0857	0.1388
		MSE	1.6309	0.0162	0.0115	0.0342
	M-H	Bias	1.6520	0.0876	0.0825	0.1111
		MSE	2.7621	0.0117	0.0142	0.0183
	M-T	Bias	1.3776	0.1100	0.0821	0.1111
		MSE	1.9417	0.0149	0.0130	0.0227
	LTS	Bias	1.4092	0.1065	0.1148	0.1262
		MSE	1.9995	0.0141	0.0183	0.0282
	LMS	Bias	0.7951	0.1509	0.1847	0.1000
		MSE	0.6746	0.0380	0.0673	0.0146

Source: Calculations based on simulation data

Table (A.4.8) Simulated Data Set

Case	x_1	x_2	x_3	y
1	0.92	0.04	0	10.05
2	-0.11	1.23	0	9.77
3	1.39	1.2	1	12.45
4	0.13	-1.12	0	8.66
5	0.25	-0.86	0	8.28
6	-1.44	0.26	0	7.96
7	2.18	0.3	1	12.65
8	-0.48	-1.65	0	7.28
9	-1.75	0.59	1	9.36
10	1.64	1.17	1	13.28
11	-0.2	0.55	0	9.05
12	0.87	0.7	1	11.31
13	-1.39	2.01	0	9.33
14	0.18	-0.7	1	9.95
15	-1.21	-1.29	0	6.89
16	-0.56	1.15	0	9.46
17	0.94	-1.72	1	9.33
18	-1.46	1.13	1	9.38
19	1.2	-1.16	0	8.9
20	-1.29	0.15	0	8.62
21	-1.7	1.04	0	9.01
22	0.17	-0.1	0	9.17
23	-1.09	-0.3	0	7.43
24	-1.27	0.15	0	7.93
25	0.06	-2.92	1	14.16
26	0.45	-1.97	1	14.54
27	8.89	0.22	0	9.52
28	6.87	-1.15	1	7.19
29	10.96	8.26	0	28.35
30	10.37	9.21	1	29.28

Source: Simulation Data Sets

Table (A. 4.9) Production and Export of Maize

Year	Production of Maize (X) (in thousand metric ton)	Export of Maize(Y) (in thousand metric ton)
1976	57.4	6.2
1977	74.1	10.3
1978	75.6	10.1
1979	124.5	20.3
1980	163.7	9.7
1981	202.8	22.2
1982	235.4	33.5
1983	304.7	17.6
1984	298.5	30
1985	294.1	21.9
1986	280.9	15.5
1987	220.1	20.8
1988	189.9	0.9
1989	190.7	13.9
1990	184.1	20
1991	188.1	41
1992	205	44.4
1993	201.4	40.4
1994	251.5	70.4
1995	270.4	62
1996	281.4	102.5
1997	303.4	50
1998	297.9	174.3
1999	343.6	88.8
2000	358.9	147.9
2001	524	90.1
2002	593.4	219.2
2003	692.9	150.9
2004	771.1	254.8
2005	903.5	90
2006	1015.8	183.3
2007	1128.1	156

Source: Central Statistical Organization (CSO), Ministry of National Planning, *Statistical Yearbook* (various issues)

Table (A.4.10) Summary of Various Regression Models Fitted to the Production and Export of Maize Data

Model	β_0	β_1	R^2	Standard Error of Estimates	*F*- Value	D.W.
Original[a]	2.700 (0.198)	0.190*** (6.158)	0.558	46.7654	37.943***	1.77
New[b]	-1.549*** (2.954)	0.559*** (6.028)	0.548	0.3695	36.330***	1.28
Cochrane-Orcutt	-	-	$\hat{\rho}$ =0.36	-	-	-
Transformed[c]	-0.948* (1.739)	0.546*** (3.688)	0.32	0.3503	13.604***	2.27

Note: (1) a. Original Model

Predictors: (Constant), *PROD*

Dependent Variable: *EXP*

b. New Model

Predictors:(Constant), ln*PROD*

Dependent Variable: ln*EXP*

c. Transformed Model

Predictors:(Constant), $PROD^*$

Dependent Variable: EXP^*

(2) Absolute values of *t* statistics in parentheses.

(3) Significant at *** 1%, ** 5%, * 10%

Source: Appendix Table (A.4.9)

Table (A.4.11) Summary of Curvilinear Regression Model Fitted to the Production and Export of Maize Data

Model	β_0	β_1	β_{11}	R^2	Standard Error of Estimates	*F*- Value	D.W.
Centered [d]	88.6880*** (8.430)	0.3008*** (5.950)	-0.0003** (2.641)	0.644	42.707	26.235***	2.197

Note: (1) d. Predictors: (Constant), (*PROD- mean*), $(PROD- mean)^2$

Dependent Variable: *EXP*

(2) Absolute values of *t* statistics in parentheses.

(3) Significant at *** 1%, ** 5%, * 10%

Source: Appendix Table (A.4.9)

Table (A.4.12) Maternal Mortality Ratio for Selected Countries

Country	Maternal Mortality ratio	Contracep-tive prevalence rate	Adult literacy rate	Birth attended by skilled health personnel	Total fertility rate	Life expecta-ncy	Physicians	Public expendit-ure on education
Canada	7	75	99	98	1.5	80.3	214	6.8
Sweden	3	78	99	100	1.7	80.5	328	7.7
Japan	6	56	99	100	1.3	82.3	198	6.3
Netherlands	6	79	99	100	1.7	79.2	315	5.7
France	8	75	99	99	1.9	80.2	337	8.2
United States	11	76	99	99	2	77.9	256	6.9
Austria	4	51	99	100	1.4	79.4	338	7.8
United Kingdom	8	84	99	99	1.7	79	230	7
New Zealand	9	75	99	100	2	79.8	237	6.5
Germany	4	75	99	100	1.3	79.1	337	8.2
Singapore	14	62	92.5	100	1.4	79.4	140	1.3
Korea	14	81	99	100	1.2	77.9	157	2.9
Kuwait	4	50	93.3	98	2.3	77.3	153	2.2
Malaysia	62	55	88.7	97	2.9	73.7	70	2.2
Thai	110	79	92.6	99	1.8	69.6	37	2.3
China	45	87	90.9	97	1.7	72.5	106	1.8
Philippines	230	49	92.6	60	3.5	71	58	1.4
Paraguay	150	73	93.5	77	3.5	71.3	111	2.6
Georgia	66	47	100	92	1.5	70.7	409	1.5
Jamaica	26	69	79.9	97	2.6	72.2	85	2.8
EL Salvador	170	67	80.6	92	2.9	71.3	124	3.5
Viet Nam	150	77	90.3	85	2.3	73.7	53	1.5
Indonesia	420	57	90.4	72	2.4	69.7	13	1
Mongolia	46	69	97.8	97	2.1	65.9	263	4
South Africa	400	60	82.4	92	2.8	50.8	77	3.5
Morocco	240	63	52.3	63	2.5	70.4	51	1.7
India	450	47	61	43	3.1	63.7	60	0.9
Cambodia	590	24	73.6	32	3.6	58	16	1.7
Myanmar	380	34	89.9	57	2.2	60.8	36	0.3
Bhutan	440	31	47	37	2.9	64.7	5	3
Pakistan	320	28	49.9	31	4	64.6	74	0.4
Bangladesh	570	58	47.5	13	3.2	63.1	26	0.9
Nepal	830	38	48.6	11	3.7	62.6	21	1.5
Uganda	550	20	66.8	39	6.7	49.7	8	2.5

Source: Human Development Report by UNDP (2007/2008)

Table (A.4.13) Summary of Original Model Fitted to the Maternal Mortality Data

Coefficients	Value of Coefficients	Standard Errors of Coefficients	*t* Statistics	Significance of *t*	Collinearity Statistics	
					Tolerance	VIF
Constant	5.802***	0.833	6.968	0.000	-	-
CPR	0.007	0.004	1.599	0.122	0.374	2.672
ALR	0.002	0.006	0.286	0.777	0.225	4.451
BABSHP	-0.010**	0.004	-2.449	0.021	0.161	6.219
TFR	-0.017	0.079	-0.214	0.832	0.323	3.095
LE	-0.046***	0.010	-4.568	0.000	0.320	3.129
PHY	-0.001	0.001	-1.462	0.156	0.288	3.475
PEOH	-0.084**	0.032	-2.648	0.014	0.366	2.734

