LINEAR REGRESSION FOR DATA HAVING MULTICOLLINEARITY, HETEROSCEDASTICITY AND OUTLIERS

BELLO ABDULKADIRI RASHEED

UNIVERSITI TEKNOLOGI MALAYSIA

LINEAR REGRESSION FOR DATA HAVING MULTICOLLINEARITY, HETEROSCEDASTICITY AND OUTLIERS

BELLO ABDULKADIRI RASHEED

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Doctor of Philosophy (Mathematics)

Faculty of Science
Universiti Teknologi Malaysia

JANUARY 2017

This thesis is dedicated to my late father (Alhaji Abdulkadiri Haruna Rasheed).

ACKNOWLEDGEMENT

Bismillahi ar-Rahman ar-Rahim. In the name of ALLAH, the Most Gracious, the Most Merciful. AlhamdulilLahi Rabbil 'Alamin. All the praises and thanks be to ALLAH Subhanahu wa Ta'ala, the Lord of the whole universe,for giving me the opportunity to go through this trial with His help and guidance. May the peace and blessings of Allah be upon The Final Massenger, Prophet Mohammad his family and companions, and those who follow his path.

First and foremost, I would like to express my sincere appreciation and warmnest thanks to my supervisor, **Assoc. Prof. Dr. ROBIAH ADNAN**, for her encouragement when I have problems, guidance to the right way, and suggestions when I am hesitate. Her invaluable help of constructive comments and useful advice throughout my research have contributed to the success of this research. Not forgotten, my appreciation to my co-supervisor, Dr. Seyed Ehsan Saffari for his support and encouragement. Without their continued support and interest, this thesis would not have been the same as presented here. An Appreciation also goes to the Universiti Teknologi Malaysia (UTM).

Finally, I would like to thank my father, who passed away in 2012, after I began my study in Malaysia. I feel sorry that I did not accompany him in his last minute. This thesis is dedicated to him. I would also like to thank my mother and beloved family members for their support, love and encouragement throughout my study. My sincere appriciation also extends to my lovely brothers ,and friends who have directly or indirectly helps and contribute to the success of this thesis.

ABSTRACT

Evaluation of regression model is very much influenced by the choice of accurate estimation method since it can produce different conclusions from the empirical results. Thus, it is important to use appropriate estimation method in accordance with the type of statistical data. Although reliable for a single or a few outliers, standard diagnostic techniques from wild bootstrap fit can fail while the existing robust wild bootstrap based on MM-estimator is not resistant to high leverage points. The presence of high leverage points introduces multicollinearity while the MM-estimator is also not resistant to the presence of multicollinearity in the data. This research proposes new methods that deal with heteroscedasticity, multicollinearity, outliers and high leverage points more effectively than currently published methods. The proposed methods are called modified robust wild bootstrap, modified robust principal component (PC) with wild bootstrap and modified robust partial least squares (PLS) with wild bootstrap estimations. These methods are based on weighted procedures that incorporate generalized M-estimator (GM-estimator) with initial and scale estimate using S-estimator and MM-estimator. In addition, the multicollinearity diagnostics procedures of PC and PLS were also used together with the wild bootstrap sampling procedure of Wu and Liu. Empirical applications of data for national growth, income per capital data of the Organisation of Economic Community Development (OECD) countries and tobacco data were used to compare the performance between wild bootstrap, robust wild bootstrap, modified robust wild bootstrap, modified robust PC with wild bootstrap and modified robust PLS with wild bootstrap methods. A comprehensive simulation study evaluates the impacts of heteroscedasticity, multicollinearity outliers and high leverage points on numerous existing methods. A selection criterion is proposed based on the best model with bias and root mean squares error for the simulated data and low standard error for real data. Results for both real data and simulation study suggest that the proposed criterion is effective for modified robust wild bootstrap estimation in heteroscedasticity data with outliers and high leverage points. On the other hand, the modified robust PC with wild bootstrap estimation and modified robust PLS with wild bootstrap estimation is more effective in multicollinearity, heteroscedasticity, outliers and high leverage points. Moreover, for both methods, the modified robust sampling procedure of Liu based on Tukey biweight with initial and scale estimate from MM-estimator tend to be the best. While the best method for data with multicollinearity, heteroscedasticity, outliers and high leverage points is the modified robust PC with wild bootstrap estimation. This research shows the ability of the computationally intense method and viability of combining three different weighting procedures namely robust GM-estimation, wild bootstrap and multicollinearity diagnostic methods of PLS and PC to achieve accurate regression model. In conclusion, this study is able to improve parameter estimation of linear regression by enhancing the existing methods to consider the problem of multicollinearity, heteroscedasticity, outliers and high leverage points in the data set. This improvement will help the analyst to choose the best estimation method in order to produce the most accurate regression model.

ABSTRAK

Penilaian model regresi sangat dipengaruhi oleh pilihan kaedah anggaran tepat kerana ia boleh menghasilkan kesimpulan yang berbeza dari hasil empirikal. Oleh itu, adalah penting untuk menggunakan kaedah yang sesuai mengikut jenis data statistik. Walaupun ia boleh dipercayai bagi satu atau beberapa titik terpencil, teknik diagnostik piawai dari cangkuk liar boleh gagal manakala teknik cangkuk liar teguh berdasarkan penganggar MM yang sedia ada juga tidak mempunyai daya tahan terhadap titik leveraj yang tinggi. Tambahan pula, kehadiran titik leveraj yang tinggi akan mewujudkan kolinearan berganda dan penganggar MM juga tidak mempunyai daya tahan untuk menangani kewujudan kolinearan berganda dalam data. Kajian ini mencadangkan kaedah baru yang menangani heteroskedastisiti, kolinearan berganda, titik terpencil dan titik leveraj yang tinggi dengan lebih berkesan berbanding dengan kaedah terkini yang pernah diterbitkan. Kaedah yang dicadangkan dipanggil sebagai penganggar cangkuk liar teguh terubahsuai, komponen utama (PC) teguh terubahsuai dengan cangkuk liar dan kuasa dua terkecil separa (PLS) teguh terubahsuai dengan cangkuk liar. Kaedah-kaedah ini adalah berdasarkan kepada prosedur wajaran yang menggabungkan penganggar M (penganggar GM) terubahsuai dengan anggaran awal dan skala menggunakan penganggar S dan penganggar MM. Di samping itu, prosedur diagnostik kolinearan berganda PC dan PLS juga digunakan bersama-sama dengan prosedur persampelan cangkuk liar Wu dan Liu. Penggunaan empirikal bagi data pertumbuhan negara, pendapatan perkapita bagi data negara-negara Organisasi Pembangunan Ekonomi Masyarakat (OECD) dan data tembakau telah digunakan untuk membanding prestasi antara cangkuk liar, cangkuk liar teguh, cangkuk liar teguh terubahsuai, PC teguh terubahsuai dengan cangkuk liar dan PLS teguh terubahsuai dengan kaedah cangkuk liar. Satu kajian simulasi yang menyeluruh dilakukan untuk menilai kesan heteroskedastisiti, kolinearan berganda, titik terpencil dan titik leveraj terhadap kaedah yang ada pada masa kini. Satu kriteria pemilihan telah dicadangkan berdasarkan model terbaik dengan nilai pincang dan ralat punca min kuasa dua yang terkecil bagi data simulasi dan ralat piawai terkecil bagi data sebenar. Keputusan bagi kedua-dua data sebenar dan simulasi menunjukkan bahawa kriteria yang dicadangkan itu adalah berkesan untuk anggaran cangkuk liar teguh terubahsuai dalam data heteroskedastisiti bersama titik terpencil dan titik leveraj. Sebaliknya, anggaran PC teguh terubahsuai dengan cangkuk liar dan anggaran PLS teguh terubahsuai dengan cangkuk liar adalah lebih berkesan dalam kolinearan berganda, heteroskedastisiti, titik terpencil dan titik leveraj tinggi. Tambahan pula, bagi kedua-dua kaedah, prosedur persampelan teguh terubahsuai Liu yang berdasarkan Tukey biweight dengan anggaran awal dan skala dari penganggar MM cenderung untuk menjadi yang terbaik. Manakala kaedah terbaik untuk data mengandungi kolinearan berganda, heteroskedastisiti, titik terpencil dan titik leveraj tertinggi adalah PC teguh terubahsuai dengan anggaran cangkuk liar. Kajian ini menunjukkan kebolehan kaedah pengiraan intensif dan daya maju menggabungkan tiga prosedur pemberat yang berbeza iaitu penganggar GM teguh, cangkuk liar dan kaedah diagnostik kolinearan berganda PLS dan PC untuk mencapai model regresi tepat. Kesimpulannya, kajian ini dapat meningkatkan anggaran parameter regresi linear dengan meningkatkan kaedah sedia ada untuk mengambil kira masalah kolinearan berganda, heteroscdastisiti, titik terpencil dan titik leveraj tinggi dalam set data. Peningkatan ini akan membantu penganalisis untuk memilih kaedah anggaran yang terbaik untuk menghasilkan model regresi yang paling tepat.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

