ROBUST DIAGNOSTIC AND PARAMETER ESTIMATION FOR MULTIPLE LINEAR AND PANEL DATA REGRESSION MODELS

MUHAMMAD SANI

IPM 2019 3

ROBUST DIAGNOSTIC AND PARAMETER ESTIMATION FOR MULTIPLE LINEAR AND PANEL DATA REGRESSION MODELS

By

MUHAMMAD SANI

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

November 2018

© COPYRIGHT UPM

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

© COPYRIGHT UPM

DEDICATION

To my parents;

My mother Hajjiya Fatima Abubakar &

Father Late Alhaji Sani Ibrahim K/soro (May his soul rest in perfect peace, Ameen)

© COPYRIGHT UPM

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

ROBUST DIAGNOSTIC AND PARAMETR ESTIMATION FOR MULTIPLE LINEAR AND PANEL DATA REGRESSION MODELS

By

MUHAMMAD SANI

November 2018

Chairman : Professor Habshah Midi, PhD
Institute : Mathematical Research

The Influential Distance (ID) is proposed to identify multiple influential observations (IOs) in linear regression. However, the method not only considered good leverage observations (GLOs) as IOs, but also takes long computational running time with high rate of swamping and masking effects. Fast Improvised Influential Distance (FIID) is proposed to overcome these shortcomings. The results indicate that FIID successfully identified and classified GLOs and IOs with less computational running time, no masking effect and smaller rate of swamping.

The presence of high leverage points (HLPs) and violation of the assumption of homoscedasticity are very common in analyzing data in linear and panel data regression models. To remedy this problems weighted least squares (WLS) based on FIID weighting method for Heteroscedasticity Consistent Covariance Matrix (HCCM) estimator is developed. The results obtained from simulation study and real data sets indicate that the proposed method is superior compared to the existing methods.

The presence of outlying observations in a data set causes heteroscedasticity in a homoscedastic data set and vice versa. To know the type of outliers that are responsible for these irregularities is very important so that appropriate measure will be taken. To bridge the gap in the literature, we have successfully proposed robust White test to detect heteroscedasticity and identifies the types of outliers that causes and hide heteroscedasticity termed heteroscedasticity-enhancing and heteroscedasticity-reducing observations (HEO and HRO), respectively. Furthermore, we proposed appropriate remedial measures for both HEO and HRO denoted by GM-FIID and ITSRWLS, respectively. The results of the simulation study show that the proposed methods are efficient and consistent than the existing methods.

© COPYRIGHT UPM

The panel data estimators for both fixed and random effect models becomes bias and causes inconsistency in variance-covariance matrix when there exist heteroscedasticity of unknown form and high leverage points in a data set. To date no research has been done to address this problem. To fill-in the gap in the literature we proposed a WLS estimation technique for both fixed and random effect model based on RHCCM estimator with FIID weighting method. The MM-Centering technique is employed instead of mean centering to reduce the effect of HLPs. The results of simulation study and real data sets indicate that weighted least squares based on FIID (WLS_{FIID}) was found to be the best method.

The classical Hausman pretest is used to choose between random and fixed effect panel data models. In the presence of heteroscedastic error variances and high leverage points (HLPs) or IOs in a data set, the right model may not be correctly identified. To the best of our knowledge no research has been done to address this issue. We proposed a robust Hausman pretest denoted as RHT_{FIID} based on FIID and Robust Heteroscedasticity Consistent Covariance Matrix (RHCCM) estimator to remedy the problem. The results of simulation and real data set indicate that the proposed method was found to perform better than the conventional Hausman pretest.

© COPYRIGHT UPM

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

DIAGNOSTIK TEGUH DAN PENGANGGARAN PARAMETER BAGI MODEL REGRESI LINEAR BERGANDA DAN MODEL REGRESI PANEL DATA

Oleh

MUHAMMAD SANI

November 2018

Pengerusi : Profesor Habshah Midi, PhD
Institute : Penyelidikan Matematik

Jarak Pengaruh (ID) dicadangkan untuk mengenal pasti cerapan berpengaruh berganda (IOs) dalam regresi *linear*. Walau bagaimanapun, kaedah ini bukan sahaja menganggap titik tuasan yang baik (GLO) sebagai IO, tetapi juga mengambil masa pengiraan yang lama dengan kadar *swamping* dan *masking* yang tinggi. Peningkatan Jarak pengauh Pantas (FIID) dicadangkan untuk mengatasi kekurangan ini. Hasilnya menunjukkan bahawa FIID telah berjaya mengenal pasti dan mengklasifikasikan GLO dan IO dengan masa pengiraan yang pendek, tiada kesan *masking* dan kadar *swamping* yang rendah.

Kehadiran titik tuasan tinggi (HLPs) dan pelanggaran terhadap andaian homoskedastisiti adalah menjadi kebiasaan dalam menganalisis data dalam model regresi *linear* dan regresi panel data. Untuk mengatasi masalah ini, pemberat kuasadua terkecil (WLS) berdasarkan kaedah pemberat FIID bagi anggaran kepada *Heteroskedastisity ConsistentCovariance Matrix* (HCCM) di cadangkan. Hasil yang diperoleh daripada kajian simulasi dan set data sebenar menunjukkan bahawa kaedah yang dicadangkan lebih unggul berbanding kaedah yang sedia ada.

Kehadiran titik terpencil dalam set data menyebabkan heteroskedastisiti dalam set data homoskedastik dan sebaliknya. Untuk mengetahui jenis titik terpencil yang bertanggungjawab terhadap penyelewengan ini adalah sangat penting supaya langkah yang sesuai boleh diambil. Untuk merapatkan jurang dalam literatur, kami telah berjaya mencadangkan ujian Putih yang teguh untuk mengesan heteroskedastisiti dan mengenalpasti jenis titik terpencil yang menyebabkan dan menyembunyikan heteroskedastisiti yangmasing-masing dinamakan penggalakkan-heteroskedastisiti

© COPYRIGHT UPM

dan pengurangan-heteroskedastisiti (HEO dan HRO) masing-masing. Selain itu, kami mencadangkan langkah pemulihan yang sesuai untuk kedua-dua HEO dan HRO yang masing-masing, digelar sebagai GM-FIID dan ITSRWLS. Hasil kajian simulasi menunjukkan bahawa kaedah yang dicadangkan adalah cekap dan konsisten daripada kaedah sedia ada.

Penganggar data panel untuk kedua-dua model kesan tetap dan rawak menjadi tidak saksama dan menyebabkan ketidakkonsistenan dalam matriks variasi-kovarians apabila terdapat heteroskedastisiti bentuk yang tidak diketahui dan titik tuasan tinggi dalam set data. Sehingga kini tiada kajian telah dilakukan untuk menangani masalah ini. Untuk mengisi jurang dalam kesusasteraan, kami mencadangkan penganggaran WLS untuk kedua-dua model kesan tetap dan rawak berdasarkan anggaran RHCCM dengan kaedah pemberat FIID. Kaedah MM-berpusatdigunakan dan bukannya purata berpusat untuk mengurangkan kesan HLPs. Hasil kajian simulasi dan set data sebenar menunjukkan bahawa kaedah kuasadua terkecil berpemberat berdasarkan FIID (WLS_{FIID}) telah dikenalpasti sebagai kaedah terbaik.

