UNIVERSITI PUTRA MALAYSIA

WEIGHTED MAXIMUM MEDIAN LIKELIHOOD ESTIMATION FOR PARAMETERS IN MULTIPLE LINEAR REGRESSION MODEL

NORAZAN BINTI MOHAMED RAMLI

FS 2008 33

WEIGHTED MAXIMUM MEDIAN LIKELIHOOD ESTIMATION FOR PARAMETERS IN MULTIPLE LINEAR REGRESSION MODEL

By

NORAZAN BINTI MOHAMED RAMLI

Thesis Submitted to the School of Graduate Studies, Universiti Putra Malaysia, in Fulfilment of the Requirements for the Degree of Doctor of Philosophy

May 2008

Dedicated to:

my husband
my parents
my children

Abstract of the thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of requirements for the degree of Doctor of Philosophy

WEIGHTED MAXIMUM MEDIAN LIKELIHOOD ESTIMATION FOR PARAMETERS IN MULTIPLE LINEAR REGRESSION MODEL

By

NORAZAN BINTI MOHAMED RAMLI

May 2008

Chairman: Associate Professor Habshah Midi, PhD
Faculty: Science

The performance of the Maximum Median Likelihood Estimator (MML) proposed by Hao (1992) is very inconsistent and sensitive to outliers, which results in biased parameter estimates. We propose Weighted Maximum Median Likelihood (WMML) estimators as alternatives. The basic idea in the WMML estimation is to isolate outliers from the majority of the observations and use only a certain number of fittest observations to estimate parameters. We study in details the performance of the WMML estimators in real and simulated data sets. The WMML estimates are consistent and can be as good as the MLE estimates in outlier free data sets and more efficient for data sets with multiple outliers than the MLE and MML estimates.

The research also develops a diagnostic method for the identification of high-leverage points. We realize that often their presence in data set gives adverse effect on the inference. We propose Diagnostic-Robust Generalized Potentials (DRGP)

technique as an alternative approach that performs well relative to current techniques.

The WMML estimators also function as indirect methods for identifying multiple outliers in data sets. Visual analysis of the WMML estimates shows that the estimators can be a reliable method to identify multiple outliers in linear regression.

We also propose Transformed Both Sides (TBS) Robust Based estimators, namely the TBS-WMML1 Based estimator, the TBS-WMML2 Based estimator and the TBS-WMML3 Based estimator for data sets with problems of outliers and non-constant variance error terms. The problem of non-constant variance error terms is also known as heteroscedastic problem. To induce homoscedasticity for data sets with outliers, TBS-Robust Based estimators are used. The resulting estimates are expected to have constant variance and the resulting model is called the TBS model with constant variance error terms. Our analysis shows that the TBS-Robust Based estimators provide estimates with lower variance than the MLE and the MML methods when both problems of outliers and non-constant variance errors exist.

The thesis also checks the variability of the WMML estimators using bootstrap procedures. Current bootstrap procedures such as Fixed-x Resampling, Random-x Resampling and Diagnostic Before Bootstrap are not robust to outliers. To accommodate this problem we propose a new bootstrap procedure, which we call as

Weighted Bootstrap with Probability (WBP). In the WBP procedure, outlying observations are attributed with low probabilities and consequently with low chances of being selected in the re-sampling process. Simulation results show that in most instances the WBP is more robust against a given number of arbitrary outliers than the current bootstrap procedures.

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia
sebagai memenuhi keperluan untuk Ijazah Doktor Falsafah

PENGANGGARAN KEBOLEHJADIAN MAKSIMUM MEDIAN BERPEMBERAT DALAM MODEL REGRESI LINEAR BERGANDA

Oleh

NORAZAN BINTI MOHAMED RAMLI

Mei 2008

Pengerusi: **Profesor Madya Habshah Midi, PhD**
Fakulti: **Sains**

Prestasi kaedah penganggaran Kebolehjadian Maksimum Median (MML) yang dicadangkan oleh Hao (1992) boleh menjadi tidak konsisten dan sensitif kepada data terpencil menyebabkan pincangan dalam anggaran parameter. Kami mencadangkan kaedah penganggaran Kebolehjadian Maksimum Median Berpemberat (WMML) sebagai alternatif. Secara asasnya, penganggaran Kebolehjadian Maksimum Median Berpemberat memisahkan data terpencil dan penganggaran hanya berasaskan data terbaik. Beberapa contoh numerik dan kajian simulasi telah dijalankan untuk menguji keteguhan kaedah WMML. Keputusan yang diperolehi menunjukkan, tanpa kehadiran titik terpencil, anggaran WMML adalah konsisten dan sebaik penganggaran MLE dan lebih efisyen berbanding MLE dan MML untuk data yang mengandungi titik terpencil berganda.