$ALM3$	–	Andrews Liu method three
$ALM1$	–	Andrews Liu method one
$ALM2$	–	Andrews Liu method two
$ALM4$	–	Andrews Liu method four
$ALM6$	–	Andrews Liu method six
$ALM5$	–	Andrews Liu method five
AV	–	asymptotic variance
$AWM3$	–	Andrews Wu method three
$AWM1$	–	Andrews Wu method one
$AWM2$	–	Andrews Wu method two
$AWM4$	–	Andrews Wu method four
$AWM5$	–	Andrews Wu method five
$AWM6$	–	Andrews Wu method six
$ARGMBWu$	–	Andrews robust generalized M-estimator wild bootstrap Wu
$ARBGMBiu$	–	Andrews robust generalized M-estimator wild bootstrap Liu
$ARBGMBWu$	–	Andrews robust generalized M-estimator wild bootstrap Wu
$ARPLSGMBWu$	–	Andrews robust partial least squares generalized M-estimator wild bootstrap Wu
$ARPLSGMBLiu$	–	Andrews robust partial least squares generalized M-estimator wild bootstrap Liu
$ARPCGMBWu$	–	Andrews robust generalized M-estimator wild bootstrap Wu
$ARPCGMBLiu$	–	Andrews robust generalized M-estimator wild bootstrap Liu
AM	–	alarmgir method
$BLUE$	–	best linear unbiased estimator
$BootWu$	–	bootstrap Liu's
$BootWu$	–	bootstrap Wu's
BP	–	breakdown point
CI	–	confidence interval
CM	–	covariance matrix

$DRLTS$	–	dynamic robust least trimmed squares
$DRGP(MVE)$	–	Diagnostic Robust Generalized Potential Minimum Volume Ellipsoid
EQP	–	equipment investment
FBP	–	finite breakdown point
GES	–	gross error sensitivity
$GM(DRGP$	–	generalized M-estimator diagnostic robust generalized potential
$GM - estimator$	–	generalized M-estimator
GDP	–	gross domestic products
$HCCME$	–	heteroscedasticity consistence covariance matrix estimator
$HLM1$	–	Huber Liu method one
$HLM2$	–	Huber Liu method two
$HLM3$	–	Huber Liu method three
$HLM4$	–	Huber Liu method four
$HLM5$	–	Huber Liu method five
$HLM6$	–	Huber Liu method six
HM	–	Huber method
$HWM1$	–	Huber Wu method one
$HWM2$	–	Huber Wu method two
$HWM3$	–	Huber Wu method three
$HWM4$	–	Huber Wu method four
$HWM5$	–	Huber Wu method five
$HWM6$	–	Huber Wu method six
$HRGMBWu$	–	Huber robust generalized M-estimator wild bootstrap Wu
$HRBGMBLiu$	–	Huber robust generalized M-estimator wild bootstrap Liu
$HRPLSGMBWu$	–	Huber robust partial least squares generalized M-estimator wild bootstrap Wu
$HRPLSGMBLiu$	–	Huber robust partial least squares generalized M-estimator wild bootstrap Wu
$HRPCGMBWu$	–	Huber robust generalized M-estimator wild bootstrap Wu
$HRPCGMBLiu$	–	Huber robust generalized M-estimator wild bootstrap Liu
IF	–	influence function
IQ	–	intelligent quation
$IRLS$	–	iterative reweighted least squares
JA	–	jacknife

LFG	–	labor force growth
LMS	–	least median squares
LQS	–	least quartile of squares
LSS	–	local shift sensitivity
LS	–	least squares
LTS	–	least trimmed squares
MAD	–	median absolute deviation
MCD	–	minimum covariance determinant
MGQ	–	modified goldfeld-qundl
MVE	–	Minimum Volume Ellipsoid
NEQ	–	non-equipment investment
$OLSVIF$	–	ordinary least squares variance inflation factor
$OECD$	–	organization of economic community development
OLS	–	ordinary least squares
PC	–	principal component
PCA	–	principal component analysis
PCR	–	principal component regression
$PLSR$	–	partial least square regression
PLS	–	partial least square
QEA	–	quartile evolutionary algorithms
$QMDPE$	–	quick maximum density power estimator
RMD	–	Robust Mahalanobis Distance
$RMSE$	–	Root Mean Squares Error
$RBootLiu$	–	robust bootstrap Liu's
$RBootLiu$	–	robust bootstrap Wu's
RE	–	relative efficient
RLS	–	reweighted least squares
RMD	–	robust Mahalanobis distance
RP	–	rejection point
RR	–	ridge regression
$RRLS$	–	robust reweighted least squares
$RVIF$	–	robust variance inflation factor
SE	–	standard error
$TLM1$	–	Tukey Liu method one
$TLM2$	–	Tukey Liu method two

$TLM4$	–	Tukey Liu method four
$TLM5$	–	Tukey Liu method five
$TLM6$	–	Tukey Liu method six
$TWM1$	–	Tukey Wu method one
$TWM3$	–	Tukey Wu method three
$TWM4$	–	Tukey Wu method four
$TWM5$	–	Tukey Wu method five
$TWM6$	–	Tukey Wu method six
$TRGMBWu$	–	Tukey robust generalized M-estimator wild bootstrap Wu
$TRGMBLiu$	–	Tukey robust generalized M-estimator wild bootstrap Liu
$TRPLSGMBWu$	–	Tukey robust partial least squares generalized M-estimator wild bootstrap Wu
$TRPLSBLiu$	–	Tukey robust partial least squares generalized M-estimator wild bootstrap Liu
$TRPCBLiuLMS$	–	Tukey robust principal component generalized M-estimator wild bootstrap Wu
$TRPCGMBLiu$	–	Tukey robust principal component generalized M-estimator wild bootstrap Liu
$TRPCGMBWu$	–	Tukey robust principal component generalized M-estimator wild bootstrap Wu
TC	–	tuning constant
$TLM2$	–	Tukey Liu method two
TM	–	Tukey method
$TWM2$	–	Tukey Wu method two
VIF	–	variance inflation factor
$VBQMDPE$	–	variable bandwidth quick maximum density power estimator
WLS	–	weighted least squares
	–	

LIST OF SYMBOLS

Ω	–	Omega
σ	–	Population variance
Σ	–	Covariance
β	–	Parameter
k	–	Bias Constant
Z	–	Principal Component
Θ	–	Theta
Γ	–	Gamma
y	–	y-axis
x	–	x-axis
B	–	bootstrap replicate
n	–	No. of Observation
ε_i	–	Residual
$\sum$	–	Summation
h	–	sub sample
p	–	No. of Explanatory Variables
$\binom{n}{h}$	–	Combination
b_i	–	Coefficients
r	–	Standardized Residuals
λ	–	Eigenvalue
V	–	Eigenvector
B	–	Bootstrap Replicate
ψ	–	Condition Number
X	–	Explanatory Variables
X'	–	Transpose of Explanatory Variables
R^2	–	Determination Coefficient
y_i	–	Response Variable
x_i	–	Explanatory Variables

min	–	Minimum
med	–	Median
$\hat{\beta}$	–	Estimate of parameter
α_2	–	Percentage
log	–	logarithm
S^2	–	variance of parameter
$\forall$	–	for all
$\sim$	–	distribute
$\geq$	–	Greater or equal
$\rightarrow$	–	Approximate
$>$	–	Greater
∞	–	Infinty
c	–	tuning constant
W	–	weight
ρ	–	Correlation
$\bar{\beta}_j$	–	Average of parameters
$\%$	–	Percentage
$\leq$	–	Least than or equal
μ	–	Mean
π	–	3.14286
	–	

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

This chapter is aimed at introducing the significance of this research. The research background will be described in Section 1.1 followed by Section 1.2 which discussed the statement of the problem. The research objective is presented in Section 1.3. In the proceeding section, it present the scope of the study. Section 1.5 and Section 1.6 will present thesis organization and the significance of the research respectively.

1.1 Research Background

The common objective in statistics is to identify an appropriate transformation idea from a sample to relate a dependent variable to a set of independent variables. Linear regression is the customary method used to mathematically model a dependent variable as a function of the independent variables. The term regression analysis is a statistical technique used in all fields of engineering, science, and management that require fitting a model to sets of data. There are several methods available in the literature to estimate the parameter in regression model. The ordinary least squares (OLS) method is the most popular method in statistics application because of its optimal properties and ease of computation. The OLS estimator was discovered independently by Gauss in 1795 and Legendre in 1805 which minimizes the sum of the squared distances for all points from the actual observation to the regression surface. From the theorem of Gauss-Markov, OLS is always the best linear unbiased estimator (BLUE). The word BLUE means that among all unbiased estimators, OLS has the minimum variance. If the error ε is assumed to be normally, independently distributed with mean 0 and variance $\sigma^2 I$, least squares is considered as the uniformly minimum variance unbiased estimator. The assumption of constant variance is one of the basic requirements of regression model. Under this assumption, inference procedures such as hypothesis tests, confidence intervals, and prediction intervals are powerful. However, if the error term ε is not normally distributed, then the OLS parameter estimates and inferences can be flawed. A common reason for the violation of this assumption is for the response variable to follow a probability distribution in which the variance is functionally related to the mean Wisnowski *et al.* (2003). This condition is known as heteroscedasticity and can also have detrimental effects on the OLS estimates of the coefficients.

