SOME STATISTICAL PROBLEMS ON CIRCULAR SIMULTANEOUS FUNCTIONAL RELATIONSHIP MODEL

FATIN NAJIHAH BINTI BADARISAM

DOCTOR OF PHILOSOPHY
(STATISTICS)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

2024

SOME STATISTICAL PROBLEMS ON CIRCULAR SIMULTANEOUS FUNCTIONAL RELATIONSHIP MODEL

FATIN NAJIHAH BINTI BADARISAM

Thesis submitted to the Centre for Graduate Studies, Universiti Pertahanan Nasional Malaysia, in fulfilment of the requirements for the Degree of Doctor of Philosophy (Statistics)

2024

ABSTRACT

This study focuses on parameter estimation and outlier detection within the circular simultaneous functional relationship model (CSFRM), considering both equal and unequal variances. The model for unequal variance is a newly proposed extension from the previous model of equal variances. Two approaches, the minimum sum (`ms`) and `polyroot` functions, are employed to approximate parameter estimates due to the complexity of the log-likelihood function. Simulation studies demonstrate reduced bias in the parameter estimates, indicating their effectiveness. Confidence intervals for model parameters are constructed using covariance matrices. Simulation results demonstrate the superiority of the Bootstrap Confidence Interval (BCI) method in constructing confidence intervals for estimated parameters. Additionally, a novel modification method is proposed for identifying single outliers in CSFRM for cases of unequal variances, utilizing the Simultaneous Functional Difference Mean Circular Error cosine (*SFDMCEc*). Simulation results show *SFDMCEc*'s robustness in outlier detection as contamination levels increase. Overall, this study presents effective techniques for parameter estimation, confidence interval and outlier detection in CSFRM, with promising performance in simulation studies for practical applications involving circular simultaneous functional relationships.

ABSTRAK

Kajian ini memberi tumpuan kepada penganggaran parameter dan pengesanan nilai pencilan dalam Model Hubungan Fungsian Bulat Serentak (CSFRM), dengan mempertimbangkan kedua-dua varian sama dan tidak sama. Model dengan varian tidak sama merupakan sambungan model baru yang dicadangkan dari sebelumnya dengan varian sama. Dua pendekatan, iaitu fungsi minimum (`ms`) dan `polyroot`, digunakan untuk menganggarkan parameter disebabkan kekompleksan fungsi log-kebarangkalian. Kajian simulasi menunjukkan pengurangan bias dalam penganggaran parameter, menunjukkan keberkesanan mereka. Selang keyakinan untuk parameter model dibina menggunakan matriks kovarians. Keputusan simulasi menunjukkan keunggulan kaedah Bootstrap Confidence Interval (BCI) dalam membina selang keyakinan untuk parameter yang dianggarkan. Tambahan pula, satu kaedah modifikasi baru dicadangkan untuk mengenal pasti nilai pencilan tunggal dalam CSFRM untuk varian tidak sama, dengan Perbezaan Ralat Bulatan Purata Model Fungsian Serentak kosinus (*SFDMCEc*). Keputusan simulasi menunjukkan ketahanan *SFDMCEc* dalam pengesanan nilai pencilan apabila tahap pengubahsuaian meningkat. Secara keseluruhan, kajian ini menyediakan teknik yang berkesan untuk penganggaran parameter, selang keyakinan, dan pengesanan nilai pencilan dalam CSFRM, dengan prestasi yang menjanjikan dalam kajian simulasi untuk aplikasi praktikal yang melibatkan hubungan fungsian bulat serentak.

ACKNOWLEDGEMENTS

First and foremost, I extend my gratitude to Allah for granting me the strength and opportunity to complete this doctoral thesis. I would like to express sincere appreciation to my supervisors, Dr. Mohd Syazwan Mohamad Anuar, Dr. Adzhar Rambli, and Dr. Mohd Faisal Saari, for their guidance and unwavering encouragement throughout this journey. I also acknowledge Prof. Dr. Abdul Ghapor Hussin for his invaluable guidance at the initial stages, which laid the foundation for my understanding of this research. Their excellent supervision enabled me to persevere and successfully complete this thesis.

