SOME PROBLEMS OF OUTLIERS IN CIRCULAR DATA

ALI H. M. ABUZAID

FACULTY OF SCIENCE
UNIVERSITY OF MALAYA
KUALA LUMPUR

2010

SOME PROBLEMS OF OUTLIERS IN CIRCULAR DATA

ALI H. M. ABUZAID

THESIS SUBMITTED FOR THE FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2010

ABSTRAK

Kajian ini mengambil kira tiga masalah nilai tersisih dalam statistik bulatan. Masalah yang pertama adalah cubaan untuk menggunakan prosedur piawai dalam mengesan nilai tersisih untuk data set yang linear dengan menganggarkan pembolehubah bulatan oleh pembolehubah linear. Ini adalah mungkin bagi nilai penumpuan parameter yang besar. Siri kajian simulasi dilaksanakan bagi menentukan nilai penumpuan parameter yang boleh diterima supaya taburan von Mises boleh dianggarkan oleh taburan normal.

Kedua adalah masalah nilai tersisih dalam sampel bulatan. Dua ujian berangka tak sejajar dicadangkan bagi mengenal pasti nilai tersisih. Ujian statistik berdasarkan penjumlahan jarak bulatan dan panjang perentas masing-masing daripada satu titik nilai ke semua titik lain pada lilitan satu bulatan. Statistik ujian taburan penghampiran yang baru telah diterbitkan. Kajian simulasi menunjukkan bahawa prestasi kedua-dua statistik tersebut lebih baik daripada ujian tak sejajar. Selain itu, satu versi plotkotak untuk data set bulatan dicadangkan. Melalui kajian simulasi, kita dapat menunjukkan kriteria perintang amat bergantung kepada ukuran penumpuan sampel bulatan.

Masalah ketiga ialah kewujudan nilai tersisih dalam model regresi bulatan. Pertamanya, kita mencadangkan satu takrif baru bagi ralat bulatan yang boleh mengenal pasti nilai tersisih dengan menggunakan pelbagai graf dan ujian-ujian berangka. Kedua, tiga ujian berangka dibangunkan untuk mengesan nilai berpengaruh berasaskan pendekatan penghapusan baris. Dua yang pertama didefinisikan menggunakan jarak bulatan antara pemerhatian dan nilai-nilai penyesuaian dengan terbitan taburan penghampiran. Ujian yang lain adalah lanjutan satu versi statistik COVR ATIO dalam

regresi linear untuk kes bulatan. Secara umum, ketiga-tiga ujian berangka menunjukkan prestasi yang baik dalam mengesan nilai berpengaruh.

Untuk ilustrasi, kita mempertimbangkan dua data set bulatan yang sebenar, yakni, set data arah pergerakan katak dan set data arah angin. Kesimpulannya, statistik yang dicadangkan di sini mampu menyelesaikan sebahagian besar masalah nilai tersisih dalam data bulatan.

ABSTRACT

This study considers three problems of outliers in circular statistics. The first problem is an attempt to use the standard outlier detection procedures for linear data set by approximating circular variables by linear variables. This is possible for large values of concentration parameter. Series of simulation studies are carried out to specify the accepted value of the concentration parameter so that the von Mises distribution can be approximated by normal distribution.

The second is the problem of outliers in circular samples. Two numerical tests of discordancy are proposed to identify outliers. The test statistics are based on the summation of circular distances and chord lengths respectively from the point of interest to all other observations on the circumference of a unit circle. The approximate distributions of the test statistics are derived. Simulation studies show that both statistics perform better than other known discordancy tests. On the other hand, a boxplot version for circular data sets is proposed. Via simulation studies, we show that the resistant criterion highly depends on the concentration of circular samples.

The third problem is the existence of outliers in the circular regression model. Firstly, we propose a new definition of circular residuals which can be used to identify outliers using various graphical and numerical tests. Secondly, three numerical tests are developed to detect influential observations based on row deletion approach. The first two are defined using the circular distance between the observed and fitted values with the derivation of the approximate distributions. The other test is an extended version of the COVRATIO statistic in linear regression to the circular case. In general, the three numerical tests perform well in detecting influential observations.

For illustration, we consider two real circular data sets, namely, the frogs' data set and the wind direction data set. In conclusion, the statistics proposed by this study are able to solve some problems of outliers in circular data.

# ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisors, Assoc. Prof. Dr. Ibrahim Bin Mohamed and Assoc. Prof. Dr. Abdul Ghapor Hussin, for their invaluable help, encouragement and supervision throughout my research.

My heartfelt thanks go to my parents, brothers and sisters for their support and encouragement at all stages of this study.

I would like to gratefully acknowledge all of the staff members of the Institute of Mathematical Sciences, University of Malaya. I would also like to thank the University of Malaya for its financial support.

I especially thank my friends; Adzhar, Ahmed, Akmal, Ayed, Hatem, Khaled, Isfahani, Mohamed, Norli and Yahia for their assistance and support during my course of study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATION

κ	Concentration parameter
$VM(\mu,\kappa)$	The von Mises distribution with mean direction μ and concentration parameter κ
C	C test of discordancy
D	D test of discordancy
M	M test of discordancy
A	A test of discordancy
$Chord$	$Chord$ test of discordancy
R	Resultant length
$\bar{R}$	Mean resultant length
$I_0(\kappa)$	The modified Bessel function of the first kind and order zero.
U_n^2	Waston test
$\bar{\theta}$	Sample mean direction
ϕ	Sample median direction
ν	Resistant constant for circular boxplot
$CIQR$	Circular interqurtile range
$\boldsymbol{r}_A$	Circular residuals
$A(\kappa)$	Ratio of Bessel functions, a measure of goodness-of-fit for circular regression model
$A^*(\kappa)$	Modified measure of goodness-of-fit for circular regression model
d_{ij}	Circular distance between two circular observations θ_i and θ_j
$N(\kappa d_{ij})$	Number of κd_{ij} exceeding the critical value of $\chi^2_{1,\alpha}$ for $i=1,\ldots,n$
r_c	Circular correlation coefficient
$MCEc$	Mean circular error in terms of cosine function
$MCEs$	Mean circular error in terms of sine function

$DMCEc$	The maximum absolute difference between $MCEc$ for full and reduced data
$DMCEs$	The maximum absolute difference between $MCEs$ for full and reduced data
λ	Contamination level
$MMCEc$	The approximate distribution of the modified mean circular error statistic in terms of cosine function
$MMCEs$	The approximate distribution of the modified mean circular error statistic in terms of sine function
$MDC_{(-j)}$	The absolute difference between $MMCEc$ statistic for full and reduced data.
$MDS_{(-j)}$	The absolute difference between $MMCEs$ statistic for full and reduced data.
$COVRATIO_{(-i)}$	The determinantal ratio of covariance matrices for reduced and full data.

CHAPTER ONE

INTRODUCTION

1.1 Background of the study

Statistical data can be classified according to their distributional topologies. Most of the data are linear type which can be represented on a real line. However, the circumference of the circle or the surface of sphere are more convenient to represent a data of directional type.

The disparate topologies of the circle and straight line are reflected in the mathematical and statistical treatments of the data. Circle is a closed curve but not for a line. From the properties of the circle, the directions close to the opposite end-points are near neighbour in a circular metric but maximally distant in linear metric.

Circular data refer to a set of observations measured by angles and distributed within $(0,2\pi]$ or $(0°,360°]$. It can be displayed on the circumference of a unit circle. Circular data are found in many scientific fields:

(i) **meteorology**: there are many circular data arising in meteorological studies such as wind and wave directions (Johnson & Wehrly, 1977; Hussin *et al.*, 2004 and Gatto & Jammalamadaka, 2007), the number of times a day at which thunderstorms occur and the frequencies of heavy rain in a year (Mardia & Jupp, 2000).

(ii) **biology**: animal navigation (Batschelet, 1981), spawning times of a particular fish (Lund, 1999).

(iii) **physics**: fractional part of atomic weights (von Mises, 1918), source of signals in the case of airplane crashes (Lenth, 1981).

(iv) **psychology**: studies of mental maps to represent the surroundings of respondents (Gordon *et al.*, 1989).

(v) **medicine**: the angle of knee flexion as a measure of recovery of orthopaedic patients (Jammalamadaka *et al.*, 1986).

(vi) **geology**: modelling the cross-bedding data (Jones & James, 1969), the direction of earthquake displacement in terms of the direction steepest decent (Rivest, 1997).

(vi) **political science:** the modelling of the casualties in the second Iraq war and suicide cases in Switzerland (Gill & Hangartner, 2009).

In general, circular data can be found whenever periodic phenomena occur. Applying the conventional linear techniques on circular data may lead to paradoxes. For example, let us consider two angles $5°$ and $355°$ as illustrated in Figure 1.1. The arithmetic mean by treating the data as linear observation is $180°$. However, the mean direction of the two directions has to be $0°$. Therefore, special statistical methods and techniques are needed to analyse circular data while taking into account the structure of circular sample space.

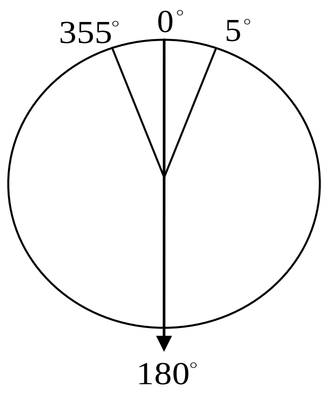


Figure 1.1: Arithmetic and geometric mean

The history of circular data emphasizes the saying “Necessity is the mother of invention”. The development of circular data analysis is a response to applied problems in several fields of science. Astronomy was the host soil for the roots of circular data when the Reverend John Mitchell FRS analysed the angular separation between stars in 1767. The second valuable contribution is in geographical context, where John Playfair in 1902 was the first man who pointed out the requirement of new and different methods to analyse circular data. He recommended the use of the resultant vector method in averaging directions.

Construction of statistical graphics for circular data goes back to the end of the first Millennium. Nightingale in 1858 proposed a circular graphical device (currently known as “rose diagram”) to present social and medical data that could save thousands of lives (Kopf, 1916 and Fisher, 1993, p.5).

The interest in circular data analysis increased gradually until circular probability distributions started to appear in the literature in 1950s. A significant development of circular data analyses occurred when Waston & Williams (1956) introduced the statistical inference about the mean direction and dispersion for samples from a von Mises distribution. Since then, the analysis of circular data has seen vigorous developments where many related books and review papers were published. The first comprehensive book was written by Mardia (1972) followed by a specialised book on circular statistics in Biology by Batschelet (1981). Historical review of directional statistics was thoroughly covered by Fisher *et al.* (1987) and Fisher (1993). On the other hand, Jupp & Mardia (1989) published the first statistical review paper concerning the directional data which summarized the developments of circular data analysis over the years.

Currently, the analysis of circular data attracts the interest of statisticians and researchers from different scientific fields due to the availability of solid foundation theory and the accessibility to this kind of data. Recently, new circular distributions have been proposed (see Siew *et al.* (2008), Gatto & Jammalamadaka (2007)). Strong interests on circular regression model have also been shown (see Downs & Mardia (2002), Hussin *et al.* (2004) and Kato *et al.* (2008)). Works on functional relationship models for circular variables have also been reported (see Hussin (1997), Bowtell & Patefield (1999) and Caires & Wyatt (2003)).

As the analyses of circular data are being developed and its applications are highly sought after, the necessity for special statistical software to analyse circular data has increased. Currently, there are a few statistical software that provide limited analyses of circular data, inter alia, Axis, Oriana and DDSTP (a Statistical Packages for the Analysis of Directional Data). Moreover, some routines are written in R/S-Plus language and are provided by Jammalamadaka & SenGupta (2001). Therefore, more algorithms and programmes need to be developed to conduct necessary analyses and simulations using appropriate packages.

However, the problem of outliers in circular data has not received enough attention. A few tests of discordancy that have been formulated but none has been shown superior over the others. The interest here is to develop new numerical and graphical tests of discordancy that are more powerful and interpretable. Similarly, the problem of outliers in circular regression models has not been mentioned in any published work. Throughout this thesis, the development of some statistics to detect outliers and influential observations in circular samples and simple circular regression model will be discussed.

1.2 Statement of the problem

Circular data is subjected to contamination with outlying observations. So far, very few published papers focusing on the detection of outlier in circular data exist. Moreover, no related study on outlier problem in circular regression has been found in the literature. Hence, in this study, we will develop new test statistics and graphical procedures to detect outliers and influential observations in circular samples and circular regression. The asymptotic distributions for some of the proposed statistics will also be derived. Further, we look at several important issues in circular regression. The first issue is to investigate the diagnostic checking for circular regression models, and the second is to explore the possibility of applying standard procedures in linear case by approximating circular variables by linear. The performance of the relevant proposed procedures will be compared.

1.3 Objectives

Based on the statement of problem above, the researcher has outlined the following objectives for this study:

1. To specify the accepted value of the concentration parameter that von Mises distribution can be approximated by normal distribution.
2. To detect possible outliers in circular samples by:
   (i) Proposing alternative statistical tests of discordancy.
   (ii) Deriving the approximate distribution of proposed tests.
3. To develop the circular boxplot and formulate its criterion to identify outliers.
4. To formulate a new definition of circular residuals for diagnostic checking purposes.

5. To propose new statistics to identify outliers and influential observations in circular regression.
6. To derive the approximate distributions of some of the proposed statistics in (5).

1.4 Significance of the study

The findings from this study will be beneficial in the following ways:

1. Contribute to the knowledge in statistics regarding the modelling of circular data and detection of outliers and influential observations.
2. Optimize the estimation of parameters in circular models by identifying outliers and influential observations.