Pra-ujian Hausman klasik digunakan untuk memilih antara model data panel kesan rawak dan tetap. Dengan kehadiran variasi ralat berheteroskedastik dan titik tuasan tinggi (HLPs) dalam set data, model yang betul mungkin tidak dapat dikenal pasti. Bagi pengetahuan terbaik kami, tiada kajian telah dilakukan untuk menangani isu ini. Kami mencadangkan pra-ujian Hausman yang teguh yang dipanggil RHT_{FIID} berdasarkan FIID dan anggaran Matriks Kovarians Konsisten Teguh Berheteroskedastik(RHCCM)untuk memperbaiki masalah tersebut. Hasil daripada simulasi dan set data sebenar menunjukkan bahawa kaedah yang dicadangkan didapati lebih baik daripada pra-ujian Hausman konvensional.

ACKNOWLEDGEMENTS

All praises and glorification are due to Allah (SWT). I thank Him for the ability, wisdom and strength bestowed on me to conduct this study. I really thank Him.

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Habshah Midi, who has taught and put me through over the years. I have benefited enormously from her continuous support and confidence throughout my research. Without her help and support, this dissertation would have been impossible, I feel truly privileged to have been her student. I am also very grateful to my committee members Associate Prof. Dr. Jayanthi A/P Arasan, Dr. Mohd Shafie Bin Mustafa for their guidance, encouragement and support. I thank all the staff of Institute for Mathematical Research for their help and support in one way or the other.

I would also like to express my gratitude to my mother Hajiya Fatima for her prayers, support and encouragement, to also my brothers, sisters and relative for their moral support and prayers. My special thanks go to my beloved wife Amina for her prayers, endurance, support and encouragement. I would not forget the patient and endurance of our beloved children Fatimah and Abubakar they have really scarifies much in order to see this dream has become a reality, may Allah bless them. Similarly, I would like to appreciate my friends, colleagues in research, housemate and well-wishers for their prayers.

Lastly, I am very grateful to TetFund for their financial support during my studies, without their support this study wouldn't have been possible. I also sincerely appreciate Institute for Mathematical Research and UPM in general for providing support and assistance in one way or the other.

© COPYRIGHT UPM

I certify that a Thesis Examination Committee has met on 19 November 2018 to conduct the final examination of Muhammad Sani on his thesis entitled "Robust Diagnostic and Parameter Estimation for Multiple Linear and Panel Data Regression Models" 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 Doctor of Philosophy.

Members of the Thesis Examination Committee were as follows:

Ibragimov Gafurjan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairman)

Mohd Rizam b 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 Qaiser Shahbaz, PhD
Professor
King Abdul Aziz University Jeddah
Saudi Arabia
(External Examiner)

RUSLI HAJI ABDULLAH, PhD
Professor and Deputy Dean
School of Graduate Studies
Universiti Putra Malaysia

Date: 13 December 2018

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

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

Jayanthi A/P Arasan, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Mohd Shafie Bin Mustafa, 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 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: Muhammad Sani, GS46046

© 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.

Signature: ______________________
Name of Chairman of Supervisory Committee: Professor Dr. Habshah Bt Midi

Signature: ______________________
Name of Member of Supervisory Committee: Associate Professor Dr. Jayanthi A/P Arasan

Signature: ______________________
Name of Member of Supervisory Committee: Dr. Mohd Shafie Bin Mustafa

© COPYRIGHT UPM

TABLE OF CONTENTS

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF TABLES

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF FIGURES

© COPYRIGHT UPM

© COPYRIGHT UPM

LIST OF APPENDICES

© COPYRIGHT UPM

LIST OF ABBREVIATIONS

BLUE	Best Linear Unbiased Estimators
CD	Cooks Distance
CHIM	Classical Heteroscedasticity Influential Measure
DRGP(MVE)	Diagnostic Robust Generalized Potential based on Minimum Volume Ellipsoid
$DRGP_{(ISE)}$	Diagnostic Robust Generalized Potential based on Index Set Equality
FE	Fixed Effect
FIID	Fast Improvised Influential Distance
FMGt	Fast Modified Generalized Studentized Residual
FMGt-$DRGP_{(ISE)}$	Fast Modified Generalized Studentized Residual with $DRGP_{(ISE)}$
GLO	Generalized Leverage Observation
GLV	Generalized Leverage Value
GM	Generalized M-estimator
GP	Generalized Potential
GM-FIID	Generalized M-estimator with Fast Improvised Influential Distance
GUM	Group Union Method
GSR	Generalized Studentized Residual
HCCM	Heteroscedasticity Consistent Covariance Matrix
HCCME	Heteroscedasticity Consistent Covariance Matrix Estimator
HEO	Heteroscedasticity-Enhancing Observation
HIO	Heteroscedasticity-Influential Observations
HLHIO	High Leverage Heteroscedasticity-Influential Observations
HLP	High Leverage Point

© COPYRIGHT UPM

HRO	Heteroscedasticity-Reducing Observation
HT	Hausman Test
ID	Influential Distance
IO	Influential Observations
ISE	Index Set Equality
RE	Random Effect
RHT_{FIID}	Robust Hausman Test based on FIID
RMD	Robust Mahalanobis Distance
RO	Regular Observations
LM	Lagrange Multiplier
LMS	Least Median Squares
LP	Leverage Point
LTS	Least Trimmed Squares
MAD	Median Absolute Deviation
MCD	Minimum Covariance Determinant
MD	Mahalanobis Distance
MM	Modify of M-estimator
MGt	Modified Generalized studentized Residual
MGt-DRGP	Modified Generalized studentized Residual with Diagnostic Robust Generalized Potential
MSGME	Multi-Stage Generalized M-estimator Estimator
MTSRWLS	Modified Two-Step Robust Weighted Least Squares
MVE	Minimum Volume Ellipsoid
OLS	Ordinary Least Squares
RHCCM	Robust Heteroscedasticity Consistent Covariance Matrix

© COPYRIGHT UPM

RMD(MVE)	Robust Mahalanobis Distance based on Minimum Volume Ellipsoid
RMD(ISE)	RMD based on Index Set Equality
RWT	Robust White Test
TSRWLS	Two-Step Robust Weighted Least Squares
VO	Vertical Outlier
WLS	Weighted Least Square
WT	White Test

© COPYRIGHT UPM

CHAPTER 1

INTRODUCTION

1.1 Background and Purposes

Regression analysis is an important statistical method for investigating the linear relationships between the response variable and one or more predictor variable(s). It was introduced by Sir Frances Galton in the nineteenth century. There are several techniques for modeling and analyzing variables in linear regression. The ordinary least squares (OLS) technique introduced by Legendre and Gauss (Maronna et al., 2006) has been generally adopted due to its simplicity and computational ease. The OLS estimates are obtained by minimizing the sum of squared errors. Under usual assumptions, that is, the distribution of the errors (residuals) is normal and the residual variances are equal (satisfied Gauss-Markov theorem) the OLS method provides the Best Linear Unbiased Estimator (BLUE) for the parameter of a linear model. Because of the convenient properties of the OLS, such as closed form solution and ease of computation, it is often applied in many fields of study such as applied sciences and Engineering. Among these convenient properties is the assumption of homogeneity of its residual variances, commonly called homoscedasticity. Rana et al. (2012), Carrol and Ruppert (1982) as well as Habshah and Bashar (2008) elucidated many different occasions where homoscedasticity assumption breaks down and resulted to heteroscedasticity (unequal residual variances). The heteroscedasticity problem has been reported by many researchers such as (Montgomery et al., 2001; Gujarati, 2003; Kutner et al., 2004; Chatterjee and Hadi, 2006; Greene, 2008; Lima et al., 2009; Habshah et al., 2013).