Kami ini juga mencadangkan satu kaedah diagnosis untuk mengenal pasti titik "leverage". Kehadiran titik "leverage" memburukkan takbiran. Kami mencadangkan kaedah Diagnosis Potensi Teritlak Teguh (DRGP) sebagai alternatif yang menunjukkan prestasi yang baik berbanding dengan beberapa kaedah yang sedia ada.

Kaedah WMML juga berfungsi sebagai kaedah tidak langsung untuk mengenal pasti titik terpencil. Analisis visual menunjukkan anggaran WMML juga boleh berfungsi sebagai satu kaedah untuk mengenal pasti titik terpencil dalam model regresi linear.

Kami juga mencadangkan kaedah Merubah Kedua Sisi Berasaskan Penganggar Teguh, iaitu TBS-WMML1, TBS-WMML2 dan TBS-WMML3 untuk data yang mempunyai masalah ralat varians yang tidak konsisten dan titik terpencil. Masalah ralat varians yang tidak konsisten ini juga dikenali sebagai ralat heteroskedasti. Kaedah Merubah Kedua Sisi Berasaskan Penganggar Teguh (TBS-Robust Based estimator) digunakan untuk mencetuskan ralat homokedastisiti untuk data yang bermasaalah titik terpencil dan ralat varians yang tidak konsisten. Anggaran yang terhasil mempunyai ralat varians yang lebih konsisten dan model yang terhasil dikenali sebagai TBS-Berasaskan Penganggar Teguh. Analisis juga menunjukkan kaedah TBS-Berasaskan Penganggar Teguh menghasilkan anggaran dengan varians yang lebih rendah berbanding kaedah MLE dan MML apabila kedua-dua masalah titik terpencil dan ralat berheteroskedasti wujud bersama.

Tesis ini juga menguji variasi penganggaran kaedah WMML menggunakan kaedah "bootstrap". Kaedah "bootstrap" yang sedia ada seperti Pensampelan Semula - x Secara Tetap (Fixed- x Resampling), Pensampelan Semula - x Secara Rawak (Random- x Resampling) dan Diagnolisis-Sebelum "Bootstrap" adalah tidak teguh apabila titik terpencil hadir. Untuk menyelesaikan masalah ini, kami mencadangkan kaedah "bootstrap" yang baru, yang kami kenali sebagai "Bootstrap" Keberangkalian Berpemberat (WBP). Mengikut kaedah WBP, titik terpencil akan menerima kebarangkalian yang rendah dan mempunyai peluang yang tipis untuk terpilih semasa proses pensampelan semula. Hasil simulasi dalam setiap kes kajian menunjukkan kaedah WBP adalah lebih teguh berbanding dengan kaedah "bootstrap" yang sedia ada.

ACKNOWLEDGEMENTS

First, I would like to express my deep gratitude and warmest thanks to my supervisor, Associate Professor Dr. Habshah Midi for the great academic help and her continuous guidance during the course of my studies. I also greatly value her judgment, friendship and encouragement that she has provided prior to my entrance to the PhD program through completion. I would like to emphasize that this thesis would not be possible without her continuous support and encouragement.

In addition, I would also like to thank my supervisory committee members, Associate Professor Dr. Kassim Haron and Dr. Zainiddin Eskuvatov for their invaluable discussions, comments, and help.

Special thanks to Professor Dr. A. M. H. Rahmatullah Imon, Statistics Professor from University of Rajshahi, Bangladesh, for his useful remarks and being the co-author of two presented papers.

I gratefully acknowledge the financial support from the Universiti Teknologi MARA (UiTM) as my main sponsor during my studies.

I would also like to extend my thanks to all members of Faculty of Information Technology and Quantitative Science, UiTM, for their kind assistance during my studies. This particularly goes to Associate Professor Dr. Adnan Ahmad, Dean of Faculty of Information Technology and Quantitative Science, for giving me permission to use the faculty laboratory facilities. I must thank Puan Nik Arni Nik Mohamad, for her help in editing the first drafts of Chapter 1 and Chapter 2 of this thesis.

