UNIVERSITI PUTRA MALAYSIA

ROBUST ESTIMATION TECHNIQUE AND
ROBUST AUTOCORRELATION DIAGNOSTIC FOR
MULTIPLE LINEAR REGRESSION MODEL WITH
AUTOCORRELATED ERRORS

LIM HOCK ANN

FS 2014 9
ROBUST ESTIMATION TECHNIQUE AND ROBUST AUTOCORRELATION DIAGNOSTIC FOR MULTIPLE LINEAR REGRESSION MODEL WITH AUTOCORRELATED ERRORS

LIM HOCK ANN

DOCTOR OF PHILOSOPHY
UNIVERSITI PUTRA MALAYSIA

2014
ROBUST ESTIMATION TECHNIQUE AND
ROBUST AUTOCORRELATION DIAGNOSTIC FOR
MULTIPLE LINEAR REGRESSION MODEL WITH
AUTOCORRELATED ERRORS

By

LIM HOCK ANN

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

August 2014
COPYRIGHT

All material contained within the thesis, including without limitation text, logos, icons, photographs and all other artwork, is copyright material of Universiti Putra Malaysia unless otherwise stated. Use may be made of any material contained within the thesis for non-commercial purposes from the copyright holder. Commercial use of material may only be made with the express, prior, written permission of Universiti Putra Malaysia.

Copyright © Universiti Putra Malaysia
DEDICATION

I would like to dedicate this dissertation work to

☞ my respectful parents, who have taught me a lot on the meaning of persistency in life.

☞ my beloved wife, for being so patient and understanding throughout my doctoral pursue. Her love has always been my greatest support and anchor in my life.

☞ my precious newborn son, Isaac. He is such a joy and pride to me and my wife.
Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Doctor of Philosophy

ROBUST ESTIMATION TECHNIQUE AND ROBUST AUTOCORRELATION DIAGNOSTIC FOR MULTIPLE LINEAR REGRESSION MODEL WITH AUTOCORRELATED ERRORS

By

LIM HOCK ANN

August 2014

Chairperson: Professor Habshah Midi, Ph.D.

Faculty: Science

Autocorrelated errors cause the Ordinary Least Squares (OLS) estimators to become inefficient. Hence, it is very essential to detect the autocorrelated errors. The Breusch-Godfrey (BG) test is the most commonly used test for detection of autocorrelated errors. Since this test is easily affected by high leverage points, the robust Modified Breusch-Godfrey (MBG) test is proposed. The results of the study indicate that the MBG test is a robust test to detect the autocorrelated errors.

Thus far, there is no specific method proposed to identify high leverage points in linear model with autocorrelated errors. Hence, the Diagnostic Robust Generalized Potentials Based on Index Set Equality (DRGP(ISE)) is proposed to close the gap in the literature. The findings indicate that DRGP(ISE) is an excellent and fast identification method to detect the high leverage in linear model with autocorrelated errors.

High leverage points have tremendous effect in regression analysis. In this study we verified that high leverage points is another cause of autocorrelation.

Not much research has been done to investigate autocorrelation-influential observations. Hence, the Robust Autocorrelation-Influential Measure based on DRGP (RAIM(DRGP)) is formulated to identify the autocorrelation-influential observations in autocorrelated data. The RAIM(DRGP) is found to do a credible job to identify the high leverage autocorrelation-enhancing and autocorrelation-reducing observations and autocorrelation-influential observations.

Cochrane-Orcutt Prais-Winsten (COPW) iterative method is the most commonly used remedial measure to rectify the autocorrelation problems. However, this procedure is extremely vulnerable in the presence of high leverage points. On the other hand, the autocorrelation may be caused by the presence of high leverage points. The Robust Cochrane-Orcutt Prais-Winsten (RCOPW) iterative method is
therefore developed. The results of the study show that RCOPW estimation is a robust remedial measure in the case when the autocorrelated data come together with the presence of high leverage points and also in the autocorrelation caused by the high leverage points.

The existing diagnostic plot does not take into the consideration of autocorrelated errors. Thus, the robust remedial of autocorrelated errors - Robust Cochrane-Orcutt Prais-Winsten (RCOPW) is incorporated in the diagnostic plot to form the Diagnostic Plot for Autocorrelation Based on Standardized Cochrane-Orcutt Prais-Winsten Residuals (DPA-RCOPW). The results based on simulated autocorrelated data and well known outlying datasets show that DPA-RCOPW successfully identifies and classifies the outlying observations according to its types precisely.

In this study, an alternative method of finding confidence intervals of regression parameters for autocorrelated data in the presence of high leverage points and for autocorrelation caused by high leverage points is proposed. The findings provide strong evidences that the Diagnostic Before Bootstrap based on Robust Cochrane-Orcutt Prais-Winsten estimate (DBB RCOPW) estimate is a robust procedure and consistently provides close answers to the actual confidence intervals of the regression parameters for data with autocorrelated errors in the presence of high leverage points and for autocorrelation caused by high leverage points.
Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk ijazah Doktor Falsafah

TEKNIK PENGANGGARAN TEGUH DAN DIAGNOSTIK AUTOKORELASI TEGUH BAGI MODEL LINEAR BERGANDA RALAT BERAUTOKORELASI

Oleh

LIM HOCK ANN

Ogos 2014

Pengerusi : Profesor Habshah Midi, Ph.D.

Fakulti: Sains


Setakat ini, tidak ada kaedah spesifik yang dicadangkan untuk mengenal pasti titik tuasan tinggi dalam model linear ralat berautokorelasi. Dengan itu, Kaedah Teguh Berdiagnostik Potensi Teritlak Berasaskan Indeks Set Kesaksamaan (DRGP(ISE)) dicadangkan untuk menutup jurang kesusasteraan ini. Hasil kajian menunjukkan bahawa DRGP (ISE) adalah kaedah yang sangat baik and pantas untuk mengesan titik tuasan tinggi dalam model linear bermasalah ralat berautokorelasi.

Titik tuasan tinggi mempunyai kesan yang amat besar dalam analisis regresi. Dalam kajian ini kita mengesahkan bahawa titik tuasan tinggi adalah penyebab lain autokorelasi.

Tidak banyak penyelidikan telah dijalankan untuk menyiasat cerapan autokorelasi-berpengaruh. Dengan itu, Pengukur Autokorerlasi-Berpengaruh Teguh berdasarkan DRGP (RAIM(DRGP)) dibina bagi mengenal pasti cerapan autokorelasi berpengaruh dalam data berautokorelasi. Didapati RAIM(DRGP) telah menunjukkan kecemerlangan untuk mengenal pasti cerapan tuasan tinggi autokorelasi-meningkat dan autokorelasi-menurun dan cerapan autokorelasi berpengaruh.

Kaedah lelaran Cochrane-Orcutt Prais-Winsten (COPW) adalah langkah pemulihan yang paling biasa digunakan untuk membetulkan masalah autokorelasi. Walau bagaimanapun, prosedur ini sangat terdedah dengan kehadiran titik tuasan tinggi.
Pada masa yang sama, masalah autokorelasi mungkin disebabkan oleh kehadiran titik tuasan tinggi. Dengan itu, Kaedah lelaran Cochrane-Orcutt Prais-Winsten Teguh (RCOPW) dicadangkan. Keputusan kajian menunjukkan anggaran RCOPW langkah pembaikan yang teguh dalam kes data berautokorelasi bersama dengan kehadiran titik tuasan tinggi dan masalah autokorelasi disebabkan oleh titik tuasan tinggi.


Dalam kajian ini, kaedah alternatif mencari selang keyakinan parameter regresi bagi data bermasalah autokorelasi dengan kehadiran titik tuasan tinggi dan autokorelasi yang disebabkan oleh titik tuasan tinggi dicadangkan. Hasil kajian menunjukkan Diagnostik Sebelum Bootstrap berdasarkan anggaran Cochrane-Orcutt Prais-Winsten Teguh (DBB RCOPW) adalah prosedur yang mantap dan konsisten dalam memberikan jawapan dekat dengan selang keyakinan sebenar parameter regresi untuk data ralat berautokorelasi dengan kehadiran titik tuasan tinggi dan autokorelasi disebabkan oleh titik tuasan tinggi.
ACKNOWLEDGEMENTS

First and foremost, I would like to give thanks to my God, who have provided me His strength and grace to throughout my doctoral pursue.