However, when the homoscedasticity assumption is violated, the OLS estimate is still unbiased, but becomes inefficient due to the inconsistency of the variance-covariance matrix of the estimate. As a consequence, the inference will become unreliable.

Heteroscedasticity occurs in the cross sectional data as well as panel data which are almost used in every field of study such as economics, finance, history, business, law, education, meteorology, medicine, biology, chemistry, engineering, physics, sociology, and psychology.

Panel data regression model is one of the most widely used models especially in economics and finance because of it advantage over cross-sectional and time-series model. Panel data referred to as data collected for many individuals over time, it has two dimensions cross sectional and times series (Baramati, 2007). It can be analyzed by running a regression over these two-dimensions using a classical least squares. The fixed effect (FE) and random effect (RE) models are the commonly used methods of analyzing panel data regression. The major difference between these two models is the definition of the unobserved time invariant variable in the data set. However, the

© COPYRIGHT UPM

problem of heteroscedasticity and influential observations (outliers in x or y direction) affects both the cross sectional and panel data estimators.

1.2 Outliers in Linear Regression

Outliers are those observations which are markedly far away from the majority of the data. Barnett and Lewis (1994) defined outliers as a set of data (or subset of observations) which appears to be inconsistent with the remainder of that set of data. There are different sources of outliers. It may be the natural feature of a population that is uncontrollable. It can result from typing error, measurement error, unusual values, transmission or copying error (Leroy and Rousseeuw, 1987). Imperfect collection of data is another source of outliers.

The existence of atypical observations which often referred to as outliers is inevitable in real data sets (Hampel et al., 1986). Rousseeuw and Van Zomeren (1990) classified outliers into high leverage points (HLPs) and vertical outliers (VOs). Presence of anomalous observations especially HLPs in a data set invalidate classical statistical inference (Hampel et al., 1986). The OLS is inefficient and produce unreliable estimates even when a single outlying observation is added or present in a data set. Hampel et al. (1986) claimed that a routine data set typically contains about 1–10% outliers and even the highest quality data set cannot be guaranteed to be free of outliers

In the case of linear regression, observations are judged as outliers on the basis of how the fitted regression equation accommodates them. Observations corresponding to excessively large residuals are treated as outliers. In OLS method, the residual mean square is generally used to estimate the variance of the errors. The residual mean sum of squares can be greatly inflated by outliers so that we may not be able to reliably estimate the variance of the errors and consequently the entire inferential procedure may be in fault. In the literatures there are several types of outliers for a regression problem. In regression problem if outlier occurs in Y direction it is called vertical outlier or residual outlier. Moreover, outlier may occur in X direction, usually called high leverage points (HLPs). The HLPs are classified into good and bad based on their effect to the model fit. Only bad leverage points influence the model fit, but not the good leverage points. The good leverage points have no effect or very little effect on the parameter estimates and may contribute to the precision of the estimates.

1.2.1 Basic Properties of Robust Estimators

Robust estimators target to provide useful information even if some of the parametric assumptions are violated. In linear regression analysis, the robust regression methods are used to produce resistance estimates, which lead to the stability in the results in the presence of unusual observations in a data set (for more details, one can refer to Huber, 1964; Hampel, 1974; Andrews, 1974; Ramsay, 1977; Simpson, 1995; Rousseeuw and Leroy,1987; Welcox, 2005; Marrona, 2006).

© COPYRIGHT UPM

The objective of a robust estimator is to provide estimates based on the information contains by the majority of the data set. Moreover, robust regression aim to fit a model based on the information in the most of the data. The most fundamental or basic properties used to measure the performance of robust estimator are; efficiency, breakdown point and bounded influence. These three properties are introduced briefly as follows.

1.2.1.1 Efficiency

Efficiency of an estimator is the measure of the degree of a robust method performance relative to least squares method under it basic assumptions. Moreover, it can similarly be expressed as a percentage of the ratio between the variance of the least squares fits for a clean data (without outliers) and the variance of the robust fit (Maronna et al. 2006). An efficient estimator is also the minimum variance unbiased estimator (MVUE) in which it attains the minimum variance for all parameter estimates. An estimator should also be precise, as measured by its statistical efficiency. The statistical efficiency of an estimator relies on the postulated distribution. For instance, the sample mean has the perfect efficiency of 100 percent at the normal distribution, but at other distributions its efficiency may become very low. According to Simpson (1995) efficiency near 90 to95 percent relative to OLS with normal random error is desirable.

1.2.1.2 Breakdown Point

The breakdown point (BP) is usually expressed as a percentage measure of resistivity of an estimator for a given amount of contamination (Hampel, 1974; Wilcox, 2005; Maronna et al., 2006). It is the smallest fraction of bad observation that can change an estimator dramatically by an arbitrary large value. In general, a large BP means that the estimator has an ability to withstand a large percentage of outliers without distraction the analysis. The breakdown point of an estimator T_x given the data matrixX_y, is,

$$\mathrm{BP}\left(T_x/X_y\right) = \min\left\{\frac{m}{n} : \underset{X_y^*}{\mathrm{SUP}}\left\|T_x\left(X_y\right) - T_x(X_y^*)\right\| = \infty\right\}$$

where the supremum is over all possible data matrix X_y^* include of $n - m$ observation and m contaminated points (Donoho and Huber, 1983; Leroy and Rousseeuw, 1987; Maronna et al. 2006). The least squares estimator has breakdown point as low as $1/n$ (which sometimes referred to as zero percent) meaning that even a single outlying observation can make an estimator of OLS to be meaningless. However, there are some robust regression estimators that have high BP of approximately 50% (meaning that up to half of the data can be contaminated and the estimator can still be useful) such as least median of square, least trimmed square, S and MM-estimator. According

© COPYRIGHT UPM

to Rousseeuw and Croux (1993) the highest possible BP is 50%, because the estimate keeps bounded when fewer than 50% of the data are replaced by outlying observation.

1.2.1.3 Bounded Influence Function

The bounded influence function (BIF) is another essential property of a robust estimator. The BIF referred to the ability of an estimator to control the amount of impact that outlying points in the X direction (i.e., high leverage points) have on model estimation (Simpson, 1995). Least squares are the most susceptible to high leverage points, but some robust methods also have unbounded influence. A study of the influence function determines whether or not an estimator has bounded influence. The influence function (IF) measures the robustness with respect to small amounts of contamination. Rousseeuw and Leroy (1987) described the IF of an estimator T_x at a distribution F in those points x_0 of the sample space where the limit exists as

$$\mathrm{IF}(x_0; T_x, F) = \lim_{\varepsilon \to \infty} \frac{T_x((1-\varepsilon)F + \varepsilon\varphi_{x_0}) - T_x(F)}{\varepsilon}.$$

where φ_{x_0} is the probability mas function of x_0. The influence function explain the bias caused by adding a few outliers at the point x_0, standardized by the amount ε of contamination (for more details, one can refer to Leroy and Rousseeuw, 1987; Simpson, 1995; Wilcox, 2005; Maronna, 2006; Andersen, 2008).

