

**ON PARAMETER ESTIMATION, CONFIDENCE
INTERVALS AND OUTLIER DETECTION FOR
SOME CIRCULAR MODELS**

MOHD SYAZWAN BIN MOHAMAD ANUAR

DOCTOR OF PHILOSOPHY (STATISTICS)

UNIVERSITI PERTAHANAN NASIONAL

MALAYSIA

2018

**ON PARAMETER ESTIMATION, CONFIDENCE INTERVALS AND
OUTLIER DETECTION FOR SOME CIRCULAR MODELS**

MOHD SYAZWAN BIN MOHAMAD ANUAR

Thesis Submitted to the Centre for Graduate Studies, Universiti Pertahanan Nasional
Malaysia, in Fulfillment of the Requirements for the Degree of Doctor of Philosophy
(Statistics)

JANUARY 2018

ABSTRACT

This study focuses on the parameter estimation, confidence interval estimation and outlier detection for several types of the circular model. The models consider in this study are Down and Mardia Circular Regression Model, Circular Functional Relationship Model (CFRM) and a new model known as the Simultaneous Circular Functional Relationship Model (CSFRM). Firstly, for the Down and Mardia (DM) Circular Regression model, we propose the roots of a polynomial function (polyroot) and minimizing the negative value of the log-likelihood function (ms) to obtain the parameter estimates. The confidence interval for this model are obtained based on the normal distribution where the parameter are estimated using polyroot or minimum sum (ms) and estimation using the bootstrap technique. Secondly, we consider the parameters estimation and confidence intervals for CFRM. We used the polyroot and ms to estimate the concentration, angular and slope parameters. The confidence intervals for the parameters are obtained based on the normal distribution where the parameters are estimated using polyroot or ms and estimation using the bootstrap technique. Thirdly, we consider the outlier problem in CFRM using two statistics namely the Functional Difference Mean Circular Error Using Cosine (FDMCEc) statistic and Functional Difference Mean Circular Error Using Sine (FDMCEs) statistic. This methods also used the row deletion method to calculate FDMCEc and FDMCEs values for the full and reduced data set. In our CFRM, we determine the cut-off points for several values of sample size and concentration parameter designed using simulation study. Lastly, a new simultaneous circular functional relationship model (CSFRM) which is an extended version of a CFRM is developed. We

want to study the relationship for more than two circular variables. We derived the maximum likelihood function of the model and the estimate the variance-covariance of parameters based on Fisher information matrix. Then, the efficiency of the model is assessed using the biasness for the circular variables using circular mean, circular distance and mean resultant length and mean, absolute estimated bias and estimated root mean square error for continuous variables respectively. Model verification and assessment of all methods and model proposed in this study are examined using the simulation study. The wind and wave direction data set used for illustration and application to real data set.

ABSTRAK

Kajian ini memberi tumpuan kepada penganggaran parameter, selang keyakinan dan titik terpencil untuk beberapa jenis model bulatan. Pertama, kami mempertimbangkan penganggaran parameter dan selang keyakinan bagi model regresi membulat,. Kaedah penganggaran parameter yang digunakan adalah penganggaran menggunakan fungsi punca polynomial (polyroot) dan meminimumkan nilai negatif bagi fungsi log-kebolehjadian (ms). Keyakinan kaedah anggaran selang untuk model ini adalah teknik berdasarkan taburan normal di mana parameter dianggarkan menggunakan fungsi polyroot atau ms dan anggaran menggunakan teknik bootstrap. Kedua, kami mempertimbangkan penganggaran parameter dan selang keyakinan untuk model hubungan berfungsi bagi pembolehubah bulatan (CFRM). Kaedah-kaedah yang digunakan untuk menganggarkan parameter dan selang keyakinan adalah sama untuk kedua-dua model. Kami menggunakan polyroot dan ms untuk menganggarkan parameter-parameter menumpu, membulat dan kecerunan. Penganggaran selang keyakinan menggunakan teknik berdasarkan taburan normal di mana parameter dianggarkan menggunakan polyroot atau ms dan anggaran menggunakan teknik yang bootstrap. Ketiga, kami mempertimbangkan masalah titik terpencil dalam CFRM menggunakan dua statistik yang perbezaan. Statistik-statistik ini dikenali sebagai Perbezaan Ralat Bulatan Purata model Fungsian kosinus (FDMCEc) dan sinus (FDMCEs). Kaedah ini juga menggunakan kaedah penyingkiran baris untuk mengira nilai-nilai FDMCEc dan FDMCEs bagi set data penuh dan yang telah disingkirkan. Cerapan terpencil dikesan melalui perbezaan jarak terbesar antara set data penuh dan set data yang telah

