UNIVERSITI PUTRA MALAYSIA

MODIFICATION OF TUKEY'S SMOOTHING TECHNIQUES FOR EXTREME DATA

QASIM NASIR HUSAIN

FS 2017 84

MODIFICATION OF TUKEY'S SMOOTHING TECHNIQUES FOR EXTREME DATA

By

QASIM NASIR HUSAIN

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

August 2017

COPYRIGHT

DEDICATIONS

Mum
Dad
Wife
Brothers
Sisters
Family

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

MODIFICATION OF TUKEY'S SMOOTHING TECHNIQUES FOR EXTREME DATA

By

QASIM NASIR HUSAIN

August 2017

Chair:Associated Professor Mohd Bakri Adam, PhD
Faculty: Science

Two Tukey's techniques which are resistant line for linear trend and resistant smoothing for non linear trend have been reviewed in this research. The new resistant line for method of dividing the batches using range dealing with ties and non ties is recommended. The determination of sample size in each batch is also being introduced.

In resistant smoothing, mathematical terms have been initiated and incorporated in this technique, the part which has been neglected by introducer of exploratory data analysis. New symmetric mean, right mean, left mean, right median, and left median have been proposed, leading to more simple process of smoothing technique. Later, the proposed methods have been used for the suggested compound smoothing techniques and hannings with simpler, faster and better smooth.

Additionally, in order to evaluate the efficiency of variant proposed techniques together with the smoothing index and extra balance test, simulation data with big data size have successfully been applied.

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

PENGUBAHSUAIAN TEKNIK PELICINAN TUKEY UNTUK DATA KEWANGAN YANG EKSTRIM

Oleh

QASIM NASIR HUSAIN

Ogos 2017

Pengerusi: Profesor Madya Mohd Bakri Adam, PhD
Fakulti: Sains

Dua teknik Tukey iaitu garis rintangan untuk haluan linear dan rintangan pelicinan untuk haluan bukan linear telah ditinjau di dalam kajian ini. Garis rintangan yang baru untuk kaedah pembahagian kelompok menggunakan julat bagi menangani masalah seri dan tidak seri dicadangkan. Manakala, penentuan bagi saiz sampel di dalam setiap kelompok turut diperkenalkan.

Bagi rintangan pelicinan, beberapa ungkapan matematik telah diterbitkan dan diselaraskan dalam teknik ini yang mana telah diabaikan oleh individu yang memperkenalkan analisis data penerokaan. Purata simetrik, purata kanan, purata kiri, median kanan dan median kiri yang baru telah dicadangkan di mana ia menjurus kepada teknik pelicinan yang lebih ringkas. Kemudian, kaedah yang dicadangkan telah digunakan di dalam teknik pelicinan kompaun dan Hanning yang menghasilkan nilai pelicinan yang lebih baik, ringkas dan cepat .

Tambahan lagi, untuk menilai tahap keberkesanan kaedah yang dicadangkan bersama-sama dengan kaedah indeks pelicinan dan ujian imbangan tambahan, simulasi data dengan sampel yang besar telah berjaya diaplikasikan.

ACKNOWLEDGEMENTS

All praise and thanks are to Almighty Allah, the most Gracious, the most Merciful. I would like to thank my supervisor Associate Professor Dr. Mohd Bakri bin Adam, as an excellent researcher to develop and discuss ideas with, who has been a great source of inspiration and also for his continuous guidance. Forever in debt to your priceless advice. Many thanks go to my co-supervisors; Associate Professor Dr. Mahendran Shitan and Dr. Anwar Fitrianto for their constant availability to motivate and give feedbacks on my thesis writing process.

I have also received a lot of input and support from my friends in the Group of Ph.D. students namely, Noor Izyan Mohammad Adnan and Nurul Nisa' Khairul Azmi. I thank them very much. My warm gratitude goes to all of INSPEM staff for the administrative matters, for providing facilities that encourage research and for the friendly services.

