UNIVERSITI PUTRA MALAYSIA

ROBUST ESTIMATION AND DETECTION OF OUTLIERS IN SIMULTANEOUS REGRESSION MODEL

OROOBA MOHSIN MAHDI

FS 2016 84

ROBUST ESTIMATION AND DETECTION OF OUTLIERS IN SIMULTANEOUS REGRESSION MODEL

By

OROOBA MOHSIN MAHDI

© COPYRIGHT UPM

Thesis submitted to the School of Graduate Studies, Universiti Putra Malaysia, in fulfillment of the Requirements for the Degree of Master of Science

December 2016

© COPYRIGHT UPM

COPYRIGHT

All material contained within the hesis, 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

© COPYRIGHT UPM

DEDICATION

- TO my respectful father and lovely mother who taught me the meaning of courage and always had confidence in me.

- To my kids, who accompanied me through the different parts of my study. Their love has always been my greatest inspiration.

© COPYRIGHT UPM

Abstract of thesis presented to the Senate of Universiti Putra Malaysia in fulfilment of the requirement for the degree of Master of Science

ROBUST ESTIMATION AND DETECTION OF OUTLIERS IN SIMULTANEOUS REGRESSION MODEL

By

OROOBA MOHSIN MAHDI

December 2016

Chairman : Prof. Habshah Midi, PhD
Faculty : Science

The Two Stage Least Squares (2SLS) method is the commonly used method to estimate the parameters of the Simultaneous Equation Regression Model (SEM). This method employs the Ordinary Least Squares (OLS) method twice. Firstly, the endogenous X variable is estimated by the OLS and secondly the parameters of the SEM are again estimated using the OLS.

It is now evident that the OLS method is easily affected by outliers. Consequently the 2SLS estimates are less efficient in the presence of outliers. Hence robust estimation methods such as the 2SMM, 2SGM6, 2SMMGM6 and 2SGM6MM are formulated to remedy this problem. These methods employ two robust methods in the first and in the second stages. The findings signify that the 2SGM6MM provides the most efficient results compared to other methods.

Since the distributions of the proposed methods are intractable, robust bootstraps methods are developed to estimate the standard errors of the estimates. The findings indicate that the 2SGM6MM bootstraps standard errors of the estimates are the smallest compared to other estimates.

The identification of high leverage points (HLPs) is very crucial because it is responsible for the drastic change in the parameter estimates of various regression models. Nonetheless, thus far no research has been done to detect HLPs in SEM. Hence, the Diagnostic Robust Generalized Potential (DRGP), Generalized Potential (GP) and Hat Matrix (w_{ii}) are incorporated with OLS, MM and the GM6 estimator in the development of diagnostic measures for the identification of HLPs in SEM. The results of the study show that the DRGPSEM based on the GM6 estimator is the most successful method in the detection of HLPs compared to other methods in this study.

© COPYRIGHT UPM

Abstrak tesis yang dikemukakan kepada Senat Universiti Putra Malaysia sebagai memenuhi keperluan untuk Ijazah Master Sains

ANGGARAN TEGUH DAN PENGESANAN TITIK TERPNCIL DI DALAM MODEL REGRESI SERENTAK

Oleh

OROOBA MOHSIN MAHDI

Disember 2016

Pengerusi : Prof. Habshah Midi, PhD
Fakulti : Sains

Kaedah Kuasa Dua Terkecil Dua Peringkat (2SLS) adalah kaedah yang biasa digunakan untuk menganggarkan parameter Model Regresi Persamaan Serentak (SEM). Kaedah ini menggunakan kaedah Kuasa Dua Terkecil Biasa (OLS) dua kali. Pertama, pembolehubah endogen X dianggarkan dengan OLS dan pada kali keduanya parameter SEM sekali lagi dianggar menggunakan OLS.