1.3 Importance and Motivation of the Study

The existence of unusual observations (high leverage point, outliers and influential observation) is very common in regression. Unfortunately, these anomalous observations are responsible for misleading conclusions about the fitting of a regression model. The diagnostic measure in regression dealing with high leverage point (outlying observations in x-direction) denoted by HLPs and outliers (outlying observations in y-direction) have very close ties with influential observations (IOs). Generally, any observation that individually or together with several other observations causes a large impact on the calculated values of various estimates (standard error, coefficients, t-values, p-values, etc) is referred to as IOs (Belsley et al., 1980). Andrews and Pregibon (1981) showed that outliers may have an influence on the parameter estimates. Chatterjee and Hadi (1986) pointed out that HLPs and outliers need not always be influential, and IOs are not necessarily be high leverage points. Since the IOs give a very bad effect on the parameter estimates, it is very imperative to identify them and their effect should be minimized.

© COPYRIGHT UPM

Imon (2005) proposed a generalized version of DFFITS based on group deletion technique denoted by GDFFITS to detect IOs but the method is not successful to correctly identify multiple IOs. Pena (2005) introduced a new idea to measure the influence of an observation based on how this observation is being influenced by the rest of the data denoted by S_i. The shortcoming of Pena's method is that it is totally different from the way of measuring the influence of observations. To quote him, ''instead of looking at how the deletion of a point or the introduction of same perturbation affects the parameters, the forecasts, or the likelihood function, we look at how each point is influenced by the others in the sample. That is, for each sample point we measure the forecasted change when each other point in the sample is deleted''. Imon et al. (2011) extend the idea of Pena to group deletion for identifying multiple IOs termed generalized version of $S_i$which is denoted M_i.

Recently, Nurunnabi et al. (2016) proposed new identification measure for IOs termed influential distance (ID) based on group detection technique for identifying multiple IOs. The technique has three major stages. The first stage identifies the suspected unusual observations to be deleted using a method termed Group Union Method (GUM), the second stage identifies HLPs and VOs, and the third stage computes the ID. This method is very good for the identification of IOs. However, the shortcoming of this method is that in the first stage it employed the union of five different detection methods (standardized studentized residual, standardized LMS residuals, leverage values or hat matrix, Cooks distance and difference in fits) for the identification of the suspected unusual observation that will form the deletion group. Some of these detection methods have been reported to have high rate of masking and swamping (for more details refer to Habshah et al., 2009). According to Hadi (1992) the choice of the initial suspected unusual observations is very important as it may lead to correct detection of the final IOs. Moreover, the computation of all these diagnostic methods takes a lot of computer times. Additionally, the ID method only identified IOs but fail to differentiate between the good leverage points and IOs. Hence, ID incorrectly detects IOs. The good leverage observations have little or no effect on the parameter estimates (Habshah and Mohammed, 2015). This has motivated us to develop another version of ID named Fast Improvised Influential Distance denoted by FIID which is relatively simple and fast to compute and also separate IOs and good leverage observations. FIID does not consider the good leverage observations as IOs.

The Heteroscedasticity Consistent Covariance Matrix (HCCM) estimator denoted by HC0 was proposed by White (1980) to remedy the problem of heteroscedasticity of unknown form in linear regression. It has been reported that the HCCM is biased in finite samples (MacKinnon and White 1985; Cribari-Neto and Zarkos 1999; Long and Ervin 2000). Later, MacKinnon and White (1985) proposed another HCCM estimator termed HC1 and HC2 to improve the efficiency of HC0. Davidson and MacKinnon (1993) slightly modified HC2 and named it HC3 which is closely approximated to jackknife estimator. Cribari-Neto (2004) proposed another HCCM estimator and named it HC4 where he adjusted the residuals by a leverage factor. Cribari-Neto et al. (2007) then proposed HC5 whereby they modified the exponent used in HC4 in order to consider the effect of maximal leverage.

© COPYRIGHT UPM

However, the HCCM estimators are biased in the presence of IOs as they are based on OLS residuals. Furno (1996) proposed another HCCM where he employed residual of WLS instead of OLS, but the weight used is hat matrix. The hat matrix is reported to be inefficient as it suffers from masking and swamping effect (Habshah et al., 2009). Similarly, least median of squares (LMS) and least trimmed squares (LTS) residuals were considered by Lima et al. (2009), but the shortcoming of their methods is that it truncated some observations which may contribute to the precision of the estimate.

Nonetheless, it is now evident that hat matrix is not successful in detecting IOs (Habshah et al. 2009). Also, the use of residual from LTS and LMS in the construction of HCCM performed poorly. Their work has inspired us to formulate a new robust estimation method based on our fast improvised influential distance (FIID) weighting method. The FIID classifies the observations into regular observation, good leverage and influential observations. The influential observations were down weighted as they are responsible for the deviation of the model fit. However, the good leverage observations were allowed to take part in the estimation as they will increase the precision of the estimate.

Bagheri and Habshah (2015) highlighted that many statistics practitioners are not aware that those HLPs that changed the multicollinearity pattern of a data is referred to as High leverage collinearity influential observations (HLCIO). The presence of high leverage points may also change the heteroscedasticity pattern of a data and mislead the conclusion of statistical analysis. It has been reported that HLPs is another source of heteroscedasticity (Rana et al., 2008; Alih and Choon, 2015). Habsshah et al. (2009) as well as Imon and Khan (2003) stated that the presence of HLPs in a data set makes the residual variances become heteroscedastic. To the best of our knowledge no research has been done to identify the type of outliers that are responsible for causing/affecting heteroscedasticity in a data set. Therefore, to fill the gap in the literature, we proposed a diagnostic method to identify the type of outliers that causes or affect heteroscedasticity problem. It is very important to identify these points as they are responsible for causing inconsistency and bias in OLS estimation. Moreover, identifying these observations is very important, to enable us to identify the actual source of the heteroscedasticity problem in order to provide appropriate remedial measure.

This thesis also addressed panel data estimation method in the presence of heteroscedasticity and HLPs. As already mentioned, the commonly used estimation strategy in panel data is either fixed effect (FE) or random effect (RE) estimation. The classical estimation method employed the OLS to the demeaned transformed data or partially demeaned transformed data for FE and RE models, respectively. The demeaned transformation commonly known as mean-centering is the transformation of panel data within each time series by mean. The OLS method is known to be very sensitive to outliers particularly HLPs, even one HLPs is enough to breakdown the estimate of the OLS (Leroy and Rousseeuw, 1987). There are few researches on robust estimation technique for panel data, such as (Bramati and Croux, 2007; Baltagi, 2008;

© COPYRIGHT UPM

Verardi and Wagner, 2011). Nevertheless, their techniques do not take into consideration the combined problem of HLPs and heteroscedasticity in panel data set. Mazlina and Habshah (2015) proposed a Within Group estimator based on robust MM and robust GM6 using robust centering method in which the data is centered by MM-estimate of location (MM centering). This robust centering approach reduces the effect of HLPs and also increases the efficiency of the estimate. The weakness of their method is that it down weights all HLPs irrespective of whether they are Good HLPs or Bad HLP.

Recently, Visek (2015) used the least weighted squares (LWS) to estimate the parameters of fixed and random effects models in panel data. He used classical centering method (mean centering) to transform the data and apply LWS, where the weight used was defined by the residual order statistic. The shortcoming of this method is that it employed mean centering which have been reported to perform poorly in the presence of HLPs (Bramati and Croux, 2007; Mazlina and Habshah, 2015). Also, it is inefficient and provide inconsistent covariance matrix in the presence of heteroscedasticity of unknown structure. This motivated us to develop a robust estimation method for both FE and RE panel data regression models in the presence of heteroscedasticity and HLPs based on FIID weighting method for robust HCCM estimator. The good leverage observation were not down weighted, instead they were allowed to take part in the estimation. Moreover, the MM centering approach was employed in our new estimation technique instead of mean centering used by Visek (2015) and median centering used by Bramati and Croux (2007).