I also extend my thanks and gratitude to the brothers in the teaching staff of the Faculty of Management and Economics, University of Tikrit, especially those of Dr. Qassim Handhal, Mr. Khalaf Mohammad Hamad and Mr. Alaa Al Azawee for their support and confidence. All pride and appreciation for them.

I would like to mention the immortal support I received from my brothers in religion (Al Suhbah) who did not hesitate to meet my needs and the needs of my family as long as I was engaged in my studies. Last but never least; to my parents, my wife, my daughters and my family, there is nothing much to say except, I do love all of you. Thanks.

I certify that a Thesis Examination Committee has met on 22 August 2017 to conduct the final examination of Qasim Nasir Husain on his thesis entitled "Modification of Tukey's Smoothing Techniques for Extreme Data" 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:

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

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

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

Sutawanir Darwis, PhD
Professor
Bandung Islamic University
Indonesia
(External Examiner)

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

Date: 26 October 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 Doctor of Philosophy. The members of the Supervisory Committee were as follows:

Mohd Bakri Adam, PhD
Associate Professor
Faculty of Science
Universiti Putra Malaysia
(Chairperson)

Mahendran Shitan, PhD
Professor
Faculty of Science
Universiti Putra Malaysia
(Member)

Anwar Fitrianto, 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:

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: Qasim Nasir Husain, GS 37561

Declaration by Members of Supervisory Committee

This is to confirm that:

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

Signature: _____
Name of Chairman of Supervisory Committee:
Associate Professor Dr. Mohd Bakri Adam

Signature: _____
Name of Member of Supervisory Committee:
Professor Dr. Mahendran Shitan

Signature: _____
Name of Member of Supervisory Committee:
Dr. Anwar Fitrianto

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EX4253H	Existent Running Medians Technique
RM4253H	Proposed Running Means Technique
IMP4253H	Improved Running Medians Technique using right running medians
SRMED357H	Proposed Symmetric Running Medians Technique
SRM357H	Proposed Symmetric Running Means Technique
INC	Increase in the value of two following observations
DEC	Decrease in the value of two following observations
S.I	Smoothing Index
EB	Extra Balance Test
TEC	Smoothing Technique
res	Residual
GEV	General Extreme Value Distribution

CHAPTER 1

INTRODUCTION

1.1 Background of Exploratory Data Analysis

The classical analysis is generally quantitative in nature, either models or analysis such as ANOVA, t test, chi-squared test and F test. The future of data analysis can include great advance, the overcoming of actual difficulties, and the providing of a great service to many fields of science and technology (Tukey, 1962).

Exploratory Data Analysis (EDA) is generally either graphical or tabular in nature with few quantitative techniques, deals with analysis first then the pattern. EDA includes scatter plots, character plots, box plots, histogram, bi histogram, probability plots and residual plots. EDA is an approach or philosophy for data analysis that appoints a diversity of mostly graphical techniques. The advantages of EDA are maximizing insight into a data set, revealing underlying construction, extracting important variables, discovering outliers and oddity, testing implicit presumptions, evolving miserly models and finding out optimal factor settings.

EDA is more an art, or even a bag of tricks, than a science. If this is so, it might be difficult or impossible to find a reasonably comprehensive philosophy of EDA. In fact, the philosophy of EDA is simple, handy and the statistics still simple. Even if the data becomes larger, the computers and programming can help.

One quality of EDA is that the techniques that present data to the eye often carry more detail than numerical summaries. The reason is that the notable features can be ordinarily extracted from visual information more quickly than from numerical information (Good, 1983). Even if data have not been gathered to be appropriate for formal statistical analysis, EDA techniques can often detect information from them (Maindonald and Braun, 2006).

1.2 Motivation

EDA, primarily treats the location and scale and provide a basic understanding of the techniques at a reduced level of mathematical sophistication (Hoaglin et al., 2006).
EDA is concerned with observational data more than with data obtained by means of a formal design of experiments. When data are obtained informally, the methods for

handling and analyzing are also often informal.