disingkirkan. Satu kajian simulasi dijalankan untuk mendapatkan nilai genting bagi beberapa nilai- nilai saiz sampel dan parameter menumpu. Akhir sekali, satu model hubungan berfungsi bagi pemboleh ubah bulatan serentak (CSFRM) yang merupakan lanjutan CFRM dibangunkan. Penganggar kebolehjadian maksimum bagi parameter dan penganggaran varians-kovarians bagi parameter dalam model serentak adalah berdasarkan Matriks Fisher Bermaklumat. Kemudian, hasil kajian dengan menggunakan simulasi menunjukkan bahawa parameter yang dianggarkan mempunyai bias yang kecil. Kaedah yang digunakan untuk menguji prestasi model bagi pemboleh ubah bulat dan pemboleh ubah bukan bulat adalah berbeza. Kesemua penilaian prestasi yang diuji di dalam kajian ini seperti kaedah-kaedah yang digunakan dan model yang dicadangkan dalam kajian ini diuji menggunakan kajian simulasi. Set data pergerakan arah angin dan ombak yang digunakan adalah sebagai ilustrasi kepada uji kaji menggunakan set data sebenar.

ACKNOWLEDGEMENT

I am grateful to Allah the most merciful. I also would like to express my sincere gratitude to my supervisors, Prof. Dr. Abdul Ghapor Hussin and Assoc. Prof. Dr. Yong Zulina Zubairi for their guidance, continuous encouragement and constant support in making this Doctor of Philosophy (Statistics) study possible. I greatly appreciate their guidance from the initial to the final level that enabled me to develop an understanding of this research thoroughly. Without their advice and assistance, it would be a lot tougher to completion.

Secondly, I acknowledge my sincere indebtedness and gratitude to my beloved parents Mohamad Anuar Harithuddin and Azizah Shaari, my mother-in-law Munariza Abd Halim, my beloved wife Nur Hidayu Ab Rahman and my beloved children Muhammad Daniyal Hakimi and Dhia Nur Qaisara for their love, dream and sacrifice throughout my life. I am really thankful for their support, patience, and understanding that makes this work possible.

Thirdly, I would like to express very special thanks to my seniors Zanariah and Iqbal for guiding me until this research has been successfully completed. Last but not least, I want to thank some of my friends especially Khairany for helping me during this study and always supporting me till the end of the study period.

Last but not least, I am grateful to UPNM for granting me the study leave and KPT for the financial assistant thus enabling me to pursue my PhD studies.

APPROVAL

I certify that an Examination Committee has met on **18 January 2018** to conduct the final examination of Mohd Syazwan Bin Mohamad Anuar on his degree thesis entitled — On Parameter Estimation, Confidence Intervals and Outlier Detection for Some Circular Models. The committee recommends that the student be awarded Doctor of Philosophy (Statistics). Members of the Examination Committee were as follows.

Muhd Zu Azhan Yahya, PhD
Professor
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Chairman)

Wan Rozita Wan Din, PhD
Associate Professor
Center for Foundation Studies
Universiti Pertahanan Nasional Malaysia
(Internal Examiner)

Azami Zaharim , PhD
Professor
Faculty of Engineering and Built Environment
University Kebangsaan Malaysia
(External Examiner)

Muhammad Qaiser Shahbaz, PhD
Professor
King Abdulaziz University
Jeddah, Kingdom of Saudi Arabia
(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.

Abdul Ghapor Hussin, PhD
Professor
Faculty of Defence Science and Technology
Universiti Pertahanan Nasional Malaysia
(Main Supervisor)

Assoc. Prof. Yong Zulina Zubairi, PhD
Associate Professor
Center for Foundation Studies in Science
Universiti Malaya
(Co-Supervisor)

UNIVERSITI PERTAHANAN NASIONAL MALAYSIA

DECLARATION OF THESIS

Author's full name : Mohd Syazwan Bin Mohamad Anuar

Date of birth : 23 April 1986

Title : On Parameter Estimation, Confidence Intervals and Outlier Detection for Some Circular Models.

Academic session : 2013/2014

I declare this thesis is classified as:

- CONFIDENTIAL (Contains confidential information under the Official Secret Act 1972)*
- RESTRICTED (Contains restricted information as specified by the organization where research was done)*
- OPEN ACCESS I agree that my thesis to be published as onlines 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 MAIN SUPERVISOR

I/C NUMBER

NAME OF MAIN SUPERVISOR

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.