My family deserves special recognition. I am most grateful to my husband, my parents and my children. I greatly appreciate their love and support throughout my life. It is to them that I dedicate this work.

I certify that an Examination Committee met on 7th May 2008 to conduct the final examination of Norazan Binti Mohamed Ramli on her Doctor of Philosophy thesis entitled "Weighted Maximum Median Likelihood Estimation and Outlier Detection in Multiple Linear Regression" in accordance with Universiti Pertanian Malaysia (Higher Degree) Act 1980 and Universiti Pertanian Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the candidate be awarded the relevant degree. Members of the Examination Committee are as follows:

Isa Bin Daud, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Noor Akma Ibrahim, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Mohd Rizam Abu Bakar, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Internal Examiner)

Anton Abdulbasah Kamil, PhD
Associate Professor
School of Mathematical Sciences
Universiti Sains Malaysia
(External Examiner)

HASANAH MOHD. GHAZALI, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Universiti Putra Malaysia and has been accepted as fulfilment of the requirement for the degree of Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Habshah Midi, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Kassim Bin Haron, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Zainiddin Eshkuvatov , PhD
Faculty of Science
Universiti Putra Malaysia
(Member)

AINI IDERIS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 14 August 2008

DECLARATION

I hereby declare that the thesis is based on my original work except for quotations and citations, which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Putra Malaysia or other institutions.

NORAZAN BINTI MOHAMED RAMLI

Date: 12 June 2008

TABLE OF CONTENTS

		Page
ABSTRACT		iii
ABSTRAK		vi
ACKNOWLEDGEMENTS		ix
APPROVAL		xi
DECLARATION		xiii
LIST OF TABLES		xix
LIST OF FIGURES		xxii
CHAPTER		
1	**INTRODUCTION**	1
	1.1 Background and Motivation for this Research	1
	1.2 Statement of the Problem	13
	1.3 Research Objectives	14
	1.4 Scope of Research	16
	1.5 Research Outline	16
2	**MAXIMUM MEDIAN LIKELIHOOD ESTIMATOR**	18
	2.1 Introduction	18
	2.2 A Review on Classical and Robust Estimators	18
	2.3 The MML Estimator	29
	2.3.1 The MML Estimation	30
	2.3.2 The Performance of the MML Estimator in Real Data Sets	35
	2.3.3 The Performance of the MML Estimator in Simulated Data Sets	43
	2.4 Summary	55
3	**WEIGHTED MAXIMUM MEDIAN LIKELIHOOD ESTIMATORS**	56
	3.1 Introduction	56
	3.2 Improving and Robustifying the MML Estimator	59
	3.2.1 High Breakdown Initial Estimate	59
	3.2.2 Robust Concentration Algorithm	60
	3.2.3 Weighting Functions	63

3.3 Improved MML Estimator 68
3.3.1 The WMML1 Estimator 69
3.3.2 Details in the WMML1 Estimation 69
3.3.3 The WMML2 Estimator 72
3.3.4 Details in the WMML2 Estimation 73
3.3.5 The WMML3 Estimator 76
3.3.6 Details in the WMML3 Estimation 76
3.4 The Performance of the WMML1, WMML2 and WMML3 Estimators on Real Data Sets 78
3.5 The Performance of the WMML1, WMML2 and WMML3 Estimators on the Simulated Data Sets 103
3.5.1 Simulation Design 103
3.5.2 Measures of Performance 104
3.5.3 Simulation Results for Clean Data Sets 107
3.5.4 Simulation Results for Data Sets with Outliers in Y 111
3.5.5 Simulation Results for Data Sets with Outliers in X 120
3.5.6 Simulation Results for Data Sets with Outliers in both X and Y 128
3.6 Summary 137

4 DIAGNOSTIC-ROBUST GENERALIZED POTENTIALS (DRGP) FOR IDENTIFYING HIGH LEVERAGE POINTS 141
4.1 Introduction 141
4.2 A Review on Diagnostic for High Leverage Points 142
4.3 Diagnostic-Robust Generalized Potentials (DRGP) 153
4.4 On the Performance of Diagnostic-Robust Generalized Potentials (DRGP) 154
4.4.1 The Performance of Diagnostic-Robust Generalized Potentials on Real Data Sets 154
4.4.2 The Performance of Diagnostic-Robust Generalized Potentials on Simulated Data Sets 165
4.5 Summary 169