Heartfelt appreciation also goes to my committee chairperson, Prof. Dr. Habshah Midi for her constant inspiration, efficient guidance, and constructive feedback rendered. I am deeply honoured to have the opportunity to complete my degree under her supervision.

I would also like to thank my internal co-supervisors, Assoc. Prof. Dr. Jayanthi Arasan and Dr. Md. Sohel Rana for all their guidance provided and also knowledge shared.

A special word of thanks to Professor Dr. Rahmatullah Imon, who is a professor of statistics from the Department of Mathematical Sciences, Ball State University, U.S.A, for his valuable time in sharing with me some of the insightful ideas during his visit to my campus.

I would like to extend my gratitude to all the wonderful people such as Mazlina, Shafie, Balqish and others. Their presence have indeed enriched my journey in completing my doctoral pursue. Appreciation also extended to all members of Graduate School and Faculty of Science, who have worked hard in creating a conducive environment for all post graduate students. I am glad to be able to graduate from this institution.

My sincere regards to all my siblings, who have continuously encourage me not to loose heart in all that I am pursuing, both mentally and also spiritually.

Last but not least, my special thanks go to my beloved wife, for standing by with me patiently with her never ending encouragement, prayers and support throughout my doctoral pursue. My newborn son, Isaac has also been very cooperative and well behaved for allowing me to complete my dissertation. Thank you.
APPROVAL

I certify that a Thesis Examination Committee has met on 26 August 2014 to conduct the final examination of Lim Hock Ann on his thesis entitled “Robust Estimation Technique and Robust Autocorrelation Diagnostic for Multiple Linear Regression Model with Autocorrelated Errors” in accordance with Universities and University Colleges Act 1971 and the Constitution of the Universiti Putra Malaysia [P.U.(A) 106] 15 March 1998. The Committee recommends that the student be awarded the Doctor of Philosophy.

Member of the Thesis Examination Committee were as follows:

Noor Akma binti Ibrahim, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

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

Abdul Ghapor bin Hussin, PhD
Professor
Universiti Pertahanan Nasional Malaysia
Malaysia
(External Examiner)

Muhammad Hanif Mian, PhD
Professor
Lahore University of Management Science
Pakistan
(External Examiner)

_______________________________
NORITAH OMAR, Ph.D.
Associate Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 19 September 2014
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 Supervisory committee were as follows:

**Habshah Midi, PhD**
Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

**Jayanthi a/p Arasan, PhD**
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

**Md. Sohel Rana, PhD**
Senior Lecturer
Faculty of Science
Universiti Putra Malaysia
(Member)

---

**BUJANG BIN KIM HuAT, Ph.D.**
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:
DECLARATION

Declaration by Graduate Student

I hereby confirm that:

• this thesis is my original work;
• quotations, illustrations and citations have been duly referenced;
• this thesis has not been submitted previously or concurrently for any other degree at any other institutions;
• intellectual property from the thesis and copyright of thesis are fully-owned by Universiti Putra Malaysia, as according to the Universiti Putra Malaysia (Research) Rules 2012;
• written permission must be obtained from supervisor and the office of Deputy Vice-Chancellor (Research and Innovation) before thesis is published (in the form of written, printed or in electronic form) including books, journals, modules, proceedings, popular writings, seminar papers, manuscripts, posters, reports, lecture notes, learning modules or any other materials as stated in the Universiti Putra Malaysia (Research) Rules 2012;
• there is no plagiarism or data falsification/fabrication in the thesis, and scholarly integrity is upheld as according to the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) and the Universiti Putra Malaysia (Research) Rules 2012. The thesis has undergone plagiarism detection software.

Signature: ___________________________ Date: ___________________________

Name and Matric No.: ___________________________________________________
Declaration by Members of Supervisory Committee

This is to confirm that:

- the research conducted and the writing of this thesis was under our supervision;
- supervision responsibilities as stated in the Universiti Putra Malaysia (Graduate Studies) Rules 2003 (Revision 2012-2013) are adhered to.

Signature: __________________
Name of Chairman of Supervisory Committee: __________________

Signature: __________________
Name of Member of Supervisory Committee: __________________

Signature: __________________
Name of Member of Supervisory Committee: __________________
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>i</td>
</tr>
<tr>
<td>ABSTRAK</td>
<td>iii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>v</td>
</tr>
<tr>
<td>APPROVAL</td>
<td>vi</td>
</tr>
<tr>
<td>DECLARATION</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xvi</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xix</td>
</tr>
<tr>
<td>LIST OF APPENDICES</td>
<td>xxi</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td>xxiii</td>
</tr>
</tbody>
</table>

## CHAPTER

### 1 INTRODUCTION

1.1 Introduction and Background of the Study 1
1.2 Importance and Motivation of the Study 1
1.3 Research Objectives 4
1.4 Significance of Study 5
1.5 Scope and Limitation of the Study 6
1.6 Outline of the Thesis 7

### 2 LITERATURE REVIEW

2.1 Introduction 9
2.2 Ordinary Least Squares Estimations 9
2.3 Violation from the Least Squares Assumptions 11
2.4 Autocorrelation in Linear Regression Model 11
2.4.1 Sources of Autocorrelation 11
2.4.2 Effects of Autocorrelation 13
2.4.3 Autocorrelation Detection Methods 14
2.4.3.1 Graphical Methods 14
2.4.3.2 Runs Test 15
2.4.3.3 Durbin-Watson $d$ Test 16
2.4.3.4 Breusch Godfrey (BG) Test 17
2.4.4 Correcting for Autocorrelation 18
2.4.4.1 When $\rho$ is Known 19
2.4.4.2 When $\rho$ is Unknown 20
2.4.4.3 Cochrane-Orcutt Iterative Method 20
2.4.4.4 Prais-Winsten Transformation 21

2.5 Introduction to Robust Estimators 21
2.5.1 Basic Concepts of Robust Estimators 22
2.5.1.1 Efficiency 22
2.5.1.2 Breakdown Point 22
2.5.1.3 Bounded Influence 23
2.5.2 Robust Estimators of Location and Scatter 23
2.5.2.1 Minimum Volume Ellipsoid (MVE) Estimator 23
2.5.2.2 Minimum Covariance Determinant (MCD) Estimator 24
2.5.2.3 Covariance Matrix Equality (CME) 25
2.5.2.4 Index Set Equality (ISE) 26
2.5.3 Robust Linear Regression 26
2.5.3.1 L-Estimator 26
2.5.3.2 Least Median of Squares (LMS) Estimator 27
2.5.3.3 Least Trimmed Squares (LTS) Estimator 27
2.5.3.4 M-Estimator 28
2.5.3.5 S-Estimator 31
2.5.3.6 MM-Estimator 31

2.6 Diagnostic Methods of Outlying Observations 32
2.6.1 High Leverage Points Diagnostic Methods 32
2.6.1.1 Three-sigma Edit Rule 33
2.6.1.2 Leverages Based on Weight Matrix 34
2.6.1.3 Hadi’s Potentials 35
2.6.1.4 Mahalanobis Distance 35
2.6.1.5 Generalized Potentials (GP) Measure 36
2.6.1.6 Diagnostic Robust Generalized Potential Based on Minimum Volume Ellipsoid (DRGP (MVE)) Measure 37
2.6.2 Vertical Outliers Diagnostic Methods 38
2.6.2.1 Standardized OLS Residuals 39
2.6.2.2 Studentized Residuals 39
2.6.2.3 Standardized LTS Residuals 39

2.7 Bootstrapping 40
2.7.1 Asymptotic Standard Errors for Roust Regression Estimators 40
2.7.2 Residuals Bootstrap 40
2.7.3 Fixed-X Bootstrapping 41
2.7.4 Diagnostic-Before Bootstrap 41
2.7.5 Stationary Bootstrap 42