Note: (1) Dependent variable: lnMMR
(2) Significant at *** 1%, ** 5%, * 10%

Source: Appendix Table (A.4.12)

Table (A.4.14) Summary of New Model Fitted to the Maternal Mortality Data

Coefficients	Value of Coefficients	Standard Errors of Coefficients	*t* Statistics	Significance of *t*	Collinearity Statistics	
					Tolerance	VIF
Constant	5.863***	0.513	11.437	0.000	-	-
LE	-0.044***	0.009	-4.927	0.000	0.452	2.215
PEOH	-0.110***	0.026	-4.263	0.000	0.588	1.702
BABSHP	-0.008***	0.003	-3.118	0.004	0.474	2.108

Note: (1) Dependent variable: lnMMR
(2) Significant at *** 1%, ** 5%, * 10%

Source: Appendix Table (A.4.12)

Table (A.4.15) Performance of Models Fitted to the Maternal Mortality Data

Model	Adjusted R^2	Standard Error of Estimates	F- Value	Significance of F
Original	0.882	0.2802	36.196	0.000
New	0.873	0.2904	76.645	0.000

Source: Appendix Table (A.4.12)

Table (A.5.1) Daily Average Number of Defects per Truck

Day	No of Defect
November 4	1.2
5	1.5
6	1.54
7	2.7
8	1.95
9	2.4
10	3.44
11	2.83
12	1.76
13	2
14	2.09
15	1.89
16	1.8
17	1.25
18	1.58
19	2.25
20	2.5
21	2.05
22	1.46
23	1.54
24	1.42
25	1.57
26	1.4
27	1.51
28	1.08
29	1.27
30	1.18
December 1	1.39
2	1.42
3	2.08
4	1.85
5	1.82
6	2.07
7	2.32
8	1.23
9	2.91
10	1.77
11	1.61
12	1.25
13	1.15
14	1.37
15	1.79
16	1.68
17	1.78
18	1.84

Source: Wei,(1990), *"Time Series Analysis Univariate and Multivariate Methods"*

Table (A.5.2) The *sacf* and *spacf* of the Truck Series

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	0.43	0.26	0.14	0.08	-0.09	-0.07	-0.21	-0.11	-0.05	-0.01
Std Err	0.15	0.15	0.17	0.18	0.19	0.19	0.19	0.19	0.19	0.19
$\hat{\phi}_{kk}$	0.43	0.09	0.00	0.00	-0.16	0.00	-0.18	0.07	0.05	0.01
Std Err	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15	0.15

Source: Appendix Table (A.5.1)

Table (A.5.3) Residual *acf* and *pacf* of the ARMA(1,0) Model

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-.044	.077	.025	.086	-.128	.048	-.205	-.023	-.014	.040
Std Err	.144	.143	.141	.139	.138	.136	.134	.132	.130	.129
$\hat{\phi}_{kk}$	-.044	.075	.032	.084	-.126	.025	-.195	-.041	.028	.044
Std Err	.149	.149	.149	.149	.149	.149	.149	.149	.149	.149

Source: Appendix Table (A.5.1)

Table (A.5.4) Residual *acf* and *pacf* of the RAB-AR(1) Model

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	-0.057	0.068	0.089	0.075	-0.098	0.141	-0.191	-0.052
Std Err	0.146	0.144	0.142	0.141	0.139	0.137	0.135	0.133
$\hat{\phi}_{kk}$	-0.057	0.065	0.097	0.082	-0.103	0.113	-0.186	-0.076
Std Err	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151
k	9	10	11	12	13	14	15	16
$\hat{\rho}_k$	0.001	0.058	-0.075	0.054	0.016	0.044	-0.093	0.023
Std Err	0.132	0.130	0.128	0.126	0.124	0.122	0.120	0.118
$\hat{\phi}_{kk}$	0.009	0.081	-0.006	0.004	0.049	0.024	-0.121	-0.028
Std Err	0.151	0.151	0.151	0.151	0.151	0.151	0.151	0.151

Source: Appendix Table (A.5.1)

Table (A.5.5) Box-Ljung Statistic of the RAB-AR(1) Model

k	Box-Ljung Statistic	
	Value	Sig. *p* value
1	0.154	0.695
2	0.377	0.828
3	0.766	0.858
4	1.049	0.902
5	1.549	0.907
6	2.615	0.855
7	4.610	0.707
8	4.763	0.783
9	4.763	0.854
10	4.965	0.894
11	5.314	0.915
12	5.501	0.939
13	5.518	0.962
14	5.650	0.975
15	6.254	0.975
16	6.293	0.985

Source: Appendix Table (A.5.1)

Table (A.5.6) Export of Maize

Year	Export (Thousand metric tons)
1976	6.2
1977	10.3
1978	10.1
1979	20.3
1980	9.7
1981	22.2
1982	33.5
1983	17.6
1984	30
1985	21.9
1986	15.5
1987	20.8
1988	0.9
1989	13.9
1990	20
1991	41
1992	44.4
1993	40.4
1994	70.4
1995	62
1996	102.5
1997	50
1998	174.3
1999	88.8
2000	147.9
2001	90.1
2002	219.2

Table (A.5.6) Export of Maize (Contd.)

Year	Export (Thousand metric tons)
2003	150.9
2004	254.8
2005	90
2006	183.3
2007	156
2008	120.3

Source: Central Statistical Organization (CSO), Ministry of National Planning, *Statistical Yearbook* (various issues)

Table (A.5.7) The *sacf* and *spacf* for Natural Logarithms of the Export of Maize

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	.672	.661	.498	.468	.380	.281	.246	.242	.184	.116
Std Err	.166	.164	.161	.158	.156	.153	.150	.147	.144	.141
$\hat{\phi}_{kk}$	.672	.383	-.070	.050	.010	-.130	.033	.135	-.071	-.126
Std Err	.174	.174	.174	.174	.174	.174	.174	.174	.174	.174

Source: Appendix Table (A.5.6)

Table (A.5.8) The *sacf* and *spacf* for the Differenced Series of Natural Logarithms of the Export of Maize

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-.557	.256	-.169	.000	.090	-.112	-.070	.105	.009	-.112
Std Err	.169	.166	.163	.160	.158	.155	.152	.149	.145	.142
$\hat{\phi}_{kk}$	-.557	-.079	-.086	-.172	.025	-.051	-.273	-.065	.106	-.202
Std Err	.177	.177	.177	.177	.177	.177	.177	.177	.177	.177

Source: Appendix Table (A.5.6)

Table (A.5.9) Residual *acf* and *pacf* of the ARIMA(1,1,0) Model

k	1	2	3	4	5	6	7	8	9	10
$\hat{\rho}_k$	-.045	-.108	-.111	-.077	.077	-.195	-.136	.149	.016	-.171
Std Err	.169	.166	.163	.160	.158	.155	.152	.149	.145	.142
$\hat{\phi}_{kk}$	-.045	-.110	-.123	-.105	040	-.231	-.188	.084	-.058	-.274
Std Err	.177	.177	.177	.177	.177	.177	.177	.177	.177	.177

Source: Appendix Table (A.5.6)

Table (A.5.10) Residual *acf* and *pacf* of the RAB-ARIMA(1,1,0) Model

k	1	2	3	4	5	6	7	8
$\hat{\rho}_k$	-.032	-.127	-.103	-.083	.066	-.190	-.138	.154
Std Err	.171	.168	.165	.162	.159	.156	.153	.150
$\hat{\phi}_{kk}$	-.032	-.128	-.114	-.112	.028	-.232	-.187	.077
Std Err	0.180	0.180	0.180	0.180	0.180	0.180	0.180	0.180
k	9	10	11	12	13	14	15	16
$\hat{\rho}_k$	.027	-.172	.041	.056	.084	-.089	-.093	-.066
Std Err	.147	.143	.140	.136	.133	.129	.125	.121
$\hat{\phi}_{kk}$	-.054	-.267	.017	-.002	-.067	-.130	-.043	-.246
Std Err	0.180	0.180	0.180	0.180	0.180	0.180	0.180	0.180

Source: Appendix Table (A.5.6)

