UNIVERSITI PUTRA MALAYSIA

ROBUST DIAGNOSTICS AND ESTIMATION IN HETEROSCEDASTIC REGRESSION MODEL IN THE PRESENCE OF OUTLIERS

MD. SOHEL RANA

IPM 2010 5

ROBUST DIAGNOSTICS AND ESTIMATION IN HETEROSCEDASTIC REGRESSION MODEL IN THE PRESENCE OF OUTLIERS

By

MD. SOHEL RANA

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

October 2010

Dedication

- To the best memory of my parents who passed away and wanted to see my dreams come true
- To my family, having unconditional love for me
- To my one beloved teacher who uplifted my life

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

ROBUST DIAGNOSTICS AND ESTIMATION IN HETEROSCEDASTIC REGRESSION MODEL IN THE PRESENCE OF OUTLIERS

By

MD. SOHEL RANA

October 2010

Chairman: Habshah Midi, PhD

Faculty: Institute for Mathematical Research

The violation of the assumption of homoscedasticity in OLS method, usually called heteroscedasticity, gravely misleads the inferential statistics. The current study has considered the situation when outliers occur in heteroscedastic data. Hence, the main focus of this research is to take remedial measures on the violation of the assumption of homoscedasticity in the presence of outliers. This thesis also concerns on the normality assumption of the errors of regression model in the presence of outliers. It is now evident that outliers have great impact on the existing normality tests, heteroscedasticity tests, and the estimators for heteroscedastic model. We propose the Robust Rescaled Moment (RRM) test for testing the normality of the regression residuals when there is an evidence of outlier(s). The results of the study signify that the RRM test offers substantial improvements over other existing tests in the presence of outliers. For the detection of heteroscedasticity in the presence of outliers, a modified version of

the classical Goldfeld-Quandt (MGQ) test is proposed which is most powerful than the classical tests of heteroscedasticity. Most statistics practitioners assume that the forms of the heteroscedastic error structures are known which may lead to inefficient estimates if it is not correctly specified. In this respect, a Leverage Based Near-Neighbor (LBNN) method is proposed, where prior information on the structure of the heteroscedastic error is not required. The findings indicate that the LBNN is very efficient for correcting the problem of heteroscedastic errors with unknown structure. We also examine the effect of outliers on the existing remedial measures of heteroscedasticity. Hence, in this thesis, a one step M-type of Robust Weighted Least Squares Method (RWLS) and the Two-Step Robust Weighted Least Squares (TSRWLS) are developed. Finally, the new robust wild bootstrap techniques which are resistant to outliers are proposed. The proposed techniques are based on the weighted residuals which incorporated the MM estimator, robust location, robust scale and the bootstrap sampling schemes of Wu (1986) and Liu (1988). All procedures, in this thesis, are examined by using real data and Monte Carlo simulation studies. The comparative studies among the classical and proposed robust methods reveal that all the proposed robust methods outperform the classical methods.

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

DIAGNOSTIK TEGUH DAN PENGANGGARAN MODEL REGRESI BER HETEROSKEDASTIK DENGAN KEHADIRAN TITIK TERPENCIL

Oleh

MD. SOHEL RANA

October 2010

Pengerusi: Habshah Midi, PhD

Fakulti: Institut Penyelidikan Matematik

Andaian homoskedastik yang tidak dipatuhi oleh kaedah OLS, biasanya dinamakan heteroskedastik, memburukkan pentakbiran statistik. Kajian terkini mempertimbangkan situasi titik terpencil berada dalam data berheteroskedastik. Oleh itu, fokus utama dalam kajian ini adalah untuk mengambil langkah yang sewajarnya bagi menangani andaian homoskedastik yang tidak dipenuhi dengan kehadiran titik terpencil. Tesis ini juga mempertimbangkan andaian normal ralat bagi model regresi dengan kehadiran titik terpencil. Bukti terkini menunjukkan bahawa titik terpencil memberikan kesan buruk keatas ujian kenormalan, ujian heteroskedastik dan penganggar bagi model heteroskedastik. Kami mencadangkan ujian Teguh Momen Berskala (RRM) bagi menguji kenormalan reja regresi apabila terdapat bukti kehadiran titik terpencil. Keputusan kajian menunjukkan bahawa ujian RRM lebih baik daripada kaedah sedia ada dalam kehadiran titik terpencil. Bagi pengenalpastian heteroskedastik dengan kehadiran titik terpencil, pengubahsuaian ujian Goldfeld klasik Quandt (MGQ) dicadangkan yang