However, real data like economic, engineering, sciences and medical data, usually does not completely satisfy the assumptions often made by researchers which result in a dramatic effect on the quality of statistical analysis. In the presence of of heteroscedasticity, the OLS estimator will remain unbiased. However, the most harmful consequence of heteroscedasticity would be the parameter covariance matrix (CM) of the OLS. As a result, the elements in the diagonal matrix that are utilized to estimate the standard errors of the regression coefficient become biased and unreliable. In addition, the t-tests for individual coefficients are generally too liberal or too conservative depending on the form of heteroscedasticity and bias in the confidence intervals. Consequently, the OLS estimator is no longer BLUE (Midi *et al.*, 2009a). There are several methods proposed in the literature to address the heteroscedasticity problem Rana *et al.* (2012), Midi *et al.* (2009c), Cribari-Neto and Ferrari (1995), , Liu *et al.* (1988),Wu (1986) and White (1980a). However, another challenging assumption that invalidates the OLS estimator is when assumption of independence between the explanatory variables is violated, thus bringing about the existance of Multicollinearity. Multicollinearity is a situation in which a set of data have two or more regressor variables that are redundant and contain similar information. The linear dependencies among the regressors can affect the model ability to estimate regression coefficients. The violation of independence assumption usually occurs right from the method of data collection. This situation can occur when the researchers only sample subspace of the region for regressor variables, or the model constraints in the population being sampled, which can also cause multicollinearity problems.

The redundant information describe what the regressor variable explains about the response variable, which is exactly the same as what the other regressor variable explains. In such situation, the two or more redundant regressor variables would be completely unreliable since the coefficient would measure the same effect of those regressor variables. The major problem that may greatly influence the estimate of regression model is when there is correlation between the regressor variables which is defined as multicollinearity. However, the problems of multicollinearity in regression model is that the variances of the parameter estimates become very large which resulted from producing wrong sign of the parameter estimate. In addition, the standard error of the estimate becomes inflated and unstable Rai *et al.* (2013). Although in the presence of multicollinearity, the OLS estimator still remains unbiased, its estimates become inefficient Midi *et al.* (2010). There is now evidence that multicollinearity problems have a great impact on regression model. Violation of the NID distribution of the error term can occur when there are one or more outliers or extreme outliers in the X-direction (high leverage outliers) in the data set. Multicollinearity problems are very common in the areas of medical, economics and management. The principal component analysis regression (PCR) and partial least squares regression (PLSR) methods, which are commonly used, become unreliable in the presence of outliers. The methods of PCR and PLSR are based on OLS method and thus outliers may affect the PCR and PLSR model Candès *et al.* (2011). The identification of multicollinearity and its remedial measures have been discussed by many standard books and a number of articles Phatak and De Jong (1997), Maitra and Yan (2008), Hubert and Branden (2003), and Ahmad *et al.* (2006).

The challenge of having outliers in the data set is another focus that this thesis is going to emphasize. Outliers, is one of the earliest statistical interests, since nearly all data sets contain outliers of varying percentages, and it continues to be one of the most important issues in regression analysis. Sometimes outliers can grossly distort the statistical analysis, at other times their influence may not be as noticeable. An outlier is an observation that is inconsistent with the remainder of the data and it is not unusual to see an average of 10% outliers in data sets for some processes Leroy and Rousseeuw (1987), Shevlyakov and Vilchevski (2001), Barnett and Lewis (1994) and Hampel *et al.* (1986). Some of the sources of outliers are errors in data entry or measurement, the inadvertent inclusion of an observation from another population or a plausible event. Outliers can also be due to genuine long tailed distributions. Outliers can be found in the response variable (y-variable) or the regressor variables (x-variables). Regardless of the origin, the outliers with respect to the regressor variables are referred to as a leverage outliers. In this thesis, two outliers ie. high leverage outliers and residuals outliers, would both be referred to as outliers. According to Midi *et al.* (2010), high leverage outliers may decrease or increase multicollinearity problem of a collinear data matrix X. A single sufficiently outlying observation in a data set can render least squares estimation approach inappropriate. Alma (2011) points out that the effect of these outlying observations both in the direction of the response and regressor variables to the regression model is that they have a strong adverse effect on the estimates and sometimes may remain unnoticed. In practice there are certain situations in which the outliers cannot be seen. But as described by Hampel *et al.* (1986), Leroy and Rousseeuw (1987) and Shevlyakov and Vilchevski (2001) the OLS estimator will perform poorly when there are multiple outliers in the data or gross error in the y-direction. This situation as described by Huber and Ronchetti (2009) ,Hadi (1992) and Rousseeuw and van Zomeren (1991) is usually as a result of masking and swamping problems. In swamping problem, the clean observations are considered as outliers Barnett and Lewis (1984). Conversely, in masking problem, the outliers are identified as clean data points.

There have been a number of literature in recent years developing the theory and practice of robust regression estimators and some have been successfully used in practice, but the results obtained still need to be improved. Typically, these estimators require significant computational resources because of nonlinear solutions or the requirement to search numerous subsets of the data to satisfy a constrained objective function. Outliers diagnostic methods and its remedial measures have been discussed in many standard books and articles. Rosseeuw and Yohai (1984), Kafadar and Morris (2002) and Tukey and Tukey (1988). The behavior of statistical data set make it very important for the researchers to identify the efficiency of an estimator in evaluating the precision of an estimated regression coefficient. In regression model, bootstrap techniques is applied when dealing with data designs that have limited experimental unit or sample size such as life human subjects. Other than that, it helps to approximate the distribution of the coefficients and the distribution of the prediction errors when the regressors are data Stine (1985) or random variable McCullough (1996). The situation of the regression model structure will differ and the Bootstrap estimate produces sub-optimal or even wrong inferential statement with inaccurate forecasting. The most common approach to this situation is the wild bootstrap technique. However, different wild bootstrap have been proposed in literature Wu (1986) and Liu *et al.* (1988). According to Rana *et al.* (2012), the existing wild bootstrap methods are computed based on OLS method and can be duly affected in the presence of outliers. The most often used robust wild bootstrap with high-breakdown estimator and high efficient is the MM-estimator introduced by Yohai (1987).

However, the robust wild bootstrap methods developed to this level have weaknesses under certain outlier scenarios or in the presence of multicollinearity. The problem with these robust wild bootstrap of MM-estimator is that they can fail if the outliers have extreme values in the regressor variables high leverage outliers) in the data. In addition, in the presence of multicollinearity, the robust wild bootstrap methods will produce sub-optimal solution. This study attemps to solve this problems by introducing a parameter estimation methods that is able to simultaneously estimate regression model in a situation when residual outliers, high leverage points and heteroscedasticity error variance are presence. Moreover, in the presence of multicollinearity, residual outliers, high leverage points and heteroscedasticity error variance in data set. The ability of a robust estimator to accommodate high leverage points is called bounded-influence estimators. On the other hand, the most pronouns diagnostics measures of multicollinearity problems, is the principal component analysis (PCA) and partial least squares (PLS) approaches. However, different classes of robust regression that are bounded-influence estimators has been proposed that simultaneously achieve all the three properties of robust estimation Krasker and Welsch (1982), Staudte and Sheather (1990), Simpson and Montgomery (1998) and Thomas (1997). One of these bounded influence estimator is the GM-estimators introduced by Krasker and Welsch (1982) and modified by Midi *et al.* (2010) to overcome the limitation of GM-estimators. However, different researchers use the objective function of Krasker and Welsch (1982) which down weights outliers with high leverage points only if the corresponding residual is large Samkar and Alpu (2010). These GM-estimators have the potential not only to identify a wide range of multiple outliers, but also to accommodate them in a model.

Many of the existing robust estimators that are bounded-influence which can be easily combined with wild bootstrap estimators, using the multicollinearity diagnostic measures of PC and PLS approach. The GM-estimator used in this research was modified in pattern of modified GM-estimator approach of Midi *et al.* (2010) which was based on the initial S-estimator introduced by Rosseeuw and Yohai (1984). The modified GM-estimator of Walker (1984) based on the initial and scale estimate of S-estimator and the MM-estimator of Yohai (1987) were also used by Samkar and Alpu (2010) to estimate the parameter of the model. The approach of Samkar and Alpu (2010) was also adopted in this research to estimate the parameters of the regression model. Application of modified robust GM-estimator wild bootstrapping method and multicollinearity diagnostic method of PC and PLS can provide stable coefficients estimates with computational ease. The goal of this thesis is to estimate the parameter of regression model in the presence of residual outliers, high leverage point and heteroscedasticity error variance. Moreover, this research estimate the parameter of regression model in the presence of multicollinearity, residuals outliers, high leverage point and heteroscedasticity error variance. However, from now on, in this thesis, the modified GM-estimator based on Krasker and Welsch (1982) is considered as modified robust wild bootstrap based on GM-estimator, modified robust wild bootstrap with PC based on GM and modified robust wild bootstrap with PC based on GM-estimator respectively.

1.2 Statement of Problem

Generally, difficulties may arise usually when researchers try to apply appropriate regression estimation techniques to estimate the regression coefficient. The traditional view is that the ordinary least squares (OLS) method estimation is robust to deviations from the assumptions of normality and thus discourage users from applying other methods. Literature have shown that the regression model diagnostics procedure produced a wrong regression fit due to the presence of residual outliers. The researchers have not often been able to properly fit the regression model as the literature of robust estimation method for parameter estimation of regression model in the presence of residual outliers are very limited. Presence of residual outliers is the most common situation in statistical data. The frequent increases of observational data with numerous residual outliers make it necessary to introduce other techniques of robust estimation methods that will handle the problems of residual outliers According to Hampel *et al.* (1986), the sources of residual outliers are errors in data entry or measurement and about 1-10% of every routine data set are residual outliers. The presence of residual outliers can result in producing invalid inferential statement, the residual outliers outliers diagnostic for linear regression model is vital in regression analysis. Some outliers diagnostic have been proposed in the literature. However it is suspect that they may not perform well in the presence of high leverage points.