3 ROBUST MODIFICATION OF BREUSCH-GODFREY TEST IN THE PRESENCE OF HIGH LEVERAGE POINTS 44
3.1 Introduction 44
3.2 Breusch-Godfrey (BG) Test 45
3.3 Modified Breusch-Godfrey (MBG) Test 46
3.4 The Distribution of MBG Statistic 47
3.5 Numerical Results 51
3.5.1 Monte Carlo Simulation Study 51
3.5.1.1 Both Positive Directions for $\beta_1$ and $\beta_2$ 52
3.5.1.2 One Positive and One Negative Direction for $\beta_1$ and $\beta_2$ 51
3.5.2 The Power of MBG Test 55
3.5.3 Numerical Examples: Time Series Data 56
  3.5.3.1 Economic Report of the President Data 56
  3.5.3.2 Boat Production Data 58
  3.5.3.3 General Road Accident Data in Malaysia 59
3.5.4 Numerical Examples: Cross Sectional Data 60
  3.5.4.1 Quality Data 60
  3.5.4.2 Burger King Nutrition Data 61
3.6 Conclusion 62

4 DIAGNOSTIC ROBUST GENERALIZED POTENTIAL BASED ON INDEX SET EQUALITY (DRGP(ISE)) FOR THE IDENTIFICATION OF HIGH LEVERAGE POINTS IN LINEAR MODEL WITH AUTOCORRELATED ERRORS
4.1 Introduction 64
4.2 High Leverage Point Detection Measures 65
  4.2.1 Leverages Based on Weight Matrix 65
  4.2.2 Hadi’s Potentials 66
  4.2.3 Robust Mahalanobis Distance 66
    4.2.3.1 Minimum Volume Ellipsoid (MVE) 67
    4.2.3.2 Minimum Covariance Determinant (MCD) 68
    4.2.3.3 Covariance Matrix Equality (CME) 69
    4.2.3.4 Index Set Equality (ISE) 76
    4.2.3.5 Computational Complexity 84
4.3 Diagnostic Robust Generalized Potential Based on Index Set Equality (DRGP(ISE)) 84
4.4 Results and Discussions 87
  4.4.1 Monte Carlo Simulation Study 87
  4.4.2 Time Series Data - Effects of Inflation and Deficits on Interest Rates Data 93
  4.4.3 Cross Sectional Data - Hawkins-Bradu-Kass Data 95
4.5 Conclusion 98

5 HIGH LEVERAGE POINTS AS NEW CAUSE OF AUTOCORRELATION
5.1 Introduction 99
5.2 Autocorrelation Caused By high Leverage Points 99
5.3 Results and Discussions 101
  5.3.1 Monte Carlo Simulation Study 101
  5.3.2 Numerical Examples: Time Series Data 103
    5.3.2.1 Water Salinity Data 103
    5.3.2.2 Fresh Detergent Data 104
  5.3.3 Numerical Examples: Cross Sectional Data 106
    5.3.3.1 Hawkins-Bradu-Kass Data 106
    5.3.3.2 Herksprung-Russel Star Data 107
5.4 Conclusion 108

xii
6 DIAGNOSTIC MEASURE OF AUTOCORRELATION INFLUENTIAL OBSERVATION BASED ON A GROUP DELETION APPROACH

6.1 Introduction 109
6.2 Robust Autocorrelation-Influential Measure Based on a Group Deletion Approach 110
6.3 Results and Discussions 113
   6.3.1 Monte Carlo Simulation Study 113
      6.3.1.1 Autocorrelation-Enhancing Influential Observations 113
      6.3.1.2 Autocorrelation-Reducing Influential Observations 115
   6.3.2 Numerical Examples: Time Series Data 117
      6.3.2.1 Fresh Detergent Data 117
      6.3.2.2 Canveg Data 119
   6.3.3 Numerical Examples: Cross Sectional Data 121
      6.3.3.1 New Artificial Data 121
      6.3.3.2 Hawkins-Bradu-Kass Data 123
6.4 Conclusion 126

7 ON THE ROBUST PARAMETER ESTIMATION FOR LINEAR MODEL WITH AUTOCORRELATED ERRORS

7.1 Introduction 127
7.2 Cochrane-Orcutt Prais-Winsten (COPW) Iterative Method 128
7.3 Robust Cochrane-Orcutt Prais-Winsten (RCOPW) Iterative Method 130
7.4 Results and Discussions 131
   7.4.1 Autocorrelated Data in the Presence of High Leverage Points 132
      7.4.1.1 Monte Carlo Simulation Study 132
      7.4.1.2 Numerical Examples: Time Series Data 137
         7.4.1.2.1 Poverty Data 137
         7.4.1.2.2 U.S. Consumption Expenditure Data 140
      7.4.1.3 Numerical Examples: Cross Sectional Data 142
         7.4.1.3.1 Traffic Delays Data 143
         7.4.1.3.2 Auction Price Data 145
   7.4.2 Autocorrelation Caused by High Leverage Points 148
      7.4.2.1 Monte Carlo Simulation Study 148
      7.4.2.2 Time Series Numerical Example - Fresh Detergent Data 153
      7.4.2.3 Cross Sectional Numerical Example – Cereal Chemical Data 154
7.5 Conclusion 156
8 ROBUST DIAGNOSTIC PLOT FOR CLASSIFYING OUTLYING OBSERVATIONS IN DATA WITH AUTOCORRELATED ERRORS

8.1 Introduction 158
8.2 Types of Observations 159
8.3 Robust Diagnostic Plot (RDP) 160
8.4 Diagnostic Plot for Autocorrelated Data 165
  8.4.1 Diagnostic Plot for Autocorrelation Based on Standardized Cochrane-Orcutt Prais-Winsten Residuals (DPA-COPW) 165
  8.4.2 Diagnostic Plot for Autocorrelation Based on Standardized Robust Cochrane-Orcutt Prais-Winsten Residuals (DPA-RCOPW) 168
8.5 Results and Discussions 170
  8.5.1 Monte Carlo Simulation Study 170
  8.5.2 Numerical Examples: Time Series Data 174
    8.5.2.1 Belgian Phone Data 175
    8.5.2.2 Water Salinity Data 177
  8.5.3 Numerical Example: Cross Sectional Data 179
    8.5.3.1 Herksprung-Russel Star Data 179
8.6 Conclusion 181

9 ROBUST BOOTSTRAP CONFIDENCE INTERVALS FOR DATA WITH AUTOCORRELATED ERRORS

9.1 Introduction 182
9.2 Bootstrapping Residuals 183
9.3 Diagnostic Before Bootstrap 184
9.4 Percentile Confidence Intervals 185
9.5 Results and Discussions 186
  9.5.1 Robust Bootstrap Confidence Intervals for Autocorrelated Data in the Presence of High Leverage Points 186
    9.5.1.1 Monte Carlo Simulation Study 187
    9.5.1.2 Time Series Data – Boat Production Data 190
    9.5.1.3 Cross Sectional Data – Cigarette Data 192
  9.5.2 Robust Bootstrap Confidence Intervals for Autocorrelation Caused by High Leverage Points 194
    9.5.2.1 Monte Carlo Simulation Study 194
    9.5.2.2 Time Series Data – Water Salinity Data 198
    9.5.2.3 Cross Sectional Data- Hawkins-Bradu-Kass Data 200
9.6 Conclusion 201

10 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS FOR FURTHER STUDIES

10.1 Introduction 202
10.2 Summary

10.2.1 The Performance of Robust Modification of Breusch-Godfrey Test in the Presence of High Leverage Points 203
10.2.2 Diagnostic Robust Generalized Potentials Based on Index Set Equality (DRGP(ISE)) for the Identification of High Leverage Points in Linear Model Model with Autocorrelated Errors 203
10.2.3 High Leverage Points as New Cause of Autocorrelation 204
10.2.4 Diagnostic Measure of Autocorrelation Influential Observation Based on a Group Deletion Approach 204
10.2.5 On The Robust Parameter Estimation for Linear Model with Autocorrelated Errors 205
10.2.6 Robust Diagnostic Plot for Classifying Outlying Observations in Data with Autocorrelated Errors 206
10.2.7 Robust Bootstrap Confidence Intervals for Data with Autocorrelated Errors 206
10.3 Conclusion 207
10.4 Areas of Future Studies 208