Plotting data and making comparisons can take the process very far and is better to do than getting a dataset and directly running a regression. Its been a disfavor to analysts and data scientists that EDA has not proceeded as a critical side of the process of dealing with data (Schutt and O'Neil, 2013).

Development of EDA

Tukey (1977) suggests that data smoothers based on running medians rather than running means would provide greater resistance to data spikes. Several forms have proposed that such smoothers might take and exhibited applications on a variety of data sequences. Tukey (1977) develops a large and diverse array of intuitively sound tools for exploring data prior to the application of confirmatory methods. The developed tools help investigators to organize data efficiently, construct compelling graphic displays, examine traditional distributional assumptions, and explore the structure of functional dependencies-all without recourse to the probabilistic assumptions basic to traditional statistical methods.

Moreover, the introduced EDA tools by Tukey (1977) provide means for remedying such common data problems as stray data values, asymmetry, nonlinearity, and location-scale interaction and unresolved problems (Leinhardt and Wasserman, 1979). Tukey (1977) focuses on paper and pencil, graph paper and tracing paper. Tukey (1977) strongly supports the idea of analyzing data can be done manually, with or without the aim of a hand-held calculator and the directly foreseeable future of hand calculation. The distinction between descriptive and inferential statistics, made by the statistician Tukey (1977) is essential. However, the confusion between the two statistical concepts can easily arise because of the special produced techniques for exploratory work, yet placed at some distance from classical statistics.

Hoaglin et al. (1983) explain and illustrate the methods used in EDA to a readership with a moderate knowledge of statistics, then it is a considerable success. EDA studies singly and in combination of "four R's": resistance, residuals, re-expression, and revelation. Resistance provides insensitivity of estimators to a change in a small portion of the data. Residuals are studied to ascertain whether the dominant and unusual features of the data have been adequately isolated and explained. Re-expression (transformation) for the data is used to promote symmetry, homogeneity and linearity. Revelation through visual displays the meets a clear need of data analysts to see behavior and pattern of a data set. Velleman and Hoaglin (2004) introduce the resistant line and the resistant smoothing as parts of nine selected EDA techniques.

The applications have demonstrated real data and provide a unified set of computer programming. The dominant character in the improvement of Velleman and Hoaglin (2004) is to focus on the results rather than to focus on mathematical aspects, also dealing with large real data and setting up facilities for disposal of boring calculations and repeated errors were present in the work. Add to this, the attempt to integrate EDA methods with other methods of data analysis using software that have already been programmed is discussed. The use of the software is a priority for Velleman and Hoaglin (2004) to showcase the features of EDA.

Shitan and Vazifedan (2011) focus on running medians technique and talk about the compound smoothing as a particularly useful method to elaborate the resistant outliers when the procedure of smoothing data is done. The study of Shitan and Vazifedan (2011) shows running medians smoother is resistant while running means smoother is not. The technique of running smoothing is applied using simple real data examples.

have developed an assistant for intelligent data exploration called Aide. Aide is a knowledge-based planning system that incrementally explores a dataset

EDA by Others

(Helsel and Hirsch, 1992) mention that, in spite of the importance of the scatter plot as one of the most familiar graphical methods for data analysis, there are some related issues should be resolved. Some of these issues are whether the relationship appears to be linear or curved, whether different groups of data lie in separate regions of the scatter plot, and whether the variability or spread is constant over the range of data. A smoothing operation as an enhancement of scatter plot enables the viewer to resolve these issues rather than using the scatter plot alone.

Ellison (1993) shows that EDA is intended to edify implicit pattern in noisy data. It is critical that the tacit structure should not be ignored completely in the process. EDA as an antecedent to formal analysis, should not be time-consuming. The feature of using the new technology in computer programming interactive graphics construction with little needed of effort for analysis.