Kini adalah jelas bahawa kaedah OLS ini mudah dipengaruhi oleh titik terpencil. Akibatnya anggaran 2SLS adalah kurang cekap dengan kehadiran titik terpencil. Oleh itu kaedah anggaran teguh seperti 2SMM, 2SGM6, 2SMMGM6 dan 2SGM6MM digubal untuk membetulkan masalah ini. Kaedah ini menggunakan dua kaedah teguh di peringkat pertama dan kedua. Dapatan menandakan bahawa 2SGM6MM memberikan hasil yang paling cekap berbanding dengan kaedah lain.

Oleh kerana taburan bagi kaedah yang dicadangkan sukar ditentukan, teguh dibangunkan untuk menganggarkan ralat piawai anggaran. Dapatan kajian menunjukkan bahawa anggaran ralat piawai bootstrap 2SGM6MM adalah yang paling kecil berbanding anggaran lain.

Pengenalpastian titik tuasan tinggi (HLP) adalah sangat penting kerana ia bertanggungjawab menyebabkan perubahan drastik anggaran parameter bagi pelbagai model regresi. Walau bagaimanapun, setakat ini tidak ada kajian yang dilakukan untuk mengesan HLP di dalam SEM. Oleh itu, Potensi Teritlak Teguh Berdiagnostik (DRGP), Potensi Teritlak (GP) dan Matriks Topi (w_{ii}) digabungkan dengan OLS, MM dan penganggar GM6 di dalam membangunkan ukuran diagnostik untuk mengenal pasti HLP di dalam SEM. Keputusan kajian menunjukkan bahawa DRGPSEM berdasarkan penganggar GM6 adalah kaedah yang paling berjaya daripada yang lain dalam pengesanan HLP berbanding kaedah lain dalam kajian ini.

© COPYRIGHT UPM

ACKNOWLEDGEMENTS

First of all, I wish to thank God who always supported me in all difficulties of my study life.

To have successful children has been one of my parent's dreams. I tried as much effort

as I could to fulfil their dreams in order to thank them sincerely for scarifying their life to grow me up.

I would like to express my deep gratitude to my master thesis advisors, Prof. Habshah and Dr.Md Sohel. I have learned many things since I became Prof. Habshah and Dr. Md. Sohel student. They spend much time for the completion of this thesis .

I could not possibly forget all the wonderful people that have offered me their friendship and have enriched my life during these master studies duration. My acknowledgement would be incomplete without mentioning my friends Hassan, Mohammed, Shelan, Ahmed and all the others who made wonderful memories for me. Thanks you all. Lastly, my special thanks to my kids and my husband whose patience is admirable for me. Without his undoubting faith, my thesis would never have been completed. My sincerely regards to my sisters, specially my elder one, my brother, who encouraged me not to miss my hope in doing my research and supported me a lot mentally.

My master studies wouldn't be possible without the scholarship granted to me by my country. Much gratitude is also due to the entire faculty of science members who created an environment in which master and PhD students can flourish. I was lucky to have the chance to be graduated from this faculty.

© COPYRIGHT UPM

I certify that a Thesis Examination Committee has met on 9 December 2016 to conduct the final examination of Orooba Mohsin Mahdi on her thesis entitled "Robust Estimation and Detection of Outliers in Simultaneous Regression Model" in accordance with the 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 Master of Science.

Members of the Thesis Examination Committee were as follows:

Leong Wah June, PhD
Associate 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)

NOR AINI AB. SHUKOR, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 22 March 2017

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

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

Sohel Rana, PhD
Senior Lecturer
Faculty of science
Universiti Putra Malaysia
(Member)

ROBIAH BINTI YUNUS, PhD
Professor and Dean
School of Graduate Studies
Universiti Putra Malaysia

Date:

© COPYRIGHT UPM

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: Orooba Mohsin Mahdi, GS41685

© COPYRIGHT UPM

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) were adhered to.