REFERENCES 210
APPENDICES 218
BIODATA OF STUDENT 252
LIST OF PUBLICATIONS 253
LIST OF PRESENTATIONS 254
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Decision Rules of Durbin-Watson $d$ Statistic</td>
<td>17</td>
</tr>
<tr>
<td>3.1</td>
<td>Mean and Variance of BG and MBG Statistics, $r^2$ of Lagrange Multiplier Quantile of BG and MBG and Theoretical Chi-Square Quantile</td>
<td>50</td>
</tr>
<tr>
<td>3.2</td>
<td>Cramér-von Mises One Sample Test for Testing the Distribution of BG and MBG Statistics</td>
<td>50</td>
</tr>
<tr>
<td>3.3</td>
<td>Anderson-Darling Test for Testing the Distribution of BG and MBG Statistics</td>
<td>51</td>
</tr>
<tr>
<td>3.4</td>
<td>The $p$-Values of BG and MBG Tests in the Simulation Study (Both Positive Directions for $\beta_1$ and $\beta_2$)</td>
<td>54</td>
</tr>
<tr>
<td>3.5</td>
<td>The $p$-Values of BG and MBG Tests in the Simulation Study (One Positive and One Negative Direction for $\beta_1$ and $\beta_2$)</td>
<td>55</td>
</tr>
<tr>
<td>3.6</td>
<td>Detection Power of BG and MBG Tests in Simulation Study</td>
<td>56</td>
</tr>
<tr>
<td>3.7</td>
<td>Autocorrelation Diagnostics for Economic Report of the President Data</td>
<td>57</td>
</tr>
<tr>
<td>3.8</td>
<td>Autocorrelation Diagnostics for Boat Production Data</td>
<td>59</td>
</tr>
<tr>
<td>3.9</td>
<td>Autocorrelation Diagnostics for General Road Accident Data in Malaysia</td>
<td>60</td>
</tr>
<tr>
<td>3.10</td>
<td>Autocorrelation Diagnostics for Quality Data</td>
<td>61</td>
</tr>
<tr>
<td>3.11</td>
<td>Autocorrelation Diagnostics for Burger King Nutrition Data</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Comparison of the Running Time for MVE,MCD,CME and ISE</td>
<td>84</td>
</tr>
<tr>
<td>4.2</td>
<td>High Leverage Points Detection by Leverages, Hadi’s Potentials and DRGP(ISE) in Simulation Study</td>
<td>88</td>
</tr>
<tr>
<td>4.3</td>
<td>High Leverage Points Detection and the Programme Running Times of DRGP(MVE) and DRGP(ISE) in Simulation Study</td>
<td>91</td>
</tr>
<tr>
<td>4.4</td>
<td>MBG $p$-Values, High Leverage Points Detection By Leverages, Hadi’s Potentials and DRGP(ISE) for Effects of Inflation and Deficits on Interest Rates</td>
<td>94</td>
</tr>
<tr>
<td>4.5</td>
<td>MBG $p$-Values, High Leverage Points Detection By Leverages, Hadi’s Potentials and DRGP(ISE) for Original and Contaminated Hawkins-Bradu-Kass Data</td>
<td>96</td>
</tr>
<tr>
<td>5.1</td>
<td>The Effect of High Leverage Points on Non-Autocorrelated Data ($n=20$)</td>
<td>102</td>
</tr>
<tr>
<td>5.2</td>
<td>Simulation Study of the Effect of High Leverage Points On Non-Autocorrelated Data</td>
<td>103</td>
</tr>
<tr>
<td>5.3</td>
<td>Autocorrelation Diagnostics for Water Salinity Data</td>
<td>104</td>
</tr>
<tr>
<td>5.4</td>
<td>Autocorrelation diagnostics for Fresh Detergent Data</td>
<td>105</td>
</tr>
<tr>
<td>5.5</td>
<td>Autocorrelation Diagnostics for Hawkins-Bradu-Kass Data</td>
<td>106</td>
</tr>
</tbody>
</table>
5.6 Autocorrelation Diagnosis for Herksprung-Russel Star Data

6.1 The Number of High Leverage Autocorrelation-Enhancing Influential Observations Detected by RAIM(DRGP) and CAIM in the Simulation Study

6.2 The Number of High Leverage Autocorrelation-Reducing Influential Observations Detected by RAIM(DRGP) and CAIM in the Simulation Study

6.3 Autocorrelation Diagnostics for Modified Fresh Detergent Data

6.4 The Performance of RAIM(DRGP) and CAIM in Modified Fresh Detergent Data

6.5 Autocorrelation Diagnostics for Modified Canveg Data

6.6 The Performance of RAIM(DRGP) and CAIM in Modified Canveg Data

6.7 Autocorrelation Diagnostics in Contaminated New Artificial Data

6.8 The Performance of RAIM(DRGP) and CAIM in Modified New Artificial Data

6.9 Autocorrelation Diagnostics for Hawkins-Bradu-Kass Data

6.10 The Performance of RAIM(DRGP) and CAIM in Hawkins-Bradu-Kass Data

6.11 Regression Model Based on OLS for Hawkins-Bradu-Kass Data

7.1 Simulation Study of the Parameters Estimates Based on COPW and RCOPW Iterative Methods

7.2 Performance of COPW and RCOPW Iterative Methods in the Original and Modified Poverty Data

7.3 Performance of COPW and RCOPW Estimations in the Original and Modified U.S. Consumption Expenditure Data

7.4 Performance of COPW and RCOPW Estimations in the Original and Contaminated Traffic Delays Data

7.5 Performance of COPW and RCOPW Estimations in the Original and Contaminated Auction Price Data

7.6 Simulation Study of the Parameters Estimates Based on OLS, COPW and RCOPW in Autocorrelation Caused by High Leverage Points

7.7 Performance of OLS, COPW and RCOPW Estimations in Autocorrelation Caused by High Leverage Points (Contaminated Fresh Detergent Data)

7.8 Performance of OLS, COPW and RCOPW Estimations in Autocorrelation Caused by High Leverage Points (Contaminated Cereal Chemical Data)