Behrens (1997) introduces the central heuristics and computational tools of EDA and compares it with CDA and exploratory statistics in general. EDA techniques are clarified using earlier published psychological data. Variations in statistical training and practice are recommended to combine these tools.

Amant and Cohen (1998) have developed an assistant for intelligent data exploration called Aide. Aide is a knowledge-based planning system that incrementally explores

a dataset. The system treats some of the strategic drawbacks of traditional soft- ware. This study describes the behavior of the system, gives a high-level illustration of the design, and discusses its experimental evaluation.

Bowman and Azzalini (2003) mention that nonparametric smoothing techniques even if they have simplest forms can be implemented relatively easily through elementary programming techniques. The present statistical computing environments are generally outfitted towards vector and matrix representations of data. An initial aim of the study is to provide simple matrix formulations of smoothing techniques which grant efficient achievements in the environment. Another target of the study is to describe the computational issues that surface when nonparametric methods are applied to large datasets.

Maindonald and Braun (2006) describe some of EDA tools such as histogram, density plot, the stem-and-leaf display, the boxplot, the scatterplot, the lowess smoother, and the trellis-style graphics that are available in the lattice package.

Hébrail et al. (2010) suggest an exploratory analysis algorithm for functional data. The method divides a set of functions into K clusters and represents each cluster by a simple prototype. The number of all segments in the prototypes, P, is chosen by the user and optimally distributed among the clusters via two dynamic programming algorithms. The empirical relevance of the method is shown on two real world datasets.

Habash et al. (2011) present an overview of the mathematical foundations for techniques in EDA for the purpose of investigating the relationships among a lot of variables in large sets of multivariate space weather data. Particularly, this study involves techniques in Principal Components Analysis (PCA) and Common Factor Analysis (CFA). This paper reveals the use of EDA in space weather studies of large multivariate data sets.

EDA techniques are considered useful tools to detect outliers through visual representations but a limitation of this direction is the scarcity of studies that concern the reliability of the visual clarification. Mogoş (2013) proposes a method that combines EDA technique, Andrews curves, with a statistical approach to be applied as automatically classification of the data.

1.3 Problem Statement

Since 1977, work in exploratory data analysis (EDA) area has appeared in technical journals and books but not so much as other literature normally scanned by sociologists. Although a preliminary version of Tukey's text was circulated in the early 1970s,

its unique organization, peculiarly personal style, large size, and lack of mathematical formations caused his methods to remain recondite and little known to non statisticians throughout this period. The book still possesses the unique style of the preliminary edition and, consequently, is a difficult book for professionals, let alone beginning students.

Recently, studies on EDA approaches are classifying the technique as one of the most simplified and exciting fields in data analysis searches due to smoothing. Many novel methods realizing EDA have been proposed. Tukey (1977) introduced the EDA techniques resistant line and resistant smoothing.

Despite the success of resistant line and resistant smoothing techniques provided by Tukey (1977) but there are still some drawbacks. These drawbacks can be summarized as follows. The exploration techniques are based on the manual handling of calculations and results that requires simple data sets to be applied. The techniques adopt simplified numerical examples and lack of operations into mathematical formations that set to process the steps programmatically. Using the exhibition method to illustrate the results and move among steps of operations. The complications of dividing the batches operation in the resistant line procedure in terms of ties issue. Using limited types of compound smoothing methods even if there exists an unlimited number of possible options. Introducing one type of Hanning to enhance the smoothing process that fits the simple data sets only.

Therefore, more investigations on resistant line and resistant smoothing of EDA structures in details should be done to further assist design considerations in order to improve their utilization in many applications. For example, Velleman and Hoaglin (2004) had initiated an investigation on the possibility of enhancing the performance of resistant line and running smoothing applications which resulted in a significant improvement in the fields of using computers to reduce the consuming time of the procedures. Unfortunately, the improvement avoids incorporating smoother structures and shows a lack of emphasis on the mathematical aspects of the techniques.