Table (A.5.11) Box-Ljung Statistic of the RAB-ARIMA(1,1,0) Model

k	Box-Ljung Statistic	
	Value	Sig. *p* value
1	0.035	0.851
2	0.607	0.738
3	0.994	0.803
4	1.252	0.870
5	1.424	0.922
6	2.894	0.822
7	3.711	0.812
8	4.762	0.783
9	4.795	0.852
10	6.234	0.795
11	6.318	0.851
12	6.487	0.890
13	6.891	0.908
14	7.366	0.920
15	7.917	0.927
16	8.218	0.942

Source: Appendix Table (A.5.6)

APPENDIX B

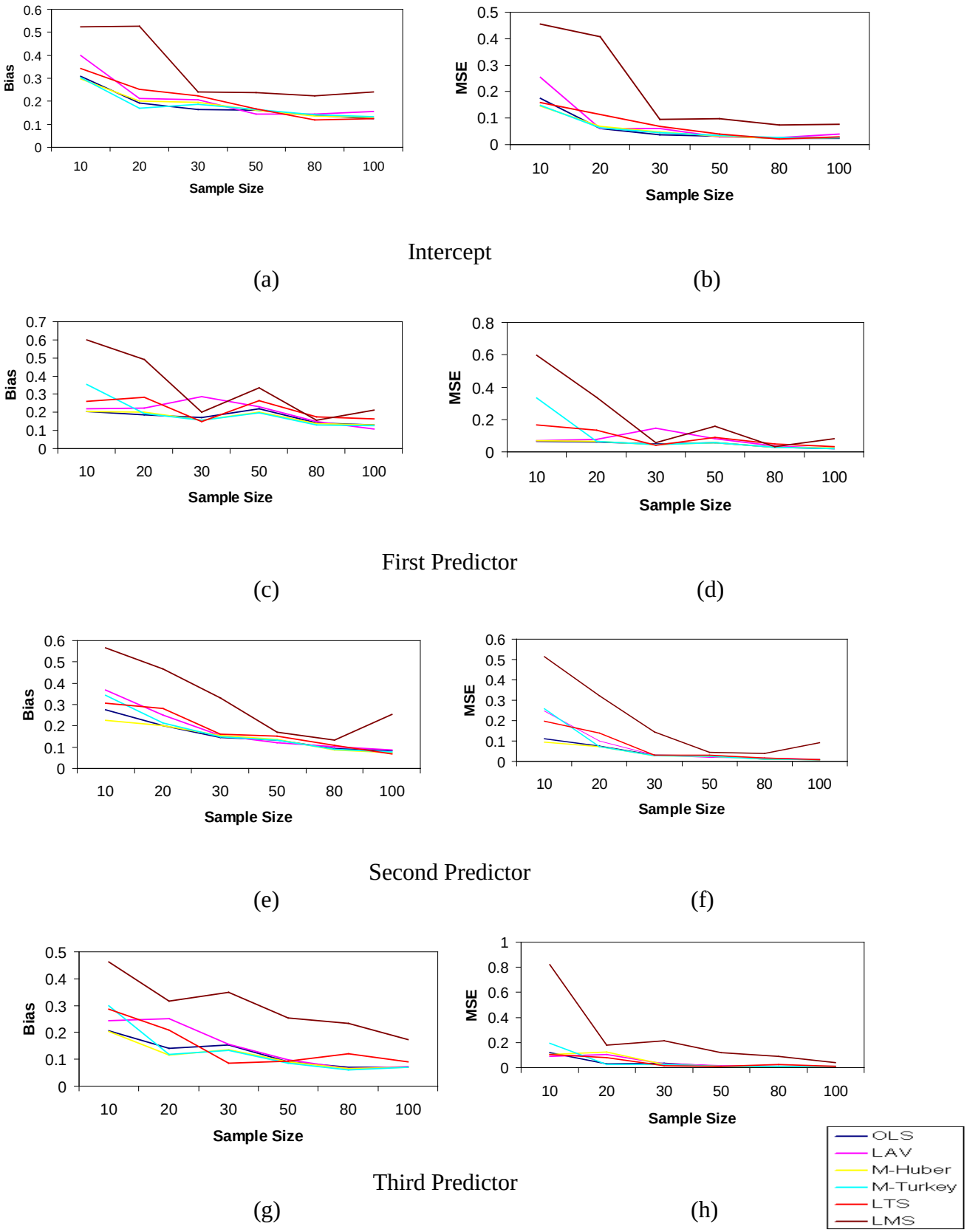
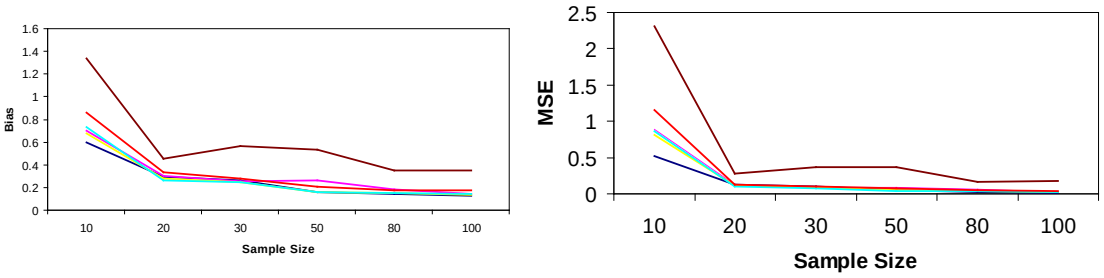
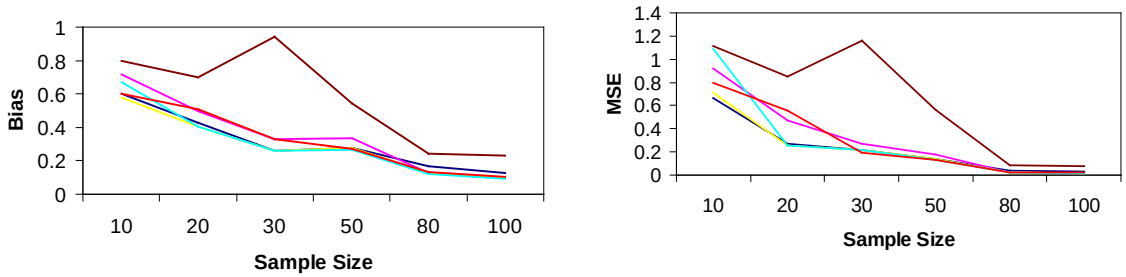


Figure B.4.1 Bias and MSE for 10 Simulations from Normal (0,1) Distribution
Source: Appendix Table (A.4.3)



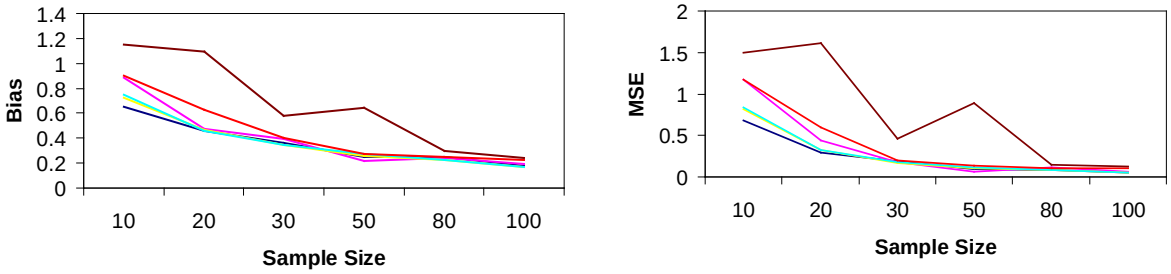
Intercept

(a) (b)



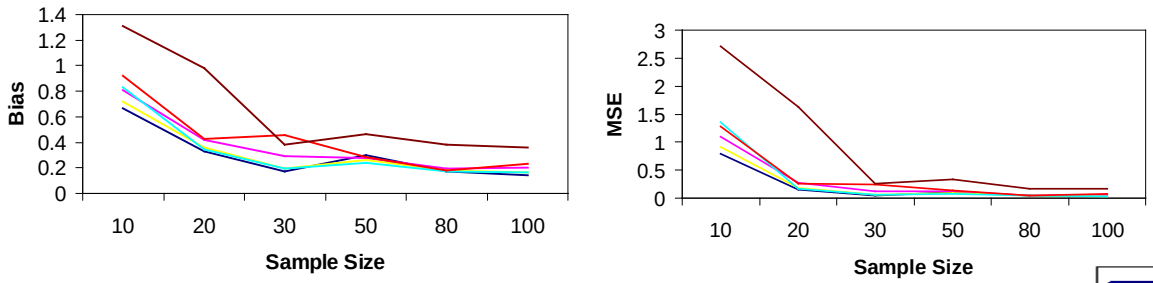
First Predictor

(c) (d)