1.5 Thesis outline

This research attempts to handle the problem of outliers in circular data and circular regression by proposing new alternative statistical techniques. The research is outlined as follows:

Chapter two provides a literature review about the circular regression models and the problem of outliers in univariate samples and regression models. The focus will be on the regression of circular variables. A review of method of identification of outliers in both linear and circular univariate samples is presented. Special discussions are on the test of discordancy in circular samples. In addition, we review some of the outlier and influential observations detection methods in linear regression which has the possibility to be extended to circular case.

Chapter three introduces descriptive statistics for circular data, such as the mean and median directions, mean resultant length, the circular variance and the standard deviation. A brief discussion is given on the von Mises distribution. Two circular data sets which are used in this study are presented.

Chapter four looks at the problem of approximating von Mises distribution by normal distribution. Two circular data sets are approximated by linear data to identify possible outliers and highlight the drawback of approximation technique.

Chapter five discusses the general effect of outliers on the summary statistics for circular data. Here, we propose two alternative tests of discordancy for circular samples. The cut-off points for both tests are obtained and the power of performance is investigated through simulation studies. We also look at the derivation of the approximate distribution of the test statistics.

Chapter six proposes a circular boxplot to label possible outliers in circular samples. Extensive simulation studies are used to find the suitable circular boxplot criterion. Special sub-routine is developed within S-Plus environment.

Chapter seven presents the development of the simple circular regression model. It proposes and tests the satisfaction of the assumptions for a new practical definition of circular residuals based on the circular distances. The diagnostic checking tools for circular regression are discussed.

Chapter eight proposes several numerical statistics to detect possible influential observations in the circular regression. Two statistics are derived based on the difference between fitted and observed values. The cut-off points and the power of performance are discussed. The approximate distributions for modified statistics are derived. Further, the idea of COVRATIO statistic in linear regression model is extended to the circular case. Via simulation, the cut-off points are obtained and the power of performance is investigated.

Chapter nine presents the general conclusion and highlights the significant contributions of this research work. Moreover, the researcher also suggests a few possibilities for extending research work on the problem of outlier in the area of circular statistics.

Lists of appendices are attached at the end of this thesis, including the wind direction data, simulation results and the S-Plus subroutines.

CHAPTER TWO

LITERATURE REVIEW

2.1 Introduction

One of the most common problems arising in any statistical analysis is the existence of some unexpected observations. Such observations are known as outliers and are not guaranteed to be a part of the phenomena under study. The problem of outlier is considered to be as old as the subject of statistics itself. Beckman & Cook (1983) and Barnett & Lewis (1984) reviewed the literature on outliers and the available approaches to deal with outliers in different areas of statistics. The earliest discussion on outliers was done by Bernoulli (1777) where he questioned the assumption of identically distributed error. The first attempt to develop an objective statistical method to deal with outliers was proposed by Peirce (1852). Later, Wright (1884) extended their works and suggested that any observation whose residual exceeds 3.37 times the standard deviation it is rejected. Extensive literature on outliers includes different definitions of outliers and a general agreement that outliers in a set of data refer to observations which appear to be inconsistent with the remaining observations.

Until the late of 1950's, there was not much development in the detection of outliers due to the absence of high speed computing facilities. It was only after the existence of high performance computing that there was interest in the outliers' problem. Lately, it has become the central focus, and must be taken into account in any data analysis to obtain better estimation of the considered models.

Beckman & Cook (1983) outlined the importance and the reasons for studying outliers as follows:

i. Special interest: Barnett (1978) described the interesting legal case of Hudlum versus Hadlum in 1949 as an example.

ii. Detection of specific alternative rare phenomena rather than estimating a common characteristic. Beckman & Cook (1983) gave an example of the changes of radiation level to locate the dropped Russian satellites in the central Canada.

iii. Diagnostic indicator to test the strength and weakness of a model. For example, the data may conform well to the model when they are transformed to the logistic scale as the transformation may lessen the effect of outliers.

iv. Accommodation of outliers to make improvement on modelling and estimation.

v. Identification of influential observations by looking at how outliers affect the estimation of parameters.

The existence of outliers in any data set makes statistical analysis difficult, where the underlying assumptions are subject to breakdown. Anscombe (1960) and Barnett (1978) stated that outliers may reflect: (i) Measurement error, (ii) Inherent variability in the population, or (iii) Execution error.

The following section discusses the development of circular regression model when the response is a circular variable. The third section reviews the outliers in linear and circular univariate samples by presenting some of the popular techniques in the detection of outliers among samples. Some of outliers' detection techniques for linear regression models are considered in the fourth section.

2.2 Circular regression models

Regression analysis is one of the most popular statistical techniques to investigate the relationship between variables. Regression of a linear variable on a set of linear explanatory variables has received wide interest from statisticians and researchers (see Montgomery & Peck (1992), Chatterjee *et al.* (2000)) whereas regression analysis when either response or explanatory variables are circular has been considered only in the past 40 years.

Circular regression is commonly occurs in many areas of application in biology, meteorology, geology and physics. Gould (1969) emphasized the necessity of analyzing circular variables by using techniques different from those appropriate for usual Euclidean type variables because the circumference is a bounded closed space. When the response is a linear variable $\boldsymbol{X}$ and the explanatory is a circular variable Θ, then the mean value of $\boldsymbol{X}$ given $\Theta = \theta$, can be simply obtained by

$$E(\boldsymbol{X} \mid \Theta = \theta) = a_0 + a\cos\theta + b\sin\theta,$$

where a_0, a and b are unknown parameters. This can be fitted by using the classical methods of linear regression models. Laycock (1975) has considered the case involving more than one circular explanatory variable.

In the following subsections we consider two other cases: the first when the response variable is circular and the explanatory variables are linear; and the second, when both response and explanatory variables are circular.

2.2.1 Regression of circular variable on linear variables

In several circumstances, one may be interested in investigating the relationship between circular and linear variables, for example, the relationship between wind direction and its speed; or the direction a has bird flown and the distance moved.

The regression of a circular variable on a set of linear variables was first discussed by Gould (1969). A regression model was proposed to predict the mean direction μ of a circular variable Θ from a set of linear covariates $\boldsymbol{X} = (\boldsymbol{x_1},..., \boldsymbol{x_k})$, where Θ follows von Mises distribution with mean direction μ and concentration parameter κ denoted by $\Theta \sim VM(\mu,\kappa)$. The proposed model is given by

$$\mu = \mu_0 + \sum_{j=1}^{k} \beta_j \boldsymbol{x_j} , \qquad j = 1,\ldots,k \qquad (2.1)$$

where μ_0 and β's are unknown parameters and $\boldsymbol{x}_j$ is a linear covariate. Model (2.1) produces various forms of the so-called "barber's pole" model in which μ conditioned on $\boldsymbol{X} = \boldsymbol{x}$ is a curve winding in an infinite number of spirals up the surface of an infinitely long cylinder. The maximum likelihood estimations (MLE) of model parameters are obtained iteratively. Gould (1969) pointed out that the estimated parameters are the local maxima and may not be the absolute maxima. He also pointed out that, for large concentration parameter κ of the response variable Θ, linear statistics is applied to fit the model from the start.

Analogous to normal theory, Mardia (1972) extended Gould (1969) model by assuming θ_i to be independently distributed as the von Mises distribution with mean μ_i and concentration parameter κ and this model is given by

$$\mu_i = \mu_0 + \beta t_i , \qquad i = 1,\ldots,n , \qquad (2.2)$$

where t_i are known numbers, while μ_0 , β and κ are unknown parameters.

Laycock (1975) discussed Model (2.1) and showed that the maximum likelihood estimates are equivalent to the least squares estimates for large sample size *n*. Moreover, Laycock pointed out that the linear statistics can be applied from the start for large sample size n or large concentration parameter κ.

Johnson & Wehrly (1978) mentioned that the MLE of model (2.1) has infinite many high peaks, which leads to ambiguously defined MLEs. Alternatively, they proposed a different class of models in which the response completes just a single spiral as x increases through its range. Further, for one explanatory variable, they suggested the use of specific model for the joint distribution of the continuous linear variable $\boldsymbol{X}$ and circular variable Θ, with a completely specified marginal distribution of linear variable $\boldsymbol{X}$, $F(x)$. The conditional distribution is given by

$$\Theta \mid \boldsymbol{X} = x \sim VM(\mu + 2\pi F(x), \kappa) .$$

Fisher & Lee (1992) emphasized the necessity to work with von Mises family of distributions because it has a measure of dispersion while for other distributions there is no natural measure of scale. Moreover, the von Mises family shares many of the properties of normal distribution. They extended Johnson & Wehrly (1978) model by assuming that the circular observations $\theta_1,..., \theta_n$ follow von Mises distribution with mean directions $\mu_1,..., \mu_n$ and concentration parameters $\kappa_1,..., \kappa_n$, respectively. They assumed that all of the concentration parameters are equals to κ and the μ's are related to the explanatory variables $\boldsymbol{X}_i$ by means of link function $g(.)$. The model is given by

$$\mu_i = \mu + g(\boldsymbol{\beta}' \boldsymbol{X}_i) , \qquad i = 1,\ldots,n ,$$

where $\boldsymbol{\beta}$ is k- vector of regression coefficients. The function $g(.)$ will map the real line to the circle, for x ranges from $-\infty$ to ∞ and $g(x)$ ranges from $-\pi$ to π and assume that $g(0)=0$. One of the practical possibilities for function $g(.)$ is given by

$$g(x) = 2\tan^{-1}\left(\mathrm{sgn}(x)|x|^{\gamma}\right).$$

The parameter γ can be estimated from the data, analogous to the estimation of Box-Cox transformation in the ordinary linear regression. When $\gamma = 0$, function $g(x)$ corresponds to a log transformation.

2.2.2 Regression of circular variable on circular variables

The first attempt to fit a regression model for circular variable on an explanatory circular variable was made by Laycock (1975) using the complex linear regression, where the model can be expressed as a conventional linear model with complex entries. Laycock (1975) pointed out that the use of his model to predict a pure direction is open to objections.

For response variable $\boldsymbol{Y}$ and explanatory variable $\boldsymbol{X}$ in which both of them follow von Mises with concentration parameter $\kappa \geq 2$, Fisher & Lee (1992) suggested that the problem can be handled satisfactorily by transforming the data to continuous linear variables.

Jammalamadaka & Sarma (1993) proposed a circular model for two circular random variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ in terms of the conditional expectation of the vector $e^{(iy)}$ given x such that

$$E\left(e^{iy} \mid x\right) = \rho(x)e^{i\mu(x)} = q_1(x) + iq_2(x),$$

where $\mu(x)$ is the conditional mean direction of y given x with conditional concentration $0 \le \rho(x) \le 1$. Equivalently, $E(\cos y \mid x) = q_1(x)$ and $E(\sin y \mid x) = q_2(x)$. Then the predicted $\hat{y}$ is obtained by $\mu(x) = \hat{y} = \tan^{-1}\dfrac{q_2(x)}{q_1(x)}$. Due to the difficulty of estimating $q_1(x)$ and $q_2(x)$ they are expressed instead in terms of their Fourier series expansions.

Rivest (1997) proposed a circular–circular regression model to predict the *y*-direction based on the rotation of the decentred *x*-angle. The model is given by

$$y = \theta(x;\beta,\alpha,r) + \varepsilon,$$

$$\theta(x;\beta,\alpha,r) = \beta + \tan^{-1}\{\sin(x-\alpha), r + \cos(x-\alpha)\} \pmod{2\pi},$$

where β and α are angles belonging to $[0,2\pi)$, r is real number and ε has a distribution with mean 0. The parameters are estimated by maximizing the average cosine residuals angles given by

$$\hat{L}(\beta,\alpha,r) = \frac{1}{n}\sum_{i=1}^{n}\cos\{y_i - \theta(x_i;\beta,\alpha,r)\}.$$

Lund (1999) proposed a regression model where the independent variables consist of one circular variable and a set of linear variables. For a circular response $\boldsymbol{Y}$, a circular predictor ϕ and a set of linear covariates $\boldsymbol{X}$, the least circular distance regression model is given by

$$y = \mu(\phi, \boldsymbol{X}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2) + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ are vectors of parameters and $\boldsymbol{\varepsilon}$ is the random circular error with mean direction 0. The parameter estimates are obtained by maximizing the average cosine residuals. The estimates are similar to the maximum likelihood estimates.

Downs & Mardia (2002) described the models which were proposed by Gould (1969), Johnson & Wehrly (1978), and Fisher & Lee (1992) as non-rotational models.

This is because these models use linear combinations of linear concomitant variables, and are relatively difficult to interpret. Moreover, the absence of any topologically appropriate method for angular scale change is another serious shortcoming.

In cases where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are circular variables with mean directions α and β respectively, Downs & Mardia (2002) applied the following mapping

$$\tan\frac{1}{2}(y-\beta)=\omega\tan\frac{1}{2}(x-\alpha),$$

where ω is a slope parameter in the closed interval [−1, 1]. The mapping defines a one-to-one relationship with a unique solution given by

$$y=\beta+2\tan^{-1}\left\{\omega\tan\frac{1}{2}(x-\alpha)\right\}.$$

They classified the regression model according to the nature of the parameters α,β and ω. The maximum likelihood estimates were derived and the properties of the model were discussed with an application to circadian biological rhythms and wind direction data.