The goal of this thesis is to explore resistant line and resistant smoothing techniques which may be utilized in EDA to overcome some of the previous drawbacks of the original design through enhancing the structure of smoothers and to enhance the focus on mathematical formulas that fit and support the importance of software in reducing the time required for data analysis.

1.4 Research Objectives

In this section, we review the research objectives and clarify the method used towards accomplishing the achieved objectives as follows.

1. To introduce the range method for dividing the batches.

2. To introduce non-symmetric running smoothing techniques.

3. To propose new symmetric compound running smoothers.

4. To introduce new mathematical Hanning forms.

5. To evaluate and compare the performance of the proposed techniques of smoothing.

1.5 Thesis Outline

The arrangement of the thesis is as follows;
In Chapter 1, an introduction and motivation of this research are written.

In Chapter 2, some backgrounds of EDA techniques such as the improvement of resistant line and resistant smoothing by Tukey (1977), Velleman and Hoaglin (2004), and others are given. We review some techniques related to the smoothing methods that will be used throughout this thesis. Furthermore, we provide a survey of the smoothing index and the measure tools; degree of smoothing and balance test.
In Chapter 3, some of the drawbacks of previous strategies are provided that need to be avoided for realistic implementation of smoothing techniques. We also highlight the method of enhanced dividing the batches and improvement of running smoothing of the research performed. This is followed by our proposed smoothing techniques, namely the symmetric and right running smoothing operations. Then, we introduced new proposed mathematical formulas of Hanning. This is followed by the analytic and the discussion on the validation of the proposed techniques. Chapter 4 discusses and summarizes the results of the proposed techniques and the comparison of the variant new proposal ways of Hanning with the existing techniques that introduced by Tukey (1977) and suggested by Shitan and Vazifedan (2011). Chapter 5, the conclusion of the study are presented with some of the limitations of our research and some future works.

BIBLIOGRAPHY

Amant, R. S. and Cohen, P. R. (1998). Interaction with a mixed-initiative system for exploratory data analysis. *Knowledge-based systems*, 10(5):265–273.

Babura, B. I., Adam, M. B., Fitrianto, A., and Rahim, A. A. (2017). Modified boxplot for extreme data. In *AIP Conference Proceedings*, volume 1842, page 030034. AIP Publishing.

Behrens, J. T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods*, 2(2)(2):131–160.

Bernholt, T., Fried, R., Gather, U., and Wegener, I. (2006). Modified repeated median filters. *Statistics and Computing*, 16(2):177–192.

Bowman, A. W. and Azzalini, A. (2003). Computational aspects of nonparametric smoothing with illustrations from the sm library. *Computational statistics & data analysis*, 42(4):545–560.

Buja, A., Hastie, T., and Tibshirani, R. (1989). Linear smoothers and additive models. *The Annals of Statistics*, pages 453–510.

Burroughs, W. J. (1978). On Running Means and Meteorological Cycles. *Weather*, 33:101–109.

Clark, M. and Anfinson, C. (2011). *Intermediate Algebra: Connecting Concepts Through Applications*. Charlie Van Wagner, USA.

Cox, N. J. and Jones, K. (1981). Exploratory data analysis. *Quantitative Geography, London: Routledge*, pages 135–143.

Ellison, A. M. (1993). Exploratory data analysis and graphic display. *Design and analysis of ecological experiments*, pages 14–45.

Evans, J. R. (1982). Running median filters and a general despiker. *Bull. Seism. Soc. Am*, 72:331–338.

Fried, R., Einbeck, J., and Gather, U. (2007). Weighted repeated median smoothing and filtering. *Journal of the American Statistical Association*, 102(480):1300–1308.

Gebski, V. and McNeil, D. (1984). A refined method of robust smoothing. *Journal of the American Statistical Association*, 79(387):616–623.

Good, I. J. (1983). The philosophy of exploratory data analysis. *Philosophy of science*, 50(2):283–295.