8.1 Performance of RDP, DPA-COPW and DPA-RCOPW in Simulation Study

8.2 Classification of the Observations by RDP, DPA-COPW and DPA-RCOPW in Belgian Phone Data
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.3</td>
<td>Classification of the Observations by RDP, DPA-COPW and DPA-RCOPW in Water Salinity Data</td>
<td>178</td>
</tr>
<tr>
<td>8.4</td>
<td>Classification of Outlying Observations by RDP, DPA-COPW and DPA-RCOPW in Herzsprung-Russel Star Data</td>
<td>181</td>
</tr>
<tr>
<td>9.1</td>
<td>Confidence Intervals of Simulated Autocorrelated Data in the Presence of High Leverage Points</td>
<td>189</td>
</tr>
<tr>
<td>9.2</td>
<td>Confidence Intervals of Boat Production Data</td>
<td>191</td>
</tr>
<tr>
<td>9.3</td>
<td>95% Confidence Intervals of Cigarette Data</td>
<td>193</td>
</tr>
<tr>
<td>9.4</td>
<td>95% Confidence Intervals of Simulation Data for Autocorrelation Caused by High Leverage Points</td>
<td>197</td>
</tr>
<tr>
<td>9.5</td>
<td>95% Confidence Intervals of Water Salinity Data</td>
<td>199</td>
</tr>
<tr>
<td>9.6</td>
<td>95% Confidence Intervals of Hawkins-Bradu-Kass Data</td>
<td>201</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Example of Index Plot of Residuals for Data with Autocorrelated Errors</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Example of Current Residuals (Res1) Versus Lagged Residuals (Res(-1)) for Data with First Lagged Autocorrelated Errors</td>
<td>15</td>
</tr>
<tr>
<td>2.3</td>
<td>Decision Zones of Durbin-Watson $d$ Statistic</td>
<td>17</td>
</tr>
<tr>
<td>3.1</td>
<td>CDF of Chi-square, BG and MBG for Sample Sizes $n=40$, 80 and 200</td>
<td>48</td>
</tr>
<tr>
<td>3.2</td>
<td>Lagrange Multiplier Quantile of BG and MBG Test Versus Theoretical Chi-square Quantile for Sample Sizes $n=40, 80$ and 200</td>
<td>49</td>
</tr>
<tr>
<td>3.3</td>
<td>Example of the 3D Plots in the Present of A High Leverage Point in $X_1$, $X_2$, and Both $X_1$ and $X_2$ Directions for Sample Size $n=20$</td>
<td>53</td>
</tr>
<tr>
<td>3.4</td>
<td>Index Plot of Residuals for Economic Report of the President</td>
<td>57</td>
</tr>
<tr>
<td>3.5</td>
<td>Current Residuals (Res1) versus Lagged Residuals (Res(-1)) for Boat Production Data</td>
<td>58</td>
</tr>
<tr>
<td>3.6</td>
<td>Index Plot of Residuals for General Road Accident Data in Malaysia</td>
<td>59</td>
</tr>
<tr>
<td>3.7</td>
<td>Index Plot of Residuals for Quality Data</td>
<td>60</td>
</tr>
<tr>
<td>3.8</td>
<td>Current Residuals (Res1) versus Lagged Residuals (Res(-1)) for Quality Data</td>
<td>61</td>
</tr>
<tr>
<td>3.9</td>
<td>Index Plot of Residuals for Burger King Nutrition Data</td>
<td>62</td>
</tr>
<tr>
<td>4.1</td>
<td>Graphical Display of the Programme Running Times of DRGP(MVE) and DRGP(ISE) for Each Level and Direction of Contaminations</td>
<td>92</td>
</tr>
<tr>
<td>4.2</td>
<td>Index Plot of Residuals for Effects of Inflation and Deficits on Interest Rates Data</td>
<td>93</td>
</tr>
<tr>
<td>4.3</td>
<td>Residuals Plot for Hawkins-Bradu-Kass Data</td>
<td>95</td>
</tr>
<tr>
<td>4.4</td>
<td>Index plot of (a) Leverages, (b) Hadi’s Potentials and (c) DRGP(ISE) for Hawkins-Bradu-Kass Data</td>
<td>96</td>
</tr>
<tr>
<td>4.5</td>
<td>Index Plot of DRGP(ISE) When an Additional HLP is Contaminated in $X_1$, $X_2$, and $X_3$ Direction for Hawkins-Bradu-Kass Data</td>
<td>95</td>
</tr>
<tr>
<td>5.1</td>
<td>Index Plot of Residuals for Water Salinity Data</td>
<td>104</td>
</tr>
<tr>
<td>5.2</td>
<td>Index Plot of Residuals for Contaminated Fresh Detergent Data</td>
<td>105</td>
</tr>
<tr>
<td>5.3</td>
<td>Residuals Plots for Hawkins-Bradu-Kass Data</td>
<td>107</td>
</tr>
<tr>
<td>5.4</td>
<td>Residuals Plots for Herksprung-Russel Star Data</td>
<td>108</td>
</tr>
</tbody>
</table>
6.1  Index Plot of DRGP(ISE) for Modified Fresh Detergent Data  118
6.2  Index Plot of DRGP(ISE) for Modified Canveg Data  119
6.3  Index Plot of DRGP(ISE) for Modified New Artificial Data  122
6.4  Index Plot of DRGP(ISE) for Hawkins-Bradu-Kass Data  124
7.1  3D Scatter Plot for Original and Contaminated Poverty Data  138
7.2  Current Residuals (Res1) Versus Lagged Residuals (Res(-1)) for Poverty Data  138
7.3  Current Residuals (Res1) Versus Lagged Residuals (Res(-1)) for U.S. Consumption Expenditure Data  141
7.4  3D Scatter Plot for the Original and Modified Traffic Delays Data  143
7.5  Index Plot of Residuals for Traffic Delays Data  144
7.6  Current Residuals (Res1) Versus Lagged Residuals (Res(-1)) for Auction Price Data  146
7.7  Index Plot of Residuals for Contaminated Fresh Detergent Data  153
7.8  Current Residuals (Res1) Versus Lagged Residuals (Res(-1)) for Contaminated Cereal Chemical Data  155
8.1  Scatter Plot of the Example of Univariate Data  160
8.2  RDP for Data in Appendix A21  162
8.3  Outlying Observations in Different Locations  163
8.4  RDP for Data in Appendix A21 Following the Further Classifications of Outlying Observations  164
8.5  Example of DPA-COPW  168
8.6  Example of DPA-RCOPW  170
8.7  3D Scatter Plot for Sample Size of 40  172
8.8  Scatter Plot of Belgian Phone Data  175
8.9  Diagnostic Plots for Belgian Phone Data  176
8.10 Diagnostic Plots for Water Salinity Data  178
8.11 Scatter Plot of Herksprung-Russel Star Data  179
8.12 Diagnostic Plots for Herksprung-Russel Star Data  180
# LIST OF APPENDICES

<table>
<thead>
<tr>
<th>Appendix</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>Original and Modified Economic Report of the President Data</td>
<td>219</td>
</tr>
<tr>
<td>A2</td>
<td>Original and Modified Indexes of Boat Production Data</td>
<td>220</td>
</tr>
<tr>
<td>A3</td>
<td>Original and Modified General Road Accident Data in Malaysia</td>
<td>221</td>
</tr>
<tr>
<td>A4</td>
<td>Original and Contaminated Indexes of Quality Data</td>
<td>222</td>
</tr>
<tr>
<td>A5</td>
<td>Original and Modified Burger King Nutrition Data</td>
<td>223</td>
</tr>
<tr>
<td>A6</td>
<td>Original and Modified Effects of Inflation and Deficits on Interest Rates Data</td>
<td>224</td>
</tr>
<tr>
<td>A7</td>
<td>Original and Modified Hawkins-Bradu-Kass Data</td>
<td>225</td>
</tr>
<tr>
<td>A8</td>
<td>Water Salinity Data</td>
<td>226</td>
</tr>
<tr>
<td>A9</td>
<td>Original and Modified Fresh Detergent Data</td>
<td>227</td>
</tr>
<tr>
<td>A10</td>
<td>Hawkins-Bradu-Kass Data</td>
<td>228</td>
</tr>
<tr>
<td>A11</td>
<td>Herksprung-Russel Star Data</td>
<td>229</td>
</tr>
<tr>
<td>A12</td>
<td>Original and Modified Fresh Detergent Data</td>
<td>230</td>
</tr>
<tr>
<td>A13</td>
<td>Original and Modified Canveg Data</td>
<td>231</td>
</tr>
<tr>
<td>A14</td>
<td>Original and Modified New Artificial Data</td>
<td>232</td>
</tr>
<tr>
<td>A15</td>
<td>Original and Modified Poverty Data</td>
<td>233</td>
</tr>
<tr>
<td>A16</td>
<td>Original and Modified U.S. Expenditure Data</td>
<td>234</td>
</tr>
<tr>
<td>A17</td>
<td>Original and Contaminated Traffic Delays Data</td>
<td>235</td>
</tr>
<tr>
<td>A18</td>
<td>Original and Modified Cigarette Data</td>
<td>236</td>
</tr>
<tr>
<td>A19</td>
<td>Original and Contaminated Fresh Detergent Data</td>
<td>237</td>
</tr>
<tr>
<td>A20</td>
<td>Original and Contaminated Cereal Chemical Data</td>
<td>238</td>
</tr>
<tr>
<td>A21</td>
<td>Example of Univariate Data</td>
<td>239</td>
</tr>
<tr>
<td>A22</td>
<td>Belgian Phone Data</td>
<td>240</td>
</tr>
<tr>
<td>A23</td>
<td>Original and Modified Indexes of Boat Production Data</td>
<td>241</td>
</tr>
<tr>
<td></td>
<td>Original and Modified Cigarette Data</td>
<td>242</td>
</tr>
<tr>
<td>---</td>
<td>-------------------------------------</td>
<td>-------</td>
</tr>
<tr>
<td>B</td>
<td>R Programming Codes</td>
<td>243</td>
</tr>
</tbody>
</table>