Hussin *et al.* (2004) extended model (2.2) for the case when both response and explanatory variables are circular, where the t_i's in (2.2) are considered to be circular while β is an integer. For any circular observations $(x_1,y_1),...,(x_n,y_n)$ of circular variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ with a linear relationship between them, the proposed model is given by

$$y=\alpha+\beta x+\varepsilon \quad (\text{mod}2\pi), \qquad (2.3)$$

where ε is a circular random error having a von Mises distribution with circular mean 0 and concentration parameter κ. One application of model (2.3) is to compare two instruments to measure the wind and wave direction. The maximum likelihood estimations of the model parameters are obtained iteratively. Model (2.3) will be discussed extensively in Chapter 7.

On the other hand, Kato *et al.* (2008) expressed the regression curve as a form of Möbius circle transformation. For an angular variable, $\boldsymbol{Y}$, and angular covariate, $\boldsymbol{X}$, which takes values on the unit circle, $\Omega = \{z \in C; |z| = 1\}$ in the complex plane, they proposed the regression curve to be:

$$y = \beta_0 \frac{x + \beta_1}{1 + \bar{\beta}_1 x} \varepsilon, \quad x \in \Omega,$$

where β_0 and β_1 are complex parameters with $\beta_0 \in \Omega$ and $\beta_1 \in C$. In this case, the angular error ε is assumed to follow wrapped Cauchy distribution, while Downs & Mardia (2002) assumed the angular error to follow von Mises distribution. Due to the attractive properties of wrapped Cauchy distribution, some desirable properties of the model have been derived.

In this study, we consider model (2.3) which is known as the simple circular regression model due to its simpler form compared to other circular regression models and its similarities to the simple linear regression. The adequacy of the model will be investigated and some of the outlier detection techniques will be extended to the circular case.

2.3 Outliers in univariate samples

The literature on the tests of outliers in univariate data is in abundance. Most of these tests are developed for the linear samples, while there are few tests available for circular data. Outliers are also expected to occur in circular data, but the identification method differs from linear case. Collett (1980) illustrated the difference between outlier problem in linear and circular cases by considering the following data set (unit in degree):

$$10, 18, 33, 48, 67, 349. \tag{2.4}$$

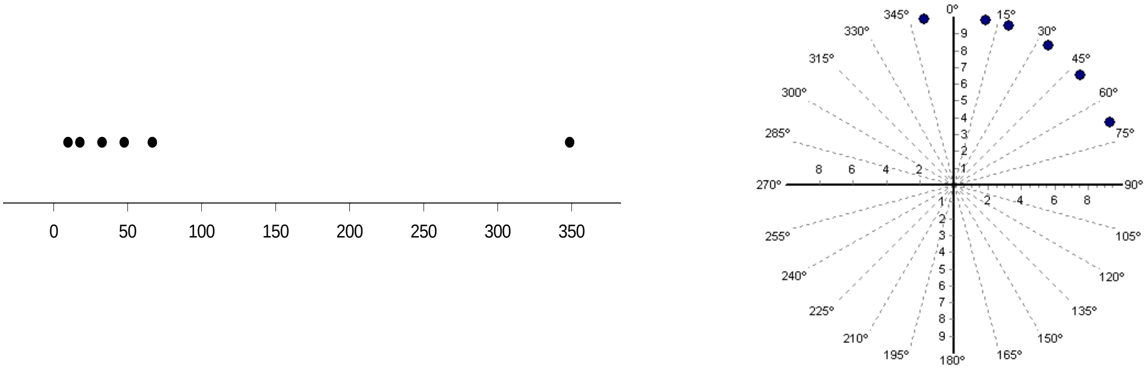


Figure 2.1: Graphical presentation of data in (2.4)

Figure 2.1 shows the plot of data in (2.4) by using linear and circular plots. If we treat the data as linear data, then we can easily identify 349 as an outlier. However, if the data are treated as circular data, then 349 is basically consistent with the rest of the observations and it is no longer an outlier.

The following subsection reviews some of the popular tests of outlier in linear univariate dataset, followed by a review of tests of outlier in circular univariate data.

2.3.1 Outlier identification in linear univariate data set

There are different methods to identify outliers in linear univariate samples, and we review some of them in this subsection.

(i) Boxplot

Boxplot plays an important rule in the exploratory data analysis. It was developed by Tukey (1977) and consists of five-number summaries: the smallest observation, first quartile Q_1, median Q_2, third quartile Q_3, and the largest observation. Boxplot is a popular tool to detect outliers in univariate linear samples based on $1.5 \times IQR$ boxplot criterion, where IQR is the interquartiles range and $IQR = Q_3 - Q_1$. In other words, any observation below $F_L = Q_1 - 1.5 \times IQR$ or above $F_U = Q_3 + 1.5 \times IQR$ is labelled as outlier, where F_L and F_U are called the lower and upper fences, respectively. Further discussion on boxplot and on developing a new boxplot for circular variables is given in Chapter 6.

(ii) The 'three – sigma' rule

Under normality assumptions, an observation x_i can be identified as an outlier if its distance to the sample mean is greater than $3 \times s$, where s is the sample standard deviation given by $s = \sqrt{\sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{n-1}}$, $\bar{x}$ is the sample mean and n is the sample size.

(iii) Dixon test

The Dixon's (Q-test) was developed by Dixon (1950, 1951). It is a simple test to examine if one (and only one) observation from a small set can be "legitimately" rejected or not. The data is ranked in ascending order $x_{(1)},..., x_{(n)}$, from a sample of size n. The statistic τ is a ratio defined as the difference of the suspect value from its nearest one divided by the range of the values. The τ statistics for the highest and the lowest values are computed and compared with critical values, respectively as follows

$$\tau_n = \frac{x_{(n)} - x_{(n-1)}}{x_{(n)} - x_{(1)}} \text{ and } \tau_1 = \frac{x_{(2)} - x_{(1)}}{x_{(n)} - x_{(1)}}.$$

Both tests and their critical values are available in many statistical tables (see Murdoch & Barnes (1998, p.27)). To avoid the problem of two outliers on the same side of the distribution, Dean & Dixon (1951) suggested taking a more elaborate approach by using different formulas for different sample sizes. They defined the various ratios based on sample size *n*. An example is given in Section 4.3.1.

(iv) Maximum normed residual (Grubbs) test

Based on the normality assumptions, Grubbs (1969) proposed a test to detect one outlier at a time in a univariate data set of a size not less than $n = 6$ under the null hypothesis that there are no outliers in the data set. The two sided version of the test statistic is given by

$$G = \frac{\max\limits_{i=1,\ldots,n} |x_i - \bar{x}|}{s},$$

where $\bar{x}$ is the sample mean and s is the sample standard deviation. On the other hand, the one sided Grubbs test is used to examine whether the maximum or the minimum values are outliers and they are given respectively by

$$G_n = \frac{x_{(n)} - \bar{x}}{s} \text{ or } G_1 = \frac{\bar{x} - x_{(1)}}{s},$$

where $x_{(n)}$ and $x_{(1)}$ are the maximum and the minimum values, respectively.

(v) Least Absolute Deviation

Wu & Lee (2006) proposed a least absolute deviation (LAD) method for the determination of the number of upper or lower outliers in normal sample by minimizing its sample mean absolute deviation.

2.3.2 Outlier identification in circular univariate data set

Outlier in the context of circular data would be defined as a set of observations which is inconsistrent with the rest of the sample. It is expected to lie far from the mean direction of the circulae sample. To date, there are a few numerical and graphical tests of discordancy in circular samples. Three of the numerical tests were suggested by Collett (1980) and the remaining one was suggested earlier by Mardia (1975). Another test based on Bayesian methods was suggested by Bagchi & Guttman (1990).

Suppose we have an independent random circular sample $\theta_1,\ldots,\theta_n$ of size n which follows a von Mises distribution, $VM(\mu,\kappa)$, with a density function given by

$$f(\theta;\mu,\kappa)=\frac{1}{2\pi I_0(\kappa)}\exp(\kappa\cos(\theta-\mu)),\quad 0\le\theta,\mu<2\pi,$$

where μ is the mean direction, κ is the concentration parameter with $\kappa\ge 0$, and $I_0(\kappa)$ is the modified Bessel function of first kind and order zero. Detail description of the distribution is given in Section 3.2.2.

To test whether or not a surprising value is an outlier in a von Mises distribution, we consider a null hypothesis that all n observations follow $VM(\mu,\kappa)$ against an alternative hypothesis that $(n-1)$ observations come from $VM(\mu,\kappa)$ and one observation comes from $VM(\mu^*,\kappa)$, where $\mu\ne\mu^*$.

The available numerical tests of discordancy in circular data are presented briefly as follows. Most of the tests use the descriptive measure for circular data; the resultant length for circular sample is given by $R = \sqrt{S^2 + C^2}$, where $S = \sum_{i=1}^{n} \sin\theta_i$ and $C = \sum_{i=1}^{n} \cos\theta_i$, $i = 1,\ldots,n$.

(i) *C* statistic

The mean resultant length of circular data set is given by $\bar{R} = \frac{R}{n}$. By omitting the *i*th observation, the mean resultant length is given by $\bar{R}_{(-i)} = \frac{R_{(-i)}}{(n-1)}$. Therefore,

$$C = \max_i \left\{ \frac{\bar{R}_{(-i)} - \bar{R}}{\bar{R}} \right\}, \tag{2.5}$$

can be considered as a test statistic. Values of C will then be compared with percentage points for different sample size n and estimated concentration parameter $\hat{\kappa}$. If C is larger than the critical value, we reject the null hypothesis so that the *i*th observation can be considered as an outlier.

(ii) *D* statistic

The D statistic uses the relative arc lengths based on ordered observations of a circular sample $\theta_{(1)},\ldots,\theta_{(n)}$. Let T_i be the arc length between consecutive observations given by $T_i = \theta_{(i+1)} - \theta_{(i)}$, $i = 1,\ldots,n-1$ and $T_n = 2\pi - \theta_{(n)} + \theta_{(1)}$. Define $D_i = \frac{T_i}{T_{i-1}}$, $i = 1,\ldots,n$ and $T_0 \equiv T_n$. Let $D_k = \frac{T_k}{T_{k-1}}$ corresponds to the greatest arc containing a single observation θ_k. Since D_k is two tailed, Collett (1980) suggested working in terms of

$$D = \min(D_k, D_k^{-1}), \tag{2.6}$$

where $0 < D < 1$. The observation θ_k can be considered as an outlier if the value of D is larger than the value of percentage points as given in Collett (1980).

(iii) *L* statistic

The *L* test is based on the maximum likelihood ratio statistic for the alternative hypothesis. The *L* statistic is given by

$$L = (R_{(-k)} + 1)\hat{\kappa}_{(-k)} - \hat{\kappa}R - n\ln\left\{\frac{I_0(\hat{\kappa}_{(-k)})}{I_0(\hat{\kappa})}\right\}, \tag{2.7}$$

where $R^2_{(-k)} = C^2_{(-k)} + S^2_{(-k)}$, $C_{(-k)}$ and $S_{(-k)}$ are the values of C and *S*, respectively, based on $(n-1)$ observations excluding θ_k ; $\hat{\kappa}$ is the maximum likelihood estimator of κ based on n observations and $\hat{\kappa}_{(-k)}$ is the maximum likelihood estimator of κ based on $(n-1)$ observations excluding θ_k .

(iv) *M* statistic

Mardia (1975) proposed a statistic of discordancy which is given by $M' = \min_i\left\{\frac{n-1-R_{(-i)}}{n-R}\right\}$. Later, Collett (1980) reformulated the M' statistic in terms of

$$M = 1 - M' = \max_i\left\{\frac{R_{(-i)} - R + 1}{n-R}\right\} = \frac{R_k - R + 1}{n-R}, \tag{2.8}$$

where $R_k = \max_i\{R_{(-i)}\}$. Collett stated the asymptotic distribution of M statistic for large value of κ. As the value of κ increases the von Mises distribution will be approximated by a simple normal distribution (see Jammalamadaka & SenGupta (2001)). On the other hand, the statistic M can be approximated by $\frac{n(b^*)^2}{(n-1)}$, where $b^* = \max_i\left\{\frac{|x_i - \bar{x}|}{\sum (x_i - \bar{x})^2}\right\}$ is the test statistic used to identify discordancy in normal data. Percentage points for b^* are given in Pearson and Hartley (1966).

The C and D statistics can be used for any circular sample, while L and M are useful only for circular samples from von Mises distribution. Collett (1980) provided the percentage points of the null distribution of the C and D statistics for various values of κ and provided the asymptotic distribution of M statistic. The distribution of the L statistic is unknown. Further, among the available tests for the detection of outliers in circular data, the most appropriate test is Mardia's test, since his test is independent of the concentration parameter of the von Mises distribution (see Upton (1993)). Alternative statistics to identify outliers in circular samples based on the circular distance or chord lengths are discussed in detail Chapter 5.

On the other hand, Jammalamadaka & SenGupta (2001) suggested several graphical techniques to explore circular samples which are summarized below.

(i) Circular distance between circular sample observations

The circular distance between any two points is taken to be the smaller of the two arc lengths between the two points along the circumferences. For any two angles ϕ and θ, the circular distance is defined as

$$d_{\circ}(\phi,\theta)=\min(\phi-\theta,2\pi-(\phi-\theta))=\pi-\left|\pi-|\phi-\theta|\right|. \tag{2.9}$$

If the circular distances between observation θ_i and its neighbours on both sides are relatively larger than the distance between other successive observations, then θ_i is considered as an outlier.

