



UNIVERSITI PUTRA MALAYSIA

**OUTLIER DETECTIONS AND ROBUST ESTIMATION METHODS FOR
NONLINEAR REGRESSION MODEL HAVING AUTOCORRELATED AND
HETEROSCEDASTIC ERRORS**



HOSSEIN RIAZOSHAMS

IPM 2010 13

**OUTLIER DETECTIONS AND ROBUST ESTIMATION METHODS FOR
NONLINEAR REGRESSION MODEL HAVING AUTOCORRELATED AND
HETEROSCEDASTIC ERRORS**

By

HOSSEIN RIAZOSHAMS



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

November 2010

Dedicated to:

My father



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

**OUTLIER DETECTIONS AND ROBUST ESTIMATION METHODS FOR
NONLINEAR REGRESSION MODEL HAVING AUTOCORRELATED
AND HETEROSCEDASTIC ERRORS**

By

Hossein Riazoshams

November 2010

Chairman: Associate Professor Habshah Midi, PhD

Faculty: INSPEM

The ordinary Nonlinear Least Squares (NLLS) and the Maximum Likelihood Estimator (MLE) techniques are often used to estimate the parameters of nonlinear models. Unfortunately, many researchers are not aware of the consequences of using such estimators when outliers are present in the data. The problems get more complex when the assumption of constant error variances or homoscedasticity is violated. To remedy these two problems simultaneously, we proposed a Robust Multistage Estimator (RME).

The heterogeneity of error variances is considered when the variances of residuals follows a parametric functional form of the predictors. Both Nonlinear model function parameters and variance model parameters must be robustified. We have incorporated the MM, the generalized MM and the robustified Chi-Squares Pseudo Likelihood function in the formulation of the RME. The results of the study reveal that the RME is more efficient than the existing methods.

The thesis also addresses the problems when the assumptions of the independent error terms are not met. We proposed a new Robust Two Stage (RTS) estimator in this regard. The proposed method is developed by incorporating the generalized MM estimator in the classical two stage estimator. The performance of the RTS is more efficient than other existing methods revealed by having the highest robustness measures.

We also proposed two outlier identification measures in nonlinear regression. The Tangent leverage, the NLLS, the M and the MM estimators are incorporated in the formulation of the first outlier identification measures. The formulation of the second measure is based on the differences between the derived robust Jacobian Leverage and Tangent leverage. Both proposed measures are very successful to identify the correct outliers.

Finally, we proposed statistics practitioners to use the formal modeling algorithms to get better inferences. We also suggest them to employ appropriate robust methods for further analysis once a correct model has been chosen. The results of the study based on real data signify that the robust estimator is more efficient indicated by lower values of standard errors when compared to the classical estimator.

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

**PENGESANAN DATA TERPENCIL DAN KAEDAH PENGANGGARAN
TEGUH BAGI MODEL TAK LINEAR YANG MEMPUYAI RALAT
BERAUTOKORELASI DAN BERHETEROSKEDASTIK.**

Oleh

HOSSEIN RIAZOSHAMS

November 2010

Pengerusi: Profesor Madya Habshah Midi, PhD

Faculty: INSPERM

Teknik Kuasa Dua Terkecil Tak Linear (NLLS) dan Penganggar Kebolehjadian Maksimum (MLE) kerap digunakan untuk menganggar parameter model tak linear. Malangnya, kebanyakan penyelidik kurang peka akan akibat penggunaan teknik tersebut jika terdapat titik terpancil pada data. Masalah ini semakin kompleks apabila andaian varians beracat tetap atau berhomoskedastisiti tidak dipatuhi. Untuk menyelesaikan dua permasalahan ini secara serentak, kami mencadangkan Penganggar Multiperingkat Teguh (RME).

Varians ralat yang berbeza dipertimbangkan apabila varians reja mematuhi bentuk fungsi parametrik beberapa pembolehubah tak bersandaran. Kedua-dua fungsi parameter bagi model tak linear dan model varians bagi parameter mesti diteguhkan. Kami menggabungkan penganggar M, penganggar MM teritlak dan juga fungsi teguh kebolehjadian *Pseudo* Khi Kuasa Dua ke dalam pembentukan PME Dapatan

daripada kajian ini menunjukkan RME lebih efisien berbanding kaedah yang sedia ada.

Kajian ini juga menyebut permasalahan apabila andaian ralat bebas tidak dipenuhi. Oleh itu kami mencadangkan Penganggar Dua Langkah Teguh (RTS) untuk permasalahan ini. Kaedah cadangan ini dibina dengan menggabungkan penganggar MM teritlak kedalam penganggar dua langkah klasik. Keupayaan RTS didapati lebih efisien berbanding kaedah terdahulu, kerana ianya mempunyai ukuran keteguhan yang tinggi.