TABLE OF CONTENTS

ABSTRACT	ii
ABSTRAK	iv
ACKNOWLEDGEMENTS	vi
APPROVAL	vii
DECLARATION OF THESIS	ix
TABLE OF CONTENTS	x
LIST OF TABLES	xvii
LIST OF FIGURES	xx
LIST OF SYMBOLS	xxii
LIST OF ABBREVIATIONS	xxv
CHAPTER 1 – INTRODUCTION	1
1.1 Research Overview	1
1.2 Motivation of Study	5
1.3 Problem Statement	6
1.4 Research Objective	7
1.5 Scope of Research	8
1.6 Limitation of Study	8
1.7 Thesis Organization	9
CHAPTER 2 – LITERATURE REVIEW	12
2.1 Introduction	12
2.2 Circular Data and Graphical Representations	12
2.3 Descriptive Statistics	15
2.3.1 Measure of Location and Measure of Concentration	15

2.3.2 Measure of Dispersion	17
2.4 Software for Circular Data	18
2.5 Distribution of Circular Data	19
2.5.1 The von Mises Distribution	20
2.6 Variance-Covariance Matrix of Parameters	23
2.7 Confidence Interval for Parameters	24
2.7.1 Confidence Interval for Concentration Parameter for Von Mises Distribution Using Von Mises Distribution	25
2.8 Circular Regression Models	31
2.8.1 Downs and Mardia Circular Regression Model	34
2.9 Error-in-Variable Models (EIVM)	39
2.10 Functional Relationship Models	41
2.11 Functional Relationship Models for Circular Variables	42
2.11.1 Circular Functional Relationship Model (CFRM)	43
2.12 Simultaneous Functional Relationship Models	50
2.13 Circular Distance	52
2.14 Outliers Detection Methods	52
2.14.1 Graphical Methods	53
2.14.2 COVRATIO Statistics	54
2.14.3 Difference Mean Circular Error Cosine (DMCEc) Statistic	55

2.14.4 Difference Mean Circular Error Sine (DMCEs) Statistic	57
2.15 Illustration of Real Data Set for Practical Example	61
2.16 Summary	62
CHAPTER 3 – METHODOLOGY	63
3.1 Introduction	63
3.2 Maximum Likelihood Estimation for Parameter in DM Circular Regression Model	64
3.2.1 Estimation for α, β and ω	66
3.2.2 Estimation for κ	67
3.2.2.1 Estimation of Concentration Parameters using <code>ms</code> function	67
3.2.2.2 Estimation of Concentration Parameters using <code>Polyroot</code> function	68
3.3 Variance-Covariance Matrix for All Parameters for DM Circular Regression Model	71
3.4 Confidence Interval for Error Concentration Parameter for DM Circular Regression Model	72
3.5 Confidence Interval for Angular Location (α, β) and Slope ω Parameters for DM Circular Regression Model	75
3.6 Maximum Likelihood Estimation for all Parameters $(\alpha, \beta, \omega$ and $\kappa)$ for CFRM	79
3.6.1 Angular and Slope Parameters Estimation using <code>ms</code> Function for CFRM	81

3.6.2 Numerical Approximation of Mean Resultant Length and Concentration Parameter	81
3.6.3 Concentration Parameter Estimation using Polyroot Function for CFRM	82
3.6.4 Concentration Parameter Estimation using ms Function for CFRM	83
3.7 Variance-Covariance Matrix for CFRM	86
3.8 Confidence Interval for Parameters (α , β , ω and κ) for CFRM	90
3.8.1 Confidence Interval for Error Concentration Parameter	91
3.8.2 Confidence Interval for Angular Location (α, β) and Slope ω Parameters	93
3.9 Circular Residuals	96
3.10 Functional Different Mean Circular Error (FDMCE) Statistics	97
3.10.1 Functional Difference Mean Circular Error Cosine Statistics (FDMCEc)	97
3.10.2 Functional Difference Mean Circular Error sine Statistics (FDMCEs)	99
3.11 Flow of Research	101
CHAPTER 4 - CONFIDENCE INTERVAL FOR THE PARAMETERS IN DM CIRCULAR REGRESSION MODEL	102
4.1 Introduction	102
4.2 Simulations Studies for Confidence Interval of Parameters in DM Circular Regression Model	103
4.2.1 Result on the Confidence Interval of Concentration Parameter, κ	105