(ii) P-P plot

P-P plot can be obtained by finding the best fitting of cumulative von Mises distribution $\hat{F}(\theta;\hat{\mu},\hat{\kappa})$ for the circular sample. Then the plot is obtained by plotting the pairs of $\left(i/(n+1), \hat{F}(\theta_{(i)};\hat{\mu},\hat{\kappa})\right)$, $i=1,\ldots,n$, where n is the sample size. Any point in P-P plot that seems not to be close enough to the diagonal line is suspected to be outliers.

(iii) Q-Q plot

Q-Q plot is obtained by plotting $(\sin(q_i/2), z_{(i)})$, where $q_i = F^{-1}\left(i/(n+1);0,\hat{\kappa}\right)$ and $z_i = \dfrac{\sin(\theta_i - \hat{\mu})}{2}$, $i=1,\ldots,n$, where $z_{(1)},\ldots,z_{(n)}$ are the ordered values of z_i's. Any points in Q-Q plot far from the diagonal line are candidate to be outliers.

(iv) Spoke plot

The spoke plot is introduced by Hussain *et al.* (2007). It consists of inner and outer rings in which lines are used to connect the pair of points (θ_i, ϕ_i) between the two directional variables $0^\circ \le \theta_i, \phi_i < 360^\circ$. The lesser number of lines crossing the inner ring indicates higher correlation.

In this thesis, alternative graphical tools analogue to the linear graphics will be developed to identify outliers in a circular sample. Further, they will be extended to detect possible outliers in circular regression based on circular residuals in Chapter 7.

2.4 Outliers and influential observations in regression models

The analysis of regression is subjected to the occurrence of the outliers. Barnett & Lewis (1984) and Belsley *et al.* (1980) discussed extensively on outliers and diagnostics checking for linear regression. Outliers which change the values of statistics of interest such as parameters estimates or variance of residuals are known as influential observations. However, there is no known published work discussing the problem of outliers in regression models for circular variables. A brief overview on influential observations in linear regression analysis is given in the following Section.

2.4.1 Outliers and influential observations in linear regression

Methods described in previous section can be used to detect outliers in regression modles using their residuals. This section reviews some of the techniques used to identify influential observations in linear regression based on row deletion approach which was developed by Belsley *et al.* (1980). It investigates the impact of deleting one row at a time from both $\boldsymbol{X}$ matrix and $\boldsymbol{Y}$ vector on the fitted values, residuals and the estimated parameter. Here, the approaches are reviewed for two reasons: The first is to use them to identify influential observations in circular regression after approximating the circular variables into linear variable. The second is to extend some of these techniques to the circular regression case.

Regression analysis is concerned with fitting models to data in which there is a single continuous response variable whose expected value depends on the values of the explanatory variables. Linear regression model is summarized by

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{Y}$ is n–vector of response, $\boldsymbol{X}$ is $n \times p$ full rank matrix of known constants, $\boldsymbol{\beta}$ is p-vector of unknown parameters and $\boldsymbol{\varepsilon}$ is n-vector of errors with the assumptions that $E(\boldsymbol{\varepsilon}) = \boldsymbol{0}$ and $V(\boldsymbol{\varepsilon}) = \sigma^2 \boldsymbol{I}_n$. The least square estimation of $\boldsymbol{\beta}$ is given by $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$, where $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $\text{cov}(\widehat{\boldsymbol{\beta}}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$. The residual sum of squares about the fitted model is given by $RSS = (\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})'(\boldsymbol{Y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$, the least squares estimator of σ^2 is an unbiased estimator and defined by $s^2 = RSS/(n-p)$. The ordinary residual vector is defined as

$$\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = (1 - \boldsymbol{H})\boldsymbol{Y},$$

where $\hat{\boldsymbol{Y}}$ is the vector of the fitted values and $\boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ is the hat matrix which is a symmetric and idempotent matrix. The matrix $\boldsymbol{H}$ contains the information on the influence of the response value $\boldsymbol{Y}_i$ on the corresponding fitted value $\hat{\boldsymbol{Y}}_i = \boldsymbol{H}_i'\boldsymbol{Y}$, where $\boldsymbol{H}_i'$ is the *i*th row of matrix $\boldsymbol{H}$. The h_{ii} is the diagonal elements of the hat matrix $\boldsymbol{H}$. Huber (1981) suggested that h_{ii} with values less than 0.2 appearing to be safe, values between 0.2 and 0.5 as being risky and values greater than 0.5, if possible, be avoided by the control of design matrix. Belsley *et al.* (1980) suggested an approximation cut-off value at 0.05 level of significant to be $(2p/n)$, where p is the number of model coefficients. Entries h_{ii} are used to identify leverage points.

Next, we look at different methods of measuring the effect of deleting one row on the estimation of parameters and their covariance, residual sum of squares, fitted values and their variances.

(i) Influence of the *i*th observation on $\boldsymbol{\beta}$ and its covariance

Let $\widehat{\boldsymbol{\beta}}_{(-i)}$ be the least square estimate of $\boldsymbol{\beta}$ when the *i*th observation is deleted. Then

$$\widehat{\boldsymbol{\beta}}_{(-i)} = (\boldsymbol{X}'_{(-\mathrm{i})}\boldsymbol{X}_{(-\mathrm{i})})^{-1}\boldsymbol{X}'_{(-\mathrm{i})}\boldsymbol{Y}_{(-\mathrm{i})}$$

where $\boldsymbol{X}_{(-i)}$ and $\boldsymbol{Y}_{(-i)}$ are obtained by removing the *i*th row in $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. The change in the estimate of the parameter vector $\boldsymbol{\beta}$ when the *i*th observation is deleted is given by

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(-i)} = \frac{(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_i' e_i}{1-h_{ii}} \quad .$$

where $\boldsymbol{X}_i$ is the ith row of $\boldsymbol{X}$ matrix. Cook (1977,1979) considered a statistic based on the confidence ellipsoids for investigating the contribution of each data point i to the least squares estimate of the parameter, $\boldsymbol{\beta}$, which is given by

$$\frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})' \boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})}{ps^2} \sim F_{p.n-p} \,.$$

In order to determine the degree of influence of the ith data point on the estimated parameter vector, $\boldsymbol{\beta}$, Cook suggested the measure of the critical nature of each data point to be

$$D_{(-i)} = \frac{(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{(-i)})' \boldsymbol{X}'\boldsymbol{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_{(-i)})}{ps^2}$$

$$= \frac{e_i^2}{ps^2}\left\{\frac{h_{ii}}{(1-h_{ii})^2}\right\}.$$

A large value of $D_{(-i)}$ indicates that the associated observation has a strong influence on the estimate of parameter vector $\hat{\boldsymbol{\beta}}$.

One of the diagnostic techniques of row deletion is to compare the estimated covariance matrix of $\hat{\boldsymbol{\beta}}$ using all available data, $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, with the estimated covariance matrix that results in deleting the ith observation, $\sigma^2(\boldsymbol{X}_{(-\mathrm{i})}'\boldsymbol{X}_{(-\mathrm{i})})^{-1}$. Belsley *et al.* (1980) suggested the comparison of the two matrices using determinantal ratio which is given by

$$COVRATIO_{(-i)} = \frac{\det\{s_{(-i)}^2[\boldsymbol{X}_{(-\mathrm{i})}'\boldsymbol{X}_{(-\mathrm{i})}]^{-1}\}}{\det\{s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\}}$$

$$= \left(\frac{s_{(-i)}}{s}\right)^{2p} \frac{1}{1-h_{ii}} \,.$$

A value of $COVRATIO_{(-i)}$ which is not near unity indicates that the ith observation is possibly influential. Hence, any data point with $|COVRATIO_{(-i)} - 1|$ close to or larger than $(3p/n)$ is possibly an influential observation.

(ii) Influence of the *i*th observation on the residuals sum of squares

Let $RSS = \sum e_i^2$ be the residual sum of squares after fitting the model using all observations, then

$$RSS = (n-p)s^2$$

$$= \boldsymbol{Y}'\boldsymbol{Y} - \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{Y}\ .$$

Similarly, after the deletion of the ith observation it becomes

$$RSS_{(-i)} = (n-p-1)s_{(-i)}^2 = (\boldsymbol{Y}'\boldsymbol{Y} - \boldsymbol{Y}_i^2) - \hat{\boldsymbol{\beta}}_{(-i)}'(\boldsymbol{X}'\boldsymbol{Y} - \boldsymbol{X}_i\boldsymbol{Y}_i)\ .$$

Hence,

$$(n-p-1)s_{(-i)}^2 = (n-p)s^2 - \frac{e_i^2}{1-h_{ii}}\ .$$

Goldsmith & Boddy (1973) and Mickey (1974) suggested that if deleting the ith observation results in the largest reduction of the residual sum of squares, then the ith point is most likely to be an influential observation. Belsley *et al.* (1980) formalized this into the following statistic

$$RESRATIO_{(-i)} = \frac{(RSS - RSS_{(-i)}) / \left[(n-p)-(n-p-1)\right]}{RSS_{(-i)} / (n-p-1)}.$$

$$= \frac{e_i^2}{s_{(-i)}^2(1-h_{ii})}\ .$$

Under the normality assumptions, the statistic RESRATIO follows F distribution with 1 and $(n-p-1)$ degrees of freedom.

(iii) Influence of deleting the *i*th observation on $\hat{Y}_i$ and its variance

Belsley *et al.* (1980) defined the change in fitted value for the *i*th row that results from deleting the *i*th observation as

$$DFFIT_i = \hat{\boldsymbol{Y}}_i - \hat{\boldsymbol{Y}}_i(-i) = \sqrt{\boldsymbol{V}_{ii}}\,\frac{e_i}{s_{(-i)}(1-h_{ii})}.$$

Belsely *et al.* (1980) suggested that a convenient size adjusted cut off for $|DFFIT_i|$ is $2\sqrt{p/n}$. The estimated variance of fitted value $\hat{\boldsymbol{Y}}_i$ when the fit is based on all observations is compared with the estimated variance when the *i*th data point is deleted.

$$\hat{V}(\hat{\boldsymbol{Y}}_i) = \hat{V}(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}}) = s^2 h_{ii}.$$

Similarly, $$\hat{V}(\hat{\boldsymbol{Y}}_i(-i)) = \hat{V}(\boldsymbol{X}_i^{'}\hat{\boldsymbol{\beta}}(-i)) = \frac{s_{(-i)}^2 h_{ii}}{1-h_{ii}}.$$

Thus, the ratio of the estimated variance is given by

$$FVARATIO = \frac{s_{(-i)}^2}{s^2(1-h_{ii})}.$$

FVARATIO provides a useful summary of changes that occur in the precision of the fitted values of the *i*th observation when the *i*th observation is deleted.

2.4.3 Outliers and influential observations in circular regression

There has been no published work related to the outliers and influential observations in the regression for circular variables or circular regression models. However, there is a specific discussion done on the diagnostics checking for certain circular regression models.

Fisher & Lee (1992) discussed the diagnostics checking for their proposed model. In the example of the distance and direction moved by small blue periwinkles, they used some diagnostics plot for residuals direction like the plot of residuals direction against the distance moved and von Mises Q-Q plot.

Lund (1999) used the von Mises Q-Q plot and proposed the Akaike information criterion (*AICC*) statistic by assuming that the error has a von Mises distribution with concentration parameter κ. The model with minimum *AICC* is deemed to be the best fit. Moreover, he assessed the goodness of fit by using $A(\hat{\kappa})$ given by the equation

$$A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\cos[y_i - \mu(\phi_i, \boldsymbol{X}_i, \hat{\beta}_1, \hat{\beta}_2)],$$

as an analogue of residual sums of squares in linear regression (*RSS*). Further, Lund (1999) touched on the available circular correlation measures (see Fisher (1993) and Mardia & Jupp (2000)) which could be applied to the observed and fitted values. Consequently, squaring these measures gives an analog to the coefficient of determination, R^2 for linear regression. For a random sample $(x_1, y_1),\ldots,(x_n, y_n)$, the simplest measure proposed by Jammalamadaka & Sarma (1988) and is given by

$$r_c = \frac{\sum_{i=1}^{n}\sin(x_i - \bar{x})\sin(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}\sin^2(x_i - \bar{x})\sin^2(y_i - \bar{y})}},$$

where $\bar{x}$ and $\bar{y}$ are the sample mean directions.

Due to the lack of tests to detect outliers or influential observations in circular data, there is a need to propose new tests statistics to deal with this problem. The following chapter discuses some of the summary statistics of circular data and describes the two circular data sets which are going to be used for the purpose of illustration in this study.

2.5 Summary

Like other type of data, circular data are subjected to contamination by outliers and influential observations. There is a need to come out with new outlier detection tests and influential observation detection procedures in both univariate circular samples and circular regression. In order to do so some approaches may be developed either by using circle properties (i.e. chord and circular distance), or extending some of popular graphical tools like boxplot, or employing row deletion approach to the circular regression case. These will be the focus of our study.

CHAPTER THREE

DESCRIPTIVE STATISTICS OF CIRCULAR DATA

3.1 Introduction

This chapter reviews some of the descriptive measures for circular data, and also describes the widely used circular distribution which is the von Mises distribution as can be seen in Chapter 2. In this study, we consider two circular data sets. The first is the direction of frogs, which was considered by Collett (1980). The next is the wind direction data modelled by Hussin *et al.* (2004).