In this thesis, the Hausman pretest for panel data models is also addressed. The classical Hausman pretest is the commonly used method in order to determine whether random or fixed effect panel data models should be used. However, in the presence of heteroscedastic error variances and HLPs or IOs in the data set, the classical Hausman test provides incorrect and misleading result. Nevertheless, the remedy of the combined problem of Heteroscedasticity and HLPs or IOs on Hausman test is still missing in the literature. This inspired us to propose a robust Hausman pretest based on our newly developed robust FE and robust RE estimation method.

1.4 Research Objectives

The aim of this thesis is to investigate the problems of heteroscedasticity of unknown form for linear regression and panel data regression models in the presence of HLPs. The classical estimation methods for a heteroscedastic model mostly are based on OLS estimates. Whereas, the OLS estimate are known to be very sensitive to HLPs. Moreover, there is strong evidence that the presence of HLPs causes heteroscedasticity in a data set. Therefore, it is important to detect these HLPs and provide a new estimation technique for a heteroscedastic model when there exist HLPs in the data set. The objectives of our research were systematically outlined as follows:

© COPYRIGHT UPM

1. To develop a new fast method for detecting influential observations in multiple linear regression.
2. To formulate a new weighting method for robust HCCM estimator based on Fast Improvised Influential Distance (FIID) and $DRGP_{ISE}$ in the presence of heteroscedasticity and IOs in multiple linear regression.
3. To formulate a new robust diagnostic methods and remedial measures for the heteroscedasticity influential observations (HIO) in multiple linear regression.
4. To establish a new robust estimation method for fixed effect (FE) and random effect (RE) panel data regression models based on FIID in the presence of heteroscedasticity and IOs.
5. To develop a robust Hausman specification test based on proposed robust estimation method for FE and RE models in the presence of heteroscedasticity and IOs in panel data regression model.

1.5 Scope and Limitation of the Study

High leverage points as a source of heteroscedasticity is relatively a new area in robust statistics. To the best of our knowledge no research has been done to identify these points. As such, no much referred real datasets in the literature.

The robust estimation technique for heteroscedastic panel data needs to be addressed, due to the advantages it possessed. That is, having two dimensionalities (cross-section and time series). Panel data can be applied in many areas of research especially economics and finance. The robust techniques are still new in panel data estimation especially the random effect model. Therefore, not much algorithms exist in literature concerning robust estimation in panel data. The most critical part is the development of the programming codes since most of the statistical software's does not have the robust function for panel data.

Due to space constraint, throughout the thesis we only report the results for p=3. However, the result for p=5, 10 are also consistent.

1.6 Outline of the Thesis

In accordance with the research objectives and the scope of the study, the contents of this thesis are organized into eight chapters. The thesis chapters are structured in such a way that the objectives are apparent and arrange in the sequence outlined.

Chapter Two: This chapter presents a brief review of the ordinary least squares estimation of regression parameters, violations from its assumptions and the basic concepts of robust regression. The diagnostic methods of high leverage points and vertical outliers were reviewed. Moreover, some existing robust regression methods for parameter estimation in the presence of HLPs and vertical outliers are also

© COPYRIGHT UPM

presented. The literature reviews on heteroscedasticity with example and its consequences and, heteroscedasticity with its usual detection and estimation techniques. Some literatures in fixed and random effect panel data estimation are presented. Finally, brief reviews of Hausman pretest for panel data are also included in this chapter

Chapter Three: This chapter briefly discussed the influential distance (ID) method for identifying multiple influential observations. The new method for the identification of multiple influential observations termed Fast Improvised Influential Distance (FIID) is presented. Several well-referred real data set and Monte Carlo Simulation study to evaluate the performance of the proposed method are presented.

Chapter Four: This chapter deals with the new proposed weighting method and estimation technique for multiple linear regression model in the presence of heteroscedasticity and high leverage points. The classical and robust Heteroscedasticity Consistent Covariance Matrix (HCCM) Estimators are presented. The new proposed robust HCCM estimator is described. The new proposed estimation technique involves classifying the observations into regular observations, good leverage points and influential observations. But, only influential observations will be down weighted. The WLS based on FIID is used to estimates the parameters. The numerical examples and simulation study are presented.

Chapter Five: This chapter is divided in to two sections;

First section: Investigate the type of outliers that are responsible for causing/affecting heteroscedasticity in a data set. Heteroscedasticity-Influential Observations (HIO) diagnostic for both homoscedastic and heteroscedastic data set is introduced. The White test and proposed robust White test are discussed. The TSRWLS and proposed MTSRWLS with real data examples and simulation study are presented.

Second section: This section provides appropriate remedial measure for Heteroscedasticity- Influential Observations (HIO). The remedial measure is based on a new version of GM6 estimator denoted as GM-FIID is presented. Lastly, simulation study and real data examples to evaluate the performance of the proposed method is presented.

Chapter Six: This chapter is divided into two sections;

First section: In this section, the fixed effect (FE) panel data estimation is briefly discussed. The demeaned centering method has been presented. The new proposed estimation technique for FE panel data model in the presence of heteroscedasticity and HLPs is introduced. Finally, the simulation study and real data examples are presented.

© COPYRIGHT UPM

Second section: deals with the random effect (RE) panel data regression model. The estimation for RE model is introduced. The partially demeaned centering is presented. Also, the new proposed estimation method for RE model in the presence of heteroscedasticity and HLPs is discussed. And lastly, the simulation study and real data examples are presented.

Chapter Seven: In this chapter, the conventional Hausman pretest for panel data model is presented. The proposed robust Hausman pretest for panel data model is introduced. Also, distribution of the proposed robust Hausman pretest is discussed. The simulation and real data examples to evaluate the proposed method are presented.

Chapter Eight: This chapter presents the summary and general conclusion of this thesis. Also, some recommendations for areas of further research has been presented

© COPYRIGHT UPM

REFERENCES

Alguraibawi, M., Midi, H., and Imon, A. H. M. (2015). A new robust diagnostic plot for classifying good and bad high leverage points in a multiple linear regression model.*Mathematical Problems in Engineering*, 1:1-12

Alih, E., and Choon, H. O. (2015). An outlier-resistant test for heteroscedasticity in linear models. *Journal of Applied Statistics*, 42(8):1617-1634.

Ameniya T. (1971). The Estimation of the Variance in a Variance-Components Model. *International Economic Review*, 12:1-13

Andersen, R. (2008). Modern methods for robust regression. *The United States of America:* Sara Miller McCune. SAGE publications.

Andrews D. and Pregibon D. (1978) Findings the outliers that matter, *J. R. Statist. Soc: Ser. B*. 40: 85–93.

Atkinson A.C. (1985) *Plots, Transformations and Regression*, Clarendon press, Oxford.

Atkinson A.C. (1986) Masking unmasked, *Biometrika*. 73: 533–541.

Atkinson A.C. and M. Riani (2000) *Robust Diagnostic Regression Analysis*, Springer, New York.

Atkinson, D. (1985). *Effects of some plant growth regulators on water use and the uptake of mineral nutrients by tree crops*. International Symposium on Growth Regulators in Fruit Production 179.