Special thanks are due to my beloved family, particularly my husband, Syaffiq, and my mother, Anita, for their unwavering support and the countless sacrifices they made, which greatly contributed to the completion of this challenging endeavor. Additionally, I express my profound gratitude to my dear son, Yusuff, whose presence served as a constant source of inspiration and motivation throughout my studies. I am deeply grateful for each of you and the pivotal roles you have played in my life.

Furthermore, I extend my appreciation to Prof. Dr. Siti Maftuhah Damio for her valuable guidance and insightful comments. Finally, I extend my heartfelt thanks to all individuals who have been directly or indirectly involved in this study. May Allah reward each of you abundantly.

APPROVAL

The Examination Committee has met on **4 April 2024** to conduct the final examination of **Fatin Najihah binti Badarisam** on his degree thesis entitled **'Some Statistical Problems on Circular Simultaneous Functional Relationship Model'.**

The committee recommends that the student be awarded the of **Doctor of Philosophy (Statistics)**.

Members of the Examination Committee were as follows.

Prof. Ts. Gs. Dr. Mohd 'Afizi Mohd Shukran
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Chairman)

Prof. Ts. Gs. Dr. Firdaus Mohamad Hamzah
Centre for Defence Foundation Studies
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

Prof. Ts. Dr. Mohd. Rashid Ab Hamid
Centre for Mathematical Sciences
Universiti Malaysia Pahang Al-Sultan Abdullah
(External Examiner)

Prof. Dr. Abu Sayed Md Al Mamun
Department of Statistics
University of Rajshahi, Bangladesh
(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.

Dr. Mohd Syazwan Mohamad Anuar
Centre for Defence Foundation Studies
Universiti Pertahanan Nasional Malaysia
(Main Supervisor)

Dr. Mohd Faisal Saari
Centre for Defence Foundation Studies
Universiti Pertahanan Nasional Malaysia
(Co-Supervisor)

Dr. Adzhar Rambli
Faculty of Computer and Mathematical Sciences
Universiti Teknologi MARA (UiTM)
(Co-Supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

DECLARATION OF THESIS

Student's full name : Fatin Najihah binti Badarisam

Date of birth : 4 January 1994

Title : Some Statistical Problems on Circular Simultaneous Functional Relationship Model

Academic session : 2023/2024

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
940104-05-5170	Dr. Mohd Syazwan Mohamad Anuar
IC/Passport No.	**Name of Supervisor/Dean of CGS/ Chief Librarian
Date:	Date:

*If the thesis is CONFIDENTAL OR RESTRICTED, please attach the letter from the organisation with period and reasons for confidentiality and restriction.
** Witness

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AEB	-	AEB Absolute Estimated Bias
BCI	-	Bootstrap Confidence Interval
cdf	-	Cumulative distribution function
CFRM	-	Circular Functional Relationship model
CI	-	Confidence interval
cov	-	Covariance
CP	-	Coverage Probability
CFRM	-	Circular Functional Relationship Model
CSFRM	-	Circular Simultaneous Functional Relationship Model
DM	-	Down and Mardia
DMCE	-	Difference Mean Circular Error
DMCEc	-	Difference Mean Circular Error Cosine Statistics
FDMCEc	-	Functional Difference Mean Circular Error Cosine Statistics
EAE	-	Estimated Absolute Error
EIV	-	Error in variables
EL	-	Expected Length
ESE	-	Estimated Standard Error
ERMSE	-	Estimated Root Mean square Error
MLE	-	Maximum Likelihood Estimation

ms	-	Minimum sum function
NACI	-	Normal Asymptotic Confidence Interval
polyroot	-	Polyroot function
SFMCEc	-	Simultaneous Functional Mean Circular Error cosine
SFDMCEc	-	Simultaneous Functional Distance Mean Circular Error cosine
VM	-	Von Mises distribution
WC	-	Wrap Cauchy Distribution
WN	-	Wrap Normal Distribution