3.2 Descriptive measures for circular data

In order to assess the main characteristics of any circular data set, we need some measures of location and dispersion, for example: mean direction, variance etc. Let $\theta_1,\ldots,\theta_n$ be observations in a random circular sample of size n from a circular population. The descriptive measures can be described as follows:

(i) The mean direction

To find the mean direction of circular random sample, we consider each observation to be a unit vector whose direction is specified by the circular angle and find their resultant vector. The mean direction is defined by the angle made by the resultant vector with horizontal line. Specifically, we have the resultant length R given by

$$R = \sqrt{C^2 + S^2} \ ,$$

where $C=\sum_{i=1}^{n}\cos\theta_i$ and $S=\sum_{i=1}^{n}\sin\theta_i$. The mean direction, $\bar{\theta}$, may be obtained by solving the equations, $\cos\bar{\theta}=\frac{C}{R}$ and $\sin\bar{\theta}=\frac{S}{R}$, where

$$\bar{\theta}=\begin{cases}\tan^{-1}(S/C), & \text{if} \quad S\geq 0, C>0,\\ \frac{\pi}{2}, & \text{if} \quad S>0, C=0,\\ \tan^{-1}(S/C)+\pi, & \text{if} \quad C<0,\\ \tan^{-1}(S/C)+2\pi, & \text{if} \quad S<0, C\geq 0,\\ \text{undefined}, & \text{if} \quad S=0, C=0.\end{cases}$$

One of the mean direction characteristics is that $\sum_{i=1}^{n}\sin(\theta_i-\bar{\theta})=0$, which is analogous to the linear case. Mean direction is sometimes called the "preferred direction".

(ii) Mean resultant length

Mean resultant length $\bar{R}$ is useful for unimodal data to measure how concentrated the data is towards the centre. It is defined by $\bar{R}=\frac{R}{n}$ and lies in the range $[0,1]$. When $\bar{R}$ is close to 1, it implies that all directions in the data set are similar, or the set of observation has a small dispersion and is more concentrated towards the centre. However, $\bar{R}=0$ does not imply that the directions are spread almost evenly around the circle, for example, any data set of the form $\theta_1,...,\theta_n,\theta_1+\pi,...,\theta_n+\pi$ has $\bar{R}\approx 0$.

(iii) The median direction

Fisher (1993) defined the median direction of circular variable as an axis (median axis) that divides the circular data into two equal groups. Consistently, Mardia & Jupp (2000) defined the median as any point ϕ, where half of the data lie in the arc $[\phi,\phi+\pi)$ and the other points are nearer to ϕ than to $\phi+\pi$. Practically, for any circular sample, Fisher (1993) defined the median direction as the observation ϕ which minimizes the summation of circular distances (2.9) to all observations, $d(\phi) = \pi - \sum_{i=1}^{n} |\pi - |\theta_i - \phi||$ for $i = 1,\ldots,n.$ Fisher's definition is used to obtain the circular median in the Oriana statistical software package.

It is interesting to note the robustness of mean direction as stated by Wehrly & Shine (1981). According to Mardia & Jupp (2000), the robustness property is due to the compactness of the circle.

(iv) The sample circular variance

The sample circular variance is defined by the quantity $V = 1 - \bar{R}$, where $0 \le V \le 1$. The smaller values of circular variance refer to a more concentrated sample.

(v) The sample circular standard deviation

The quantity $\mathrm{v} = \sqrt{-2\log(1-V)}$ is defined as the sample circular standard deviation with $0 < \mathrm{v} < \infty$, where V is the sample circular variance. The reason for defining the circular standard deviation in this way, rather than as the square root of the sample circular variance is to obtain some reasonable approximations for proportion of von Mises distribution, provided that the distribution is not too dispersed (see Fisher 1993, p.54).

3.3 The von Mises distribution

The von Mises distribution is introduced by von Mises (1918) to study the deviations of measured atomic weight from integral values. It is the most common distribution considered for unimodal samples of circular data. The von Mises distribution has been extensively discussed where many inference techniques have been developed. It is denoted by $VM(\mu,\kappa)$, where μ is the mean direction and κ is the concentration parameter. The probability density function for the von Mises distribution is given by

$$f(\theta;\mu,\kappa)=\frac{1}{2\pi\, I_0(\kappa)}\exp\{\kappa\cos(\theta-\mu)\}, \qquad 0<\theta,\mu\le 2\pi \text{ and } \kappa\ge 0,$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero, and it is given by

$$I_0(\kappa)=\frac{1}{2\pi}\int_0^{2\pi}\exp(\kappa\cos\theta)\,d\theta=\sum_{r=0}^{\infty}\left(\frac{\kappa}{2}\right)^{2r}\left(\frac{1}{r!}\right)^{2}.$$

Jammalamadaka & SenGupta (2001) summarized some of von Mises density properties, which are

(i) it is symmetrical about the mean direction μ,

(ii) it has a mode at μ, and

(iii) it has antimode at $(\mu\pm\pi)$.

Best & Fisher (1981) gave the maximum likelihood estimates of the concentration parameter κ as follows:

$$\hat{\kappa}=\begin{cases}2\bar{R}+\bar{R}^3+\frac{5}{6}\bar{R}^5, & \text{if} \quad \bar{R}<0.53,\\ -0.4+1.39\bar{R}+\frac{0.43}{(1-\bar{R})}, & \text{if} \quad 0.53\le\bar{R}<0.85,\\ (\bar{R}^3-4\bar{R}^2+3\bar{R})^{-1}, & \text{if} \quad \bar{R}\ge 0.85.\end{cases}$$

Further, the estimation of the concentration parameter depends on the sample size.

Watson (1961) proposed a test for goodness of fit U_n^2 of any given circular distribution (i.e. von Mises distribution). Let $F_\kappa(\theta)$ be the distribution function of the von Mises distribution which is given by $z_i = F_\kappa(\theta_i) = \{2\pi I_0(\kappa)\}^{-1} \int_0^{\theta_i} \exp(\kappa \cos\phi)\, d\phi$.

The test statistic U_n^2 is given by

$$U_n^2 = \sum_{i=1}^{n} z_i^2 - \sum_{i=1}^{n} \left(\frac{c_i z_i}{n} \right) + n \left[\frac{1}{3} - \left(\bar{z} - \frac{1}{2} \right)^2 \right]. \tag{3.1}$$

where $\bar{z} = \frac{1}{n} \sum_{i=1}^{n} z_i$ and $c_i = 2i - 1$. The critical values were supplied by Stephens (1964).

3.4 Frogs direction data

The data set have been selected from a series of experiments conducted to investigate the homing ability of the northern cricket frogs described by Ferguson *et al.* (1967). A total of 14 frogs were collected from the mud flats of an abandoned stream meander near Indianola, Mississippi. After 30 hours enclosure within a dark environmental chamber, the frogs were released and the directions taken were recorded as given below:

$$104^\circ, 110^\circ, 117^\circ, 121^\circ, 127^\circ, 130^\circ, 136^\circ, 145^\circ, 152^\circ, 178^\circ, 184^\circ, 192^\circ, 200^\circ, 316^\circ.$$

Some of the descriptive statistics for the data are given in Table 3.1. The mean and median directions are close to each other which is around 145°. The data seem to be not highly concentrated where the estimate of the concentration parameter is around 2. Figure 3.1 shows the frogs data distributed on the circumference of the circle. The plot of circular histogram for frogs direction data is displayed in Figure 3.2. The solid straight line shows the mean direction and the arched line indicates the 95% confidence interval of mean direction. The data have one mode which is close to the mean direction while there is one bar directed to the opposite of mean direction which is suspected to be an outlier.

Table 3.1: Some descriptive measures for frogs direction data

Descriptive measure	value
Mean direction, $\bar{\theta}$	145.974°
Median, ϕ	145°
Mean resultant length, $\bar{R}$	0.725
Variance, $\boldsymbol{V}$	0.274
Standard deviation, $\hat{\text{v}}$	45.931°
Concentration parameter, $\hat{\kappa}$	2.18

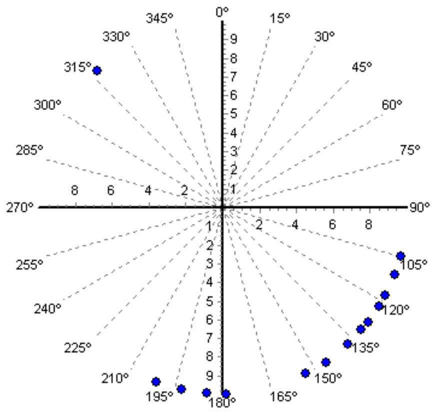


Figure 3.1: Circular plot of frogs data

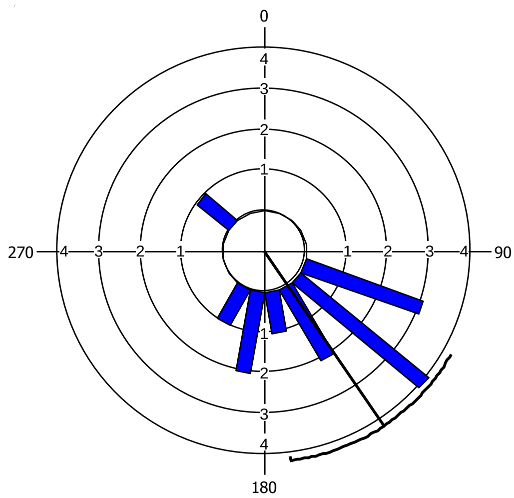


Figure 3.2: Circular histogram for frogs data

Figure 3.3 shows the Q-Q plot for the frogs data. The quantiles are close to straight line except one point. To test the goodness-of-fit, U_n^2, statistic given by (3.1) is used. The test statistic $U_n^2 = 0.066$, is smaller than the critical value 0.101 at 0.05 significant level. Thus, we may conclude that the frogs data follows von Mises distribution.

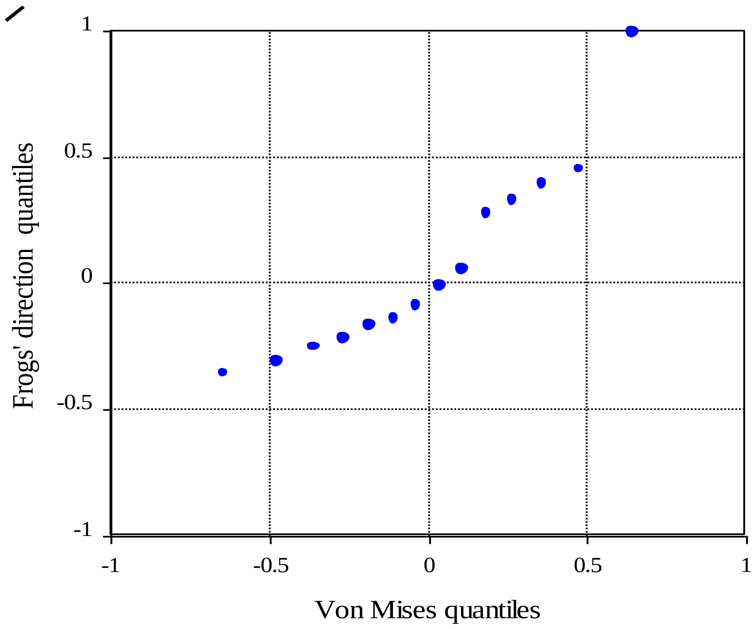


Figure 3.3: Q-Q plot for frogs data

3.5 Wind direction data

In this section we introduce briefly the HF (High frequency) radar system and anchored wave buoy techniques for measuring the ocean wind direction followed by a description of the data. The interest here is to identify the outliers in the data.

(i) The HF radar system

HF radar is a tool for synoptic on-line mapping of surface current fields and the spatial distribution of the wave directional spectrum. The HF radar used to collect the data was developed by UK Rutherford and Appleton Laboratories and subsequently by Marex Ltd and The Marconi Radar Company. The system is pulse radar that uses high frequency (24.4-27 MHz) radio frequency to map surface current patterns over a large area of ocean.

(ii) The Anchored wave buoy

It is often used to evaluate the performance of other wind or wave measuring systems. Older models measure vertical motion at a single point. Typical wave buoys additionally measure the lope of the sea surface in two directions at the same points.

3.5.1 Data description

The data were collected along the Holderness coastline (the Humberside coast of the North Sea, United Kingdom) by using HF radar system and wave buoys. The deployment began in October 1994, as a part of an experiment studying the transport of sediment away from the coast. The following information is assumed:

(i) There is temporal stationary over the period of measurements,

(ii) There is spatial stationary over the area of measurements.

(iii) The different techniques are measured independently.

The wind direction is the direction of the local wind which blows across the sea surface and along the coast where the HF radar system and anchored wave buoys are deployed. The full data set is quoted from Hussin (1997) and is given in Appendix (A.1) which consists of time (in days) when the data and the directions (in radians) were recorded. There were 129 measurements recorded by HF radar and anchored wave buoy respectively over the period of 22.7 days.