5 GRAPHICAL VISUALIZATION OF THE WMML LOGLIKELIHOOD VALUES TO IDENTIFY MULTIPLE OUTLIERS IN LINEAR REGRESSION 171
5.1 Introduction 171
5.2 Visual Analysis on the WMML1, WMML2, and WMML3 Loglikelihood Values for Identifying Multiple Outliers 173
5.3 Some Numerical Examples from Real Data Sets 174
5.4 Summary 201

6	**TRANSFORM BOTH SIDES ROBUST BASED METHODS FOR DATA WITH PROBLEMS OF OUTLIERS AND HETEROSCEDASTIC ERRORS**	202
	6.1 Introduction	202
	6.2 Heteroscedasticity Problem	203
	6.3 Transform Both Sides (TBS) Methodology	205
	6.4 Transform Both Sides Robust Based Method (TBS- Robust Based Method)	207
	6.4.1 TBS-WMML1 Based Method	208
	6.4.2 TBS- WMML2 Based Method	211
	6.4.3 The TBS- WMML3 Based Method	213
	6.5 The Performance of the TBS - Robust Based Methods on Real and Simulated Data Sets with Problems of Outliers and Heteroscedastic Errors	213
	6.5.1 Real Data Sets	214
	6.5.2 Simulated Data Sets	238
	6.5.3 Simulation Results	240
	6.6 Summary	247
7	**WEIGHTED BOOTSTRAP WITH PROBABILITY**	250
	7.1 Introduction	250
	7.2 A Review on Bootstrap Methods	252
	7.2.1 Fixed - x Resampling or Bootstrap 1 Method	252
	7.2.2 Random - x Resampling or Bootstrap 2 Method	254
	7.2.3 Diagnostic Before Bootstrap	255
	7.3 Weighted Bootstrap with Probability (WBP)	258
	7.3.1 Procedures for the Weighted Bootstrap with Probability	259
	7.4 Some Numerical Examples	262
	7.4.1 WBP Estimates for Belgium International Phone Calls Data	263
	7.4.2 WBP Estimates for Stackloss Data	270
	7.4.3 WBP Estimates for Simulated Data Sets	276
	7.5 The WBP Standard Errors	280
	7.6 Percentiles and Bias-Corrected, Accelerated (BCa) Confidence Intervals	296
	7.6.1 Bootstrap Percentile Intervals	296
	7.6.2 Bias-Corrected, Accelerated (BCa) Percentile Intervals	296
	7.7 A Simulation Study	298
	7.7.1 Simulation Results for Clean Data Sets	299
	7.7.2 Simulation Results for Data Sets with Y Outliers	300
	7.7.3 Simulation Results for Data Sets with X Outliers	301

7.7.4 Simulation Results for Data Sets with both X and Y Outliers 301
7.8 Summary 311

8 CONTRIBUTIONS, CONCLUSION AND SUGGESTIONS FOR FURTHER RESEARCH 312
8.1 Introduction 312
8.2 Contributions of the Study 313
8.2.1 The Performance Study of the MML Estimator 313
8.2.2 The WMML1, the WMML 2 and the WMML 3 Estimators 313
8.2.3 Diagnostic-Robust Generalized Potentials (DRGP) for Identifying High Leverage Points 315
8.2.4 Visual Analysis of the WMML1, the WMML2, and the WMML3 Estimates for Identifying Multiple Outliers 316
8.2.5 The TBS-WMML1 Based, the TBS-WMML2 Based and the TBS-WMML3 Based Methods 316
8.2.6 Weighted Bootstrap with Probability (WBP) 316
8.3 Conclusions 317
8.4 Suggestions for Further Research 318

REFERENCES 320
APPENDICES 328
A1 Pilot Plant Data 329
A2 Belgium International Calls Data 330
A3 Hertzsprung Russel Stars Data 331
A4 Stackloss Data 333
A5 Salinity Data 334
A6 Coleman Data 335
A7 Wood Specific Gravity Data 336
A8 Road Construction Data 337
A9 Minitab Tree Data 338
A10 Body and Brain Weight Data 339
A11 Education Expenditure Data 340
B1 Simulation Results of 1000 Simulations (replicates) for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$ with Outliers in Y 342
B2 Simulation Results of 1000 Simulations (replicates) for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$ with Outliers in X 348
B3 Simulation Results of 1000 Simulations (replicates) for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$ with Outliers in Both X and Y 354