Second Predictor

(e) (f)



Third Predictor

(g) (h)

Figure B.4.2 Bias and MSE for 10 Simulations from Logistic (0,1) Distribution
Source: Appendix Table (A.4.4)

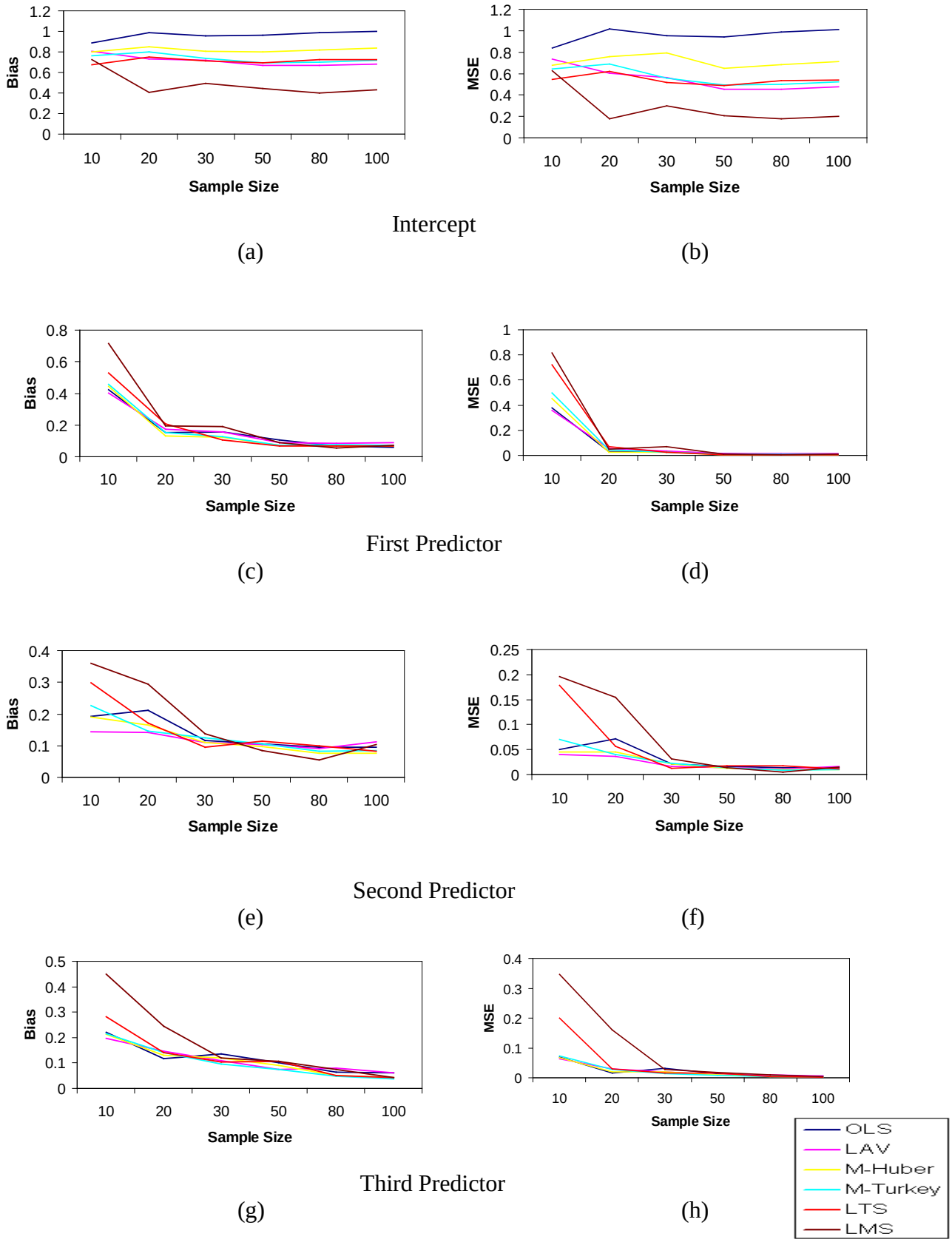
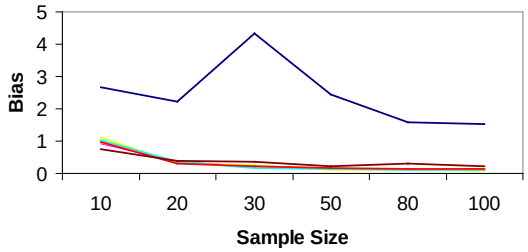
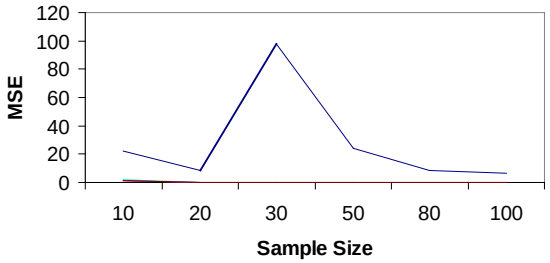


Figure B.4.3 Bias and MSE for 10 Simulations from Exponential (1) Distribution

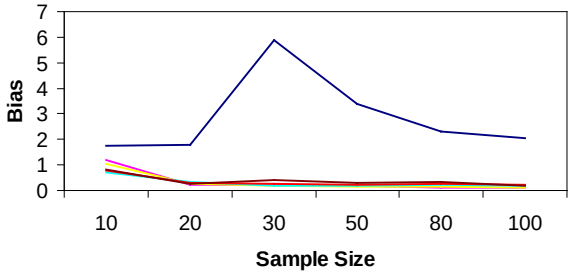
Source: Appendix Table (A.4.5)

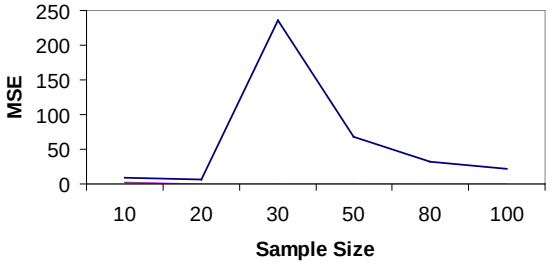




Intercept

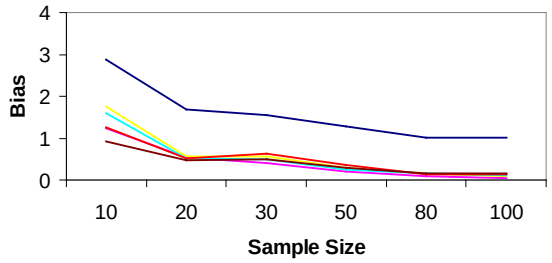
(a) (b)

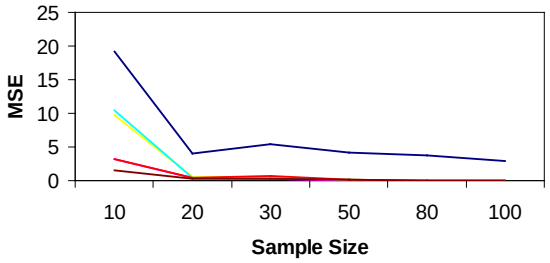




First Predictor

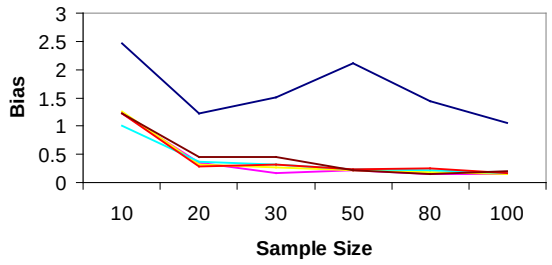
(c) (d)