3.5.2 Descriptive statistics

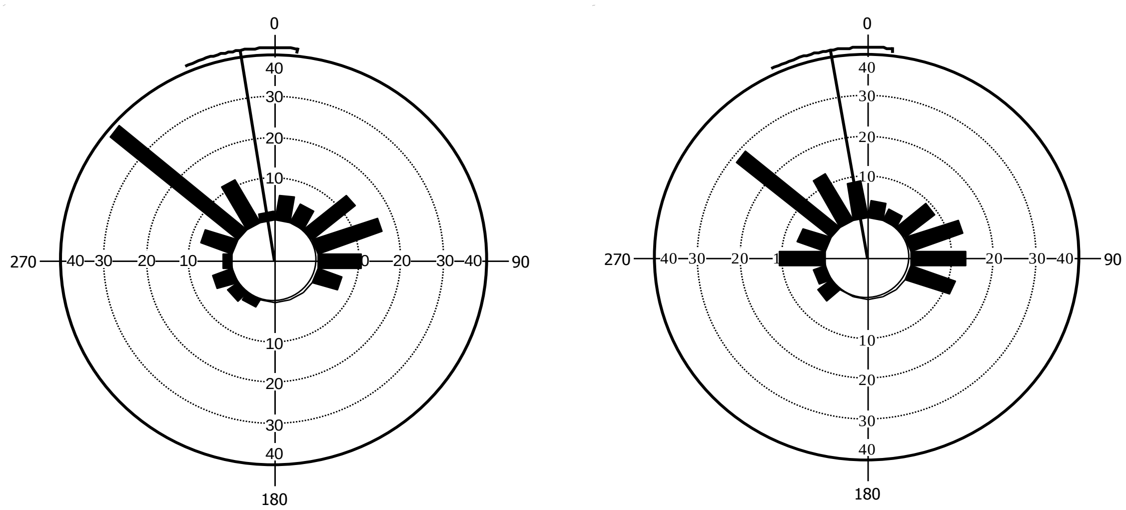
Some of the descriptive statistics for wind direction data are given in Tables 3.2 (in radians). Both summary statistics for the measurements which are recorded by HF radar and anchored wave buoy are almost similar. The concentration is small and the mean resultant length is around 0.4.

Table 3.2: Some descriptive measures for wind direction data

Descriptive Measure	HF radar	Anchored buoy
Mean direction, $\bar{\theta}$	6.127	6.116
Median, ϕ	5.713	5.842
Mean resultant length, $\bar{R}$	0.444	0.411
Variance, $\boldsymbol{V}$	0.556	0.588
Standard deviation, $\hat{v}$	1.274	1.333
Concentration parameter, $\hat{\kappa}$	0.999	0.902

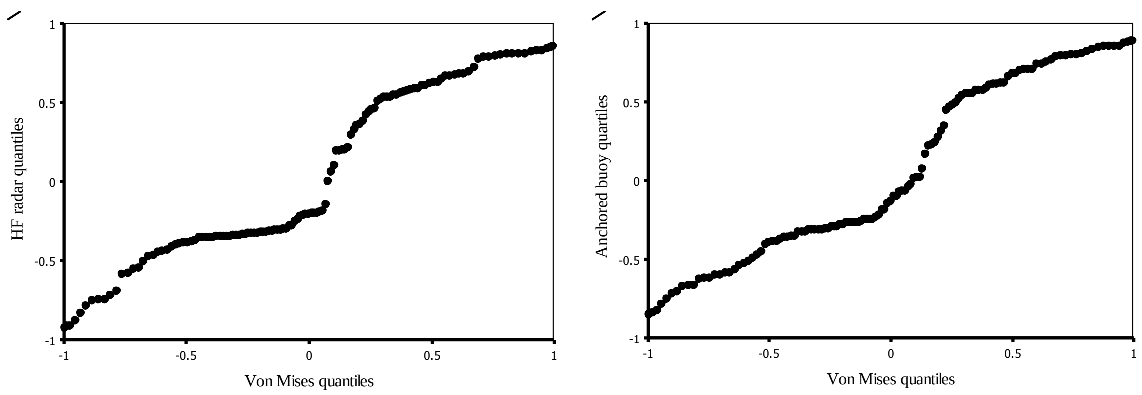
Figure 3.4 shows the plot of simple circular histograms for direction data measured by HF Radar and Anchored buoy. It seems that the two measurements have similar distribution. Figure 3.5 shows the Q-Q plot for HF radar and anchored buoy directions data, where most of the quantiles are close to the straight line. This suggests that the data follow von Mises distribution. Figures 3.4 and 3.5 indicate the closeness between the measurements of the two techniques.

Figure 3.6 shows the ordinary scatter plot of wind direction data. The measurements are in radians where the scale is artificially broken at 0 (or equivalently 2π) radians. Two points seem to be outliers at the left top of the plot. However, they are actually consistent with the rest of the observations as they are close to other observations at the right top or left bottom due to the closed range property of the circular variable. Further, there is a linear relationship between HF radar system and anchored wave buoy measurements. Figure 3.7 shows the spoke plot of wind direction data. The inner ring represents the measurements by HF radar while the outer ring represents the measurements of the anchored buoy. Since almost all the lines do not cross the inner circle it means that the data are highly correlated with estimated correlation parameter $\hat{r}_c = 0.952$. Further, there are only two lines crossing the inner ring, which are associated with observations number 38 and 111. This indicates that the pairs corresponding to the two observations are inconsistent.



(a) Measurements by HF radar (b) Measurements by anchored buoy

Figure 3.4: Circular histograms for wind direction data



(a) Measurements by HF radar (b) Measurements by anchored buoy

Figure 3.5: Q-Q plot for wind direction data

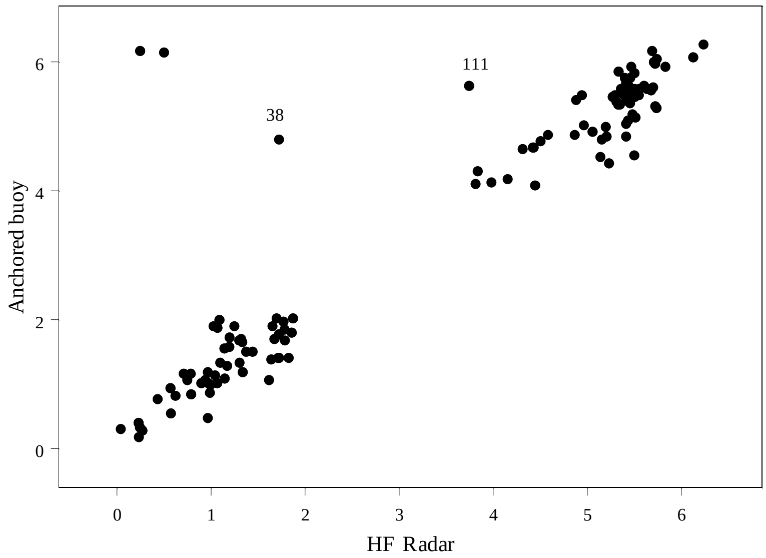


Figure 3.6: Scatter plot of wind direction data measured by both techniques

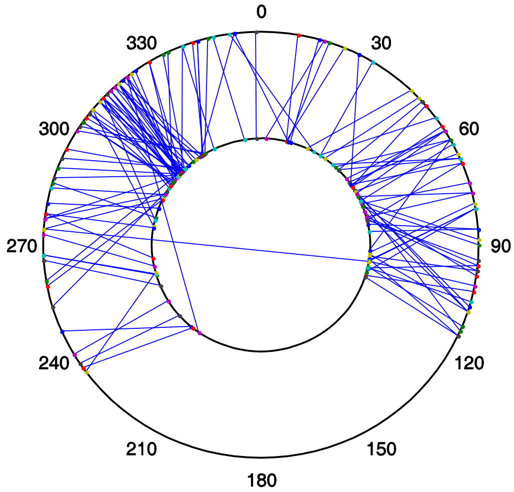


Figure 3.7: Spoke plot of wind direction data measured by both techniques

3.6 Summary

Two data sets were described in this chapter and will be used for illustration purposes through out this thesis. Based on the explanatory analysis, we expect to identify one outlier in frogs data and two outliers in the wind direction data.

CHAPTER FOUR

LINEAR APPROXIMATION OF CIRCULAR VARIABLES AND DETECTION OF OUTLIERS

4.1 Introduction

The approximation of circular variables by linear variables enables us to use linear statistical tests. Several authors had state that a sample from the von Mises distribution with large concentration parameter κ can be treated as linear sample (see Mardia (1975), Fisher & Lee (1992) and Jammalamadaka & SenGupta (2001)).

This chapter reviews one of the important theorems in this aspect and discusses how large the concentration parameter κ should be in order to approximate circular data to linear. Two circular data sets (frogs and wind direction data) are considered. Consequently, we apply the linear outlier detection procedures on both sets of data.

4.2 The approximate distribution for von Mises samples with large concentration parameter

This section reviews the proof of the approximate distribution for von Mises samples with large concentration parameter κ and discusses the value of the concentration parameter in order to approximate von Mises samples by normal distribution.

4.2.1 The proof of the approximate distribution

Jammalamadaka & SenGupta (2001, p.41) stated the proof that for a circular random variable θ from a von Mises distribution with mean direction μ and concentration parameter κ, that is $\theta \sim VM(\mu,\kappa)$, $\beta = \sqrt{\kappa}(\theta-\mu)$ can be approximated by standard normal distribution as $\kappa \to \infty$, that is,

$$\beta = \sqrt{\kappa}(\theta-\mu) \sim N(0,1). \tag{4.1}$$

The proof of (4.1) is given below. The von Mises density function is given by

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi I_0(\kappa)}\exp\{\kappa\cos(\theta-\mu)\}, \qquad 0<\theta,\mu \le 2\pi,\ \kappa \ge 0.$$

For large concentration parameter κ, the modified Bessel function of the first kind and order zero can be estimated by $I_0(\kappa) \approx \frac{\exp(\kappa)}{\sqrt{2\pi\kappa}}$.

Let $\beta = \sqrt{\kappa}(\theta-\mu)$, then $\frac{\beta}{\sqrt{\kappa}} = (\theta-\mu)$ and let $\delta(\beta) = \mu + \frac{\beta}{\sqrt{\kappa}}$. Using the change-of-variable technique

$$g(\beta) = \left|\frac{\partial\theta}{\partial\beta}\right| f(\delta(\beta)),$$

$$= \frac{1}{\sqrt{\kappa}}\frac{\exp\left(\kappa\cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{2\pi I_0(\kappa)}.$$

By substituting the approximated value of $I_0(\kappa)$, we get

$$g(\beta) = \frac{\exp\left(\kappa\cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{2\pi\frac{\exp(\kappa)}{\sqrt{2\pi\kappa}}\sqrt{\kappa}}$$

$$= \frac{\exp\left(\kappa\cos\left(\frac{\beta}{\sqrt{\kappa}}\right)\right)}{\sqrt{2\pi}\exp(\kappa)}.$$

For large κ and hence small $\frac{\beta}{\sqrt{\kappa}}$, $\cos(\theta-\mu)=\cos\left(\frac{\beta}{\sqrt{\kappa}}\right)$. From the Taylor series expansion, we know that $\cos\theta \approx 1-\frac{\theta^2}{2}$. Hence,

$$\cos\left(\frac{\beta}{\sqrt{\kappa}}\right) \approx 1-\frac{\beta^2}{2\kappa}.$$

Therefore,

$$g(\beta) \approx \frac{\exp\left(\kappa\left(1-\frac{\beta^2}{2\kappa}\right)\right)}{\sqrt{2\pi}\exp(\kappa)},$$

$$=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{\beta^2}{2}\right).$$

This is the probability density function of the standard normal distribution.

□

In the following subsection we discuss how large the concentration parameter of von Mises sample should be for it to be approximated by the normal distribution.

4.2.2 The size of the concentration parameter

We use simulation to define how large the concentration parameter κ should be in order to be approximated by the normal distribution. A total of 13 different sample sizes $n=5$, 10(10)100, 120 and 150 and twenty one values of concentration parameter, $\kappa =$ 0.001, 0.01 0.1, 0.5, 1, 1.5, 2(1)10, 12, 15, 17, 20 are considered. Simulation studies are carried out with a fixed mean direction at $\pi/4$ radians as the value of β in (4.1) does not depend on the mean direction. For each combination of sample size n and concentration parameter κ, 3000 samples are generated from the von Mises distribution. For each generated sample, β is calculated. The Kolmogorov-Smirnov test is used to assess the goodness of fit at three levels of significance α = 0.1, 0.05 and 0.01. Then, the percentages of the generated samples which follow standard normal distribution are calculated as tabulated in Table 4.1. For each κ, the percentage points as given in the first, second and third rows correspond to the α = 0.1, 0.05 and 0.01 level of significance respectively. The following results are observed:

(i) As expected, the percentage is a decreasing function of the significant level α.

(ii) For $\kappa < 5$, the percentage of generated samples which are correctly approximated by normal distribution is a decreasing function of the sample size $5 \leq n \leq 150$, but constant for $\kappa \geq 5$.

(iii) For all considered sample size *n*, the percentage is a decreasing function of concentration parameter $\kappa \leq 0.5$ but an increasing function for $\kappa \geq 1$.

(iv) At 0.05 level of significance, for small sample size ($n < 20$) and $\kappa > 2$, more than 96% of the generated samples are well approximated by standard normal distribution. Further, for any sample size *n* with concentration parameter $\kappa > 3$, more than 96% approximated by standard normal distribution. For any considered sample size *n* with concentration parameter $\kappa > 4$, it is found that more than 98% of the generated samples are approximated by the standard normal distribution. Further, for $\kappa > 10$ the percentage is almost 100%.

Generally, based on the simulation results we may conclude that, any sample generated from von Mises distribution with concentration parameter $\kappa > 4$ can be considered to be normally distributed. For samples with size ($n < 20$), the concentration parameter $\kappa > 2$ is considered large enough so that the data can be approximated by normal.

