ABSTRACT

This study is on modelling circular data and proposed several methods of detecting outliers of the model considered in the study. In particular, the linear functional relationship model is considered where the maximum likelihood parameters and the covariance matrix are derived for the case where the error concentration parameters are unequal since previous study only considered the parameter estimation for the case of equal error concentration. The parameter estimate of *x* variable is improved by applying the iterative procedure in which the incidental parameter stops accumulating after the values converge to a finite number. The parameter estimate of the concentration parameter is estimated by using modified Bessel function in which it is expanded to become a cubic function. Monte Carlo simulation study shows that the proposed parameter estimations give a small bias with mean resultant length near to 1 and small estimated root mean square errors that indicate an adequacy of the estimation. Next, some methods to detect the presence of outlier in circular data are discussed. Previous studies only considered outlier detection methods for equal error concentration case. Therefore, in this thesis, all outlier detection methods take into account of both equal and unequal error concentration parameters in the model. The first proposed method in outlier detection is by using the determinantal equation of the covariance matrix, called *covratio*, in which the covariance matrix is based on the one derived in the first part of the study. Another two proposed methods used to detect the outlier are by using the difference mean circular errors. The two trigonometric functions are the *cosine* (*FDMCEC*) and *sine* (*FDMCES*) functions. The cut-off equations are derived based on the 5% upper percentile of the simulation study for each method for 95% confident level. The feasibility of all of the methods is assessed by the power of performance in Monte Carlo simulation studies when outlier is planted in the data. The results from the simulation study suggest that the power of performance for all three outlier detection methods achieves the maximum percentage which is 100% as the level of contamination increases. Hence, this suggests the feasibility of the method used in outlier detection.

ABSTRAK

Kajian ini adalah mengenai pemodelan data bulatan dan mencadangkan beberapa kaedah untuk mengesan titik terpencil. Secara khususnya, model hubungan fungsian linear dipertimbangkan di mana parameter kemungkinan maksimum dan matriks kovarians diperolehi untuk kes di mana parameter tumpuan ralat tidak sama memandangkan kajian sebelum ini hanya mempertimbangkan kes parameter tumpuan ralat adalah sama. Anggaran parameter bagi pembolehubah *x* ditingkatkan dengan menggunakan prosedur iteratif di mana parameter sampingan berhenti terkumpul setelah nilai-nilai itu berkisar kepada nombor terhingga. Parameter tumpuan dianggarkan dengan menggunakan fungsi Bessel yang diubahsuai di mana ia diperluas untuk menjadi fungsi paduan. Kajian simulasi Monte Carlo menunjukkan bahawa anggaran parameter yang dicadangkan memberi bias yang kecil dengan nilai min paduan dan penganggar ralat min kuasa dua yang kecil menunjukkan kecukupan anggaran. Seterusnya, beberapa kaedah untuk mengesan kehadiran titik terpencil dalam data bulatan akan dipertimbangkan di mana semua kaedah mengambil kira parameter kepekatan ralat sama rata dan tidak sama rata dalam model tersebut. Kaedah pertama yang dicadangkan dalam pengesanan titik terpencil adalah dengan menggunakan persamaan determinantasi matriks kovarians, yang dipanggil *covratio*, di mana matriks kovarians berdasarkan pada yang diperoleh di bahagian pertama kajian. Dua lagi kaedah yang dicadangkan yang digunakan untuk mengesan titik terpencil adalah dengan menggunakan perbezaan min kesilapan pekeliling. Kedua-dua fungsi trigonometri adalah fungsi kosinus (*FDMCEC*) dan sinus (*FDMCES*). Persamaan pemotongan diperoleh berdasarkan

peratus atas 5% kajian simulasi untuk setiap kaedah untuk tahap yakin 95%. Kemungkinan semua kaedah dinilai oleh kuasa prestasi dalam kajian simulasi Monte Carlo apabila titik terpencil dimasukkan ke dalam data. Hasil daripada kajian simulasi mencadangkan bahawa kekuatan prestasi untuk ketiga-tiga kaedah pengesanan terluar mencapai peratusan maksimum sebagai tahap peningkatan pencemaran dengan demikian mencadangkan kemungkinan kaedah tersebut.

ACKNOWLEDGEMENT

Alhamdulillah, I am thankful to Allah the Most Merciful. He has helped me all the time during my thick and thin, provides me helps beyond my expectations and indeed, the best of all providers. I would like to express my sincere and deepest gratitude to Prof. Dr. Abdul Ghapor Hussin and Assoc Prof. Dr. Yong Zulina Zubairi for guiding me throughout this entire study. Their meticulous guidance, continuous encouragement and constant support from the beginning of this study have really helped me to make this study to the completion. I would like to thank Assoc Prof. Dr. Wan Rozita too.