This problem has inspired the development of a new robust method for residuals outliers diagnostic that is resistant to high leverage point called outliers. However, the situation becomes worse when there is heteroscedasticity in the data. Presence of residual outliers and high leverage point together with heteroscedasticity will invalidate the model parameter by producing wrong statistical inferences. Not many works have been developed which focused on the issues when residual outliers, high leverage points and heteroscedasticity occur at the same time. Therefore there is a need to develop a new robust method for linear regression which is resistant to residual outliers, high leverage points, and heteroscedasticity errors. However, the situation becomes more complex when high leverage points, and multicollinearity occur simultaneously with heteroscedasticity in the data. In this regard, it is important to investigate their impact on linear regression model. Based on our knowledge, regression model with heteroscedasticity errors in the presence of residual outliers and high leverage points and heteroscedasticity errors in the presence of outliers, high leverage point, and multicollinearity were not given much attention in the literature. Thus, that is the reasons why this research is of interest and the contribution will be very significant in the field of statistics. The problems mentioned above present a challenging to the regression user. On the one hand, extra effort is often required to get the appropriate methods that will handle these problems in order to obtain better parameter estimates. In this thesis the wild bootstrap techniques proposed by Wu (1986) and Liu *et al.* (1988) which are efficient both under homoscedasticity and heteroscedasticity of unknown form will be used to estimate the model parameter. But these wild bootstrap methods are based on OLS and hence the estimator can be affected in the presence of outliers. Under this scenario the robust wild bootstrap methods which are resistant to outliers is introduced. The robust wild bootstrap methods introduce by Rana *et al.* (2012) are based on MM-estimator and our investigation revealed that the robust wild bootstrap techniques are resistant to presence of residuals outliers but not resistant to leverage points, Simpson (1995a). These motivate the introduction of a new robust wild robust bootstrap technique that is not sensitive to high leverage points. Based on this understanding of the limitation of robust wild bootstrap methods, the GM-estimator .which was described by Wilcox Rand (2005) and Andersen (2008) as highly efficient and bounded influence estimator will be consider in this study. The GM-estimator will be applied in two different techniques. The first technique involves the GM-estimator based on the initial estimate of high efficient and high break down point of MM-estimators. While the second technique involve the GM-estimator based on the initial estimate of high efficient and high break down point of S-estimator. Moreover, the robust wild bootstrap techniques are resistant to residuals outliers and heteroscedasticity but not resistant to multicollinearity. This is another challenge faced by the robust wild bootstrap and no work in the literature addresses the combined problems of multicollinearity and heteroscedasticity in the presence of residuals outliers and high leverage points using the wild bootstrap approach. The multicollinearity diagnostic method of PC and PLS with a bounded influence GM-estimator which was introduced by Krasker and Welsch (1982) will be explored.

1.3 Research objectives

1. To standardize three weighting procedures of Tukey bisquare, Huber and Andrews sin weighted function that will provides same efficient and rejection point which will be used for comparisons of the proposed methods in multiple linear regression model.
2. To develop two alternative robust wild bootstrap estimation techniques for multiple linear regression model that will include residual outliers, high leverage points and heteroscedastic error variance. The first method will be achieved by

combining GM-estimator of Krasker and Welsch (1982) with initial estimate of MM-estimator and wild bootstrap of Wu's and Liu's with three different weighting procedures. while the second method is by substituting the initial estimate of S-estimator.

3. To propose new approaches of modified robust wild bootstrap estimation techniques for multiple linear regression model in the presence of residual outliers, high leverage points and heteroscedastic error variance with three different weighting procedures and then compared it with the existing methods.
4. To formulate a new modified robust wild bootstrap method for multiple linear regression model with principal component and partial least squares procedures for handling the multicollinearity problem in the presence of residual outliers, high leverage points and heteroscedasticity error with different weighting procedures.
5. To compare the performance of modified robust wild bootstrap estimation techniques based on principal component and modified robust wild bootstrap estimation techniques based on partial least squares approach for handling the multicollinearity problem in the presence of residual outliers, high leverage points and heteroscedasticity error.

This study aims at developing a robust regression methods that can remedy the situation for the violation of homoscedasticity and uncorrelation assumption of linear regression model in the presence of residual outliers and high leverage points. There are five primary objectives that address the research problem.

1.4 The Scope of the study

This research focuses on the robust estimation techniques, wild bootstrap, residual outliers, high leverage points, multicollinearity and heteroscedasticity in regression model, performance of the published techniques and also those proposed in this research. The selection of the techniques are limited to those that are promising and often referred to and those that performed well in the literature. For this research, the diagnostics procedures of residuals outliers, high leverage points, homoscedasticity, multicollinearity and heteroscedasticity identification will be given more emphasis. The robust estimation techniques, wild bootstrap method and multicollinearity diagnostics measures with three different weighting procedures are limited to multiple linear regression models. The nonlinear regression and generalized linear models are not considered, although, many of the concepts explored in this research can be easily extended to those classes of models. Real data and simulation studies are the primary tool used to accomplish the objectives outlined in Section 1.3. In most cases, the simulation studies are set up as design instruments to gain the maximum performance of each estimation method. In this thesis, there are enough replicates for the wild bootstrap procedures to get a clear indication of the performance of each estimator. The simulation study of 2, 5 and 10 covariance variables will be used to generate the random sample that will be used to measure the performance of the existing and proposed methods. Different levels of heteroscedasticity, percentage of residual outliers, high leverage points, and degree of multicollinearity are also generated from the simulation procedures. The real data containing relevant problems in different fields is used to illustrate the advantages of the proposed methods developed in this thesis. The performance measures of each estimator were done based on their bias and root mean squares error.

1.5 Thesis Organization

This thesis is organized as follows. Chapter 1 briefly provides an overview of homoscedasticity, heteroscedasticity, multicollinearity, residual outliers and high leverage points in regression analysis and the need of bounded influence of GM-estimator methods Krasker and Welsch (1982) , principal component analysis, partial least squares and robust wild bootstrap. The classical bootstrap, wild bootstrap and ordinary least squares are discussed and research objectives are well defined. The scope is discussed in detail. The research contribution is stated at the end of this chapter. Chapter 2 highlighted the literature review on the significant impact of multiple regression, violation of assumption of constant variance, violation of assumption of independent, residual outliers, high leverage points, robust estimation methods, redescent M-estimation weighted function and the sampling procedure of wild bootstrap estimation methods of Wu's and Liu's. In Chapter 3, the strengths and limitation of these methods are discussed. The proposed modified robust wild bootstrap, modified robust PC with wild bootstrap and modified robust PLS with wild bootstrap methods using three different weight are also introduced in Chapter 3. Chapter 4 discussed the application of three different weighting procedures of modified robust wild bootstrap based on S-estimators and modified robust wild bootstrap based on MM-estimator. Chapter 4, also discussed the application of modified robust PC and PLS with wild bootstrap based on S-estimators and modified robust PC and PLS with wild bootstrap based on MM-estimators. The application of the existing robust wild bootstrap based on MM-estimators and wild bootstrap based on OLS estimator procedure are also presented in Chapter 4. Chapter 4 also contains the numerical comparisons that characterized the performance between the proposed methods and the existing methods. In order to rigorously describe the performance of the proposed methods with different weighting procedures, in Chapter 5, the performance of modified robust wild bootstrap methods and modified robust wild bootstrap of PC and PLS methods will be compared. The comparative analysis will be performed with wild bootstrap based on the OLS and the robust wild bootstrap methods of MM-estimator based on Wu's and Liu's using the case study. The conclusion of the research findings, discussion and recommendations of additional improvement for future research are done in Chapter 6.

1.6 Significance of the study

The wild bootstrap and robust wild bootstrap will provide the model that will produce the best parameter estimate of regression model. Existing wild bootstrap and Robust wild bootstrap handle the problems of heteroscedasticity and heteroscedasticity with residual outliers. The robust wild bootstrap of Generalize M-estimators (GM-estimators) which could conduct a simultaneous assessment on heteroscedasticity, residual outliers and high leverage points, as well as heteroscedasticity, multicollinearity, residual outliers and high leverage points, will cause such critical decision and conclusion be made with more precision and confidence. Previous studies showed that existing heteroscedasticity and outliers measures have their strengths and weaknesses. Most of the existing wild bootstrap method only focus on the problems of heteroscedasticity or heteroscedasticity with outliers. Thus, they are limited to one or two problems in the data sets. In contrast this study addresses three and four problems in the data sets. The behavior of each estimator was examined using simulation study, real data and modified real data. This could provide some guidelines so that researchers would be more conscious of which estimator or procedure to used especially when it involves parameter estimation in the presence of heteroscedasticity, residuals outliers and leverage points, or when heteroscedasticity, multicollinearity, residuals outliers and high leverage points, are present. It has been revealed that economic data usually contain heteroscedasticity, residual outliers and leverage points. Thus economic data will be used to illustrate the advantage of using the proposed methods. In this research, medical data is also used as most contains heteroscedasticity, multicollinearity, residual outliers and high leverage points, This data described the annual rates for varieties of domestic cigarettes according to their tar, nicotine, and carbon monoxide content. The main difference between this data and the previous data is the presence of multicollinearity. The efficiency of the model in the presence of heteroscedasticity, multicollinearity, residual outliers and high leverage points, is examined through comparative study. This will provide another beneficial guideline in deciding whether a particular estimation method is appropriate to be applied in parameter estimation of regression model whenever the historical data contains heteroscedasticity, multicollinearity, residual outliers and high leverage points.