Table 4.1: The percentage of samples correctly approximated by standard normal distribution

κ	α	n												
		5	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**120**	**150**
	0.1	78.4	61.1	43.3	32.7	23.6	20.6	17.6	14.4	12.3	9.1	7.7	6.1	3.5
0.001	0.05	84.1	67.7	50.2	38.0	28.6	25.3	21.9	17.9	15.2	12.3	10.5	7.8	4.7
	0.01	95.0	78.3	62.7	48.6	39.3	34.3	30.1	25.6	23.1	18.9	15.8	12.6	7.9
	0.1	78.1	61.5	41.2	32.4	25.4	18.7	17.4	13.5	11.7	9.4	7.6	5.2	2.5
0.01	0.05	83.2	68.6	47.0	38.1	29.7	23.3	21.3	17.2	14.9	12.4	10.2	7.2	3.4
	0.01	94.4	81.2	59.2	49.6	40.0	33.0	29.2	24.7	20.9	18.8	15.2	11.8	6.2
	0.1	76.3	58.2	32.9	25.5	17.1	14.1	11.4	7.4	5.4	4.6	3.9	2.2	1.2
0.1	0.05	82.0	65.4	39.4	31.3	21.4	17.8	14.5	9.9	7.3	6.4	5.3	3.0	1.9
	0.01	94.0	79.2	52.3	42.4	30.1	25.5	20.7	15.2	13.2	11.3	8.6	5.7	2.9
	0.1	72.9	49.7	22.0	7.4	2.5	0.9	0.5	0.0	0.0	0.0	0.0	0.0	0.0
0.5	0.05	79.2	59.2	30.8	12.5	5.2	1.5	0.7	0.1	0.1	0.0	0.0	0.0	0.0
	0.01	92.4	74.9	48.8	26.1	14.4	5.7	2.6	0.8	0.5	0.2	0.0	0.0	0.0
	0.1	75.8	57.6	28.6	11.5	3.6	1.2	0.3	0.1	0.0	0.0	0.0	0.0	0.0
1	0.05	80.5	67.5	39.7	19.4	7.9	2.9	1.0	0.3	0.0	0.0	0.0	0.0	0.0
	0.01	93.5	82.3	61.7	42.7	23.5	12.5	6.0	2.4	0.7	0.2	0.2	0.0	0.0
	0.1	83.3	71.2	47.7	32.1	18.7	10.1	5.7	2.2	1.0	0.5	0.1	0.0	0.0
1.5	0.05	86.5	80.1	61.5	45.3	29.6	18.1	11.8	5.8	3.0	1.7	0.7	0.1	0.0
	0.01	95.6	91.2	81.8	70.2	56.3	42.7	31.8	22.9	14.0	9.2	5.7	1.7	0.2
	0.1	88.9	81.4	69.1	56.7	42.9	32.9	25.5	17.3	12.5	8.4	5.4	1.8	0.2
2	0.05	91.4	88.9	79.9	70.0	58.4	47.0	39.9	29.9	23.5	17.2	12.5	5.6	1.2
	0.01	97.2	96.6	92.6	88.2	81.9	74.3	70.0	59.4	52.7	46.5	36.0	23.4	9.2
	0.1	94.5	93.0	89.8	86.6	83.3	78.2	76.5	71.8	66.2	62.4	58.8	49.1	35.4
3	0.05	96.5	96.5	94.6	93.0	91.1	88.3	86.6	83.1	80.9	78.4	74.9	65.9	54.3
	0.01	98.8	99.5	98.6	98.7	98.0	97.5	97.0	95.8	96.0	93.9	93.6	89.2	83.8
	0.1	97.1	96.8	96.4	96.0	95.0	94.7	93.9	93.7	93.0	91.2	90.4	88.0	83.5
4	0.05	98.5	98.7	98.4	98.2	98.0	97.9	97.8	97.7	97.2	96.8	96.2	96.1	95.6
	0.01	99.8	99.8	99.7	99.7	99.6	99.6	99.6	99.5	99.5	99.4	99.4	99.3	98.8
	0.1	97.9	98.3	98.8	98.4	98.8	98.1	98.4	98.1	97.8	97.9	97.7	97.4	97.1
5	0.05	98.9	99.4	99.6	99.5	99.5	99.5	99.3	99.4	98.9	99.5	99.3	99.1	99.1
	0.01	99.8	99.9	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9	100.0	99.9
	0.1	98.6	99.5	99.3	99.6	99.4	99.4	99.4	99.4	99.3	99.3	99.3	99.3	99.1
6	0.05	99.5	99.9	99.8	99.9	99.8	99.9	99.9	99.8	99.7	99.8	100.0	99.8	99.9
	0.01	99.9	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	99.4	99.3	99.6	99.7	99.8	99.4	99.7	99.7	99.7	99.6	99.7	99.6	99.5
7	0.05	99.8	99.9	99.8	99.9	99.9	99.8	100.0	99.9	99.8	99.9	99.9	99.9	99.9
	0.01	100	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	99.6	99.7	99.9	99.8	99.8	99.9	99.8	99.8	99.9	99.9	99.9	99.9	99.8
8	0.05	100	99.9	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100	100	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Table 4.1, continued.

		n												
κ	α	**5**	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**	**120**	**150**
	0.1	99.6	99.9	99.8	99.8	99.9	99.8	99.9	100.0	99.9	99.8	99.9	99.8	99.9
9	0.05	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	99.9	99.9	99.9	99.8	99.9	99.8	99.9	99.9	100.0	100.0	99.9	99.9	100.0
10	0.05	99.9	100.0	100.0	99.9	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	100.0	100.0	100.0	99.8	99.9	99.9	99.9	99.9	99.9	100.0	100.0	100.0	99.8
12	0.05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	100.0	99.9	100.0	99.9	99.9	99.9	100.0	100.0	99.8	99.9	99.9	99.9	99.9
15	0.05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	100.0	99.9	99.9	100.0	99.9	99.9	99.9	99.9	99.9	100.0	99.9	99.9	99.8
17	0.05	100.0	100.0	100.0	100.0	100.0	99.9	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.1	100.0	99.9	100.0	100.0	99.9	100.0	99.9	99.9	100.0	99.9	99.8	100.0	99.8
20	0.05	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	0.01	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0

4.3 Illustrative examples

This section considers two data sets which are described in Chapter 3, namely, the frogs data and wind direction data. Both of them are approximated as linear data set to highlight the necessity of circular tests of discordancy.

4.3.1 Frogs data

The sample size of frog data is 14 with an estimated concentration parameter $\hat{\kappa} = 2.18$. It comes from von Mises distribution as discussed in Section 3.3. Since the sample is less than 20 and the concentration parameter is greater than 2 as discussed in Subsection 4.2.2, frogs data can be treated as a linear sample. Further test based on the Kolmogorov-Smirnov test is conducted to check the normality assumption of the data set. The resulting value of the test statistic is 0.1861 with the p-value 0.5 which suggest that the data comes from normal distribution. Next, four different techniques are used to detect possible outliers. Their implication and results are discussed here.

(i) Boxplot

Figure 4.1 shows the ordinary boxplot for frogs direction data which is obtained by S-Plus statistical software package. There is only one line outside the whiskers corresponding to observation 316. Thus, observation value 316 can be considered as an outlier.

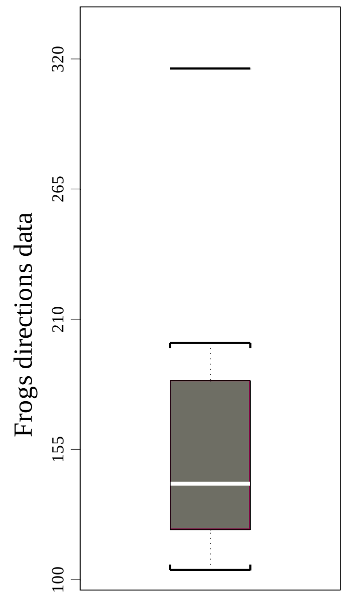


Figure 4.1: Boxplot for frogs direction

(ii) The three – sigma rule

The standard deviation of frog data is 55.24 and its mean is 158. According to the three-sigma rule any data point out of the range of $\bar{x} \pm 3s$ i.e. (-7.72, 323.72) is considered as an outlier. None of the observation values is located outside the mentioned interval. Therefore, the frogs data is free from outliers.

(iii) Dixon test

Dixon test is used to examine whether the minimum or maximum observation is an outlier. Since the sample size is 14, by following the rule suggested by Dean & Dixon (1951), we use r_{22} formula. The test statistic for the minimum value is given by

$$r_{22} = \frac{x_{(3)} - x_{(1)}}{x_{(n-2)} - x_{(1)}} = \frac{117 - 104}{192 - 104} = 0.148.$$

The test statistic for the maximum value is

$$r_{22} = \frac{x_{(n)} - x_{(n-2)}}{x_{(n)} - x_{(3)}} = \frac{316 - 192}{316 - 117} = 0.623.$$

The associated critical value is 0.501. Therefore, the maximum observation with value 316 is identified as an outlier. After the exclusion of observation 316, we re-examine the rest of data. The sample size has now reduced to 13. Thus, the formula of r_{22} is reused. The test statistic for the minimum value is $r_{22} = \frac{117 - 104}{184 - 104} = 0.163$ and for the maximum value is $r_{22} = \frac{200 - 184}{200 - 117} = 0.193$. The corresponding critical value is 0.521, which is larger than Dixon statistic for either minimum or maximum values. Thus, there is no more observations identified as outliers. Observation 316 is the only outlier identified in the data.

(iv) Maximum normed residual (Grubbs) test

The Grubbs statistic for frog data is 2.86. The critical value at $\alpha = 0.05$ level of significance is 2.36 (See Table 1, Grubbs (1969)). Therefore, the associated observation 316 is identified as an outlier.

4.3.2 Wind direction data

The wind direction data described in Section 3.5 consists of two sets of readings of wind direction using two different instruments. By treating the data as linear we may then fit the simple linear regression model to see the linear relationship between the two readings. We may identify possible outliers by using row deletion approach. S-Plus statistical software package is used to fit the data. The scatter plot for wind direction data is given in Figure 3.6 and there are four points far from the straight line. The output of fitting wind direction data by using simple linear regression model is displayed in Figure 4.2. The results show that, both coefficients are significant and the fitted model is given by

$$\hat{y}_i = 0.5398 + 0.8996\, x_i\,.$$

The coefficient of determination is R^2= 0.8515, while the F-statistic is 728.5 with the p-value equals zero. This suggests that the model fits the data well. The value of Durbin – Watson statistic (D-W) can be calculated and equals 1.77. The upper bound of D-W statistic at 0.05 significant level is 1.72. Thus, we conclude that the residuals are uncorrelated.

```
                          *** Linear Model ***
Call: lm(formula = Anchored ~ Radar, data = wind,na.action = na.exclude)
Residuals:
     Min      1Q    Median     3Q    Max
 -0.9512 -0.3572 -0.05984 0.1335 5.395
Coefficients:
              Value Std. Error t value Pr(>|t|)
(Intercept)  0.5398  0.1354     3.9880  0.0001
      Radar  0.8996  0.0333    26.9906  0.0000

Residual standard error: 0.8002 on 127 degrees of freedom
Multiple R-Squared: 0.8515
F-statistic: 728.5 on 1 and 127 degrees of freedom, the p-value is 0

Analysis of Variance Table

Response: Anchored

Terms added sequentially (first to last)
           Df   Sum of Sq   Mean Sq   F Value   Pr(F)
        R   1   466.4941   466.4941  728.494     0
Residuals  127  81.3250    0.6404
```

Figure 4.2: Simple linear regression model for wind direction data

Figure 4.3 gives the residual plot of the fitted model. It can be seen that most of residual points are distributed around zero. There are four points with numbers 38, 100, 109 and 111 located far from the other points with residual values 2.69, 5.14, 5.4 and 1.71, respectively. It is obvious that the values of observations with numbers 100 and 109 are closer to 2π rather than 0. Thus, if these residuals are treated as circular residuals, then observation with numbers 100 and 109 are consistent with the rest of observations.

It is expected that, observation with numbers 100 and 109 will be identified as outliers by using linear techniques. Figure 4.4 illustrates the Q-Q plot of the residuals. Most of the points are close to the straight line except for four points with numbers 38, 100, 109 and 111. Next, we explore the wind direction data further using different methods via row deletion approach.

(i) The hat matrix, h_{ii}

The plot of h_{ii} against the index of observation is displayed in Figure 4.5. The cut-off points is $(2p/n) = (2/129) = 0.031$. None of the values exceeds the cut off points. Thus, there is no leverage points suggested.

(ii) Cook statistic, $D_{(-i)}$

Figure 4.6 displays Cook statistic, $D_{(-i)}$ against the index of observation. Observations number 100 and 109 have very high values, which are very influential on the estimate of β. There are some other observations that have shorter spikes.