LIST OF SYMBOLS

θ	-	An angle on the circle (A circular data)
e	-	Angular error
B	-	Bootstrap sample sizes
d	-	Circular distance
ε	-	Circular random error for dependent variable
δ	-	Circular random error for independent variable
κ	-	Concentration parameter of independent variable
ν	-	Concentration parameter of dependent variable
$\hat{\kappa}$	-	Estimator of Concentration parameter
$\tilde{\kappa}$	-	Corrected estimator of concentration parameter
y	-	Dependent (response) variable
Y	-	Dependent random variable for a functional relationship model
$\hat{\mu}$	-	Estimate value of mean direction
I	-	Fisher information matrix
$\hat{e}$	-	Fitted error
x	-	Independent (predictor) variable
X	-	Independent random variable for a functional relationship model
μ	-	Mean direction
$\bar{R}$	-	Mean resultant length of circular statistics
ρ	-	Mean resultant length (precision parameter)

$A(\kappa)$	-	Mean resultant length for von Mises distribution
$I_p(\kappa)$	-	Modified bessel function of the first kind and order p
$I_o(\kappa)$	-	Modified bessel function of the first kind and order zero
s	-	Number of simulations
q	-	Number of true values of κ falls into the confidence interval
$\hat{Y}$	-	Predicted (fitted) values
λ	-	Ratio of error concentration parameter in a circular simultaneous functional relationship model
R	-	Resultant length
v	-	Sample circular standard deviation
V	-	Sample circular variance
R	-	Sample mean resultant length
n	-	Sample sizes
ω	-	Slope parameter
α	-	Angular parameter
β	-	Angular parameter

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Research Background

Statistical data exhibits unique distributional topology. Linear datasets can be represented on a straight line, while circular data can be visualized using the circumference of a unit circle. It is important to note that statistical theories for linear and circular data differ from each other. Observations of circular data are typically measured in degrees $\left(0^{\circ},360^{\circ}\right]$ or radians $(0,2\pi]$, with a single circular observation represented by a point on a circle with a unit radius.

The distinction between linear and circular data can be effectively illustrated by visualizing the same data set on a linear and a circular plot, as shown in Figures 1.1 and 1.2, respectively.

$$0^{\circ},60^{\circ},120^{\circ},180^{\circ},240^{\circ},300^{\circ},360^{\circ},720^{\circ}$$