Name of Chairman of Supervisory Committee: _____

Signature: _____

Name of Member of Supervisory Committee: _____

© COPYRIGHT UPM

TABLE OF CONTENTS

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF TABLES

© COPYRIGHT UPM

LIST OF FIGURES

© COPYRIGHT UPM

LIST OF ABBREVIATIONS

ASE	Asymptotic Standard Error
AMSEE	Average Mean Square Error of Estimation
BLUE	Best Linear Unbiased Estimators
CVIF	Classical Variance Inflation Factor
CI	Condition Indices
CN	Condition Number
CEV	Contaminated Explanatory Variable
CP	Contaminated Point
DRGP (MVE)	Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid
GM-estimator	Generalized M-estimator
GP	Generalized Potentials
HLCIM	High Leverage Collinearity-Influential Measure
HLCIM (DRGP)	High Leverage Collinearity-Influential Measures based on DRGP
ILD	Interstitial Lung Disease data set
IWLS	Iterative Weighted Least Squares
IRLS	Iteratively Reweighted Least Squares
LAD	Least Absolute Deviations
LAR	Least Absolute Residuals
LAV	Least Absolute Values
LMS	Least Median of Squares
LTS	Least Trimmed Squares
MC	Magnitude of Contamination
MD	Mahalanobis Distance
MSE	Mean Square Errors
MAD	Median Absolute Deviation
MCD	Minimum Covariance Determinant
MSAE	Minimum Sum of Absolute Errors
MVE	Minimum Volume Ellipsoid
MGM1	Modified GM-estimator 1
MGM2	Modified GM-estimator 2
MGM3	Modified GM-estimator 3

© COPYRIGHT UPM

MADN	Normalized Median Absolute Deviation
OLS	Ordinary Least Squares method
RR^2 (MM)	Robust coefficient determinations (R^2) based on MMestimator
RR^2 (MGM3)	Robust coefficient determinations (R^2) based on Modified GM- estimator 3
RCI (MCD)	Robust Condition Indices based on MCD
RCN	Robust Condition Number
RCN (MCD)	Robust Condition Number based on MCD
RD	Robust Distance
RMD (MVE)	Robust Mahalanobis Distance based on the Minimum Volume Ellipsoid
RVDP	Robust Variance Decomposition Properties
RVDP (MCD)	Robust Variance Decomposition Properties based on MCD
RVIF	Robust Variance Inflation Factors
RVIF (MCD)	Robust Variance Inflation Factors based on MCD
VDP	Variance Decomposition Proportions
VIF	Variance Inflation Factors
WLS (LMS)	Weighted Least Squares based on LMS
WLS	Weighted Least Squares regression

© COPYRIGHT UPM

CHAPTER 1

INTRODUCTION

1.1 Background and Purposes

Regression analyses are very essential for examining the operative associations between many variables, such that the dependent or response variables are predictable from single or multiple predictors or descriptive variables (Kutner et al., 2005). Regression analyses involve model construction, parameters approximations (estimates) and predictions. Of the known methods, the ordinary least-squares method (OLS), represents the most widespread and prevalent estimating criteria. The structure of OLS is grounded on the minimisation of the sums of squared deviations, particularly when the errors distribution are normal. The OLS estimators are the best-case linear non-biased estimating rules, or BLUE. Technically, OLS estimators present the least variances of all potential linear measures. Thus, of all the neutral estimating rules, the OLS generates the smallest variances in estimation schemes so long as regular assumptions about their error terms are not contravened. As well, the maximum likelihood estimators or MLE equal the OLS estimator, given these preconditions. The OLS method can generate erratic estimates if assumptions about error normality do not hold (Ryan, 1997). With violations in any one of the assumptions, the OLS method will exhibit high sensitivity and thus lesser reliability as a means of parameters estimation. Moreover, the OLS estimating rules do not operate robustly against outlying data and exhibit an exceptionally low breakdown point equal to 1/*n*, whereby *n* equals the sampling size (Maronna, 1976).

Some statisticians are unaware of the violations of normality assumptions about error terms that may be caused by at least one observed divergence, i.e. outliers. It was claimed by Belsley et al., (1980) that observational significance is highest in observations which alone or in copious combination have the greatest effect on the calculated values of diverse estimates. Barnett and Lewis (1994) pointed out that outlier is a data point observed to be significantly distanced from the majority in a dataset.