Bagheri, A. (2011). "*Robust Estimation Methods And Robust Multicollinearity Diagnostics For Multiple Regression Model in the Presence of High Leverage Collinearity-Influential Observations.*" Unpublished doctorial thesis. University Putra Malaysia .

Bagheri, A. and Habshah M. (2015). "Diagnostic plot for the identification of high leverage collinearity-influential observations." *SORT* 39(1): 51-70.

Balestra P. and Nerlove M. (1966) Pooling cross-section and time series data in the estimation of a dynamic model: the demand for natural gas. *Econometrica* 34: 585-612

Baltagi B.H. (2005). *TheEconometrics of Panel Data*. John Wiley & Sons, New York.

Baltagi B.H., Jung B.C. and Song S.H. (2009) *Testing for Heteroscedasticity and Serial Correlation in a Random Effect Panel Data Model*, Center for Policy Research. Syracuse University, Syracuse, New York.

© COPYRIGHT UPM

Baltagi, B. (2008). *Econometric analysis of panel data*, John Wiley & Sons.

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

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

Box, G.E.P. and Cox, D.R. (1964). An analysis of transformations. *Journal of theRoyal Statistical Society*. 26: 211– 252

Bramati, M. C. and C. Croux (2007). Robust estimators for the fixed effects panel data model.*The econometrics journal* 10(3): 521-540.

Breusch, T. and Pagan, A. (1979). A simple test for heteroscedasticity and random coefficient variation, *Econometrica*, 47: 1287–1294.

Brownlee K. (1965) *Statistical Theory and Methodology in Science and Engineering*, Wiley, New York

Brown, M.B. and Forsythe, A. B. (1974). Robust tests for equality of variances. *Journal of the American Statistical Association*. 69: 364–67.

Cai, Z., Hurvich, C.M. and Tsai, C.L. (1998). Score tests for heteroscedasticity inwavelet regression, Biometrika. 85: 229–234

Campbell, N. A., Lopuhaä, H. P. and Rousseeuw, P. J. (1998) . On the calculation of a robust S-estimator of a covariance matrix. *Statistics in Medicine*. 17(23): 2685-2695.

Carroll, R.J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linearmodels.*Annals of Statistics*. 10: 429–441.

Carroll, R.J. and Ruppert, D. (1988), *Transformation and Weighting in Regression*,New York:Chapman and Hall.

Carroll, R.J. and Ruppert, D. (1984). Power Transformations when fitting theoretical models to data. *Journal of the American Statistical Association*. 79: 321–328.

Chatterjee S. and A.S. Hadi (1986) Influential observations, high leverage points, and outliers in regression, *Statist. Sci*.1: 379–393.

Chatterjee, S. and A. S. Hadi (2006). *Regression Analysis by Example*, Fourth Edition.

Chatterji, S. and A. Hadi (1988). *Sensitivity Analysis in Linear Regression*, Wiley, New York.

© COPYRIGHT UPM

Chesher, A. and Jewitt, I. (1987). The bias of a heteroskedasticity consistentcovariance matrix estimator.*Econometrica*. 55: 1217– 1222.

Choulakian, V., Lockhart, R. A., and Stephens, M. A. (1994). Cramér- von Mises statistics for discrete distributions. *Canadian Journal of Statistics,* 22(1), 125-137.

Coakley, C. W. and T. P. Hettmansperger (1993). A bounded influence, high breakdown, efficient regression estimator.*Journal of the American Statistical Association* 88(423): 872-880.

Cook R.D. (1977) Detection of influential observations in linear regression, *Technometrics.* 19: 15–18.

Cook R.D. (1998) *Regression Graphic: Ideas for Studying Regression Through Graphi*cs, Wiley, New York.

Cook R.D. and S. Weisberg (1982) *Residuals and Influence in Regression*, Chapman and Hall, London.

Cook, R.D. and Weisberg S. (1983). Diagnostics for heteroscedasticity inregression. *Biometrika*.70: 1–10.

Cribari-Neto, F. and Zarkos, S. (1999). Bootstrap methods for heteroskedastic regression models: evidence on estimation and testing. *Econometric Reviews*, 18:211-228.

Cribari-Neto, F. (2004). Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis* 45: 215-233.

Cribari-Neto, F., Souza, T.C. and Vasconcellos, K.L.P. (2007). Inference under heteroskedasticity and leveraged data. *Communications in Statistics-Theory and Methods* 36: 1877-1888

Croux, C. and Haesbroeck G. (2003). Implementing the Bianco and Yohai estimator for logistic regression. *Computational Statistics and Data Analysis* 44(1): 273:295

Crowder, D., Hand, D. (1990) *Analysis of repeated measures*. Chapman and Hall, London,

Daniel, C. and Wood, F.C. (1980). *Fitting Equations to Data*, 2nd Edition. NewYork: Wiley.

Davidson, R. and MacKinnon, J.G. (1993). *Estimation and Inference in Econometrics*. New York: Oxford University Press.

© COPYRIGHT UPM

Dhhan, W., et al. (2015). Non-sparse ϵ-insensitive support vector regression for outlier detection.*Journal of Applied statistics* 42(8): 1723-1739.

Diblasi, A. and Bowman, A.W. (1997). Testing for constancy of variance.Statistics and Probability,*Letters*. 33: 95–103.

Donoho, D. L. and P. J. Huber (1983). The notion of breakdown point. *A festschrift for erich l. lehmann* 157184.

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

Edgeworth, F. Y. (1887). On observations relating to several quantities. *Hermathena.* 6:279-285

Efron, B. (1982). The Jackknife, the Bootstrap and Other Resampling Plans.Philadelphia. PA: *Society for Industrial and Applied Mathematics.*

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

Epps, T. W., and Epps, M. L. (1977). The robustness of some standard tests for autocorrelation and heteroskedasticity when both problems are present. Econometrica: *Journal of the Econometric Society*, 45: 745-753.

Furno, M. (1996). Small sample behavior of a robust heteroskedasticity consistent covariance matrix estimator. *Journal of Statistical Computation and Simulation* 54: 115-128.

Furno, M. (1997). A robust heteroskedasticity consistent covariance matrix estimator. A *Journal of Theoretical and Applied Statistics* 30: 201-219.

Glaser, R.E. (1976). Exact critical values for Bartlett's test for homogeneity ofvariances. *Journal of American Statistical Association*. 71: 488–490.

Glejser, H. (1969). A new test for heteroskedasticity. *Journal of AmericanStatistical Association.* 64: 316–323.

Goldfeld, S.M. and Quandt, R.E. (1965). Some tests for homoskedasticity.J*ournal of the American Statistical Association.* 60: 539–547

Gray, J. B. (1985). Graphics for regression diagnostics. *Proceedings of Statistical Computing Section*, pp. 102-107.

Greene, W. (2008) *Econometric Analysis*; New York: Pearson

Greene, W.H. (1993) *Econometric Analysis*, Prentice–Hall, ISBN 0-13-013297-7.

© COPYRIGHT UPM

Grieser, W. D., & Hadlock, C. J. (2015). Panel data estimation in finance: paran consistency and the standard error sideshow. Working Paper.

Grunfeld Y. (1958) *The Determinants of Corporate Investment.* Unpublished Ph.D. dissertation, Department of Economics, University of Chicago

Gujarati, D. and D. Porter (2002). *Basic econometrics,* McGraw-Hill/Irwin.