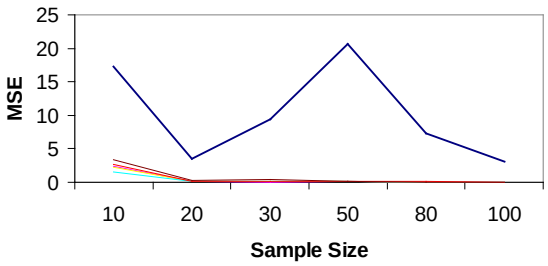




Second Predictor

(e) (f)





Third Predictor

(g) (h)

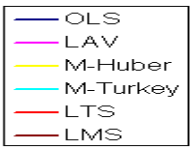


Figure B.4.4 Bias and MSE for 10 Simulations from Cauchy (0,1) Distribution
Source: Appendix Table (A.4.6)

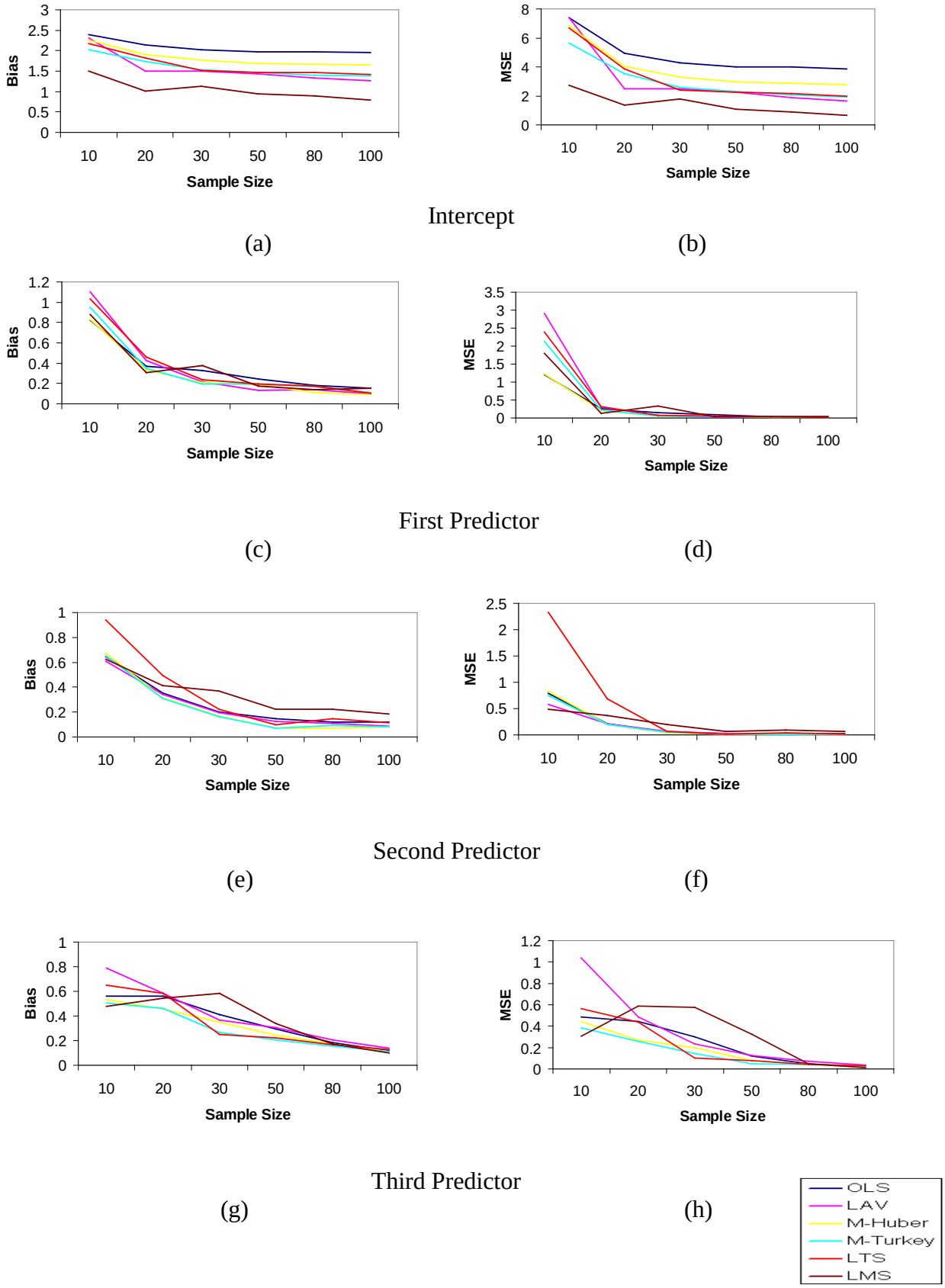
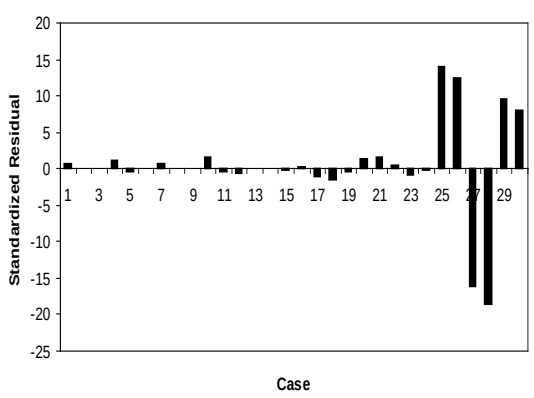
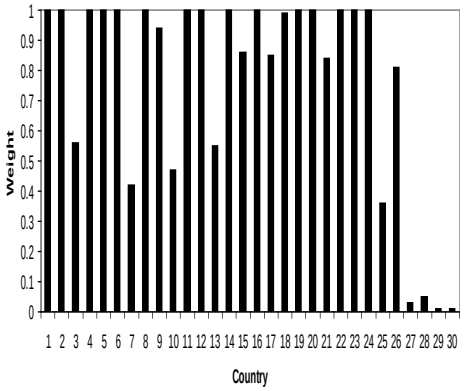


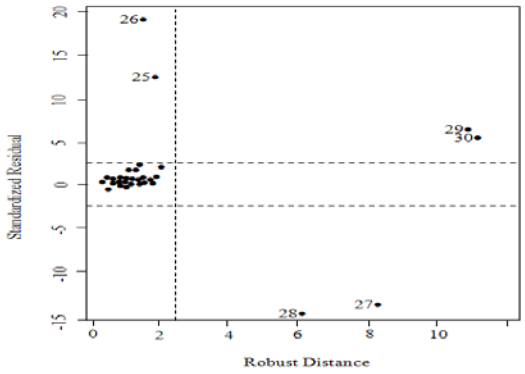
Figure B.4.5 Bias and MSE for 10 Simulations from Gamma (1,0.5) Distribution
Source: Appendix Table (A.4.7)



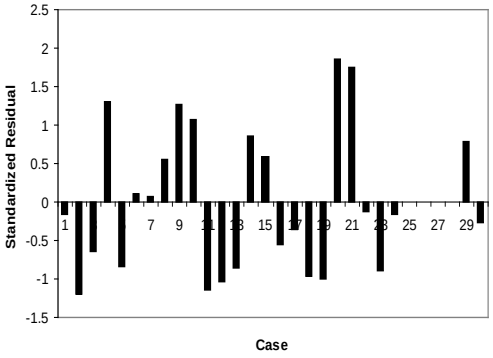
(a)



(b)



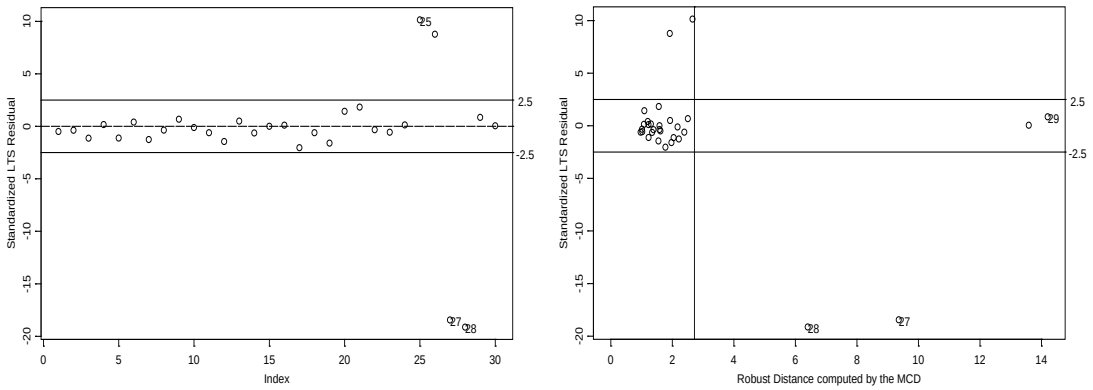
(c)