Many outlier types manifest as problems in regressions. An observation is assessed as a residual outlier to the degree it fails to be accommodated in a best-fit regression. This explains why data points which corresponds to a sizeable residuals are considered as residual outliers. Outliers can take place in three directions. Rosseeuw and Zomeren (1990) defined an outlier category in the X-direction, which is referred to as high leverage point (HLP). In regression analyses, it is usually necessary to identify HLPs, as an observation far-removed from the bulk of independent variables that has more effect on a model's fitting. HLPs are not only distant from the bulk of predictors; rather, these also diverge appreciably from regression lines (Belsley et al.,(1980) Hocking and Pendelton, (1983) and Rousseeow and Leroy, (1987)). Another outlier category appears in the Y-direction or the vertical outlier, an

© COPYRIGHT UPM

outlying data point with a sizable squared residual from fitting. The third outlier type appears in both X- and Y-directions concurrently.

Prior to remedy the outliers, it must be ascertained that the dataset is actually subject to the problem. Effective detections of outliers may result in effective remedies. The majority of outlier detection techniques are grounded on OLS procedures which exhibit great sensitivity to outlying observations. In any case, a single outlier is sufficient to breakdown an OLS estimation (Rousseeuw and Leroy, 1987).

As implied, simultaneous equation models or SEM is a schematic category centred on procedurally generated datasets which are dependent on several interactive equations jointly creating on observations.

In contrast to single-equation model, where dependent or variables (*y*) functionally vary with independent or variables (*x*), additional variables (*y)* operate as independent variables in SEM computations. These added variables are mutually and concurrently established by the system's equations.

These models are subject to the non-independent behaviours of certain descriptive variables which appear endogenously in other equations and the error terms, resulting in biases and discrepancies in the estimating. Thus, statistical practitioners tend to employ reduced forms wherein such internally caused variables occur as functions of pre-determined variables. These incorporate both exogenous and lagged-endogenous variables which are independent of random errors. There exist effective methods for achieving consistent estimation of the system's coefficients. The most frequently employed is the Instrumental Variables approach, or IV. With the availability of many instrumental variables, these are merged through first-stage regressions and then re-utilised in the second-stage regressions, in a technique known as two-stage least-squares, or 2SLS (else TSLS). However, in the event of a non-normal error, these would be unproductive. The use of estimated parameter sets in subsequent phases (Gao et al., 2008) only worsens the problem. We have an on-going need to discover estimation procedures of greater robustness and reliability for simultaneous equation models (Mishra, 2008).

1.2 Significance and Purpose of the Study

Simultaneous equation models are among the most useful econometric models. For estimating the coefficient in these models, the two-stage least-squares technique (2SLS) is traditionally employed as it is readily computed. But the existence of single or numerous outsized outliers in datasets can undermine 2SLS estimations. Several authors claimed that atypical data points normally comprise 1–10% of actual datasets (Hampel et al., 1986; Wilcox, 2005). The presence of HLPs in observations impact 2SLS estimates more seriously than outlying y-variable data points. Simultaneous regression estimators consequently exhibit breakdowns when outliers

© COPYRIGHT UPM

are present, as just one outlier is sufficient to undermine estimates (Rousseeuw and Leroy, 1987).

Unfortunately, to date, not much work has been focus on the parameter estimations of simultaneous equation model in the presence of outliers. The 2SLS method is the most widely used method to estimate the parameters of the SEM. In the 2SLS method, the OLS method is used to estimate the endogenous X variable and then again the OLS is used to estimate the parameters of the models. It is now evident that the OLS is very sensitive to outliers and produces very poor result. To remedy this problem (Insha Allah (2006), Hassan (2012), Ahmed et al., (2013) propose to use the 2SM whereby the M estimator is employed to estimate the endogenous X variable and then again the M estimator is used to estimate the parameters of the SEM. However, the M estimator has shortcoming in which it has very low breakdown point and not robust against HLPs. Therefore, to acquire effective parametric estimates, we suggest employing efficient robust methods in obtaining the estimates for the parameters of SEM. Pena and Yohai, (1995) stated that HLP occurrences account for outlying data points being swamped and masked in simultaneous regressions. Since HLPs have a drastic effect on the parameter estimates, their effect need to be reduced. The Generalized – M estimator (GM6) is very efficient in reducing the effect of HLPs. In this respect we incorporated the GM6 estimator to firstly estimate the endogenous *X* variable. Subsequently we suggest using the high–breakdown and high efficiency MM estimator to estimate the parameters of the SEM. No such attempt has been done and we anticipate that it will give good results. Additionally, we suggest to investigate a variety of combination of robust methods such as 2SMM, 2GM6, 2SMMGM6, 2 SGM6MM for the estimations of parameters in the first and the second stage.