B4	Coverage Probabilities for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$: 1000 Simulations (replicates), Outliers in Y	360
B5	Coverage Probabilities for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$: 1000 Simulations (replicates), Outliers in X	366
B6	Coverage Probabilities for $\hat{\beta}_0$, $\hat{\beta}_2$ and $\hat{\beta}_3$: 1000 Simulations (replicates), Outliers in X and Y	380
C1	Densitiy Plots of $\hat{\beta}_3$ from 1000 Simulations for Data Sets with Outliers in Y	378
C2	Densitiy Plots of $\hat{\beta}_3$ from 1000 Simulations for Data Sets with Outliers in X	381
C3	Densitiy Plots of $\hat{\beta}_3$ from 1000 Simulations for Data Sets with Outliers in both X and Y	384

COMPUTER CODES FOR THE PROPOSED METHODS (IN CD FORM) 387
AWARDS 388
BIODATA OF STUDENT 391
LIST OF PUBLICATIONS 392

LIST OF TABLES

Table		Page
1.1	Artificial data set with an outlier in the response variable	5
1.2	Artificial data set with an outlier in the explanatory variable	6
2.1	MML estimates from Pilot Plant data (original data set)	37
2.2	MML estimates from contaminated Pilot Plant data (two outliers are planted in the Y variable)	38
2.3	MML estimates from contaminated Pilot Plant data (two outliers are planted in the X variable)	38
2.4	MLE and MML estimates for Hertzsprung Russel Star data	40
3.1	Standardized Tuning Constants for the Re-descending Weighting Functions (Hampel et al., 1986)	68
3.2	A summary of real data sets	79
3.3	Estimates for Pilot Plant Data	82
3.4	Estimates for Belgium International Phone Calls data	85
3.5	Estimates for Hertzsprung-Russel Stars data	88
3.6	Estimates for Stackloss data	92
3.7	Estimates for Salinity data	95
3.8	Coefficient estimates for Coleman data	98
3.9	Standard error estimates for Coleman data	99
3.10	Coefficient estimates for Wood Specific Gravity data	101
3.11	Standard error estimates for Wood Specific Gravity data	102
3.12	Factors and Levels for Simulation Studies	104
3.13	Simulation results for $\hat{\beta}_1$, based on 1000 simulations (replications) for data sets with outliers in Y	114
3.14	Coverage Probabilities for $\hat{\beta}_1$, based on 1000 simulation (replications) for data sets with outliers in Y	118
3.15	Simulation results for $\hat{\beta}_1$, based on 1000 simulations (replications) for data sets with outliers in X	121
3.16	Coverage Probabilities for $\hat{\beta}_1$, based on 1000 simulation (replications) for data sets with outliers in X	123
3.17	Simulation results for $\hat{\beta}_1$, based on 1000 simulations (replications) for data sets with outliers in X and Y	130
3.18	Coverage Probabilities for $\hat{\beta}_1$, based on 1000 simulation (replications) for data sets with outliers in X and Y	132
4.1	Leverages (w_{ii}), Hadi's Potentials (p_{ii}), Generalized Potentials (p_{ii}^{GP}), DRGP1 (p_{ii}^{DRGP1}) and DRGP2 (p_{ii}^{DRGP2}) for Stackloss data	156
4.2	Table 4.2: Leverages (w_{ii}), Hadi's Potentials (p_{ii}), Generalized Potentials (p_{ii}^{GP}) DRGP1 (p_{ii}^{DRGP1}) and DRGP2 (p_{ii}^{DRGP2}) for Hawkins-Bradu-Kass data	161