This research attempts to fill this gap by examining whether the proposed methods in the presence of heteroscedasticity, residuals outlier and high leverage points measures such as RWGMBWu, and RWGMBLiu, can produce better parameter estimation of regression models. The proposed methods are based on the initial estimate of MM-estimator approach and the initial estimate of S-estimator techniques. Moreover, this research also examine the proposed methods in the presence of multicollinearity, heteroscedasticity, residual outliers and high leverage points, measures that include RPCGMBWu, RPCGMBLiu, RPLSGMBWu, and RPLSGMBLiu methods. The proposed methods are also based on two approaches. The first approach is GM-estimator Krasker and Welsch (1982) based on the initial estimate of MM-estimator and the second is GM-estimator based on the initial estimate of S-estimator. The approaches are expected to produce better parameter estimation of regression models. The existing methods considered for comparison are BootWu, BootLiu, RBootWu and RBootLiu. Simulation study is very important in order to identify the situation where the proposed methods performed best. However, according to Rana *et al.* (2012) the existing heteroscedasticity measures are based on OLS, so he introduced the robust heteroscedasticity measures which are resistant to heteroscedasticity and residual outliers. In contrast, the data in this research contains heteroscedasticity, residual outliers and high leverage points, as well as heteroscedasticity, multicollinearity, residual outliers and high leverage points. Moreover, the best model was also assessed through different measurement criteria, specifically in terms of bias and root mean squares errors (RMSE). Thus, it may provide a better conclusion in finding the most appropriate estimation method to estimate the economic and medical data. In the future, this proposed robust wild bootstrap method can be applied to other areas of research that contain data of such situations. Hence, it could establish a direction in research on linear regression model.

REFERENCES

Abbott, B., Abolins, M., Acharya, B., Adam, I., Adams, D., Adams, M., Ahn, S., Aihara, H., Alves, G., Amos, N. *et al.* (1998). Direct measurement of the top quark mass by the DØ collaboration. *Physical Review D*. 58(5), 052001.

Abdul-Wahab, S. A., Bakheit, C. S. and Al-Alawi, S. M. (2005). Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. *Environmental Modelling & Software*. 20(10), 1263–1271.

Adnan, N., Ahmad, M. H. and Adnan, R. (2006). A comparative study on some methods for handling multicollinearity problems. *Matematika*. 22(2), 109–119.

Adnan, R., Mohamad, M. N. and Setan, H. (2003). Multiple outliers detection procedures in linear regression. *Matematika*. 19(1), 29–45.

Aguilera, A. M., Escabias, M. and Valderrama, M. J. (2006). Using principal components for estimating logistic regression with high-dimensional multicollinear data. *Computational Statistics & Data Analysis*. 50(8), 1905–1924.

Ahmad, M. H., Adnan, R. and Adnan, N. (2006). A comparative study on some methods for handling multicollinearity problems. *Matematika*. 22, 109–119.

Al-Hassan, Y. M. (2010). Performance of a new ridge regression estimator. *Journal of the Association of Arab Universities for Basic and Applied Sciences*. 9(1), 23–26.

Alma, Ö. G. (2011). Comparison of robust regression methods in linear regression. *International Journal of Contemporary Mathematical Sciences*. 6(9), 409–421.

Andersen, R. (2008). *Modern methods for robust regression*. 152. Sage.

Andrews, D. F. (1974). A robust method for multiple linear regression. *Technometrics*. 16(4), 523–531.

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*. 2(2), 123–146.

Atkinson, A. C. (1985). *Plots, transformations, and regression: an introduction to graphical methods of diagnostic regression analysis*. Clarendon Press Oxford.

Bagheri, A. and Midi, H. (2011). On the performance of robust variance inflation factors. *International Journal of Agricultural and Statistical Sciences*. 7(1), 31–45.

Barnett, V. and Lewis, T. (1984). Outliers in statistical data. *Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics, Chichester: Wiley, 1984, 2nd ed.* 1.

Barnett, V. and Lewis, T. (1994). *Outliers in statistical data.* vol. 3. Wiley New York.

Beaton, A. E. and Tukey, J. W. (1974). The fitting of power series, meaning polynomials, illustrated on band-spectroscopic data. *Technometrics.* 16(2), 147–185.

Belsey, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity.* John Wiley.

Belsley, D. A. (1991). *Conditioning diagnostics.* Wiley Online Library.

Belsley, D. A., Kuh, E. and Roy, E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. *J. Wiley.*

Ben-Gal, I. (2005). Outlier detection. In *Data Mining and Knowledge Discovery Handbook.* (pp. 131–146). Springer.

Beran, R. (1988). Prepivoting test statistics: a bootstrap view of asymptotic refinements. *Journal of the American Statistical Association.* 83(403), 687–697.

Beringer, J., Arguin, J., Barnett, R., Copic, K., Dahl, O., Groom, D., Lin, C., Lys, J., Murayama, H., Wohl, C. *et al.* (2012). Review of particle physics. *Physical Review D.* 86(1).

Berkowitz, J. and Kilian, L. (2000). Recent developments in bootstrapping time series. *Econometric Reviews.* 19(1), 1–48.

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

Box, G. E. and Meyer, R. D. (1985). *Analysis of Unreplicated Factorials Allowing for Possibly Faulty Observations.* Technical report. DTIC Document.

Brem, R. B. and Kruglyak, L. (2005). The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proceedings of the National Academy of Sciences of the United States of America.* 102(5), 1572–1577.

Brown, R. L., Durbin, J. and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 149–192.

Brownstone, D. and Valletta, R. (2001). The bootstrap and multiple imputations: harnessing increased computing power for improved statistical tests. *Journal of Economic Perspectives*, 129–141.

Bühlmann, P. (1995). The blockwise bootstrap for general empirical processes of stationary sequences. *Stochastic Processes and their Applications*. 58(2), 247–265.

Burkhauser, R., Holden, K. C. and Myers, D. A. (1986). Marital disruption and poverty: The role of survey procedures in artificially creating poverty. *Demography*. 23(4), 621–631.

Çamdevýren, H., Demýr, N., Kanik, A. and Keskýn, S. (2005). Use of principal component scores in multiple linear regression models for prediction of Chlorophyll-a in reservoirs. *Ecological Modelling*. 181(4), 581–589.

Candès, E. J., Li, X., Ma, Y. and Wright, J. (2011). Robust principal component analysis? *Journal of the ACM (JACM)*. 58(3), 11.

Carroll, R. J. and Ruppert, D. (1988). *Transformation and weighting in regression*. vol. 30. CRC Press.

Chesher, A. and Smith, R. J. (1997). Likelihood ratio specification tests. *Econometrica: Journal of the Econometric Society*, 627–646.

Choi, E. and Hall, P. (2000). Bootstrap confidence regions computed from autoregressions of arbitrary order. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 62(2), 461–477.

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American statistical association*. 74(368), 829–836.

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

Cook, R. D. (1977). Detection of influential observation in linear regression. *Technometrics*, 15–18.

Cook, R. D. and Weisberg, S. (1983). Diagnostics for heteroscedasticity in regression. *Biometrika*. 70(1), 1–10.

Costa, C., Menesatti, P., Spinelli, R. *et al.* (2012). Performance modelling in forest operations through partial least square regression. *Silva Fennica*. 46(2), 241–252.

Cribari-Neto, F. (1997). Econometric programming environments: GAUSS, Ox and S-PLUS. *Journal of Applied Econometrics*. 12(1), 77–89.

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics & Data Analysis*. 45(2), 215–233.

Cribari-Neto, F. and Ferrari, S. L. (1995). Second order asymptotics for score tests in generalised linear models. *Biometrika*. 82(2), 426–432.

Cribari-Neto, F., Ferrari, S. L. and Cordeiro, G. M. (2000). Improved heteroscedasticity-consistent covariance matrix estimators. *Biometrika*. 87(4), 907–918.

Cribari-Neto, F., Frery, A. C. and Silva, M. F. (2002). Improved estimation of clutter properties in speckled imagery. *Computational statistics & data analysis*. 40(4), 801–824.

Cribari-Neto, F. and Galvao, N. M. (2003). A class of improved heteroskedasticity-consistent covariance matrix estimators. *Communications in Statistics-Theory and Methods*. 32(10), 1951–1980.

Cribari-Neto, F., Zarkos, S. G. *et al.* (1999). R: Yet another econometric programming environment. *Journal of Applied Econometrics*. 14(3), 319–329.

Davidson, J. (2002). A model of fractional cointegration, and tests for cointegration using the bootstrap. *Journal of Econometrics*. 110(2), 187–212.

Davidson, R. and Flachaire, E. (2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*. 141(1), 141–166.

Davison, A. C., Hinkley, D. V. and Young, G. A. (2003). Recent developments in bootstrap methodology. *Statistical Science*, 141–157.

De Jong, S. (1993). SIMPLS: an alternative approach to partial least squares regression. *Chemometrics and intelligent laboratory systems*. 18(3), 251–263.

De Long, J. B. and Summers, L. H. (1990). *Equipment investment and economic growth*. Technical report. National Bureau of Economic Research.

Dibbern, J., Goles, T., Hirschheim, R. and Jayatilaka, B. (2004). Information systems outsourcing: a survey and analysis of the literature. *ACM SIGMIS Database*. 35(4), 6–102.

Diblasi, A. and Bowman, A. (1997). Testing for constant variance in a linear model. *Statistics & probability letters*. 33(1), 95–103.

Dodge, Y. (1987). An introduction to LI-norm based statistical data analysis. *Computational Statistics & Data Analysis*. 5(4), 239–253.

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

Drapper, N. R. and Smith, H. (1998). *Applied regression analysis*. John Wiley & Sons.

Duggan, M. *et al.* (2010). ECON 420-01 Econometrics.