It is now evident that HLPs greatly influence the parameter estimations of SEM. Thus the detections of HLPs are important. The use of Hadi s' potential (Hadi,(1992)) can assist in detecting single leverage points, but the method is ineffective in recognising multiple HLPs as a result of swamping and masking effect (Rousseeuw and Leroy, (1987); Ruppert and Simpson, (1990); Imon, (1996); Imon, (2005), Habshah et al., (2009)).

Prior to remediation against HLPs, it must be determined if the dataset is actually subject to the problem. Effective detections of HLPs can result in effective remedies. The majority of detection methods for HLPs are based on OLS techniques that exhibit great sensitivity to HLPs. Imon (1996) proposed the generalised potential approach, or GP, for diagnosing and identifying multiple HLPs. The conceptual advantage of generalised potentials is in the extension of a single-case deletion to a group-case deletion. However, Habshah et al. (2009) showed that the GP method is not highly viable in detecting HLPs owing to its inefficiency in selecting for the initialised base sub-set, which is still subject to masking influences. Habshah et al., (2009) developed the diagnostic robust generalised potentials methods, or DRGP, to resolve the issue, which has been proven to be very effective in diagnosing for HLPs.

© COPYRIGHT UPM

To the best of our knowledge, no work has been focussed on the detection of HLPs in SEM. To close this gap in the literature, we attempt to develop technique of identification of HLPs in SEM. The propose method first requires estimating the endogenous variable X by using GM6 estimator and then adopt the DRGP of Habshah et al. (2009) to detect HLPs.

1.3 Research Objectives

The primary purpose of this study is to examine the outliers' problem in simultaneous regression models. Classical techniques of diagnosis of HLPs, and estimation of the model parameters are usually based on ordinary least-squares (OLS) estimates, even though the approach is not resilient in handling outliers' problem. Thus, it is vital to improve the classical estimation procedures to be more robust in the presence of outlying data. In this research, we only focussed on two simultaneous equations model. Towards these ends, the primary goals of our investigation are presented in the following:

1- To develop a new method of estimating the common independent variable (endogenous X variable) in SEM.
2- To propose new estimation method for estimating the parameters of the simultaneous equations models in the presence of outliers.
3- To develop robust bootstraps technique to estimate the standard errors of the proposed robust estimates in Sem.
4- To develop a new method of identifying high leverage points in simultaneous equation model.

1.4 Thesis Outline

In line with the objectives and scope of this research, the subjects of the thesis are arranged in six sections. After the introduction, the various sections are ordered such that research goals are clearly presented in the outlined sequence.

Chapter Two. This segment presents a concise survey of the literature on the least-squares estimation technique and the issues of violating its causal assumptions, e.g. departures of normality and the problematic existence of outliers. The nature of outlying data, HLPs, and their diagnosis are covered. In addition, the fundamental concepts of reliable regression analyses as well as certain robust estimation criteria are similarly discussed, along with explanations of efficiencies, breakdown points, and bounded influences. The impacts and estimation and diagnostic procedures are likewise emphasised. At the end of the section, concise appraisals of reliable methods for detecting outliers and estimating for simultaneous regressions equation are offered.

© COPYRIGHT UPM

Chapter Three. In this segment, we suggest reliable robust two-stage techniques for parameter estimations of SEM in the presence of HLPs. Monte Carlo simulation study and real data are used to evaluate the performances of various two-stage procedures. The robust bootstrap standard errors of the parameter estimates are exhibited.