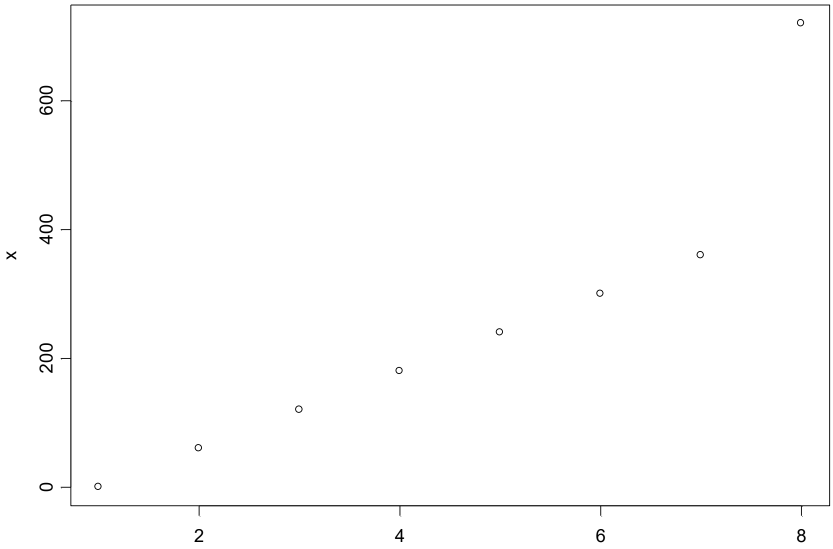

Figure 1.1 Linear plot

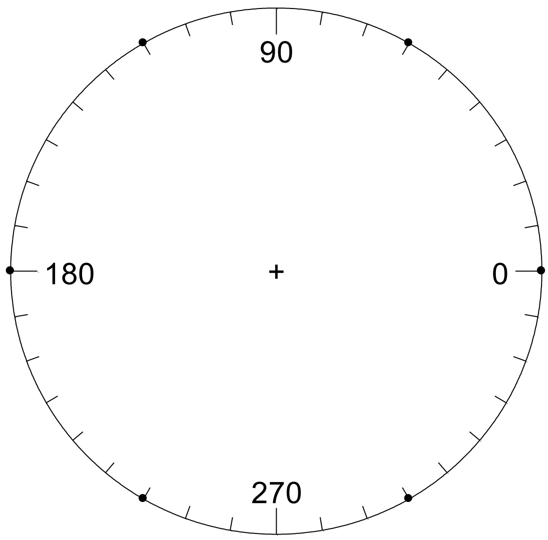

Figure 1.2 Circular plot

Figure 1.1 implies that if the data are treated as linear, the observations are evenly spaced from one another, suggesting a linear trend. However, as shown in Figure 1.2, only six points are in the circular plot when the data are viewed as a sample of circular data with observations ranging from 0° to 360°. Note that the value of circular data falls between $\left(0^{\circ}, 360^{\circ}\right]$ or between $\left(0, 2\pi\right]$ radians. Consequently, 0°, 360°, and 720° are all positioned at the same location as the observation value of 0°. Furthermore, the observation of 720° can be detected as an outlier if the data are represented as linear. However, if the data are treated as circular, the observation of 720° is consistent with the rest of the observations and is not considered an outlier.

In recent years, the analysis of circular variables or directional data has garnered significant interest due to its many practical applications across various scientific disciplines. These disciplines include physics, medicine, geology, meteorology, and astronomy. Earlier, in 1918, von Mises utilized directional data in physics to analyze the fractional part of atomic weights. Circular data has also been applied in a medical setting to aid in the recovery of orthopedic patients, accessed by measuring the angle of knee flexion (Jammalamadaka et al., 1986). Additionally, geologists consider directional data in modeling cross-bedding patterns (Jones & James, 1969) and earthquake displacement directions (Rivest, 1997). Meteorologists have used circular data to study wind directions (Johnson & Wehrly, 1978; Hussin et al., 2004; Gatto & Jammalamadaka, 2007). More recently, Ahmad et al. (2020) applied directional data in astronomy to investigate a new criterion for the visibility of the crescent moon.

The theories and the statistical approaches of circular data have evolved over time. However, they can be enhanced and refined in many statistical areas. Adcock first introduced the error in variables (EIV) problem in 1878, and this topic has drawn a lot of attention from other researchers (see Preece and Baines (1978); Chan and Mak (1979); Brown (1982); Fuller (1987); Cheng (2006); Patriota (2011); Mohammadi et al. (2012)). Some examples include Satari et al. (2014), who proposed the circular functional relationship model for circular variables (CFRM) and study the error in variables model (EIVM) which involves the study of the relationship between two circular variables. The CFRM model is an extension model of Down and Mardia (2002), DM circular regression model. Anuar (2018) then proposed the extension model to circular simultaneous functional relationship model (CSFRM) for equal variances. To the best of the author's knowledge, the study of confidence intervals for parameter estimates in CSFRM model for equal variances have not been published in any referred literature. Therefore, in this study attempted to propose the construction of confidence intervals of CSFRM for equal variances.

Next, this study proposes extending of the CSFRM for equal variances to CSFRM for unequal variances. This model focuses more on investigating directional data in the error-in-variables model (EIVM). Therefore, the focus in this study is on proposing a simultaneous model that can utilize the directional data. This model has a high tendency to be useful for studying the relationship of the directional data in the future for more than two circular variables. The study also considers the derivation of the variance-covariance matrix to construct the confidence intervals for the parameter estimates based on two different methods, and the methods are compared to each other.