Edgeworth, F. Y. (1887). On observations relating to several quantities. *Hermathena*. 6(13), 279–285.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *The annals of Statistics*, 1–26.

Efron, B. and Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. vol. 38. SIAM.

Eicker, F. (1967). Limit theorems for regressions with unequal and dependent errors. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, vol. 1. 59–82.

Eicker, F. *et al.* (1963). Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *The Annals of Mathematical Statistics*. 34(2), 447–456.

Erbacher, R. F., Walker, K. L. and Frincke, D. A. (2002). Intrusion and misuse detection in large-scale systems. *Computer Graphics and Applications, IEEE*. 22(1), 38–47.

Escabias, M., Aguilera, A. M. and Valderrama, M. J. (2004). Principal component estimation of functional logistic regression: discussion of two different approaches. *Journal of Nonparametric Statistics*. 16(3-4), 365–384.

Farrar, D. E. and Glauber, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *The Review of Economic and Statistics*, 92–107.

Feng, X., He, X. and Hu, J. (2011). Wild bootstrap for quantile regression. *Biometrika*. 98(4), 995–999.

Ferguson, T. S. (1961). Rules for rejection of outliers. *Revue de l'Institut International de Statistique*, 29–43.

Flachaire, E. (1999). A better way to bootstrap pairs. *Economics Letters*. 64(3), 257–262.

Flachaire, E. (2002). Bootstrapping heteroskedasticity consistent covariance matrix estimator. *Computational Statistics*. 17(4), 501–506.

Flachaire, E. (2005). Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. *Computational Statistics & Data Analysis*. 49(2), 361–376.

Fornell, C. and Robinson, W. T. (1983). Industrial organization and consumer satisfaction/dissatisfaction. *Journal of Consumer Research*, 403–412.

Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*. 8(15), 1–27.

Freedman, D. A. and Peters, S. C. (1984). Bootstrapping a regression equation: Some empirical results. *Journal of the American Statistical Association*. 79(385), 97–106.

Freund, R. and Wilson, W. (1998). Regression analysis: statistical modeling of a response variableAcademic Press. *New York.*

Furno, M. (1997). A robust heteroskedasticity consistent covariance matrix estimator. *Statistics: A Journal of Theoretical and Applied Statistics.* 30(3), 201–219.

Garibaldi, U. and Penco, M. (1985). Probability theory and physics between bernoulli and laplace: The contribution of jh lambert (1728-1777). In *Proc. Fifth National Congress on the History of Physics*, vol. 9. 341–346.

Gentleman, J. and Wilk, M. (1975). Detecting outliers. II. Supplementing the direct analysis of residuals. *Biometrics*, 387–410.

Gladwell, M. (2008). *Outliers: The story of success.* Hachette UK.

Godfrey, L. G. and Orme, C. D. (2004). Controlling the finite sample significance levels of heteroskedasticity-robust tests of several linear restrictions on regression coefficients. *Economics Letters.* 82(2), 281–287.

Goldfeld, S. M. and Quandt, R. E. (1965). Some tests for homoscedasticity. *Journal of the American statistical Association.* 60(310), 539–547.

Gonçalves, S. and Kilian, L. (2007). Asymptotic and bootstrap inference for AR (?) processes with conditional heteroskedasticity. *Econometric Reviews.* 26(6), 609–641.

Greene, W. H. (2000). Econometric analysis (International edition).

Grissom, R. J. (2000). Heterogeneity of variance in clinical data. *Journal of consulting and clinical psychology.* 68(1), 155.

Gujarati, D. N. and Porter, D. (2009). Basic Econometrics Mc Graw-Hill International Edition.

Habshah, M., Norazan, M. and Rahmatullah Imon, A. (2009). The performance of diagnostic-robust generalized potentials for the identification of multiple high leverage points in linear regression. *Journal of Applied Statistics.* 36(5), 507–520.

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

Hadi, A. S. (1994). A modification of a method for the detection of outliers in multivariate samples. *Journal of the Royal Statistical Society. Series B (Methodological)*, 393–396.

Hadi, A. S., Imon, A. and Werner, M. (2009). Detection of outliers. *Wiley Interdisciplinary Reviews: Computational Statistics.* 1(1), 57–70.

Hadi, A. S. and Simonoff, J. S. (1993). Procedures for the identification of multiple outliers in linear models. *Journal of the American Statistical Association*. 88(424), 1264–1272.

Hadi, Y., Essannouni, F. and Thami, R. O. H. (2006). Video summarization by k-medoid clustering. In *Proceedings of the 2006 ACM symposium on Applied computing*. ACM, 1400–1401.

Halunga, A., Orme, C. D. and Yamagata, T. (2011). A heteroskedasticity robust Breusch-Pagan test for contemporaneous correlation in dynamic panel data models. *University of Manchester Economics Discussion Paper Series*. (1118).

Hampel, F. (1997). What can the foundations discussion contribute to data analysis? And what may be some of the future directions in robust methods and data analysis? *Journal of Statistical Planning and Inference*. 57(1), 7–19.

Hampel, F., Ronchetti, E., Rousseeuw, P. and Stahel, W. (1986). Robust statistics, J. *Wiley& Sons, New York*.

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

Hampel, F. R. (1974). The influence curve and its role in robust estimation. *Journal of the American Statistical Association*. 69(346), 383–393.

Han, K.-H. and Kim, J.-H. (2004). Quantum-inspired evolutionary algorithms with a new termination criterion, H ε gate, and two-phase scheme. *Evolutionary Computation, IEEE Transactions on*. 8(2), 156–169.

Härdle, W., Horowitz, J. and Kreiss, J.-P. (2003). Bootstrap methods for time series. *International Statistical Review*. 71(2), 435–459.

Harrison, M. J. and McCabe, B. P. (1979). A test for heteroscedasticity based on ordinary least squares residuals. *Journal of the American Statistical Association*. 74(366a), 494–499.

Harvey, A. C. and Phillips, G. D. A. (1974). A comparison of the power of some tests for heteroskedasticity in the general linear model. *Journal of Econometrics*. 2(4), 307–316.

Hawkins, D. M. (1980). *Identification of outliers*. vol. 11. Springer.

Hawkins, S., He, H., Williams, G. and Baxter, R. (2002). Outlier detection using replicator neural networks. In *Data warehousing and knowledge discovery*. (pp. 170–180). Springer.

Helland, I. S. (1988). On the structure of partial least squares regression. *Communications in statistics-Simulation and Computation*. 17(2), 581–607.

Higgins, C. A., Duxbury, L. E. and Irving, R. H. (1992). Work-family conflict in the dual-career family. *Organizational Behavior and Human Decision Processes*. 51(1), 51–75.

Hinkley, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics*. 19(3), 285–292.

Ho, P. and Fung, D. (1991). Error performance of multiple symbol differential detection of PSK signals transmitted over correlated Rayleigh fading channels. In *Communications, 1991. ICC'91, Conference Record. IEEE International Conference on*. IEEE, 568–574.

Hoaglin, D. C., Mosteller, F. and Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. vol. 3. Wiley New York.

Hocking, R. R. (2013). *Methods and applications of linear models: regression and the analysis of variance*. John Wiley & Sons.

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

Horn, S. D., Horn, R. A. and Duncan, D. B. (1975). Estimating heteroscedastic variances in linear models. *Journal of the American Statistical Association*. 70(350), 380–385.

Horowitz, J. L. (1997). Bootstrap methods in econometrics: theory and numerical performance. *Econometric Society Monographs*. 28, 188–222.

Horowitz, J. L., Lobato, I. N., Nankervis, J. C. and Savin, N. (2006). Bootstrapping the Box–Pierce Q test: a robust test of uncorrelatedness. *Journal of Econometrics*. 133(2), 841–862.

Hössjer, O. (1992). On the optimality of S-estimators. *Statistics & probability letters*. 14(5), 413–419.

Huber, P. J. (1968). Robust confidence limits. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*. 10(4), 269–278.

Huber, P. J. (1973). Robust regression: asymptotics, conjectures and Monte Carlo. *The Annals of Statistics*, 799–821.

Huber, P. J. and Ronchetti, E. (2009). Robust Statistics, Hoboken. *NJ: Wiley. doi.* 10(1002), 9780470434697.

Huber, P. J. *et al.* (1964a). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*. 35(1), 73–101.

Huber, P. J. *et al.* (1964b). Robust estimation of a location parameter. *The Annals of Mathematical Statistics*. 35(1), 73–101.

Hubert, M. and Branden, K. V. (2003). Robust methods for partial least squares regression. *Journal of Chemometrics*. 17(10), 537–549.

Hubert, M., Rousseeuw, P. J. and Van Aelst, S. (2008). High-breakdown robust multivariate methods. *Statistical Science*, 92–119.

Hulland, J. (1999). Use of partial least squares (PLS) in strategic management research: A review of four recent studies. *Strategic management journal*. 20(2), 195–204.

Imon, A. and Ali, M. M. (2005). Bootstrapping regression residuals. *Journal of the Korean Data and Information Science Society*. 16(3), 665–682.

Jackson, J. E. (2005). *A user's guide to principal components*. vol. 587. John Wiley & Sons.

Ji, H., Ramsey, M. R., Hayes, D. N., Fan, C., McNamara, K., Kozlowski, P., Torrice, C., Wu, M. C., Shimamura, T., Perera, S. A. *et al.* (2007). LKB1 modulates lung cancer differentiation and metastasis. *Nature*. 448(7155), 807–810.