Habshah M., Norazan MR and Rahmatullah Imon AHM (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.

Habshah M. and Bashar A. T. (2008). The performance of robust estimator on linear regression model having both continuous and categorical variables with heteroscedastic error. *Malaysian journal of mathematical science*, 2(1): 25-48

Habshah Midi, Md. Sohel Rana, A. H. M. Rahmatullah Imon (2009). The performance of robust weighted least squares in the presence of outliers and heteroscedastic errors. *Wseas transactions on mathematics*, 7(8): 351-361

Habshah Midi, Sohel Rana, and A.H.M.R. Imon (2013). On a Robust Estimator in Heteroscedastic Regression Model in the Presence of Outliers, *Proceedings of the World Congress on Engineering*. Vol I, WCE 2013, July 3 - 5, London, U.K

Habshah, M. (2000). Heteroscedastic nonlinear regression by using Tanh PhiFunction. *Sains Malaysiana*. 29: 103–118.

Habshah M., Rana, S., and Imon, A. H. M. (2014). Two-step robust estimator in heteroscedastic regression model in the presence of outliers. *Economic computation & economic cybernetics studies & research*, 48(3).

Habshah M. and A. Mohammed (2015). The Identification of Good and Bad High Leverage Points in Multiple Linear Regression Model.*Mathematical Methods and System in Science and Engineering.*

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

Hadi A.S. and J. Simonoff (1993) Procedure for the identification of outliers in linear models, *Journal ofthe American Statistical Association.* 88: 1264–1272.

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., et al. (1986). Linear models. robust estimation. Robust Statistics: The Approach Based on Influence Functions. 307-341.

© COPYRIGHT UPM

Harvey, A.C. (1976). Estimating regression models with multiplicativeheteroscedasticity. *Econometrica*. 44: 461–465.

Hausman J.A. (1978) Specification tests in econometrics, *Econometrica* 46, 1251–1271

Hausman, J. and Palmer, C. (2012) Heteroscedasticity-robust inference in finite samples. *Economics Letters* 116 (2012) 232–235

Hawkins, D.M., Bradu, D. and Kass, G.V.(1984). Location of several outliers in multiple regression data using elemental sets. *Technometrics.* 26:197-208

Hill, R. W. (1977). *Robust regression when there are outliers in the carriers*, Harvard University.

Hinkley, D.V. (1977). Jackknifing in unbalanced situations. *Technometrics*.19:285–292.

Horn, S.D., Horn, R.A. and Duncan, D.B. (1975). Estimating heteroscedasticvariances in linear model. *Journal of American Statistics Association*. 70:380–385.

Hossain, M. G., Zyroul, R., Pereira, B. P., and Kamarul, T. (2012). Multiple regression analysis of factors influencing dominant hand grip strength in an adult Malaysian population. *Journal of Hand Surgery (European Volume)*, 37(1), 65-70.

Hsiao C. (2003) *Analysis of Panel Data*, 2nd edition. Cambridge University Press.

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

Huber, P. and E. Ronchetti (1981). *Robust Statistics,* ser. Wiley Series in Probability and Mathematical Statistics. New York, NY, USA, Wiley-IEEE 52: 54.

Huber, P. J. (2005). *Robust statistics*, John Wiley & Sons.

Imon, A. (2002). Identifying multiple high leverage points in linear regression.*Journal of Statistical Studies* 3: 207-218.

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

Imon, A.H.M.R. (2009). Deletion residuals in the detection of heterogeneity ofvariances in linear regression. *Journal of Applied Statistics*, 36: 347–358.

Imon A.H.M.R (2002) Identifying multiple high leverage points in linear regression, *Journal of Statistical Studies.* 32: 73–90.

© COPYRIGHT UPM

ImonA.H.M.R. and Khan M.A.I. (2003) A solution to the problem of multicollinearity caused by the presence of multiple high leverage points, *Int. J. Stat. Sci*. 2:37–50

Imon, A.H.M.R., A.A.M. Nurunnabi, and M. Nasser (2011) A diagnostic measure for influential observations in linear regression, *Commun. Stat. Theory Methods.* 40(7): 1169–1183.

Judge, G.G., Griffiths W.E., Hill, R.C., Lutkepohl H. and Lee T.C. (1985). *The Theory and Practice of Econometrics*. Wiley, New York.

Kabaila P., Mainzer R. and Farchione D (2015) The impact of Hausman pretest, applied to panel data, on the coverage propoability of confidence intervals. *Econometrics Letters* 131, 12-15

Kauermann, G. and Carroll, R.J. (2001). A note on the efficiency of sandwichcovariance matrix estimation. *Journal of the American Statistical Association.*96: 1387–1396.

Kennedy, P. E. (1985). A suggestion for measuring heteroskedasticity. *Communications in Statistics-Simulation and Computation*, 14(4), 845-851.

Kleiber C. and Zeileis A (2008) *Applied Econometrics with R*. New York

Kleijnen, J. P. C., Cremers, P., and Van Belle, F. (1985). The power of weighted and ordinary least squares with estimated unequal variances in experimental design. *Communications in Statistics-Simulation and Computation*, 14(1), 85-102.

Koenker, R.W and Bassett, G. (1982). Robust tests for heteroscedasticity based on regression quintile. *Econometrica*. 50: 43–61

Krasker, W. S. and R. E. Welsch (1982). Efficient bounded-influence regression estimation.*Journal of the American Statistical Association* 77(379): 595-604.

Kutner, M. H., et al. (2004). *Applied linear regression models*, McGraw-Hill/Irwin.

Leroy, A. M. and P. J. Rousseeuw (1987). *Robust regression and outlier detection*. Wiley Series in Probability and Mathematical Statistics, New York: Wiley, 1987.

Lim, H. A. and H. Midi (2016). Diagnostic Robust Generalized Potential Based on Index Set Equality (DRGP (ISE)) for the identification of high leverage points in linear model.*Computational Statistics* 3(31): 859-877.

© COPYRIGHT UPM

Lima, V.M.C., Souza, T,C., Cribari-Neto, F. and Fernandes, G.B. (2009). Heteroskedasticity- robust inference in linear regressions. *Communications in Statistics-Simulation and Computation* 39: 194-206

Long, J.S. and Ervin, L.H. (2000). Using heteroscedasticity consistent standard errors in the linear regression model. *The American Statistician* 54: 217-224.

MacKinnon, J.G. and White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics* 29: 305-325.

Mallows, C. L. (1975). *On some topics in robustness*. Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ.

Manuel F. and Colin V. (2010) Fixed, Random, or something in between? A variant of Hausman's specification test for panel data estimators. *Econometric Letters* 107, 327-329

Maronna, R. A., et al. (2006). "Wiley Series in Probability and Statistics." Robust Statistics: Theory and Methods: 404-414.

Martin M.A., S. Roberts, and L. Zheng (2010) Delete-2 and delete-3 Jackknife procedures for unmasking regression, *Aust.N.Z. J. Stat*. 52: 45–60.

Mazlina M. A. and Habshah M. (2015). Robust centering in the fixed effect panel data model.*Pakistan Journal of Statistics* 31(1).

Menjoge R. and R. Welsch (2010) A diagnostic method for simultaneous feature selection and outlier identification in linear regression, *Comput. Stat. Data Anal*. 54: 3212–3226.

Montgomery, D. C., Peck, E. A. and Viving, G.G. (2001). *Introduction to linear regression Analysis.* 3rd edition. New York: John Wiley and sons.