xxii
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>OLS</td>
<td>Ordinary Least Squares</td>
</tr>
<tr>
<td>BLUE</td>
<td>Best Linear Unbiased Estimators</td>
</tr>
<tr>
<td>MSE</td>
<td>Mean Square Errors</td>
</tr>
<tr>
<td>SSE</td>
<td>Sum of Squares Errors</td>
</tr>
<tr>
<td>SSR</td>
<td>Sum of Squares Regression</td>
</tr>
<tr>
<td>MVUE</td>
<td>Minimum Variance Unbiased Estimator</td>
</tr>
<tr>
<td>IF</td>
<td>Influence Function</td>
</tr>
<tr>
<td>MVE</td>
<td>Minimum Volume Ellipsoid</td>
</tr>
<tr>
<td>MCD</td>
<td>Minimum Covariance Determinant</td>
</tr>
<tr>
<td>CME</td>
<td>Covariance Matrix Equality</td>
</tr>
<tr>
<td>ISE</td>
<td>Index Set Equality</td>
</tr>
<tr>
<td>LAV</td>
<td>Least Absolute Values</td>
</tr>
<tr>
<td>MSAE</td>
<td>Minimum Sum of Absolute Errors</td>
</tr>
<tr>
<td>LAR</td>
<td>Least Absolute Residuals</td>
</tr>
<tr>
<td>LAD</td>
<td>Least Absolute Deviations</td>
</tr>
<tr>
<td>LMS</td>
<td>Least Median of Squares</td>
</tr>
<tr>
<td>LTS</td>
<td>Least Trimmed Squares</td>
</tr>
<tr>
<td>MAD</td>
<td>Median Absolute Deviation</td>
</tr>
<tr>
<td>MADN</td>
<td>Normalized Median Absolute Deviation</td>
</tr>
<tr>
<td>WLS</td>
<td>Weighted Least Squares</td>
</tr>
<tr>
<td>IWLS</td>
<td>Iterative Weighted Least Squares</td>
</tr>
<tr>
<td>RLS</td>
<td>Reweighted Least Squares</td>
</tr>
<tr>
<td>IRLS</td>
<td>Iteratively Reweighted Least Squares</td>
</tr>
<tr>
<td>SD</td>
<td>Standard Deviation</td>
</tr>
<tr>
<td>MD</td>
<td>Mahalanobis Distance</td>
</tr>
<tr>
<td>RMD</td>
<td>Robust Mahalanobis Distance</td>
</tr>
<tr>
<td>RMD (MVE)</td>
<td>Robust Mahalanobis Distance based on the Minimum Volume Ellipsoid</td>
</tr>
<tr>
<td>ASE</td>
<td>Asymptotic Standard Error</td>
</tr>
<tr>
<td>CLRM</td>
<td>Classical Linear Regression Model</td>
</tr>
<tr>
<td>LM</td>
<td>Lagrange Multiplier</td>
</tr>
<tr>
<td>MA</td>
<td>Moving Average</td>
</tr>
<tr>
<td>AR</td>
<td>Autoregressive</td>
</tr>
<tr>
<td>BG</td>
<td>Breusch-Godfrey</td>
</tr>
<tr>
<td>MBG</td>
<td>Modified Breusch-Godfrey</td>
</tr>
<tr>
<td>COPW</td>
<td>Cochrane-Orcutt Prais-Winsten</td>
</tr>
<tr>
<td>RCOGW</td>
<td>Robust Concrane-Orcutt Prais-Winsten</td>
</tr>
<tr>
<td>GP</td>
<td>Generalized Potentials</td>
</tr>
<tr>
<td>Acronym</td>
<td>Description</td>
</tr>
<tr>
<td>-----------------</td>
<td>-----------------------------------------------------------------------------</td>
</tr>
<tr>
<td>DRGP (MVE)</td>
<td>Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid</td>
</tr>
<tr>
<td>DRGP (ISE)</td>
<td>Diagnostic Robust Generalized Potential based on Index Set Equality</td>
</tr>
<tr>
<td>RAIM(DRGP)</td>
<td>Robust Autocorrelation-Influential Measure based on DRGP(ISE)</td>
</tr>
<tr>
<td>AEIO</td>
<td>Autocorrelation Enhancing-Influential Observations</td>
</tr>
<tr>
<td>ARIO</td>
<td>Autocorrelation Reducing-Influential Observations</td>
</tr>
<tr>
<td>CAIM</td>
<td>Classical Autocorrelation-Influential Measure</td>
</tr>
<tr>
<td>RDP</td>
<td>Robust Diagnostic Plot</td>
</tr>
<tr>
<td>DPA-COPW</td>
<td>Diagnostic Plot for Autocorrelation Based on Standardized Cochrane-Orcutt Prais-Winsten Residuals</td>
</tr>
<tr>
<td>DPA-RCOPW</td>
<td>Diagnostic Plot for Autocorrelation Based on Standardized Robust Cochrane-Orcutt Prais-Winsten Residuals</td>
</tr>
<tr>
<td>DBB</td>
<td>Diagnostic Before Bootstrap</td>
</tr>
<tr>
<td>DBB OLS</td>
<td>Diagnostic Before Bootstrap based on OLS</td>
</tr>
<tr>
<td>DBB COPW</td>
<td>Diagnostic Before Bootstrap based on COPW</td>
</tr>
<tr>
<td>DBB RCOPW</td>
<td>Diagnostic Before Bootstrap based on RCOPW</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Introduction and Background of the Study

Linear regression is widely used in all areas of human efforts. It was the primary regression analysis to be studied rigorously. Modeling and analysis using linear regression is comparatively easier than non-linear regression as the properties of parameters estimate is easier to be determined in linear regression. It has become a traditional practice to regress linear regression models using the predominant Ordinary Least Squares (OLS) estimator. The reason for the universally acceptance of OLS is because of its computational simplicity. However, the OLS estimate has its optimum properties only when all the underlying model assumptions are met. Unfortunately, in reality the assumption of random and uncorrelated errors is always violated. The classical model assumes that the error term relating to any observation is not influenced by the error term relating to any other observation. However, the errors might be correlated with the previous errors which means that \( E(u_i, u_j) \neq 0 \) or \( \text{cov}(u_i, u_j) = 0 \) for \( i \neq j \). Although autocorrelated errors do not cause any biasness in the OLS coefficients estimates, but the OLS coefficients estimates become less efficient in the presence of autocorrelated errors. The standard errors of the parameters estimate tend to be underestimated and this lead to misleading conclusion about the statistical significance of the estimated regression coefficients.

On the other hand, the OLS estimate which minimizes the sum of squared between the fitted values and the observed responses in the dataset is obviously affected by high leverage points. Research done by Harter (1974) confirmed that squaring of the residual causes the least square becomes extremely weak to the presence of high leverage points. Thus, it caused the violation from the least squares assumption. At the same time, routine dataset cannot be guaranteed free from outlying observations such as outliers and high leverage points. It is a necessity to introduce the robust methods in linear regression to address both autocorrelation and high leverage points problems.

1.2 Importance and Motivation of the Study

Autocorrelation violates the important properties of the OLS estimates (White and Brisbon, 1980). The parameters estimates obtained by the OLS estimation procedure no longer the Best Linear Unbiased Estimators (BLUE) in the sense that we are able to obtain the parameters estimate with lower standard errors. As the result, the usual \( t \) and \( F \) test of significance are no longer convincing as the tests tend to be statistically significant when in fact it is not. In addition, the coefficient of determination, \( R^2 \) becomes inflated, the estimators would look more accurate as compared to its actual values. In short, the existence of autocorrelated errors will most likely causing the wrong conclusions about the statistical significance of the estimated regression coefficients.
(Gujarati and Porter, 2009). Therefore, detection of autocorrelation problems is very critical. Breusch-Godfrey (BG) test (Breusch, 1978; Godfrey, 1978) is the most general test to detect the presence of autocorrelated errors in economics. However, this test is based on OLS estimate which is not robust, the poor performance of BG test is anticipated in the presence of high leverage points. High leverage points may be defined as the data points which are bulky different from the rest of the data points in X-direction. Many robust literatures have pointed out that high leverage points have great impact on the OLS estimates. (Habshah et al., 2009; Rana et al., 2008; Norazan, 2008). This motivates us to develop a robust autocorrelation detection method which shall perform equally good as BG test for detecting the autocorrelation problems in clean time series and cross sectional datasets. At the same time, it can detect the autocorrelation problems in the contaminated high leverage time series and cross sectional datasets. This is certainly a first attempt in statistics to develop robust autocorrelation detection technique which is resisting of the influence of high leverage points.