Chapter Four. In this segment, the detection of HLPs methods based on OLS, MM and GM6 of the SEM are discussed. The proposed methods are DRGP, GP and w_{ii} which is based on the OLS, MM and GM6 estimates respectively. Certain real datasets and a Monte Carlo simulation study are used to evaluate the performance of our recommended approaches.

Finally, Chapter Five. This chapter provides a summary and detailed discussion of the results, contributions, and recommendations for further research.

© COPYRIGHT UPM

REFERENCES

Andrews, D. F. (1974). A robust method for multiple linear regression. Technometrics.16:523-531.

Atkinson, A.C. (1982). Regression diagnostics, transformations and constructed variables (with discussion). Journal of the Royal Statistical Society, Series B, 44, 1-36.

Armstrong, R. D. and Kung, M. T. (1978). Least absolute values estimates for a simple linear regression problem. Applied Statistics. 27: 363-366.

Barnett, V. and Lewis, T. (1994). Outliers in Statistical Data. 3rd edition. New York: Wiley.

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

Belsley, D.A., Kuh, E. and Welsch, R.E. (1980). Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.

Chatterjee, S. and Hadi, A.S. (1988). Sensitivity Analysis in Linear Regression. NewYork:Wiley.

Chatterjee, S. and Hadi, A.S. (2006). Regression Analysis by Example. 4th edition. New York: Wiley.

Cook, R. D. (1977). Detection of influential observation in linear regression. Technometrics. 19:15-18.

Donoho, D. L. (1982). Breakdown Properties of Multivariate Location Estimators. Unpublished Ph.D. thesis, Harvard University, The American United States.

Draper, N. R. and Smith, H. (1998). Applied Regression Analysis. New York:Wiley.

Ellenberg, J.H. (1976). Testing for a single outlier from a general linear regression. Biometrics. 32: 637-645.

Greene, W. H. (2008). Econometric analysis. 6th edition. Upper Saddle River. New Jersey: Prentice Hall.

Habshah, M., Norazan, M.R. and Imon, A.H.M.R. (2009). The performance of Diagnostic-Robust Generalized Potentials for the identification of multiple high leverage points in linear regression. Journal of Applied Statistics. 36(5): 507-520.

Hadi, A. S. (1992). A new measure of overall potential influence in linear regression. Computational and Statistical Data Analysis. 14:1-27.

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

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). Robust Statistics. New York: Wiley.

© COPYRIGHT UPM

Hocking, R.R. & Pendelton, O.J. (1983). The regression dilemma,Communications in Statistics-Theory and Methods 12: 497-527.

Huber, P. J. (1964). Robust estimation of location parameters. Annals of Mathematical Statistics. 35:73–101.

Huber, P. J. (1973). Robust regression: asymptotic, conjectures, and Monte Carlo. The Annals of Statistics, 1: 799-821.

Huber, P.J. (1981). Robust statistics, Wiley:New York.

Huber, P.J (2003). Robust Statistics, Wiley, New York, USA.

Imon, A.H.M.R. (2002). Identifying multiple high leverage points in linear regression. Journal of Statistical Studies. Special Volume in Honour of Professor Mir Masoom Ali. 3: 207–218.

Imon, A.H.M.R. (2005). Identifying multiple influential observations in linear regression. Journal of Applied Statistics. 32: 929-946.

Imon A.H.M.R (2005a) A Stepwise Procedure for th Identification of Multiple Outliers and High Leverage Points in Linear Regression. Pakistan Journal of Statistics, 21, 71-86.

Insha Ullah, M. F. Qadir and S. Ali (2006), Insha's redescending M-estimator for robust regression: A comparative study, Pak. J. Stat. Oper. Res. II (2), 135-144.