I would like to convey a special appreciation to my parents, Mdm Kamariah Kassim and Mr Mokhtar Aman, my favourite aunty Ms Aminah Aman, my sister Nurul Adibah and my brother Affiq Ashraff for their love and prayers throughout my life. I am really thankful for their support, patience and understanding towards me.

Thirdly, I would like to exclusively acknowledge my research team especially Dr Rossita, Mdm Hafizah, Mdm Liyana, Mdm Azuraini, Dr Norli, Dr Adilah, Dr Siti Fatimah, Dr Syazwan, Dr Amel and Mr Iqbal for sharing me their help, supports and knowledge that guide me to make this thesis possible.

Last but not least, I would like to thank Pusat Pengurusan Penyelidikan & Inovasi UPNM for supporting me financially, Pusat Asasi Sains UM and Institut Sains Matematik UM for providing me the workplace and facilities to carry out my simulation during this study.

APPROVAL

The Examination Committee has met on **2nd July 2019** to conduct the final examination of **Nurkhairany Amyra binti Mokhtar** on her degree thesis entitled '**Outlier Detection Methods of Unreplicated Linear Functional Relationship Model for Circular Variables**'.

The committee recommends that the student be awarded the **Doctor of Philosophy (Statistics)**. Members of the Examination Committee were as follows.

Prof. Dr. Aidy bin Ali
Faculty of Engineering
Universiti Pertahanan Nasional Malaysia
(Chairman)

Assoc. Prof. Dr. Ummul Fahri binti Abdul Rauf
Centre for Defence Foundation Studies
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

Prof. Dr. Azami bin Zaharim
Faculty of Engineering and Built Environment
Universiti Kebangsaan Malaysia
(External Examiner)

Prof. Dr. Muhammad Qaiser Shahbaz
Faculty of Sciences
King Abdul Aziz University
(External Examiner)

APPROVAL

This thesis was submitted to the Senate of Universiti Pertahanan Nasional Malaysia and has been accepted as fulfilment of the requirements for the degree of Doctor of Philosophy (Statistics).

The members of the Supervisory Committee were as follows.

Prof. Dr. Abdul Ghapor Hussin

Faculty of Defence Science and Technology

Universiti Pertahanan Nasional Malaysia

(Main Supervisor)

Assoc. Prof. Dr. Yong Zulina Zubairi

Centre for Foundation Studies in Science

Universiti Malaya

(Co-Supervisor)

Assoc. Prof. Dr. Wan Rozita Wan Din

Centre for Defence Foundation Studies

Universiti Pertahanan Nasional Malaysia

(Co-Supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

DECLARATION OF THESIS

Student's full name : Nurkhairany Amyra binti Mokhtar
Date of birth : 18 August 1991
Title : Outlier Detection Methods of Unreplicated Linear Functional Relationship Model for Circular Variables
Academic Session : 2018/2019

I hereby declare that the work in this thesis is my own except for quotations and summaries which have been duly acknowledged.

I further declare that this thesis is classified as:

- ☐ **CONFIDENTIAL** (Contains confidential information under the Official Secret Act 1972)*
- ☐ **RESTRICTED** (Contains restricted information as specified by the organisation where research was done)*
- ☐ **OPEN ACCESS** I agree that my thesis to be published as online open access (full text)

I acknowledge that Universiti Pertahanan Nasional Malaysia reserves the right as follows.

1. The thesis is the property of Universiti Pertahanan Nasional Malaysia.
2. The library of Universiti Pertahanan Nasional Malaysia has the right to make copies for the purpose of research only.
3. The library has the right to make copies of the thesis for academic exchange.

____________________	____________________
Signature	**Signature of Supervisor/ Dean of CGS/ Chief Librarian
910818-04-5356	Prof. Dr. Abdul Ghapor bin Hussin
IC/Passport Number	**Name of Supervisor/ Dean of CGS/ Chief Librarian
Date:____________________	Date:____________________

Note : * If the thesis is CONFIDENTIAL OR RESTRICTED, please attach the letter from the organisation stating the period and reasons for confidentiality and restriction.
** Witness

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Circular data cannot be analysed by standard linear statistical methods. This is because of the wrapped around nature of the data. However, there are some selected periodical and directional statistical distributions that can be used for modelling circular data such as uniform, cardioid, wrapped normal, and von Mises distribution (Heckenbergerova et al., 2015). Circular data happens in many evidences in our real life, in various scientific fields such as earth sciences, meteorology, biology, physics, psychology, and medicine (Pandolfo et al., 2018).

One of specific examples of the studies in circular data is when Ferguson et al. (1967) used circular data to investigate the homing ability of the northern cricket frog *Acris crepitans*, in which the orientation of the frog was recorded. Gareth (2007) built a time series model for circular data with applications to protein conformational angles. In 2015, Masseran identified a distribution of wind direction for Kudat station in Malaysia (Masseran, 2015).