Johnson, R. A., Wichern, D. W. *et al.* (1992). *Applied multivariate statistical analysis*. vol. 4. Prentice hall Englewood Cliffs, NJ.

Kafadar, K. and Morris, M. (2002). Data-based detection of potential terrorist attacks. *White Paper*.

Kallel, R., Cottrell, M. and Vigneron, V. (2002a). Bootstrap for neural model selection. *Neurocomputing*. 48(1), 175–183.

Kallel, R., Cottrell, M. and Vigneron, V. (2002b). Bootstrap for neural model selection. *Neurocomputing*. 48(1), 175–183.

Kline, P. and Santos, A. (2012). A score based approach to wild bootstrap inference. *Journal of Econometric Methods*. 1(1), 23–41.

Knaub Jr, J. R. (1996). Weighted Multiple Regression Estimation for Survey Model Sampling. *InterStat*.

Koenker, R. (1981). A note on studentizing a test for heteroscedasticity. *Journal of Econometrics*. 17(1), 107–112.

Koutsofios, E. E., North, S. C. and Keim, D. A. (1999). Visualizing large telecommunication data sets. *Computer Graphics and Applications, IEEE*. 19(3), 16–19.

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

Kupek, E., Petrou, S., Vause, S. and Maresh, M. (2002). Clinical, provider and sociodemographic predictors of late initiation of antenatal care in England and Wales. *BJOG: An International Journal of Obstetrics & Gynaecology*. 109(3), 265–273.

Laitinen, E. K. (2006). Partial least squares regression in payment default prediction. *Investment Management and Financial Innovations*. 3(1), 64–77.

Lemonte, A. J., Cribari-Neto, F. and Vasconcellos, K. L. (2007). Improved statistical inference for the two-parameter Birnbaum–Saunders distribution. *Computational Statistics & Data Analysis*. 51(9), 4656–4681.

Leroy, A. M. and Rousseeuw, P. J. (1987). Robust regression and outlier detection. *Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1987*. 1.

Lim, C., Sen, P. K. and Peddada, S. D. (2010). Statistical inference in nonlinear regression under heteroscedasticity. *Sankhya B*. 72(2), 202–218.

Liu, H., Shah, S. and Jiang, W. (2004). On-line outlier detection and data cleaning. *Computers & chemical engineering*. 28(9), 1635–1647.

Liu, R. Y. *et al.* (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics*. 16(4), 1696–1708.

Long, J. S. and Ervin, L. H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician*. 54(3), 217–224.

Lorber, A., Wangen, L. E. and Kowalski, B. R. (1987). A theoretical foundation for the PLS algorithm. *Journal of Chemometrics*. 1(1), 19–31.

MacKinnon, J. G. (2002). Bootstrap inference in econometrics. *Canadian Journal of Economics/Revue canadienne d'économique*. 35(4), 615–645.

MacKinnon, J. G. (2006). Bootstrap methods in econometrics*. *Economic Record*. 82(s1), S2–S18.

Maitra, S. and Yan, J. (2008). Principle component analysis and partial least squares: Two dimension reduction techniques for regression. *Applying Multivariate Statistical Models*. 79.

Mallows, C. L. (1975). On some topics in robustness. *Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ*.

Mammen, E. (1993). Bootstrap and wild bootstrap for high dimensional linear models. *The Annals of Statistics*, 255–285.

Marasinghe, M. G. (1985). A multistage procedure for detecting several outliers in linear regression. *Technometrics*. 27(4), 395–399.

Maronna, R., Martin, D. and Yohai, V. (2006). *Robust statistics*. John Wiley & Sons, Chichester. ISBN.

Marquardt, D. W. and Snee, R. D. (1975). Ridge regression in practice. *The American Statistician*. 29(1), 3–20.

Massy, W. F. (1965). Principal components regression in exploratory statistical research. *Journal of the American Statistical Association*. 60(309), 234–256.

McCabe, B. (1986). Testing for heteroscedasticity occuring at unknown points. *Communications in Statistics-Theory and Methods*. 15(5), 1597–1813.

McCullough, B. (1996). Consistent forecast intervals when the forecast-period exogenous variables are stochastic. *Journal of Forecasting*. 15(4), 293–304.

Midi, H. and Bagheri, A. (2010). Robust multicollinearity diagnostic measure in collinear data set. In *Proceedings of the 4th international conference on Applied mathematics, simulation, modelling*. World Scientific and Engineering Academy and Society (WSEAS), 138–142.

Midi, H., Bagheri, A. and Imon, A. (2010). The Application of Robust Multicollinearity Diagnostic Method Based on Robust Coefficient Determination to a Non-Collinear Data. *Journal of Applied Sciences*. 10(8), 611–619.

Midi, H., Rana, S. and Imon, A. (2009a). The performance of robust weighted least squares in the presence of outliers and heteroscedastic errors. *WSEAS Transactions on Mathematics*. 7(8), 351–361.

Midi, H., Shabbak, A., Talib, B. A. and Hassan, M. N. (2009b). Multivariate control chart based on robust estimator. *Journal of Quality Measurement and Analysis JQMA*. 5(1), 17–33.

Midi, H., Uraibi, H. S. and A Talib, B. (2009c). Dynamic robust bootstrap method based on LTS estimators. *European Journal of Scientific Research*. 32(3), 277–287.

Miller, R. G. and Halpern, J. W. (1980). Robust estimators for quantal bioassay. *Biometrika*. 67(1), 103–110.

Mohammadi, M., Midi, H., Arasan, J. and Al-Talib, B. (2011). High breakdown estimators to robustify phase II multivariate control charts. *Journal of Applied Sciences*. 11(3), 503–511.

Moore, D. S., Dawson, L. G. and McCabe, G. P. (2005). *Introduction to the Practice of Statistics Excel Manual with Macros*. Macmillan.

Mosteller, F. and Tukey, J. W. (1977). Data analysis and regression: a second course in statistics. *Addison-Wesley Series in Behavioral Science: Quantitative Methods.*

Muhlbauer, A., Spichtinger, P. and Lohmann, U. (2009). Application and comparison of robust linear regression methods for trend estimation. *Journal of Applied Meteorology and Climatology.* 48(9), 1961–1970.

Muller, H.-G. and Zhao, P.-L. (1995). On a semiparametric variance function model and a test for heteroscedasticity. *The Annals of Statistics*, 946–967.

Myers, R. H., Montgomery, D. C., Vining, G. G. and Robinson, T. J. (2012). *Generalized linear models: with applications in engineering and the sciences.* vol. 791. John Wiley & Sons.

Neter, J., Wasserman, W. and Kutner, M. H. (1989). Applied linear regression models.

Nonneman, W. and Vanhoudt, P. (1996). A further augmentation of the Solow model and the empirics of economic growth for OECD countries. *The Quarterly Journal of Economics*, 943–953.

Noonan, R. and Wold, H. (1980). PLS Path Modelling with Latent Variables: Analysing School Survey Data Using Partial Least Squares-Part II 1, 2. *Scandinavian Journal of Educational Research.* 24(1), 1–24.

Norazan, M., Midi, H. and Imon, A. (2009). Estimating regression coefficients using weighted bootstrap with probability. *WSEAS Transactions on Mathematics.* 8(7), 362–371.

Ohtani, K. and Toyoda, T. (1980). Estimation of regression coefficients after a preliminary test for homoscedasticity. *Journal of Econometrics.* 12(2), 151–159.

Okafor, R. (1990). A biweight approach to estimate currency exchange rate: The nigerian example. *Journal of Applied Statistics.* 17(1), 73–82.

Pavlou, P. A. and Chai, L. (2002). What Drives Electronic Commerce across Cultures? Across-Cultural Empirical Investigation of the Theory of Planned Behavior. *J. Electron. Commerce Res.* 3(4), 240–253.

Peña, D. and Prieto, F. J. (2012). Multivariate outlier detection and robust covariance matrix estimation. *Technometrics.*

Phatak, A. and De Jong, S. (1997). The geometry of partial least squares. *Journal of Chemometrics.* 11(4), 311–338.

Pison, G., Rousseeuw, P. J., Filzmoser, P. and Croux, C. (2003). Robust factor analysis. *Journal of Multivariate Analysis.* 84(1), 145–172.

Politis, D. N. *et al.* (2003). The impact of bootstrap methods on time series analysis. *Statistical Science*. 18(2), 219–230.

Prescott, P. (1975). An approximate test for outliers in linear models. *Technometrics*. 17(1), 129–132.

Rahmatullah Imon, A. (2009). Deletion residuals in the detection of heterogeneity of variances in linear regression. *Journal of Applied Statistics*. 36(3), 347–358.

Rai, P. K., Pareek, S. and Joshi, H. (2013). Regression Analysis of Collinear Data using rk Class Estimator: Socio-Economic and Demographic Factors Affecting the Total Fertility Rate (TFR) in India. *Journal of Data Science*. 11(2), 323–342.

Rajab, J. M., MatJafri, M. and Lim, H. (2013). Combining multiple regression and principal component analysis for accurate predictions for column ozone in Peninsular Malaysia. *Atmospheric Environment*. 71, 36–43.

Rana, M. S., Midi, H. and Imon, A. R. (2008). A robust modification of the Goldfeld-Quandt test for the detection of heteroscedasticity in the presence of outliers. *Journal of mathematics and Statistics*. 4(4), 277.

Rana, S., Midi, H. and Imon, A. (2012). Robust Wild Bootstrap for Stabilizing the Variance of Parameter Estimates in Heteroscedastic Regression Models in the Presence of Outliers. *Mathematical Problems in Engineering*. 2012.