Table		Page
4.3	Identification of multiple high-leverage cases (average to the nearest whole number) based on 10000 simulations	167
6.1	The MLE, MML, WMML1, WMML2 (Tanh) and WMML3 estimates for Road Construction data before and after TBS	217
6.2	The WMML2 (Tanh) estimates for the Minitab Tree data	225
6.3	The WMML1, WMML2 (Tanh), and WMML3 estimates for Body and Brain Weight data before and after the TBS	231
6.4	The WMML2 (Tanh) estimates for Education Expenditure data before and after the TBS	237
6.5	Parameter estimates together with their associated standard errors (in brackets) for heteroscedastic data sets with outliers in Y and sample of size 40	243
6.6	Parameter estimates together with their associated standard errors (in brackets) for heteroscedastic data sets with outliers in X and sample of size 40	244
6.7	Parameter estimates together with their associated standard errors (in brackets) for heteroscedastic data sets with outliers in both X and Y and sample of size 40	245
7.1	95% confidence intervals of bootstrapped estimates for Belgium International Phone Calls data	266
7.2	95% confidence interval of bootstrapped estimates for Stackloss data	273
7.3	95% confidence intervals of bootstrapped estimates (based on 1000 simulations) for data sets with Y outliers and sample of size 40	277
7.4	95% confidence intervals of the bootstrapped estimates (based on 1000 simulations) for data sets with X outliers and sample of size 40	278
7.5	95% confidence intervals of the bootstrapped estimates (based on 1000 simulations) for data sets with both X and Y outliers and sample of size 40	279
7.6	Monte Carlo (MC), Bootstrap 1, Diagnostics-Before Bootstrap (DIAG), Weighted Bootstrap with Probability (WBP) and Exact Standard Errors for $\hat{\beta}_1$ obtained from the WMML1, WMML2 (Tanh) and WMML3 with Outliers in Y	282
7.7	Monte Carlo (MC), Bootstrap 1, Diagnostics-Before Bootstrap (DIAG), Weighted Bootstrap with Probability (WBP) and Exact Standard Errors for $\hat{\beta}_3$ obtained from the WMML1, WMML2 (Tanh) and WMML3 with Outliers in Y	283
7.8	Monte Carlo (MC), Bootstrap 1, Diagnostics-Before Bootstrap (DIAG), Weighted Bootstrap with Probability (WBP) and Exact Standard Errors for $\hat{\beta}_1$ obtained from the WMML1, WMML2 (Tanh) and WMML3 with Outliers in X	284
7.9	Monte Carlo (MC), Bootstrap 1, Diagnostics-Before Bootstrap (DIAG), Weighted Bootstrap with Probability (WBP) and Exact Standard Errors for $\hat{\beta}_3$ obtained from the WMML1, WMML2 (Tanh) and WMML3 with Outliers in X	285

Table		Page
7.10	Monte Carlo (MC), Bootstrap 1, Diagnostics-Before Bootstrap (DIAG), Weighted Bootstrap with Probability (WBP) and Exact Standard Errors for $\hat{\beta}_1$ obtained from the WMML1, WMML2 (Tanh) and WMML3 with Outliers in X and Y	286
7.11	Monte Carlo (MC), Bootstrap 1, Diagnostics-Before Bootstrap (DIAG), Weighted Bootstrap with Probability (WBP) and Exact Standard Errors for $\hat{\beta}_3$ obtained from the WMML1, WMML2 (Tanh) and WMML3 with Outliers in X and Y	287
7.12	Coverage probabilities and average width for $\hat{\beta}_1$ of WMML2 (Tanh) for data set with outliers in Y	303
7.13	Coverage probabilities and average width for $\hat{\beta}_3$ of WMML2 (Tanh) for data set with outliers in Y	304
7.14	Coverage probabilities and average width for $\hat{\beta}_1$ of WMML2.(Tanh) for data set with outliers in X	305
7.15	Coverage probabilities and average width for $\hat{\beta}_3$ of WMML2.(Tanh) for data set with outliers in X	306
7.16	Coverage probabilities and average width for $\hat{\beta}_1$ of WMML2 (Tanh) for data set with outliers in both X and Y	307
7.17	Coverage probabilities and average width for $\hat{\beta}_3$ of WMML2 (Tanh) for data set with outliers in both X and Y	308
7.18	Coverage probabilities for the WBP estimates: 20% Outliers in Y	309
7.19	Coverage probabilities for the WBP estimates: 20% Outliers in X	310