Greiff, W. R. (1998). A theory of term weighting based on exploratory data analysis. In *Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 11–19.

Grznar, J., Booth, D. E., and Sebastian, P. (1997). A robust smoothing approach to statistical process control. *Journal of chemical information and computer sciences*, 37(2):241–248.

Habash, K. L., Franz, A., and Stevenson, A. (2011). On the Application of Exploratory Data Analysis for Characterization of Space Weather Data Sets. *Advances in Space Research*, 47:2199–2209.

Haffajee, A., Socransky, S., and Goodson, J. (1983). Comparison of different data analyses for detecting changes in attachment level. *Journal of clinical periodontology*, 10(3):298–310.

Hardle, W., Steiger, W., et al. (1994). Optimal median smoothing. *IEEE Transactions on Image Processing*, 3:324–327.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. CRC press, New Jersey.

Hébrail, G., Hugueney, B., Lechevallier, Y., and Rossi, F. (2010). Exploratory analysis of functional data via clustering and optimal segmentation. *Neurocomputing*, 73(7):1125–1141.

Helsel, D. R. and Hirsch, R. M. (1992). *Statistical methods in water resources*. Elsevier.

Himes, J. H. and Hoaglin, D. C. (1989). Resistant cross-age smoothing of age-specific percentiles for growth reference data. *American Journal of Human Biology*, 1(2):165–173.

Hoaglin, C. D., Mosteller, F., and Tukey, J. (2006). *Exploring Data Tables, Trends, and Shapes*. Wiley & Sons, New York.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (1983). *Understanding robust and exploratory data analysis*. Wiley.

Hyndman, R. J. (2016). Cran task view: Time series analysis.

Jankowitz, M. D. (2007). *Some statistical aspects of LULU smoothers*. PhD thesis, Stellenbosch: University of Stellenbosch.

Johnstone, L. M. and Velleman, P. F. (1985). The Resistant Line and Related Regression Methods. *Journal of the American Statistical Association*, 80:1041–1054.

Klassen, K. J. (1997). *Simultaneous management of demand and supply in services*. University of Calgary, Canada.

Leinhardt, S. and Wasserman, S. S. (1979). Exploratory data analysis: An introduction to selected methods. *Sociological methodology*, 10:311–365.

Lovie, P. (2005). *Resistant Line Fit*. John Wily and Sons, Chichester.

Maindonald, J. and Braun, J. (2006). *Data analysis and graphics using R: an example-based approach*. Cambridge University Press.

Mallows, C. (1979). Some theoretical results on tukeys 3r smoother. In *Smoothing techniques for curve estimation*, pages 77–90. Springer.

Mallows, C. (2006). *Smoothing Techniques for Curve Estimation*. Springer, Heidelberg.

Mogoş, B. (2013). Exploratory data analysis for outlier detection in bioequivalence studies. *Biocybernetics and Biomedical Engineering*, 33(3):164–170.

Myatt, G. J. (2007). *Making Sense of Data: A Practical Guide to Exploratory Data Analysisand Data Mining*. Wiley & Sons, New Jersey.

Ney, H. (1981). A dynamic programming technique for nonlinear smoothing. In *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP'81*.

Rabiner, L., Sambur, M., and Schmidt, C. (1975). Applications of a nonlinear smoothing algorithm to speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 23(6):552–557.

Ramsay, J. O., Hooker, G., and Graves, S. (2009). *Functional Data Analysis with R and MATLAB*. Springer, Boston.

Sargent, J. and Bedford, A. (2010). Improving australian football league player performance forecasts using optimized nonlinear smoothing. *International Journal of Forecasting*, 26(3):489–497.

Schlittgen, R. (1991). Resistant decomposition of economic time series. *Empirica*, 18:4763.

Schutt, R. and O'Neil, C. (2013). *Doing data science: Straight talk from the frontline*. O'Reilly Media, Inc., Sebastopol, CA.