When the OLS estimate is applied for fitting the linear regression line, the resulting residuals are function of the leverages and true errors. The masking effect occurs when the high leverage points pull the fitted regression line in a way that the fitted residuals corresponding to that high leverage points. Similarly, the swamping effect happens when the residuals corresponding to inliers are too large to cause the case to be declared as high leverage cases. Pena and Yohai (1995) pointed out that high leverage points are the cause of masking and swamping of data points in linear regression. Therefore, identifying high leverage points in the data is very essential before any inferential is made. Although much works have been done on the identification of high leverage points in linear regression such as leverages method, Hadi’s Potential (Hadi, 1992), Mahalanobis Distance (Mahalanobis, 1936) and Diagnostic-Robust Generalized Potentials (Habshah et al., 2009) but no specific method was proposed to identify the high leverage points in linear regression with autocorrelated errors. In this thesis, we would like to take up the challenge to find out the most reliable approach in identifying high leverage points in linear regression with autocorrelated errors.

The recent researches done by Bagheri et al. (2012) and Riazoshams et al. (2010) have further confirmed that high leverage points have tremendous effect on the OLS estimates. However, the effect of high leverage points in data with autocorrelated errors has not been fully discussed. No study is done to justify the autocorrelation in time series and cross sectional data is due to the presence of high leverage points. This literature gap motivates us to go a step further to verify that the high leverage points are the cause of autocorrelation in time series and cross sectional data.

Bagheri et al. (2012) proposed a novel method for collinearity-influential observation diagnostic measure based on group deletion approach to measure the contribution of each observation towards the collinearity in the dataset. However, to the best of our knowledge, no research has been done to study the autocorrelation-influential observations diagnostic measures in linear model. The existing diagnostic measure only focused on time series model where the observations are viewed in the time domain. An
observation is omitted and the resultant effect on the interested statistic values is noted. Observations which give relatively large changes in the calculated values are deemed to be the influential observations. This diagnostic measure can only be applied to time series model. Since no diagnostic measure has been proposed to evaluate the autocorrelation-influential observations in linear model, in this thesis we take the initiative to develop a novel robust diagnostic measure for identification of autocorrelation-influential observations in linear model to close the gap in the literature.

On the other hand, high leverage points are discovered as a new source of autocorrelation, it may be considered to be a special case of the autocorrelation-enhancing influential observations. It is reasonable to conclude that autocorrelation-influential measure which observes the influential effect of an observation at a time may not be efficient in the presence of high leverage points as high leverage points have unduly effect on the classical estimates. In addition, an autocorrelated dataset may change its nature to a non-autocorrelated dataset in the presence of high leverage points. To our knowledge, nothing has yet been done to diagnose autocorrelation reducing-influential points. It is also interesting to find out whether all the autocorrelation-influential observations are caused by high leverage points and also whether all the high leverage points in the autocorrelated data are the high leverage autocorrelation-influential observations. These further encouraged us to develop a novel robust diagnostic measure for identification of autocorrelation-enhancing and reducing-influential observations for linear model with autocorrelated errors in the presence of high leverage points.

This thesis also addresses the parameter estimation of linear model with autocorrelated errors. A large number of novel works in the literatures about the parameter estimation of linear model with autocorrelated errors. Cochrane-Orcutt Prais-Winsten iterative method (COPW) iterative method (Prais and Winsten, 1954) is the most popular remedial measure in econometrics to obtain estimators with the optimum Best Linear Unbiased Estimators (BLUE) properties. However, the COPW iterative method is based on the OLS estimate which is expected to be easily affected by high leverage points. The shortcoming of COPW iterative procedure has inspired us to develop a robust parameter estimation method to get rid both the autocorrelation and high leverage points problems in the time series and cross sectional datasets. To the best of our knowledge, this is indeed the first attempt to remedy the autocorrelation problems in the presence of high leverage points. At the same time, we also examined the usefulness of this proposed robust parameter estimation in rectifying the autocorrelation caused by high leverage points. The proposed robust parameter estimation is indeed working well in rectifying both autocorrelation and high leverage points problems. This is also another new discovery in statistics to remedy the autocorrelation caused by high leverage points.

According to Hampel et al. (1986), a normal dataset usually contains about 1 to 10 percent outlying observations. There is no guarantee that the high quality data will be free from outlying observations. The outlying observations in univariate dataset with autocorrelated errors may be detected by visual inspection of scatter plot. However, the identification of outlying observations based on scatter diagram is not convincing enough. In addition, the graphical method does not work in high dimensional datasets.
Hubert et al. (2008) also pointed out that the outlying observations are more likely to occur in datasets with many variables. Thus, we need specify statistical method to identify the outlying observations. Many outlying observations detection methods are available in the literatures (Mishra, 2008; Maronna et al., 2006; Rocke and Woodruff, 1996; Kashyap and Maiyuran, 1993). However, not much studies have been carried out in classifying outlying observations according to its inference locations. Although Hubert et al. (2008) have proposed a robust diagnostic plot of classifying outlying observations. However, the method proposed does not take into the consideration of autocorrelated errors in time series and cross sectional data. The autocorrelation problems remain as it is without any concern. In the autocorrelated dataset, the residuals are correlated with the previous errors which means $E(u_i, u_j) \neq 0$ for $i \neq j$. An observation may be far from the bulk, but due to the autocorrelated errors, it may not really an outlying observation in the autocorrelated data. This inspires us to design a first ever exclusive diagnostic plot which incorporates the corrective action of autocorrelation to classify the outlying observations according to it types in the presence of autocorrelated errors in time series and cross sectional data. Since the outlying observations are presence in the dataset, the robust methods must be incorporated in the procedures of designing this comprehensive diagnostic plot.

Confidence interval is one of the favorite topics in linear regression analysis. It is used to indicate the reliability of an estimate. The classical confidence interval is constructed based on the sample finding. Thus, it is too obviously affected by the sample with unusual observations. At the same time, the distribution assumptions need to be made for the classical approach of finding the confidence interval. In contrast, bootstrap methods have a practical point that it does not require normality assumption of the parameters estimate. At the same time, it also enjoys the benefit of not requiring any theoretical calculations to estimate the standard errors of complicated model. This encourages us to find an alternative ways of finding confidence interval of regression parameters using bootstrap methods which do not subject to the statistical distribution requirement and applicable in unwell behaved dataset. The focus here is on the linear model with autocorrelation problems. We have seen that high leverage points have tremendous effect on the parameters estimate. The study here discusses the robust bootstrapping alternative approaches of finding the confidence intervals of regression parameters for data with autocorrelated errors in the presence of high leverage points. Autocorrelation may be due to the presence of high leverage points. Thus, in this study, some robust bootstrapping alternatives of finding the confidence intervals of regression parameters for autocorrelation due to the presence of high leverage in time series and cross sectional datasets are also examined.

### 1.3 Research Objectives

The main purpose of this thesis is to investigate the autocorrelation problems in linear regression model. Currently, the diagnostic and estimation methods dealing with autocorrelated errors are based on OLS estimates. Unfortunately, OLS estimate is easily effected by high leverage points. It will be a big success in statistics if we can have robust identification and estimation methods for autocorrelated data in the presence of...
high leverage points. Moreover, the autocorrelation may be caused by the presence of high leverage points. It will be interesting to have the autocorrelation correction measures to remedy the existence of autocorrelation because of the presence of high leverage points. Nevertheless, identification of autocorrelation influential observations is very essential in linear regression analysis. A comprehensive approach has yet to be developed to identify the autocorrelation influential observations in the presence of high leverage points. In addition, detection and classification of outlying observations is an interesting area in robust statistics. It would be great if we could have customised methods for identifying and classifying outlying observations in data with autocorrelated errors. Moreover, robust alternative approach of finding the confidence interval for regression coefficients in autocorrelated data is also an interesting area to be explored.