menunjukkan bahawa ujian MGQ paling berkuasa daripada ujian heteroskedastik klasik. Kebanyakan pengamal statistik mengandaikan bentuk taburan ralat berheteroskedastik diketahui, yang akan menghasilkan pengganggran yang tidak cekap sekiranya andaian ini tidak dipatuhi. Maka, kaedah Tuasan Berdasarkan Penghampiran Kejiranan (LBNN) telah dicadangkan, di mana maklumat asal berkenaan struktur ralat berheteroskedastik tidak diperlukan. Dapatan kajian menunjukkan bahawa kaedah LBNN sangat cekap dalam memperbaiki masalah taburan ralat berheteroskedastik dengan struktur yang tidak diketahui. Kami juga menguji kesan titik terpencil ke atas pengukuran perbaikan bagi heteroscedastik yang sedia ada. Maka dalam tesis ini, kami membangunkan Kaedah Satu Langkah jenis M-bagi kaedah Kuasadua Terkecil Berpemberat Teguh (RWLS) dan Dua langkah Kuasadua Terkecil Berpemberat Teguh (TSRWLS). Akhir sekali, teknik baru *wild bootstrap* teguh yang kalis titik terpencil dicadangkan. Teknik ynag dicadangkan berdasarkan reja berpemberat yang menggabungkan penganggar MM, lokasi teguh, skala teguh dan skema persampelan bootsrap Wu (1986) dan Liu (1988). Semua prosedur dalam tesis ini dikaji dengan menggunakan data sebenar dan simulasi Monte Carlo. Perbandingan antara kaedah klasik dan kaedah teguh yang dicadangkan, menunjukkan bahawa semua kaedah teguh yang dicadangkan lebih baik daripada kaedah klasik.

ACKNOWLEDGEMENTS

Words cannot express how grateful I am to my supervisor, Dr. Habshah Midi, who has taught me through her activities. Her willingness to share her ideas in research problems and, the energy and time she put in advising my thesis work is highly appreciated. I have benefited enormously from her continuous support and confidence throughout my research. Without her help and support, this dissertation would have been impossible. Moreover, her advice and her remarks have proven to be very useful and simulating. I also greatly value her friendship, kindness and elegant personality. I feel truly privileged to have been her student.

I acknowledge my internal co-supervisors Dr. Isthrinayagy Krishnarajah and Dr. Basher Abdul Aziz Majeed Al-Talib, senior lecturers of my institute, for their help. I am indebted to my external co-supervisor Dr. A. H. M. Rahmatullah Imon, associate professor of statistics, Department of Mathematical Sciences, Ball State University, U.S.A, who has helped me a lot by responding to my constant volley of electronic messages regarding my research problems. I am truly grateful that I have such a great mentor.

Special thanks to Dr. S. K. Sarker, fellow researcher of my institute, for his important suggestions and cooperation in my research work. His valuable words always inspired me so much. I am very grateful to Dr. Kudus for helping me in S-Plus programming.

I gratefully acknowledge the moral supports of my friends and their continuous encouragement. I remember my friends Arezoo, Sanizah, Balqish, Hassan, Hossen, Kourash, Ashkan, Vello, Ng, Jahurul, Habib, Shafiq and Rashidul. Thank you to all my friends for all of your generosity and kindness.

Finally, I would like to thank Universiti Putra Malaysia for the financial support. My sincere thanks are extended to all the staff of the Institute for Mathematical Research (INSPEM), UPM, for their cordial assistance during this research work.

I certify that a thesis Examination Committee has met on 22-10-2010 to conduct the final examination of Md. Sohel Rana on his thesis entitled "Robust Diagnostics and Estimation in Heteroscedastic Regression Model in the Presence of Outliers" 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 of Statistics.

Members of the Examination Committee are as follows:

Mohd Rizam Abu Bakar, PhD
Associate Professor
Faculty of Science,
Universiti Putra Malaysia
43400 UPM Serdang, Malaysia
(Chairman)

Noor Akma Ibrahim, PhD
Associate Professor
Institute for Mathematical Research,
Universiti Putra Malaysia
43400 UPM Serdang, Malaysia
(Internal Examiner)

Jayanthi a/p Arasan, PhD
Senior Lecturer
Faculty of Science,
Universiti Putra Malaysia
43400 UPM Serdang, Malaysia
(Internal Examiner)

Mir Masoom Ali, PhD
Emeritus Professor
Ball State University
Mauncie, Indiana 47306-0409 USA
(External Examiner)

SHAMSUDDIN SULAIMAN, PH.D
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

This thesis was submitted to the Senate of Univirsiti 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
Institute for Mathematical Research
Universiti Putra Malaysia
(Chairman)

Isthrinayagy Krishnarajah, PhD
Senior Lecturer
Faculty of science
Universiti Putra Malaysia
(Member)

Bashar Abdul Aziz Majeed Al-Talib, PhD
Senior Lecturer
Faculty of science
Universiti Putra Malaysia
(Member)

A. H. M. Rahmatullah Imon
Associate Professor
Ball State University
Muncie, IN 47306, U.S.A.
(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 at any other institution.

_____

MD. SOHEL RANA

Date: 22 October 2010

CONTENTS