Kami juga mencadangkan dua pengukuran pegesanan data terpicil dalam regresi tak linear. Tuasan Tangen digabungkan bersama penganggar NLLS, penganggar M dan penganggar MM ke dalam formulasi pengukuran pegesanan titik terpicil yang pertama. Formulasi kedua pula berasaskan pada perbezaan diantara tuasan teguh Jacobian yang diterbitkan dan tuasan Tangen. Kedua-dua pengukuran cadangan ini sangat berkesan dalam menentukan titik terpicil yang sebenar.

Akhir sekali, kami mencadangkan kepada pengamal statistik untuk menggunakan pemodelan algoritma yang formal untuk mendapatkan inferens yang lebih baik.

Kami juga mencadangkan mereka menggunakan kaedah teguh yang sesuai untuk melakukan analisis lanjutan setelah model yang betul dipilih. Hasil keputusan kajian berdasarkan pada data sebenar, menunjukkan bahawa penganggar teguh adalah lebih efisien, ditunjukkan dari nilai ralat piawai yang rendah berbanding nilai penganggar klasik.

ACKNOWLEDGEMENTS

With the merciful of God that gave me strength to understand science.

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 and encouraging, during the course of my study.

In addition I would like to thank my supervisory comity members, Dr. Bakri Adam, Dr. Ibragimof Gafourjon, for their valuable discussions and help.

Special thanks to Professor Olimjon Sh. Sharipov. Statistics Professor from Institute of Mathematics and Information Technology Uzbek Academy of Science, Tashkent, Uzbekistan, for his useful remarks and being the co-author of our best presented paper.

I gratefully acknowledge the financial support from Universiti Putra Malaysia, during my study, special thanks to Associate Professor Dr. Habshah Midi, and Dr. Bakri Adam, also for their financial support.

I would like to thanks, my Lecturers and Colleagues, and friends, Dr. Hossein Marzban, Economic Department, Shiraz University, Dr Ahmad Parsian Department of statistics, Shiraz University, Prof Javad Behboodian, Shiraz University, Benjamat Hanchana, velo suthar, ali edalati, Sarkhosh Sedighic for their kindly encouraging me to continue to PHD study. And special thanks to my sister Fariba Riazoshams for her helpful advises during my PHD study. And special appreciation to Amir ali,

arsalan sedaghat pisheh, Mohsen Solhdoost, Mohammed Ashyraf, for editing my thesis.



I certify that an Examination Committee met on 3rd Nov 2010 to conduct the final examination of Hossein Riazoshams on his Doctor of Philosophy thesis entitled "OUTLIER DETECTIONS AND ROBUST ESTIMATION METHODS FOR NONLINEAR REGRESSION MODEL HAVING AUTOCORRELATED AND HETEROSCEDASTIC ERRORS" in accordance with Universiti Putra Malaysia (Higher Degree) Act 1980 and Universiti Putra Malaysia (Higher Degree) Regulations 1981. The Committee recommends that the student be awarded the Doctor of Philosophy.

Members of the Examination Committee were as follows:

Dr Isa bin Duad, PhD

Associate Professor.

Faculty of Science

Universiti Putra Malaysia

(Chairman)

Kassim bin Haron, PhD

Profesor Madya Dr.

Faculty of Science

Universiti Putra Malaysia

(Internal Examiner)

Mohd Rizam Abu Bakar, PhD

Profesor Madya Dr.

Faculty of Science

Universiti Putra Malaysia

(Internal Examiner)

Zuhair A. Al-Hemyari, PhD

Profesor College of Art and Science

Universiti of Nizwa. Sultanate of Oman.

(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 University Putra Malaysia and has been accepted as fulfilment of requirement for the degree of Doctor of Philosophy. The members of Supervisory Committee were as follows.

Habshah Midi, PhD

Associate Professor
Institute for Mathematical Research
Universiti Putra Malaysia
(Chairman)

Bakri Adam, PhD

Assistant Professor
Institute for Mathematical Research
Universiti Putra Malaysia
(Member)

Ibragimov Garfounjan, PhD

Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

HASANAH MOHD GHAZALI, PhD

Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

DECLARATION

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

HOSSEIN RIAZOSHAMS

Date:



COPYRIGHT UPM

Table of Contents

	Page
ABSTRACT	iii
ABSTRAK	v
ACKNOWLEDGEMENTS	vii
APPROVAL	ix
DECLARATION	xi
LIST OF TABLES	xv
LIST OF FIGURES	xvii
ABREVIATIONS	xx
1 INTRODUCTION	1
1.1 Background	1
1.2 Motivation of Study	2
1.3 Research Objectives	6
1.4 Significance of Study	7
1.5 Scope and Limitation of the Study	8
1.6 Outline of the Thesis	10
2 LITERATURE REVIEW	12
2.1 Introduction	12
2.2 Nonlinear Models	13
2.3 Robust Regression	14
2.4 Parameter Estimation	18
2.5 Outlier Detection	24
2.6 Heteroscedasticity of Variance	26
2.7 Autocorrelated Errors	28
2.8 Modeling Nonlinear Data	30
3 MATHEMATICAL BACKGROUND	32
3.1 Introduction	32
3.2 Nonlinear Model	32
3.2.1 The Ordinary Least Square (NLLS)	33
3.2.2 Modified Gauss-Newton (Constant Variance)	35
3.2.3 The Curvature Measure of Nonlinearity	37
3.3 The Generalized Least Squares Estimator	41
3.4 Robust MM-Estimator	42
3.5 Generalized M and MM-Estimators	44
3.6 Optimization	44
3.6.1 Iterative Methods	46
3.6.2 Convergence Criteria	48
3.6.3 Combined Algorithm	50
3.6.4 Robust M-estimator	50
3.6.5 The Generalized M-Estimator	52
3.7 Some Mathematical Notation	52

4	OUTLIER DETECTION MEASURE BASED ON TANGENT LEVERAGE AND NLLS, M AND MM ESTIMATOR	54
	4.1 Introduction	54
	4.2 Estimation Methods	55
	4.3 The Outlier Detection Measures	56
	4.3.1 Studentized and Deletion Studentized Residuals	58
	4.3.2 Hadi's Potential	59
	4.3.3 Elliptic Norm (Cook Distance)	60
	4.3.4 Difference in Fits	60
	4.3.5 Atkinson's Distance	61
	4.4 Numerical Example	61
	4.5 Simulation Study	63
	4.6 Conclusion	68
5	A NEW DIRECT OUTLIER DETECTION MEASURE BASED ON THE DIFFERENCE BETWEEN THE ROBUST LEVERAGE AND TANGENT LEVERAGE	78
	5.1 Introduction	78
	5.2 Jacobian Leverage	79
	5.3 The Generalized and Jacobian Leverage for M-Estimator	82
	5.4 Outlier Detection Measures based on Jacobian Leverage, Robust Jacobian Leverage and M & MM Estimators	86
	5.5 Numerical Example To Assess the Performance of the New Measures Based on Jacobian and Robust Jacobian Leverages	87
	5.6 Simulation Study To Assess the Performance of the New Measures Based on Jacobian and Robust Jacobian Leverages	88
	5.7 Robust Jacobian Leverage and Local Influences	98
	5.8 Numerical Example To Assess the Performance of <i>DLev</i> Measure	103
	5.9 Simulation Study To Assess the Performance of <i>DLev</i> Measure	104
	5.10 Conclusion	111
6	ROBUST ESTIMATOR OF NONLINEAR REGRESSION WITH HETEROSCEDASTIC ERROR VARIANCES	112
	6.1 Introduction	112
	6.2 Model Parameter Estimation	116
	6.3 Variance Modeling and Estimation	117
	6.4 Robust Multistage Estimate	122
	6.5 Numerical Optimization (RME)	130
	6.6 Numerical Example	133
	6.7 Monte Carlo study	137
	6.8 Conclusion	138
7	THE PERFORMANCE OF ROBUST TWO STAGE ESTIMATOR IN NONLINEAR REGRESSION WITH AUTOCORRELATED ERROR	142
	7.1 Introduction	142
	7.2 Nonlinear Autocorrelated Model	144
	7.3 The Classical Two Stage Estimator	146

7.4	Robust Two Stage Estimator	147
7.5	Numerical Example	149
7.6	Monte Carlo Simulation	152
7.7	Conclusion	156
8	A NONLINEAR MODELING APPROACH WHEN THE DATA HAVING HETEROSCEDASTIC ERRORS	166
8.1	Introduction	166
8.2	Commonly Used Modeling Techniques.	167
8.2.1	Heteroscedastic Error	168
8.3	A Formal Technique for Model selection	171
8.3.1	Variance Modeling and Estimation	173
8.4	Applying the Commonly Used Model Selection Technique for Chicken Growth Data	174
8.5	Applying Formal Model Selection Technique (Bunke (1995 a and b)) to Chicken Growth Data	180
8.6	Further Analysis After The Model Selection Result	187
8.7	Conclusion	188
9	GENERAL SUMMARY CONCLUSIONS AND RECOMMENDATION FOR FUTURE RESEARCH	189
9.1	Introduction	189
9.2	Optimization	190
9.3	Contribution of the Study	190
9.3.1	Outlier Detection Measure Based on Tangent Leverage and NLLS, M and MM Estimator	190
9.3.2	A New Direct Outlier Detection Measure Based on the Difference Between the Robust Leverage and Tangent Leverage	191
9.3.3	Robust Estimator of Nonlinear Regression with Heteroscedastic Error Variances	192
9.3.4	The Performance of Robust Two Stage Estimator in Nonlinear Regression with Autocorrelated Error	192
9.3.5	Real Data Modelling	193
9.4	Recommendation for Further Research	194
	REFERENCES	195
	BIODATA OF STUDENT	204
	LIST OF PUBLICATIONS	205