The main objectives of this research can be outlined systematically as follows:

1. To formulate a robust autocorrelation diagnostic method and to develop a reliable high leverage identification technique for linear model with autocorrelated errors in the presence of high leverage points.
2. To develop a diagnostic measure of autocorrelation influential observation which can successfully distinguish the autocorrelation-enhancing and autocorrelation-reducing observations for linear model with autocorrelated errors in the presence of high leverage points.
3. To develop a robust parameter estimation method of autocorrelated data in the presence of high leverage points and autocorrelation caused by high leverage points.
4. To construct a diagnostic plot which is able to identify and classify the outlying observations according to their inferential locations in data with autocorrelated errors.
5. To develop a robust bootstrapping alternative approach of finding the confidence intervals of the regression coefficients of autocorrelated data in the presence of high leverage points and autocorrelation caused by high leverage points.

1.4 Significance of Study

Linear regression is used extensively in many areas of studies such as business, engineering, education, medicine and social science. It has many practical applications. The foremost application of linear regression is to make a prediction of the dependence variable based of the fitted model. Linear regression models are often fitted using the OLS estimator. The OLS estimates have optimum properties if all the underlying model assumptions are met. Unfortunately, in reality the assumption of random and uncorrelated errors is always violated. On the other hand, the OLS estimates is not a robust estimates, it is easily effected by high leverage points. Many researchers are unaware of violation of autocorrelation and the effect of high leverage points on the linear regression parameters estimates. The robust autocorrelation diagnostic and estimation methods developed in this thesis are working well in good and contaminated autocorrelated data. Their excellence performances were verified by the assessments done by Monte Carlo simulation study together with some real time series and cross sectional datasets.
This research also pointed out that the high leverage points are the cause of autocorrelation problems. Therefore, the identification of high leverage points in linear regression is very crucial before any remedial action is taken. A credible diagnostic measure was also developed for identifying autocorrelation–influential observations in autocorrelated dataset in the presence of high leverage points. The diagnostic measure working excellently in detecting all the autocorrelation enhancing and reducing influential observations and other autocorrelation influential observations which are not the high leverage points.

In this research, a comprehensive diagnostic plot was also designed for the first time in statistics specifically to identify and classify the outlying observations according to their inferential location in autocorrelated data. The designed diagnostic plot performs superb in identifying and classifying the outlying observations according to their types in autocorrelated data.

Robust alternative approach of finding the confidence intervals of regression parameters in autocorrelated data was also proposed in this study. For all these discoveries, we expect there will be a good application for researchers and industry experts in the future.

1.5 Scope and Limitation of the Study

Robust statistics is still a new area in statistics. Thus, not many statistical software are equipped with robust statistics applications. For the existing software with robust statistics applications, the applications are not really diversified. Most of the time, there is no direct method to get the solution of the desired robustified method. Writing our own programming codes are most of the time required in this case. Although we may get the desired results, but we cannot guarantee that the programming codes are perfect without mistake.

Again, since the robust statistics is a newly developed field of statistics studies, not many well referred outlying datasets are available in the literature for discussion purpose. Not to mention that the outlying datasets with autocorrelation problems. Thus, the same datasets are used repeatedly in this thesis for difference objectives of study.

Alciaturi et al. (2005) proposed the use of the autocorrelation function with lag 1 residual in model selection. Following their suggestion, in this thesis we only focus autocorrelation problems at first-order autoregressive AR(1).

There are many existing robust estimators such as S-estimator, M-estimator, Least Median Squares estimator and etc. In this study, we concentrate only on MM-estimator because it is a bounded influence estimator has high breakdown point (50 percent) and high efficiency (approximately 95 percent) relative to the OLS under the Gauss-Markov assumptions. The MM-estimator is incorporated into the existing procedures in the formulation of robustified methods in the topics of the study.
1.6 Outline of the Thesis

In accordance with the objectives and the scope of the study, the contents of this thesis are organized in such a way that the research objectives are apparent and are conducted in the sequence outlined.

**Chapter Two:** This chapter presents a brief literature review of the OLS estimations of linear regression parameters and the violations from least squares assumptions. The review on autocorrelation problems and its consequences, diagnostic methods, remedial actions and the sources of autocorrelation problems are also discussed. Moreover, basic concepts of robust regression and some important existing robust regression methods are also highlighted. Diagnostic methods of outlying observations are also reviewed. Finally, bootstrapping methods are discussed briefly.

**Chapter Three:** This chapter presents the failure of autocorrelation diagnostic using the Breusch-Godfrey (BG) test developed by Breusch (1978) and Godfrey (1978) in the presence of high leverage points in time series and cross sectional data. The BG test is then robustified by incorporating the high efficient and high breakdown point MM-estimator (Yohai, 1978) in the BG test procedure. The merit of using the Modified Breusch-Godfrey (MBG) test is studied through Monte Carlo simulation, time series and cross sectional datasets.

**Chapter Four:** In this chapter we suggests the Diagnostic Robust Generalized Potential Based on Index Set Equality (DRGP(ISE)) for identifying high leverage points in linear regression with autocorrelated errors. The advantages of using this proposed method is supported by the evidence from the Montle Carlo simulation and real time series and cross sectional datasets.

**Chapter Five:** This chapter investigates high leverage observations as a cause of autocorrelation. Study through Monte Carlo simulation and some well-referred time series and cross sectional datasets were supported the finding that the existence of autocorrelation was due to the presence of high leverage points.

**Chapter Six:** In this chapter we propose to use the Robust Autocorrelation-Influential Measure based on DRGP (RAIM(DRGP)) to identify the autocorrelation-influential observations in autocorrelated data in the presence of high leverage points. The merit and the excellent performance of RAIM(DRGP) is assessed by using Monte Carlo simulation experiments and so well-known datasets.

**Chapter Seven:** This chapter deals with the development of robust parameters estimation to address the autocorrelation and high leverage points problems. Data with autocorrelated errors may be contaminated by the high leverage points. On the other hand, autocorrelation may be due to the presence of high leverage points. The Concrane-Orcutt Prais-Winsten (COPW) iterative method performs miserably in correcting autocorrelation problems in the presence of high leverage points in time series and cross sectional datasets. The Robust Concrane-Orcutt Prais-Winsten (RCOPW) iterative
method is then proposed to remedy both autocorrelation and high leverage points problems. The performance of RCOPW procedure is evaluated by using Monte Carlo simulation experiments and real datasets.

Chapter Eight: This chapter discussed the disadvantages of Robust Diagnostic Plot (RDP) proposed by Hubert et al. (2008) in identifying and classifying the outlying observations in data with autocorrelated errors. In this chapter we designed a comprehensive plot which is able to identify and classify the outlying observations according to its inferential locations accurately for data with autocorrelated errors. The plot is called Diagnostic Plot for Autocorrelation Based on Standardized Robust Cochrane-Orcutt Prais-Winsten Residuals (DPA-RCOPW). It is a plot of Standardized Robust Residuals obtained by Robust Cochrane-Orcutt Pais-Winsten (RCOPW) iterative method versus the leverages computed from Diagnostic Robust Generalized Potentials based on Index Set Equality (DRGP(ISE)). The excellency of DPA-RCOPW is tested using Monte Carlo simulation and some famous robust statistics datasets.

Chapter Nine: This chapter introduced a robust alternative of finding confidence intervals of regression parameters for autocorrelation data in the presence of high leverage points and autocorrelation caused by high leverage points. The Diagnostic Before Bootstrap (DBB) is incorporated in the bootstrapping residuals based on Robust Cochrane-Orcutt Prais-Winsten (RCOPW) procedure to form the DBB RCOPW confidence intervals. The DBB RCOPW confidence intervals constantly provide fairly close intervals to the benchmark confidence intervals for autocorrelation data in the presence of high leverage points and autocorrelation due to the presence of high leverage points.

Chapter Ten: This chapter provides summary and detailed discussions of the thesis conclusions. Some areas of future studies are also tabulated.
REFERENCES