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

Krasker,W.S.(1986).Two –stage bonded-influence estimators for simultaneous-equations Models.J.Bus Econom.Statist.4,437-444,

Krasker,W.S.and RE.Welsch(1985).Resistant estimation for simultaneous-equations models using weighted instrumental variables.Econometrica53,1475-1488

Kutner, M.H., Nachtsheim, C.J., Neter, J. and Li, W. (2005). Applied Linear Regression Models. 5th edition. New York: MacGRAW-Hill.

Kuh,E.and R.E.Welsch(1980).Econometric models and their assessment for policy:some new diagnostics applied to the translog energy demand in manufacturing In:S.Gass,Ed.,Proc.Workshop on Validation and Assessment Issues of Energy Models.National Bureau of Stanard,Washington DC,445-475

Kim, M. G. (2004). Sources of high leverage in linear regression model. *Journalof Applied Mathematics and Computing*.16(1- 2): 509-513.

Maronna, R. A. (1976).Robust M-Estimators of Multivariate Location and Scatter. The Annals of Statistics. 4: 51–67.

Maronna, R.A., Martin, R.D. and Yohai, V.J. (2006). Robust Statistics Theory and Methods. New York: Willy and sons.\

Meintanis S.G., Donatos G.S., 1997. A comparative study of some robust methods for coefficient-estimation in linear regressionComputational Statistics and Data Analysis, 23, 525–540.

Midi, H. and Jaafar, A. (2004). The residual plot for a non-linear regression modelwith the presence of outliers and heteroscedastic errors. Jurnal Teknologi.41(c): 11–26.

Mosteller, F. and Tukey , J. W. (1977) . Data Analysis and Regression. Reading.MA: Addison- Wesley Publishing Company.

Norazan, M.R. (2008). Weighted Maximum Median Likelihood Estimation for Parameters in Multiple Linear Regression Model, Unpublished Ph. D.Thesis, University Universiti Putra Malaysia, Malaysia.

Peña, D. and Yohai, V.J. (1995). The detection of influential subsets in linear regression by using an influence matrix. Journal of Royal Statistical Society. B 57: 18–44.

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

Rousseeuw, P.J. and Leroy, A.M. (1987). Robust Regression and Outlier Detection. New York: Wiley.

Rousseeuw, P. J. (1984). Least median of squares regression. Journal of theAmerican Statistical Association. 79: 871–880.

Rousseeuw, P. J. and Croux, C. (1993). Alternatives to the median absolute values. Journal of the American Statistical Association. 88: 1273-83.

Rousseeuw P.J. and Van Driessen, K. (1999). A fast algorithm for the minimum covariance determinant estimator. Technometrics. 41:212–223.

Rousseeuw, P. and Van Zomeren, B. (1990).Unmasking multivariate outliers and leverage points. Journal of American Statistical Associations. 85: 633-639.

Rousseeuw, P. J. and Yohai, V. (1984). Robust regression by means of S-estimators, Robust and Nonlinear Time series Analysis. Lecture Notes in Statistics. 26: 256-272.

Ramsay, J.O.(1977). A comparative study of several robust estimates of slope, intercept, and scale in linear regression, Journal of American Statistical Associations.72:608-615.

Ryan, T. P. (1997). Modern Regression Methods. NewYork: Wiley.

Simpson, J. R. (1995). New Methods and Comparative Evaluations for Robust and Biased-Robust Regression Estimation. Unpublished Ph.D. thesis, Arizona State University, The United States of America.

Uraibi, H. S. (2009) .Dynamic Robust Bootstrap Algorithm for Linear Model Selection Using Least Trimmed Squares, Unpublished M.Sc. thesis, Universiti Putra Malaysia, Malaysia.

© COPYRIGHT UPM

Vellman, P.F. and Welsch, R.E. (1981). Efficient computing of regression diagnostics. American Statistician. 27: 234-242.

Wilkinson L. (1999), Statistical Methods in psychology,journals American Psychologist, 54, 594-604.

Weisberg, S. (1980). Applied Linear Regression. New York: Wiley.

Yohai, V.J. (1987). High breakdown point and high efficiency robust estimates for regression. The Annals of Statistics. 15: 642-656.

© COPYRIGHT UPM