Rao, P. V., Deng, P., Maddala, R., Epstein, D. L., Li, C.-Y. and Shimokawa, H. (2005). Expression of dominant negative Rho-binding domain of Rho-kinase in organ cultured human eye anterior segments increases aqueous humor outflow. *Mol Vis*. 11, 288–297.

Rasmussen, J. L. (1989). Parametric and non-parametric analysis of groups by trials design under variance-covariance inhomogeneity. *British Journal of Mathematical and Statistical Psychology*. 42(1), 91–102.

Reinartz, W., Krafft, M. and Hoyer, W. D. (2004). The customer relationship management process: Its measurement and impact on performance. *Journal of marketing research*. 41(3), 293–305.

Rey, W. J. (1983). *Introduction to robust and quasi-robust statistical methods*. Springer.

Rocke, D. M. (1996). Robustness properties of S-estimators of multivariate location and shape in high dimension. *The Annals of statistics*, 1327–1345.

Rocke, D. M. and Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*. 91(435), 1047–1061.

Rosseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimates. *Lecture Notes in Statistics*. 26, 256–272.

Rousseeuw, F. H. E. R. P., Hampel, F., Ronchetti, E. and Stahel, W. (1986). Robust statistics: the approach based on influence functions. *J. Wiley, New York*.

Rousseeuw, P. and Yohai, V. (1984). Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis*. (pp. 256–272). Springer.

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

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

Rousseeuw, P. J. (1994). Unconventional features of positive-breakdown estimators. *Statistics & Probability Letters*. 19(5), 417–431.

Rousseeuw, P. J. and Leroy, A. M. (2005). *Robust regression and outlier detection*. vol. 589. John Wiley & Sons.

Rousseeuw, P. J. and Van Driessen, K. (2006). Computing LTS regression for large data sets. *Data mining and knowledge discovery*. 12(1), 29–45.

Rousseeuw, P. J. and Van Zomeren, B. C. (1990). Unmasking multivariate outliers and leverage points. *Journal of the American Statistical Association*. 85(411), 633–639.

Rousseeuw, P. J. and van Zomeren, B. C. (1991). Robust distances: simulations and cutoff values. In *Directions in Robust Statistics and Diagnostics*. (pp. 195–203). Springer.

Ryan, T. P. (2008). *Modern regression methods*. vol. 655. John Wiley & Sons.

Salibian-Barrera, M. and Zamar, R. H. (2002). Bootstrapping robust estimates of regression. *Annals of Statistics*, 556–582.

Samkar, H. and Alpu, O. (2010). Ridge regression based on some robust estimators. *Journal of Modern Applied Statistical Methods*. 9(2), 17.

Sarkar, S. K., Midi, H. and Rana, M. (2011). Detection of outliers and influential observations in binary logistic regression: An empirical study. *Journal of Applied Sciences*. 11(1), 26–35.

Schervish, M. J. (2012). *Theory of statistics*. Springer Science & Business Media.

Shao, J. (1988). On resampling methods for variance and bias estimation in linear models. *The Annals of Statistics*, 986–1008.

Sharma, A. and Paliwal, K. K. (2007). Fast principal component analysis using fixed-point algorithm. *Pattern Recognition Letters*. 28(10), 1151–1155.

Shevlyakov, G. L. and Vilchevski, N. O. (2001). *Robustness in Data Analysis: criteria and methods*. Walter de Gruyter.

Simpson, D. G. and Chang, Y.-C. I. (1997). Reweighting approximate GM estimators: asymptotics and residual-based graphics. *Journal of Statistical Planning and Inference*. 57(2), 273–293.

Simpson, J. R. (1995a). *New methods and comparative evaluations for robust and biased-robust regression estimation*. Technical report. DTIC Document.

Simpson, J. R. (1995b). *New methods and comparative evaluations for robust and biased-robust regression estimation*. Technical report. DTIC Document.

Simpson, J. R. and Montgomery, D. C. (1998). A robust regression technique using compound estimation. *Naval Research Logistics (NRL)*. 45(2), 125–139.

Smyth, G. K., Huele, A. F. and Verbyla, A. P. (2001). Exact and approximate REML for heteroscedastic regression. *Statistical modelling*. 1(3), 161–175.

Sousa, S., Martins, F., Alvim-Ferraz, M. and Pereira, M. C. (2007). Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling & Software*. 22(1), 97–103.

Stanley, T., Doucouliagos, H., Giles, M., Heckemeyer, J. H., Johnston, R. J., Laroche, P., Nelson, J. P., Paldam, M., Poot, J., Pugh, G. *et al.* (2013). Meta-analysis of economics research reporting guidelines. *Journal of Economic Surveys*. 27(2), 390–394.

Staudte, R. and Sheather, S. (1990). Robust estimation and testing: Wiley series in probability and mathematical statistics. *New York*.

Steel, P. and Fay, R. E. (1995). Variance estimation for finite populations with imputed data. In *Proceedings of the Section on Survey Research Methods, American Statistical Association*. Citeseer.

Stigler, G. J. (1973). General economic conditions and national elections. *The American Economic Review*, 160–167.

Stine, R. A. (1985). Bootstrap prediction intervals for regression. *Journal of the American Statistical Association*. 80(392), 1026–1031.

Stone, M. and Brooks, R. J. (1990). Continuum regression: cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal components regression. *Journal of the Royal Statistical Society. Series B (Methodological)*, 237–269.

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

Statistician. 42(2), 152–154.

Swallow, W. H. and Kianifard, F. (1996). Using robust scale estimates in detecting multiple outliers in linear regression. *Biometrics*, 545–556.

Thomas, P. R. (1997). Modern regression methods. *J. Wiley, New York, Chichester*.

Tietjen, G. L., Moore, R. and Beckman, R. (1973). Testing for a single outlier in simple linear regression. *Technometrics*. 15(4), 717–721.

Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *Soft Computing*. 8(8), 527–533.

Tukey, J. W. (1953). SECTION OF MATHEMATICS AND ENGINEERING: SOME SELECTED QUICK AND EASY METHODS OF STATISTICAL ANALYSIS*. *Transactions of the New York Academy of Sciences*. 16(2 Series II), 88–97.

Tukey, J. W. (1960). Conclusions vs decisions. *Technometrics*. 2(4), 423–433.

Tukey, J. W. and Tukey, P. A. (1988). Computer graphics and exploratory data analysis: An introduction. *The Collected Works of John W. Tukey: Graphics: 1965-1985*. 5, 419.

Uraibi, H. S. (2009). *Dynamic Robust Bootstrap Algorithm for Linear Model Selection Using Least Trimmed Squares*. Ph.D. Thesis. Universiti Putra Malaysia.

VARGAS N, J. A. (2003). Robust estimation in multivariate control charts for individual observations. *Journal of Quality Technology*. 35(4), 367–376.

Víšek, J. Á. (2011). Consistency of the least weighted squares under heteroscedasticity. *Kybernetika*. 47(2), 179–206.

Weisberg, S. (2005). *Applied linear regression*. vol. 528. John Wiley & Sons.

Werner, M. (2003). *Identification of multivariate outliers in large data sets*. Ph.D. Thesis. Citeseer.

White, H. (1980a). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica: Journal of the Econometric Society*, 817–838.

White, R. H. (1980b). Stable isotope studies on the biosynthesis of lipoic acid in Escherichia coli. *Biochemistry*. 19(1), 15–19.

White, R. L. (1985). Synchrotron emission from chaotic stellar winds. In *Radio Stars*. (pp. 45–46). Springer.

Wilcox, D. C. *et al.* (1998). *Turbulence modeling for CFD*. vol. 2. DCW industries La Canada, CA.

Wilcox, N. T. (2010). *A comparison of three probabilistic models of binary discrete choice under risk*. Technical report. working paper.

Wilcox, R. R. (1996). Confidence intervals for two robust regression lines with a heteroscedastic error term. *British Journal of Mathematical and Statistical Psychology*. 49(1), 163–170.

Wilcox Rand, R. (2005). *Introduction to robust estimation and hypothesis testing*. Elsevier academic Press.

Willems, G. and Van Aelst, S. (2005). Fast and robust bootstrap for LTS. *Computational statistics & data analysis*. 48(4), 703–715.

Wisnowski, J. W. (1999). *Multiple outliers in linear regression: advances in detection methods, robust estimation, and variable selection*. Technical report. DTIC Document.

Wisnowski, J. W., Simpson, J. R., Montgomery, D. C. and Runger, G. C. (2003). Resampling methods for variable selection in robust regression. *Computational Statistics & Data Analysis*. 43(3), 341–355.

Wißmann, M., Toutenburg, H. *et al.* (2007). Role of categorical variables in multicollinearity in the linear regression model.

Wold, S., Ruhe, A., Wold, H. and Dunn, W., III (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*. 5(3), 735–743.

Wu, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *the Annals of Statistics*, 1261–1295.

Yang, Y.-C., Ciarletta, A. B., Temple, P. A., Chung, M. P., Kovacic, S., Witek-Giannotti, J. S., Leary, A. C., Kriz, R., Donahue, R. E., Wong, G. G. *et al.* (1986). Human IL-3 (multi-CSF): identification by expression cloning of a novel hematopoietic growth factor related to murine IL-3. *Cell*. 47(1), 3–10.

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

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

Zhu, H., Ibrahim, J. G., Tang, N., Rowe, D. B., Hao, X., Bansal, R. and Peterson, B. S. (2007). A statistical analysis of brain morphology using wild bootstrapping. *Medical Imaging, IEEE Transactions on*. 26(7), 954–966.