4.2.2 Result on the Confidence Interval of Angular Parameters, α	107
4.2.3 Result on the Confidence Interval of Angular Parameters, β	109
4.2.4 Result on the Confidence Interval of Slope Parameter, ω	111
4.3 Discussion	113
4.4 Application using Real Data	115
4.5 Summary	117
CHAPTER 5 - CONFIDENCE INTERVALS FOR PARAMETERS IN CIRCULAR FUNCTIONAL RELATIONSHIP MODEL	118
5.1 Introduction	118
5.2 Simulation Studies for Confidence Interval of Parameters in CFRM	119
5.2.1 Result on the Confidence Interval of Concentration Parameter, κ	120
5.2.2 Result on the Confidence Interval of Angular Parameters, α	122
5.2.3 Result on the Confidence Interval of Angular Parameters, β	124
5.2.4 Result on the Confidence Interval of Slope Parameter, ω	126
5.3 Discussion	128
5.4 Application using Real Dataset	130
5.5 Summary	132
CHAPTER 6 - STATISTICS FOR DETECTION OF OUTLIER	133
6.1 Introduction	133
6.2 Percentiles Points of Test Statistics	133

6.3 Discussion	139
6.4 Application using Real Dataset	144
6.5 Summary	148
CHAPTER 7 - A NEW CIRCULAR SIMULTANEOUS FUNCTIONAL RELATIONSHIP MODEL FOR EQUAL VARIANCES	149
7.1 Introduction	149
7.2 A New Circular Simultaneous Functional Relationship Model (CSFRM)	150
7.3 Parameter Estimation for CSFRM	153
7.3.1 Parameter Estimation for Angular and Slope Parameters for CSFRM	154
7.3.2 Parameter Estimation for Concentration Parameter for CSFRM	156
7.4 Variance-Covariance Matrix for CSFRM	158
7.5 Result on the Simulation Studies for CSFRM	164
7.6 Application using Real Dataset	173
7.7 Summary	174
CHAPTER 8 – CONCLUSION	175
8.1 Summary	175
8.2 Contributions	178
8.3 Further Research	180
REFERENCES	181
APPENDIX A: Circular Data Set	187

APPENDIX B: SPlus Code for the Confidence Interval Estimation for DM Circular Regression Model	191
APPENDIX C: SPlus Code for the Confidence Interval Estimation for Circular Functional Relationship Model	202
APPENDIX D: SPlus Code for the Outlier Detection in Circular Functional Relationship Model	215
APPENDIX E: SPlus Code for the New Circular Simultaneous Functional Relationship Model	220
BIODATA OF STUDENT	228
LIST OF PUBLICATIONS AND PAPER PRESENTED	229

LIST OF TABLES

	Page
Table 2.1 The study on development of confidence interval for parameters of von Mises distribution	30
Table 2.2 The study on development of circular-circular regression model	37
Table 2.3 The study on development of functional relationship model for circular variables	46
Table 2.4 Study on development for simultaneous functional relationship function	51
Table 2.5 The study on development of outlier detection	59
Table 4.1 Coverage Probability and expected length for various values of concentration parameter κ and sample sizes n $(\alpha=\pi/4, \beta=\pi/4 \text{ and } \omega=0.5)$	106
Table 4.2 Coverage Probability and expected length for various values of angular parameter α , and sample sizes n ($\alpha=\pi/4, \beta=\pi/4 \text{ and } \omega=0.5$)	108
Table 4.3 Coverage Probability and expected length for various values of angular parameters β , and sample sizes n ($\alpha=\pi/4, \beta=\pi/4 \text{ and } \omega=0.5$)	110
Table 4.4 Coverage Probability and expected length for various values of slope parameter ω , and sample sizes n ($\alpha=\pi/4, \beta=\pi/4 \text{ and } \omega=0.5$)	112
Table 4.5 Summary of the Performance for NACI and BCI as n increases	114
Table 4.6 Parameters estimation of the Humberside wind direction data (WAVE72 and W125) with the circular regression model	116
Table 5.1 Coverage Probability and expected length for various values of concentration parameter κ and sample sizes n . $(\alpha=\pi/4, \beta=\pi/4 \text{ and } \omega=0.5)$	121