Mosteller, F. and Tukey, J. W. (1977). Data Analysis And regression: A Second Course in Statistics. Reading, MA: Addison-Wesley.

Müller, HG. and Zhao, P.L. (1995). On a semiparametric variance function modeland a test for heteroscedasticity. *Annals of Statistics*. 23: 946–967.

Mundlak Y. (1961) Empirical production function free of management bias, *Journal of Farm Economics*43, 44–56

Myers, R H. and Milton, J.S. (1991). *A First Course in the Theory of Linear StatisticalModels*. Boston: PWS-Kent Publishing Co.

© COPYRIGHT UPM

Nurunnabi A.A.M., A.S. Hadi, and A.H.M.R. Imon, (2014) Procedures for the identification of multiple influential observations, *Journal of Applied Statistics* 41(6): 1315–1331.

Nurunnabi A.A.M., M. Nasser, and A.H.M.R. Imon, (2016) Identification of multiple outliers, high leverage points and influential observations in linear regression. *Journal of Applied Satatistics*. 43(3): 509-525.

Park, R.E. (1966). Estimation with heteroscedastic error terms. *Econometrica*. 34:888

Patrik G (2010) The impact of a Hausman pretest on the size of a hypothesis test: The panel data casse. *Journal of Econometrics* 156, 337-343

Peña, D. and V. J. Yohai (1995). The detection of influential subsets in linear regression by using an influence matrix.*Journal of the Royal Statistical Society*. Series B (Methodological) 57: 145-156.

Peña, D. (2005) A new statistic for influence in linear regression, *Technometrics*. 47: 1–12

Pindyck, S.R and Rubinfeld, L.D. (1997). *Econometric Models and EconometricForecasts*, 4th Edition. New York: Irwin/McGraw-Hill.

Rahman, M., Pearson, L. M., and Heien, H. C. (2006). A modified anderson-darling test for uniformity. *Bulletin of the Malaysian Mathematical Sciences Society*, 29(1).

Ramsey, J.B., (1969). Tests for specification error in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society*. Series B, 31: 350–371.

Rasheed, B. A., Adnana, R., Saffarib, S. E., and Dano Patia, K. (2014). Robust weighted least squares estimation of regression parameter in the presence of outliers and heteroscedastic errors. *J. Technol.*, 71(1): 11-18

RanaM.S., H. Midi, and A.H.M.R. Imon, (2008). A Robust Modification of the Goldfeld-Quandt Test for the Detection of Heteroscedasticity in the Presence of Outliers, *Journal of mathematics and Statistics,* 4(4), 2008, pp. 277-283.

Rana, Md.S., Midi, H. and Imon A.H.M.R. (2012). Robust wild bootstrap for stabilizing the variance of parameter estimates in heteroscedastic regression models in the presence of outliers, *Mathematical Problem in Engeering*

Richard, A. J., and Dean, W. W. (2002). *Applied multivariate statistical analysis*. London: Prenticee Hall, 265.

Robinson, P.M. (1987). Asymptotically efficient estimation in the presence ofheteroscedasticity of unknown form. *Econometrica*. 55: 875–891.

© COPYRIGHT UPM

Rocke, D. M., & Woodruff, D. L. (1996). Identification of outliers in multivariate data. *Journal of the American Statistical Association*, 91(435),1047-1061

Rohayu MS (2013) A robust estimation method of location and scale with application in monitoring process variability. Ph.D. Thesis, Universiti Teknologi Malaysia, Malaysia (unpublished)

Rousseeuw, P. J. (1984). Least median of squares regression.*Journal of the American Statistical Association* 79(388): 871-880.

Rousseeuw, P. J. (1985). Multivariate estimation with high breakdown point. *Mathematical statistics and applications* 8: 283-297.

Rousseeuw, P. J. and C. Croux (1993). Alternatives to the median absolute deviation.*Journal of the American Statistical Association* 88(424): 1273-1283.

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

Rousseeuw, P. J. and B. C. Van Zomeren (1990). Unmasking multivariate outliers and leverage points.*Journal of American Statistical Association* 85(411): 633-639.

RousseeuwP. J. and Leroy, A. M. (1987). *Robust regression and outlier detection. Wiley Series in Probability and Mathematical Statistics*, New York: Wiley, 1987.

Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41(3):212–223

Rousseeuw P.J. and A. Leroy (2003) *Robust Regression and Outlier Detection*, Wiley, New York.

Rousseeuw P. J. and B. van Zomeren (1990) Unmasking multivariate outliers and leverage points, *J. Amer. Statist. Assoc.* 85: 633–639.

Ruppert, D. and D. G. Simpson (1990). Unmasking multivariate outliers and leverage points: comment.*Journal of the American Statistical Association* 85(411): 644-646.

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

Simpson, D. G., et al. (1992). On one-step GM estimates and stability of inferences in linear regression.*Journal of the American Statistical Association* 87(418): 439-450.

Simpson, J. R. (1995). New methods and comparative evaluations for robust and biased-robust regression estimation, Air Force Inst of Tech Wright-Patterson AFB OH.

© COPYRIGHT UPM

Simpson, J. R. and D. C. Montgomery (1998). The development and evaluation of alternative generalized m- estmation techniques.*Communications in Statistics-Simulation and Computation* 27(4): 999-1018.

Stock, J. H., and Watson, M. W. (2008). Heteroskedasticity- robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1), 155-174.

Srikantan, K.S. (1961). Testing for the single outlier in a regression mode *Sankyā* Series A, 23: 251–260

Stromberg, A. J., Hawkins, D. M., and Hössjer, O. (2000). The least trimmed differences regression estimator and alternatives.*Journal of the American Statistical Association* 95(451): 853-864.

Tukey, J.W. (1977). *Exploratory Data Analysis*. Cy: Addison-Wesley Publishers

Verardi, V. and J. Wagner (2011). Robust estimation of linear fixed effects panel data models with an application to the exporter productivity premium.*Jahrbücher für Nationalökonomie und Statistik* 231(4): 546-557.

Verbeek, Marno (2004) *A Guide to Modern Econometrics,* 2nded., Chichester: John Wiley & Sons

Víšek, J.A. (2011) Consistency of the least weighted squares under heteroscedasticity. *Kybernetika,* vol. 47, 179–206.

Víšek, J. Á. (2015). Estimating the Model with Fixed and Random Effects by a Robust Method.*Methodology and Computing in Applied Probability* 17(4): 999-1014.

Wallace, T.D. and Hussain A. (1969). The use of error components models in combining cross-section and time-series data, *Econometrica*37, 55–72

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

Welsch, R. E. (1980). *Regression sensitivity analysis and bounded-influence estimation*. In Evaluation of econometric models (pp. 153-167). Academic Press.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica* 48: 817-838

Wilcox, R. R. (2005). Introduction to robust estimation and hypothesis testing (Statistical Modeling and Decision Science), Academic press.

Wooldridge, J. M. (1995). Selection corrections for panel data models under conditional mean independence assumptions.*Journal of econometrics* 68(1): 115-132.

© COPYRIGHT UPM

Wooldridge, J.M., (2002) *Econometric Analysis of Cross Section and Panel Data.* MIT Press, Cambridge, London.

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

Yule, G.U. and Kendall, M.G. (1953). An Introduction to the Theory of Statistics.London: Charles Griffin & Company.

© COPYRIGHT UPM