(d)

Figure B.4.6 **Simulated Data Set Using the RDL_1 Procedure: (a) plot of the standardized residuals; (b) plot of weights; (c) diagnostic plot and (d) least squares residuals without cases 25, 26, 27, and 28**

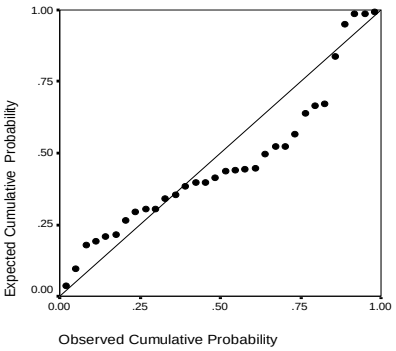
Source: **Appendix Table (A.4.8)**



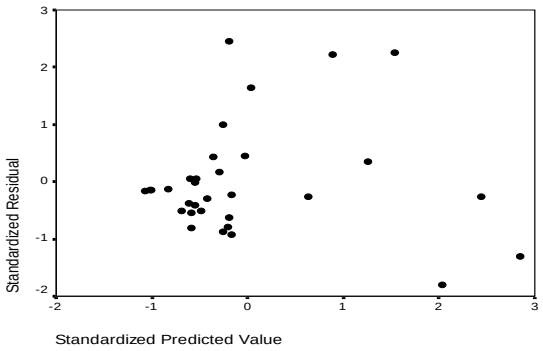
(a) (b)

Figure B.4.7 Simulated Data Set Using the LTS Robust Procedure: (a) plot of the standardized residuals; and (b) diagnostic plot

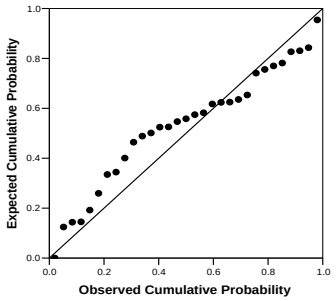
Source: Appendix Table (A.4.8)



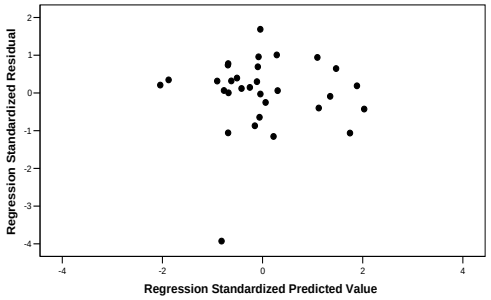
(a)



(b)



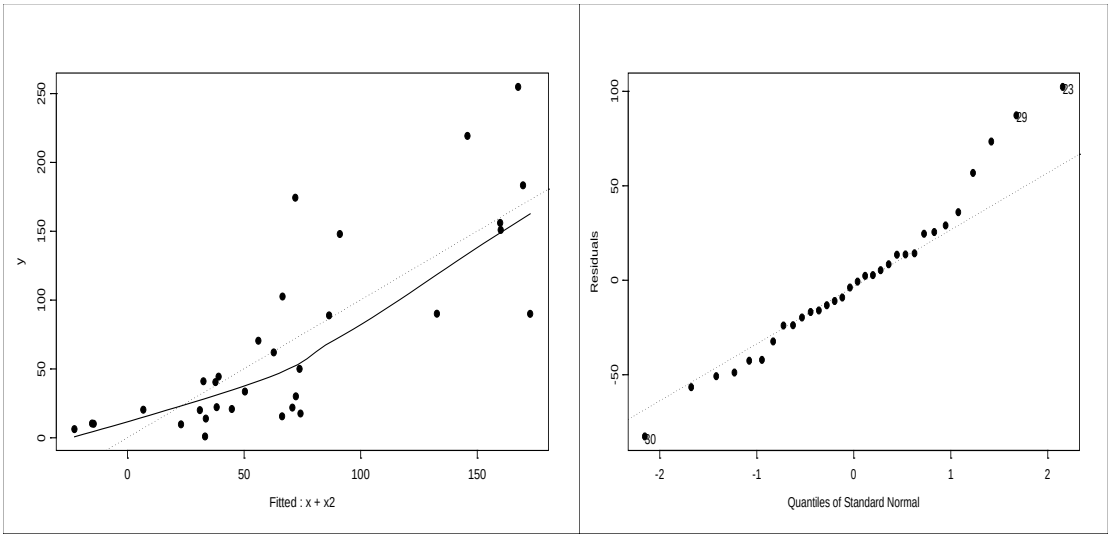
(c)



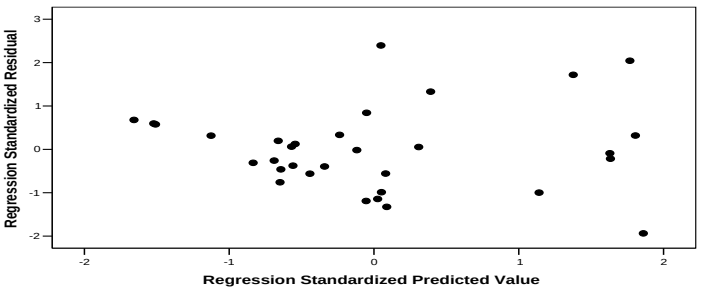
(d)

Figure B.4.8 **Production and Export Maize Data Set Using the OLS: (a) normal probability plot and (b) standardized residual versus predicted value of original model; (c) normal probability plot and (d) standardized residual versus predicted value of transformed model**

Source: **Appendix Table (A.4.9)**



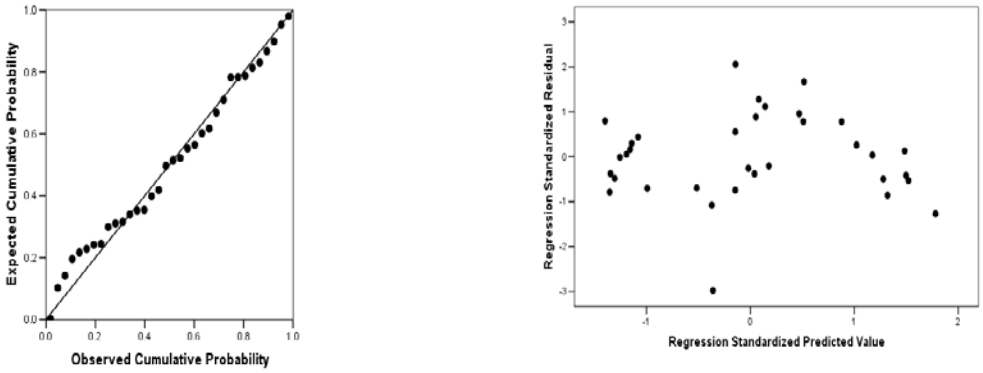
(a) (b)



(c)

Figure B.4.9 Curvilinear Model Fitted to the Production and Export Maize Data Set Using the OLS: (a) scatter plot with the fitted line; (b) quantiles standard normal plot; and (c) standardized residual versus predicted value

Source: Appendix Table (A.4.9)



(a) (b)

Figure B.4.10 **Maternal Mortality Data Set Using the OLS: (a) normal probability plot; and (b) standardized residual versus predicted value of new model**