Table 5.2	Coverage Probability and expected length for various values of angular parameter, α and sample sizes n ($\alpha=\pi/4$, $\beta=\pi/4$ and $\omega=0.5$)	123
Table 5.3	Coverage Probability and expected length for various values of angular parameters, β and sample sizes n ($\alpha=\pi/4$, $\beta=\pi/4$ and $\omega=0.5$)	125
Table 5.4	Coverage Probability and expected length for various values of slope parameter, ω and sample sizes n ($\alpha=\pi/4$, $\beta=\pi/4$ and $\omega=0.5$)	127
Table 5.5	Summary of Confidence Interval for all Parameters in Circular Functional Relationship Model Based of Coverage Probability	129
Table 5.6	Summary of Confidence Interval for all Parameters in Circular Functional Relationship Model Based of Expected Length	129
Table 5.7	The summarize of the estimation for real data set	131
Table 6.1	The simulated cut-off points for FDMCEc statistic sorted by sample size	135
Table 6.2	The simulated cut-off points for FDMCEc statistic sorted by concentration parameter	136
Table 6.3	The simulated cut-off points for FDMCEs statistic sorted by sample size	137
Table 6.4	The simulated cut-off points for FDMCEs statistic sorted by concentration parameter	138
Table 7.1	Simulation Results of $\hat{\alpha}_1$ (True value $\alpha_1 = 0.785$) for Simultaneous Model	166
Table 7.2	Simulation Results of $\hat{\alpha}_2$ (True value $\alpha_2 = 0.524$) for Simultaneous Model	167
Table 7.3	Simulation Results of $\hat{\beta}_1$ (True value $\hat{\beta}_1 = 0.785$) for Simultaneous Model	169
Table 7.4	Simulation Results of $\hat{\beta}_2$ (True value $\hat{\beta}_2 = 0.524$) for Simultaneous Model	170

Table 7.5	Simulation Results of $\hat{\omega}$ (True value $\hat{\omega} = 0.5$) for Simultaneous Model	171
Table 7.6	Simulation Results of \hat{k} for several (True value $\hat{k} = 5, 10, 15, 20$) for Simultaneous Model	172
Table 7.7	Parameters estimation of the Humberside wind direction data (Simul49) with the new circular simultaneous functional relationship model	174

LIST OF FIGURES

	Page
Figure 2.1 A graphical representation of circular data	13
Figure 2.2 A graphical representation of linear data	14
Figure 2.3 $VM(n = 20, \mu = 0, \kappa = 3)$	22
Figure 2.4 $VM(n = 20, \mu = 0, \kappa = 5)$	23
Figure 2.5 $VM(n = 20, \mu = 0, \kappa = 10)$	23
Figure 6.1 95% Cut-off points for FDMCEs and FDMCEc for different values of κ at fixed $n = 20$	140
Figure 6.2 95% Cut-off points for FDMCEs and FDMCEc for different values of κ at fixed $n = 40$	140
Figure 6.3 95% Cut-off points for FDMCEs and FDMCEc for different values of κ at fixed $n = 60$	141
Figure 6.4 95% Cut-off points for FDMCEs and FDMCEc for different values of κ at fixed $n = 80$	141
Figure 6.5 95% Cut-off points for FDMCEs and FDMCEc for different values of n at fixed $\kappa = 5$	142
Figure 6.6 95% Cut-off points for FDMCEs and FDMCEc for different values of n at fixed $\kappa = 10$	142
Figure 6.7 95% Cut-off points for FDMCEs and FDMCEc for different values of n at fixed $\kappa = 15$	143

Figure 6.8	95% Cut-off points for FDMCEs and FDMCEc for different values of n at fixed $\kappa = 20$	143
Figure 6.9	Scatter plot of wind direction data measured by HF radar system and anchored buoy	145
Figure 6.10	FDMCEs values for the wind direction data	146
Figure 6.11	FDMCEc values for the wind direction data	146

LIST OF SYMBOLS

κ	Concentration parameter
$\hat{\kappa}$	Estimate value of concentration parameter
\mathbf{x}	A unit vector in the plane
θ	An angle on the circle (A circular data)
z	A unit complex number on a circle
R	Resultant length
\bar{R}	Sample Mean resultant length
$\bar{\theta}$	Sample Mean direction
μ	Mean direction
$\hat{\mu}$	Estimate value of Mean direction
ρ	Mean resultant length (precision parameter)
$A(\kappa)$	Mean resultant length for von Mises distribution
n	Sample sizes
$\tilde{\theta}$	Sample Median direction
V	Sample circular variance
v	Sample circular standard deviation
$\hat{\delta}$	Sample circular dispersion

$d_0(\alpha)$	Circular mean deviation
\bar{D}_0	Circular mean difference
B	Bootstrap sample sizes
y	Dependent (response) variable
x	Independent (predictor) variable
e	Angular error
\hat{e}	Fitted error
ε	Circular random error for dependent variable
δ	Circular random error for independent variable
λ	Ratio of error concentration parameters in a functional relationship model
d	Circular distance
d_{ij}	Distance between observation i and j
$\bar{\kappa}$	Means of concentration parameter
$I_p(\kappa)$	Modified Bessel Function of the First Kind and Order p
$I_0(\kappa)$	Modified Bessel Function of the First Kind and Order zero
v	Dependent (response) variable for Down and Mardia model
u	Independent (predictor) variable for Down and Mardia model