Seo, S. (2006). *A review and comparison of methods for detecting outliers in univariate data sets*. PhD thesis, University of Pittsburgh.

Shitan, M. and Vazifedan, T. (2011). *Exploratory Data Analysis for Almost Anyone*. Universiti Putra Malaysia Press, Serdang., Malaysia.

Tsaknakis, H. and Papantoni-Kazakos, P. (1988). Outlier resistant filtering and smoothing. *Information and Computation*, 79(2):163–192.

Tukey, J. (1977). *Exploratory Data Analysis* . Addison-Wesley, Reading Massachusetts.

Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1):1–67.

Velleman, P. F. (1977). Robust nonlinear data smoothers: Definitions and recommendations. *Proceedings of the National Academy of Sciences*, 74(2):434–436.

Velleman, P. F. (1980). Definition and comparison of robust nonlinear data smoothing algorithms. *Journal of the American Statistical Association*, 75(371):609–615.

Velleman, P. F. (1982). Applied nonlinear smoothing. *Sociological Methodology*, 13:141–177.

Velleman, P. F. and Hoaglin, D. C. (2004). *Applications, Basics and Computing of Exploratory Data Analysis*. Duxbury Press.

Wainer, H. and Velleman, P. (2008). Looking at blood sugar. *Chance*, 21(4):56–61.

Walfgang, P. (1984). Exploring Business Cycles Using Running Medians. *Computational Statistics & Data Analysis*, 2:5170.

West, I., Coote, G., and Gauldie, R. (1992). Data smoothing techniques applied to proton microprobe scans of teleost hard parts. *International Journal of PIXE*, 2(03):313–323.

Zeng, L. (1995). The optimal degree of smoothing in equipercentile equating with postsmoothing. *Applied Psychological Measurement*, 19(2):177–190.

LIST OF PUBLICATIONS

Journal articles:

Qasim, N.H. , Adam, M.B., Shitan, M. and Fitrianto, A. 2017. Exploratory Extreme Data Analysis for Farmer Mac Data. *Malaysian Journal of Mathematical Sciences*. 11(S) February: 1 – 16.

Qasim, N.H. , Adam, M.B., Shitan, M. and Fitrianto, A. 2016. Extension of Tukeys Smoothing Techniques *Indian Journal of Science and Technology*. 9(28):1–17.

UNIVERSITI PUTRA MALAYSIA
STATUS CONFIRMATION FOR THESIS/PROJECT REPORT AND COPYRIGHT
ACADEMIC SESSION: 2016/2017

TITLE OF THE THESIS/PROJECT REPORT:
MODIFICATION OF TUKEY'S SMOOTHING TECHNIQUES FOR EXTREME DATA

NAME OF STUDENT: QASIM NASIR HUSAIN

I acknowledge that the copyright and other intellectual property in the thesis/project report belonged to Universiti Putra Malaysia and I agree to allow this thesis/project report to be placed at the library under the following terms:

1. This thesis/project report is the property of Universiti Putra Malaysia.
2. The library of Universiti Putra Malaysia has the right to make copies for educational purposes only.
3. The library of Universiti Putra Malaysia is allowed to make copies of this thesis for academic exchange.

I declare that this thesis is classified as:

*Please tick(✓)

☐	CONFIDENTIAL	(contain confidential information under Official Secret Act 1972).
☐	RESTRICTED	(Contains restricted information as specified by the organization/institution where research was done).
☐	OPEN ACCESS	I agree that my thesis/project report to be published as hard copy or online open acces.

This thesis is submitted for:

☐	PATENT	Embargo from _____until _____. (date) (date)

Approved by:

_____ _____

(Signature of Student) (Signature of Chairman of Supervisory Committee)
New IC No/Passport No.:A2063713 Name: **Associated Professor. Dr. Mohd Bakri Adam**

Date: Date:

[Note: If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization/institution with period and reasons for confidentially or restricted.]