Source: **Appendix Table (A.4.12)**

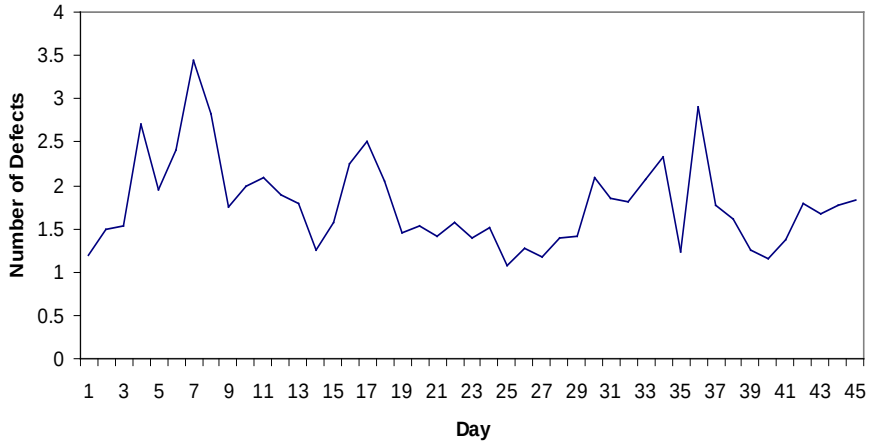


Figure B.5.1 Daily Average Number of Truck Manufacturing Defects
Source: Appendix Table (A.5.1)

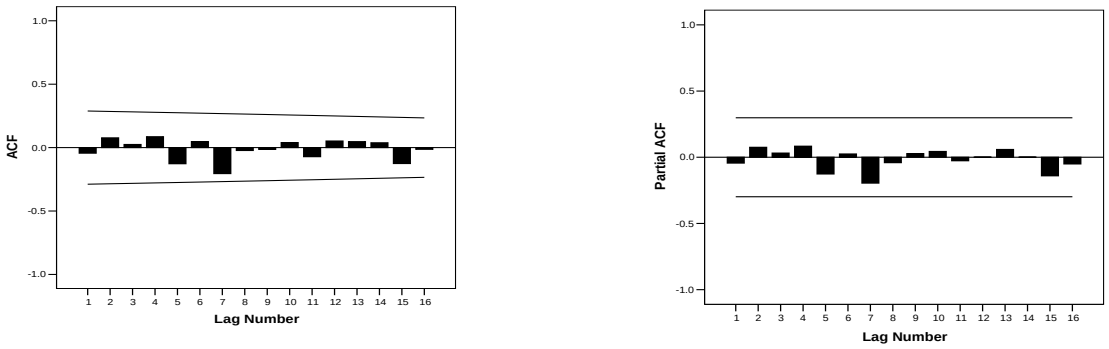


Figure B.5.2 The *sacf* and *spacf* for the Residual of the Fitted AR(1) Model

Source: Appendix Table (A.5.3)

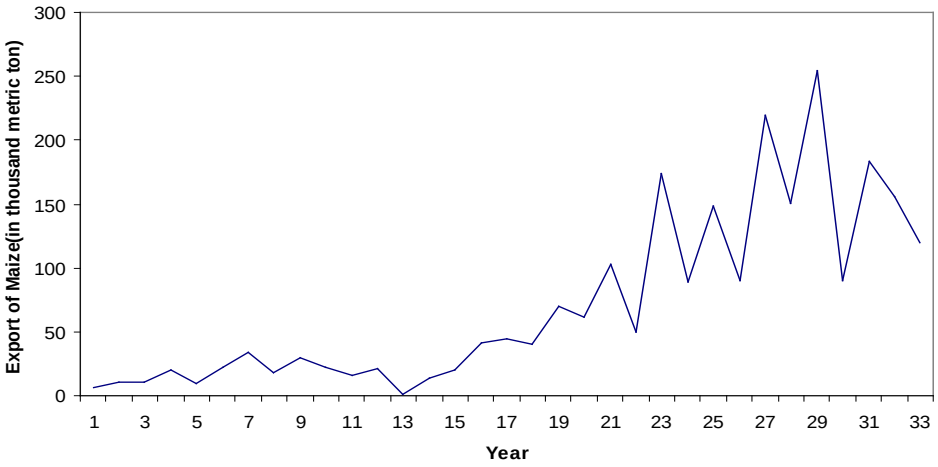


Figure B.5.3 The Yearly Export of Maize from 1976 to 2008
Source: Appendix Table (A.5.6)

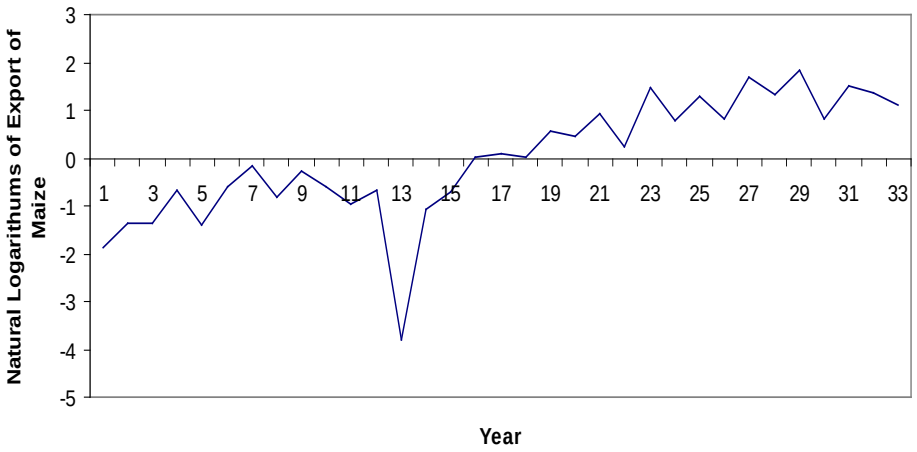


Figure B.5.4 Natural Logarithms of the Export Maize from 1976 to 2008

Source: Appendix Table (A.5.6)

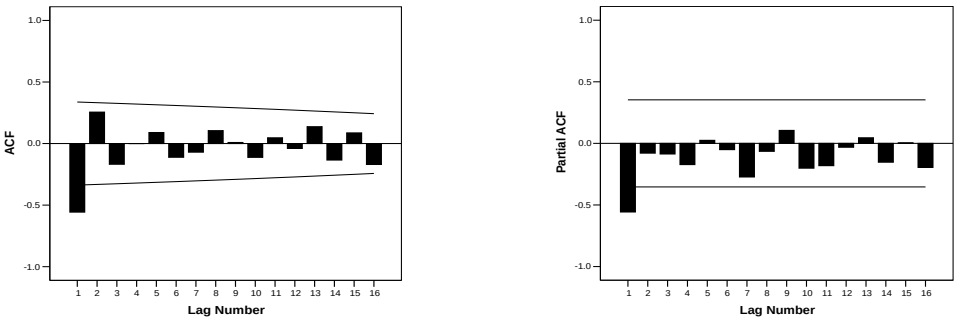


Figure B.5.5 The *sacf* and *spacf* for the Differenced Natural Logarithms of the Export Maize

Source: Appendix Table (A.5.8)

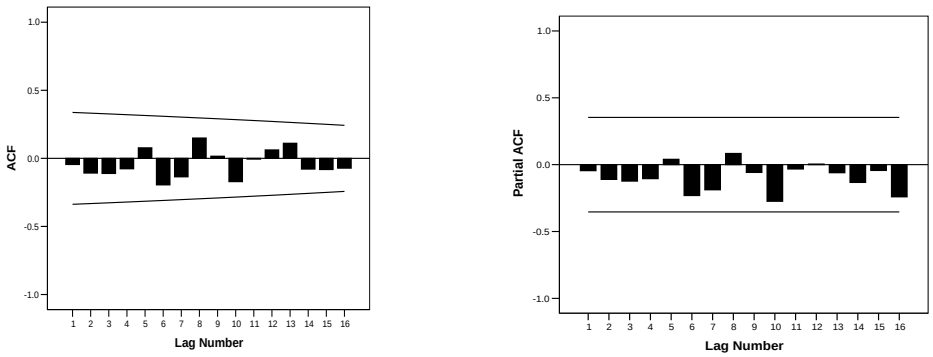


Figure B.5.6 The *sacf* and *spacf* for the Residual of the Fitted ARIMA (1,1,0) Model

Source: Appendix Table (A.5.9)

APPENDIX C
HISTORY OF ROBUST ESTIMATION (1885-1920)

In the eighteenth century, the word "robust" was used to refer to someone who was strong, yet boisterous, crude, and vulgar. By 1953 when Box first gave the word its statistical meaning, the evolution of language had eliminated the negative connotation: robust meant simply strong, healthy, and sufficiently tough to withstand life's adversities. The subject of robust inference, just like the word "robust," has a long and varied history. It is the aim of this present study to examine a part of this history and its relationship to current work.