LIST OF FIGURES

Figure		Page
1.1	A plot of MLE fit with an outlier in the y - direction.	5
1.2	A plot of MLE fit with a bad leverage point.	7
1.3	A scatter plot of Hertzsprung Russel Star data.	8
1.4	Index plot of MLE loglikelihoods for Hertzsprung Russel Stars data.	9
2.1	Box-plots comparing the 1000 MLE and MML re-calculated estimates from Stackloss data.	42
2.2	A scatter plot representing an artificial data set of size 20 with 10% outliers in *Y*.	45
2.3	A scatter plot representing an artificial data set of size 20 with 10% outliers in *X*.	46
2.4	A scatter plot representing an artificial data set of size 20 with 10% outliers in both X and *Y*.	46
2.5	Box-plots representing the MML estimates for $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, and $\hat{\beta}_3$ based on 1000 simulations (replicates).	49
2.6	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 20 and outliers (0%, 10% and 20%) in *Y*.	50
2.7	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 20 and outliers (0%, 10% and 20%) in X..	50
2.8	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 20 and outliers (0%, 10% and 20%) in both X and *Y*.	51
2.9	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 40 and outliers (0%, 10% and 20%) in *Y*.	51
2.10	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 40 and outliers(0%, 10% and 20%) in *X*.	52
2.11	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 40 and outliers (0%, 10% and 20%) in both X and *Y*.	52
2.12	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 100 and outliers (0%, 10% and 20%) in *Y*.	53

Figure		Page
2.13	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 100 and outliers (0%, 10% and 20%) in X.	53
2.14	Density plots for comparing the recalculated $\hat{\beta}_3$ for the MML estimates based on 1000 simulations (replicates), for samples of size 100 and outliers (0%, 10% and 20%) in both X and Y.	54
3.1	A plot of weight = $w(u)$ versus u for Huber's Weighting Function.	66
3.2	A plot of weight = $w(u)$ versus u for Bisquare's Weighting Function.	66
3.3	A plot of weight = $w(u)$ versus u for Hampel's Weighting	67
3.4	A plot of weight = $w(u)$ versus u for Tanh's Weighting Function.	67
3.5	A scatter plot Pilot Plant data.	81
3.6	Box-plots for the MLE, MML, WMML1, WMML2 and WMML3 estimates based on 1000 re-calculations of $\hat{\beta}_0$ and $\hat{\beta}_1$ for Pilot Plant data.	83
3.7	Box-plots for the MLE, MML, WMML1, WMML2 and WMML3 estimates based on 1000 re-calculations of $\hat{\beta}_0$ and $\hat{\beta}_1$ for Belgium International Phone Calls data.	86
3.8	The MLE, MML and WMML2 (Tanh) fitted lines for the Number of International Phone Calls from Belgium in the year 1950 – 1973.	87
3.9	Box-plots for the MLE, MML, WMML1, WMML2 and WMML3 estimates based on 1000 re-calculations of $\hat{\beta}_0$ and $\hat{\beta}_1$ for Hertzsprung Russel Stars data.	89
3.10	Figure 3.10: Box-plots for the MLE, MML, WMML1, WMML2 and WMML3 estimates based on 1000 re-calculations of $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$, and $\hat{\beta}_3$ for Stackloss data.	91
3.11	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated clean data sets of size 20 were used.	109
3.12	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated clean data sets of size 40 were used.	109
3.13	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated clean data sets of size 100 were used.	110
3.14	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 20 were used. These data sets were contaminated with 10% outliers in Y.	115
3.15	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 20 were used. These data sets were contaminated with 20% outliers in Y.	115

Figure		Page
3.16	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 40 were used. These data sets were contaminated with 10% outliers in Y.	116
3.17	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 40 were used. These data sets were contaminated with 20% outliers in Y.	116
3.18	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 100 were used. These data sets were contaminated with 10% outliers in Y.	117
3.19	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 100 were used. These data sets were contaminated with 20% outliers in Y.	117
3.20	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 20 were used. These data sets were contaminated with 10% outliers in X.	125
3.21	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 20 were used. These data sets were contaminated with 20% outliers in X.	125
3.22	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 40 were used. These data sets were contaminated with 10% outliers in X.	126
3.23	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 40 were used. These data sets were contaminated with 20% outliers in X.	126
3.24	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 100 were used. These data sets were contaminated with 10% outliers in X.	127
3.25	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 100 were used. These data sets were contaminated with 20% outliers in X.	127
3.26	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 20 were used. These data sets were contaminated with 10% outliers in both X and Y.	134
3.27	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 20 were used. These data sets were contaminated with 20% outliers in both X and Y.	134
3.28	Density plots representing the $\hat{\beta}_1$s, estimated using different methods. 1000 simulated data sets of size 40 were used. These data sets were contaminated with 10% outliers in both X and Y.	135