Simon Newcomb (1886) provided the first sound, modern approach to robust estimation, including the first use of mixtures of normal densities as representing heavy-tailed distributions. Percy Daniell (1920) should be credited with the first mathematical analysis of the class of estimators which are linear functions of order statistics, including the derivation of the optimal weighting functions for estimating scale and location parameters (the so-called "ideal" linear estimators) and the first mathematical treatment of the trimmed mean.

Scientists have been concerned with what we would call "robustness" — insensitivity of procedures to departures from assumptions, particularly the assumption of normality— for as long as they have been employing well-defined procedures, perhaps longer. For example, in the first published work on least squares, Legendre (1805) explicitly provided for the rejection of outliers:

> If among these errors are some which appear too large to be admissible, then those equations which produced these errors will be rejected, as coming from too faulty experiments, and the unknowns will be determined by means of the other equations, which will then give much smaller errors.

Yet most of the early work in mathematical statistics was obsessed with "proving" the method of least squares, either starting with the assumption that the sample mean is the best estimate of the mean and deriving the normal distribution, as Gauss did in his first proof in 1809, or starting with the Central Limit Theorem, as did Laplace in 1811. The first mathematical work on robust estimation seems to have been that of Laplace (1818) on the distribution of the median.

The next statistical problem connected with robust estimation to receive mathematical treatment was the rejection of outliers. In 1852, the first proposal of a

criterion for the determination of outliers was published by Benjamin Peirce, the Harvard mathematician-astronomer and father of logician-philosopher C. S. Peirce. Peirce's paper and most others on this subject were not really about robust estimation, as their authors did not concern themselves with the properties of the resulting estimators; rather, they implicitly assumed that after the outlier test was performed the estimation could be done with no thought given to what had gone before, nor what information might be lost. This narrowness of view did not go unnoticed at the time. The first paper proposing an outlier criterion (Peirce, 1852) was soon followed by the first paper criticizing the use of outlier criteria (Airy, 1856).

But techniques other than simply "reject outliers, then use the sample mean" were also employed. A variety of weighted means had been used prior to 1885. For example, in 1763 James Short (an English astronomer and noted manufacturer of telescopes) had estimated the sun's parallax based on observations of the transit of Venus of 1761 by averaging three means: the sample mean, the mean of all observations with residuals less than one second, and the mean of those with residuals less than half a second. The median and the midrange had appeared even earlier (Eisenhart, 1971).

By the last half of the nineteenth century, weighted least squares had become a standard topic in the literature of the theory of errors, and it was a frequent practice (at least in astronomical investigations) to weight observations differently, depending upon the scientist's (often subjective) estimate of the "probable error" of the observation. The estimate of the probable error was supposed to be based solely on external evidence: scientists were warned of the possible biases if the magnitude of the observation was allowed to influence its weight but it is doubtful that this advice was faithfully adhered to.

Other estimators were proposed in this period. In particular, De Morgan (1847) had outlined a scheme for discounting the more extreme observations. This method, more fully developed by Glaisher (1873), amounted to starting with the sample mean, and then assigning different probable errors to the different observations based on the value of the likelihood function at those observations, and iterating this process. Glaisher's estimate was criticized by both Stone (1873) and Edgeworth (1883), who both (independently) proposed an alternative based on looking for a maximum of the likelihood function (without assuming equal probable errors). Edgeworth later became disenchanted with this alternative (Edgeworth, 1887).

At about this time, Francis Galton was making much use of the median Galton (1875), although his motivation was less suspicion of the normal distribution, which he considered a good representation of many real phenomena, than an appreciation of the simplicity, ease of calculation, and ease of interpretation of the median. Many of these same features were also cited in the apparently independent work of Fechner (1878). Also, various formulas for index numbers were developed during this period; these included weighted averages and geometric means, each designed for a specific purpose.

However, it can still be said that by 1885, the conventional wisdom (but by no means the unanimous view) was that for purposes of estimation, the cautious use of the sample mean was recommended—sometimes weighted, sometimes after discarding outliers, but still the sample mean.

Simon Newcomb appears to have been the first to introduce a mixture of normal densities as a model for a heavy-tailed distribution, and to exploit this model to get an estimator of location which was more robust than the sample mean. (Francis Galton and Karl Pearson had modeled measurements of natural populations by normal mixtures about the same time, but with a completely different object in mind, namely to demonstrate how a single population could be broken down into components.) While Newcomb's name may be unfamiliar to present day statisticians, it should not be so to astronomers, applied mathematicians, and economists.

In Newcomb paper (1886), after criticizing the overuse of outlier criteria and presenting his mixture model, he proceeded to develop an estimator upon the principles of Bayesian decision theory that gave "less weight to the more discordant observations." Adopting squared errors as a loss function (Newcomb's word for loss was "evil"), he demonstrated that in general the posterior mean minimizes the expected mean square error, and he suggested the following procedure. 1) Calculate the residuals based on the sample mean, and, using trial and error, fit a mixture of a finite number of normal densities with zero means to these residuals. 2) Take this fitted mixture and, considering the location family it generates, estimate the desired mean by the posterior mean with respect to a uniform prior, given the original observations. Newcomb realized that this procedure presented practical difficulties and gave a number of simplifying approximations to arrive at a usable estimator. He illustrated its use with the data on the transits of Mercury.

Later in the nineteenth century, Galton (1875) and particularly Edgeworth (1885, 1887, and 1888), touted the use of the median in situations where heavier tails than the normal could be expected. Specifically, Edgeworth (1888) used Laplace's results to conclude that the median may well be better than the mean when the population distribution is one of Newcomb's mixtures of normal distributions. Also, Edgeworth (1886) seems to have been the first to realize that the median may possess an advantage over the sample mean for serially correlated data.

More complicated linear estimators began to appear in 1889, when Galton (1889) suggested estimating the mean and standard deviation of a normal distribution by what amounts to taking

$$\hat{\mu} = \frac{\zeta_p X^{(nq)} - \zeta_q X^{(np)}}{\zeta_p - \zeta_q},$$

$$\hat{\sigma} = \frac{X^{(np)} - X^{(nq)}}{\zeta_p - \zeta_q},$$

where ζ_p and ζ_q are the p and q percentiles of the standard normal distribution, $X^{(np)}$ and $X^{(nq)}$ are the sample $100p$ and $100q$ percentiles, and p and q are arbitrary but fixed ($0 < p < q < 1$). In 1899 in a long paper on the multivariate normal distribution and its applications, Sheppard proved the joint asymptotic normality of Galton's estimators when the population is normal. He also showed the joint asymptotic normality of $X^{(np)}$ and $X^{(nq)}$, and gave analogues to $\hat{\mu}$ and $\hat{\sigma}$ based on any finite number of sample percentiles (Sheppard, 1899).

Sheppard's paper also represented the first attempt since Laplace to optimize performance within a class of linear functions of order statistics. He both showed how the best choice (for normal population) of p and q can be made and found that which linear combination of the three quartiles has the smallest asymptotic variance (again for normal populations). Such functions of the three quartiles had been considered earlier by Edgeworth (1893), who neglected the quartiles' correlation and erroneously claimed the estimator with weights in proportions 5: 7: 5 to be superior to the sample mean for normal populations.

The next mathematical work to appear on order statistics was Karl Pearson's (1902) examination of the Galton difference problem. In this paper, which was inspired by an inquiry of Galton's (1902) as to the most suitable proportion between the values of first and second prizes, Pearson gave the joint density of any two

consecutive order statistics and found their expected difference. He proposed on another occasion to consider the application of Galton's problem to a new theory for the rejection of outlying individuals.

In 1920, a remarkable paper appeared in the *American Journal of Mathematics* (the journal Simon Newcomb co-founded) by the English mathematician P. J. Daniell. This paper, "Observations Weighted According to Order," has been all but totally overlooked since its publication. It could in fact be claimed that Daniell was at least thirty years ahead of his time, for it took that long for his major results to be rediscovered. While his paper itself is a model of clarity and rigor, its relevance to modern work is such that it merits a short summary, in his own notation.