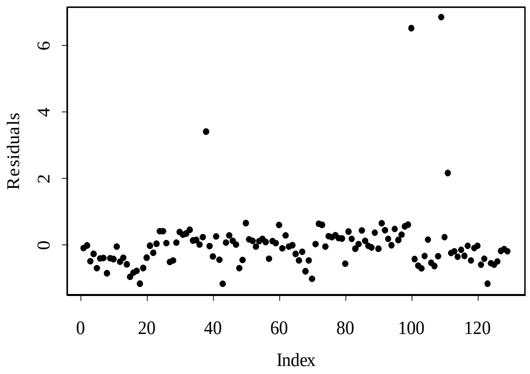


Figure 4.3: Ordinary residuals plot of wind data

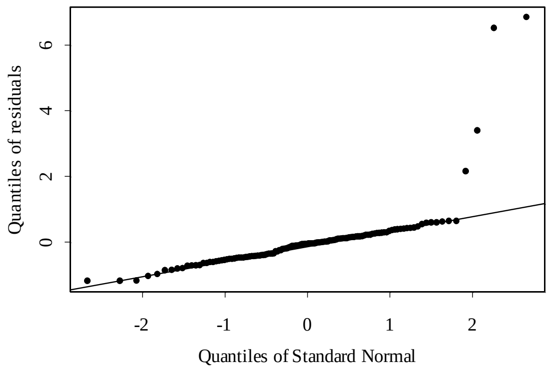


Figure 4.4: Q-Q plot for the residuals of wind data

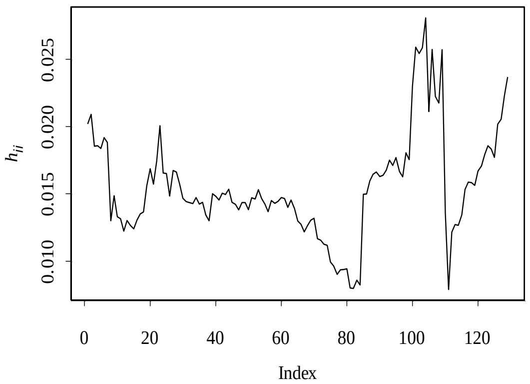


Figure 4.5: Hat matrix, h_{ii} of wind data

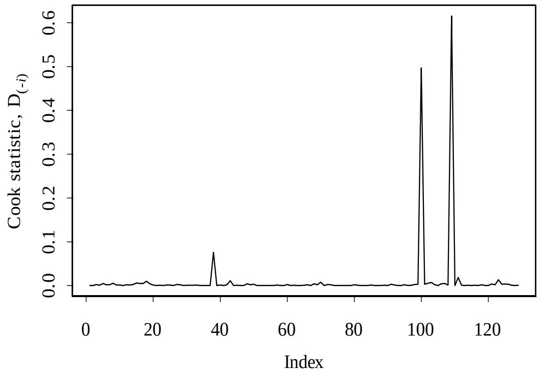


Figure 4.6: Cook statistic, $D_{(-i)}$ of wind data

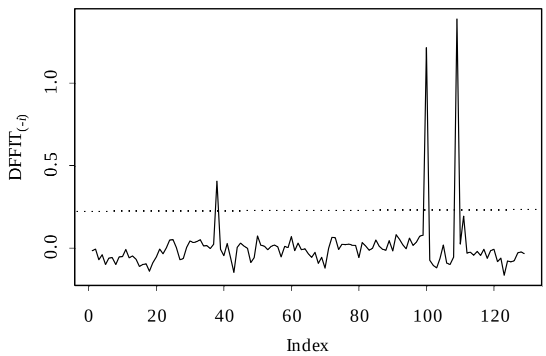


Figure 4.7: $DFFIT_{(-i)}$ statistic of wind data

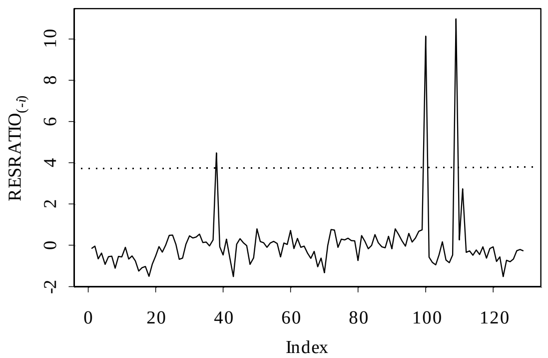


Figure 4.8: $RESRATIO_{(-i)}$ statistic of wind data

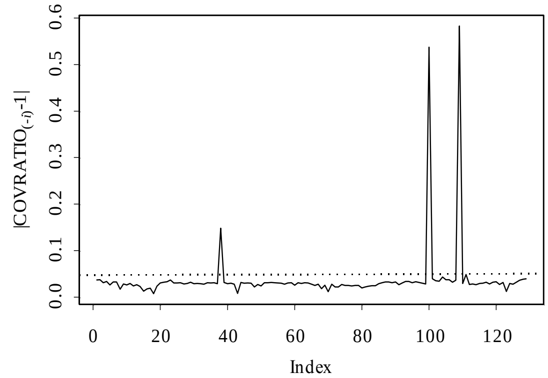


Figure 4.9: $|COVRATIO_{(-i)} - 1|$ statistic of wind data

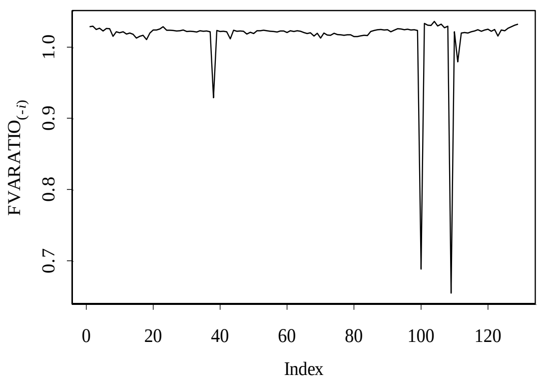


Figure 4.10: $FVRATIO_{(-i)}$ statistic of wind data

(iii) DFFIT statistic

A display of DFFIT statistic against the index of observations is given in Figure 4.7. Three observations with numbers 38, 100 and 109 exceed the cut-off point $2\sqrt{2/129} = 0.249$ given by the dotted line. Thus, they are candidates to be influential observations.

(iv) RESRATIO statistic

Figure 4.8 displays the plot of RESRATIO statistic against the index of observation. Only two observations with numbers 100 and 109 are larger than $F_{0.05,1,127} = 3.92$ given by the dotted line.

(v) COVRATIO statistic

Figure 4.9 displays the COVRATIO statistic against the index number. Statistics values corresponding to observations number 38, 100 and 109 exceed the cut–off point of $(3p/n) = (6/129) = 0.0465$.

(vi) FVARATIO statistic

There are four different points which are relatively large compared to the rest. Figure 4.10 displays the FAVRATIO statistic against the index of observation.

4.4 Summary

Based on the simulation study in Section 4.3, it is found that any sample generated from von Mises distribution with concentration parameter $\kappa > 4$ can be approximated by normal distribution. Further, for small sample size ($n < 20$) samples can be approximated by normal if the concentration parameter is $\kappa > 2$.

In the frogs data, three tests have identified observation 316 as an outlier, while the other test failed to do so. Note that, the use of linear discordancy tests for circular data highly depends on the mean direction of the circular samples. If the mean direction is close to the boundary of the circular variable (i.e. 0 or equivalently 2π radians), then most of the linear discordancy tests will perform poorly and are sometimes completely wrong. Consider the following simulated sample from von Mises distribution with mean 0 and concentration parameter $\kappa = 6.5$ given as follows:

$$6^\circ, 13^\circ, 18^\circ, 21^\circ, 34^\circ, 36^\circ, 43^\circ, 324^\circ, 338^\circ, 339^\circ, 345^\circ, 346^\circ, 351^\circ, 351^\circ, 353^\circ. \qquad (4.2)$$

The *IQR* for (4.2) is 317.96°. Based on the ordinary boxplot criterion it is impossible to identify any of the points as an outlier. Similar conclusion can be drawn for other tests, where the standard deviation is 165.08° and the mean is 192.17°. None of the values is located outside the interval ($-303.06^\circ, 687.417^\circ$). These shortcomings motivate us to develop alternative tests of discordancy for circular samples.

On the other hand, in the wind direction data the tests have wrongly identified two of the points as outliers and influential observations although they are consistent with the rest of the observations. Thus, there is a strong need to study alternative procedures to detect outlier in circular regression data.

In the following chapters, we propose alternative numerical and graphical tests to detect possible outliers in univariate and bivariate circular data.

CHAPTER FIVE

ALTERNATIVE TESTS OF DISCORDANCY

IN CIRCULAR DATA

5.1 Introduction

Circular data are subjected to one or more outlying observations. Discordancy tests in circular samples are different from those used in linear case, due to the bounded closed space of circular variables. Section 2.3 discusses these differences and reviews the available numerical and graphical tests to identify outliers in circular samples.

The existence of any outliers in circular sample affects its summary statistics. Jammalamadaka and SenGupta (2001) defined circular distance between any two points as the smaller of two arc lengths between the points along circumference. Hence, the circular distance between the mean direction $\bar{\theta}$ and each observation θ_i is defined as:

$$d_i = \min(\theta_i - \bar{\theta}, 2\pi - (\theta_i - \bar{\theta})) = \pi - \left|\pi - \left|\theta_i - \bar{\theta}\right|\right|.$$

Collett (1980) suggested that an observation with the maximum value of d_i will be a candidate of being an outlier.

The resultant length for circular data is given by $R = \sqrt{S^2 + C^2}$. Omitting any observation θ_i from a circular sample may increase the resultant length. It can be shown that

$$R_{(-i)}^2 = (C - \cos\theta_i)^2 + (S - \sin\theta_i)^2 = R^2 + 1 - 2R\cos(\theta_i - \bar{\theta}),$$

where $R^2_{(-i)}$ is the sample resultant length by omitting the *i*th observation θ_i. Therefore, as an observation gets further away from the mean direction, the value of $(\theta_i - \bar{\theta})$ increases, and the value of $\cos(\theta_i - \bar{\theta})$ decreases from 1 to -1. Similarly, the value of $R^2_{(-i)}$ increases from $(R-1)^2$ to $(R+1)^2$.

Collett (1980) mentioned that the identification of outlier in circular data is highly dependent on the concentration parameter κ, whereby, it is easier to identify an outlier in high concentrated circular samples than those with smaller concentration.

This chapter introduces three numerical tests for detection of outliers in circular samples. The first two are based on the circular distances and the chords' length between circular observations respectively. The next section is on the approximate distribution of the proposed statistics.

5.2 Alternative tests of discordancy in circular data

In this section, we propose two alternative methods of identifying outliers in circular data. The idea is based on the fact that circular data are distributed on the circumference of a circle. Thus, it is appropriate to use the properties of the circle. The first statistic is developed based on the summation of circular distances between circular observations while the second method is formulated based on the summation of the chords' length between the circular observations.

5.2.1 *A* statistic

Suppose $\theta_1,...,\theta_n$ are (*i.i.d*) circular sample located on the circumference of a unit circle. Rao (1969) defined the circular distance between θ_i and θ_j as

$$d_{ij} = 1 - \cos(\theta_i - \theta_j), \qquad i, j = 1,\ldots,n \qquad (5.1)$$

where d_{ij} is a monotone increasing function of $(\theta_i - \theta_j)$ and $d_{ij} \in [0,2]$. The summation of all circular distances of the point of interest θ_j to all other points is given by

$$D_j = \sum_{i=1}^{n}(1 - \cos(\theta_i - \theta_j)), \quad i = 1,\ldots,n.$$

If the observation θ_j is an outlier (i.e. it lies far from the rest of the observations), then the value of D_j will increase. Thus, the average circular distance given by $\frac{D_j}{n-1}$ can be used to identify possible outliers in the circular sample. The proposed statistic is given by

$$A = \max_j \left\{ \frac{D_j}{2(n-1)} \right\}, \quad j = 1,\ldots,n, \qquad (5.2)$$

where $A \in [0,1]$ is a linear measure. The average circular distance is divided by 2 in order to standardize the values of statistic *A*. The proposed statistic is based on the relative decrease in the summation of circular distances by omitting the point of interest θ_j.

An alternative definition of circular distance in terms of angles is given in (2.9). This alternative definition was used by Jammalamadaka & SenGupta (2001, p.218) for initial identification of outliers. Thus, alternative statistics of discordancy in circular samples may be defined based on this statistic. The summation of all circular distances from the observation θ_j to all other observations is given by

$$D_j^* = \sum_{i=1}^{n} (\pi - |\pi - |\theta_i - \theta_j||), \quad j = 1, \ldots, n,$$

and a reasonable statistic can be simply given by

$$A^* = \max_j \left\{ \frac{D_j^*}{(n-1)} \right\}, \quad j = 1, \ldots, n,$$

where $A^* \in [0, \pi]$. Statistic A^* is expected to have similar performance as statistic *A*.

In the following subsection an alternative statistic is proposed based on the summation of chord lengths passing through each observation in the circular sample.

5.2.2 Chord statistic

Here, the interest is to develop an alternative test of discordancy in circular data based on the geometrical properties of the chord of a circle. A chord is a segment that connects two distinct points on a circle circumference. The length of a chord between two points θ_i and θ_j can be calculated according to the formula

$$crd(\theta_{ij}) = 2r \sin \frac{\theta_{ij}}{2r},$$

where *r* is the radius length and θ_{ij} is the smallest angle between θ_i and θ_j which can be calculated using equation (2.9) such that

$$\theta_{ij} = d_\circ(\theta_i - \theta_j) = \pi - |\pi - |\theta_i - \theta_j||, \; \theta_{ij} \in [0, \pi].$$

In circular data, a unit circle (i.e. $r = 1$) is used to display the observations. Suppose there are *n* points $\theta_1, \ldots, \theta_n$ located on the circumference of a unit circle. Let B_j be the summation of all chords lengths pass through observation θ_j and given by

$$B_j = \sum_{i=1}^{n} crd(\theta_{ij}) = 2\sum_{i=1}^{n} \sin\frac{\theta_{ij}}{2}, \quad i = 1,\ldots,n. \tag{5.3}$$

For example, suppose n = 4 and let $j = 1$. Then B_1 is the sum of all chord lengths starting from θ_1 to the points θ_2, θ_3 and θ_4, as illustrated in Figure 5.1.

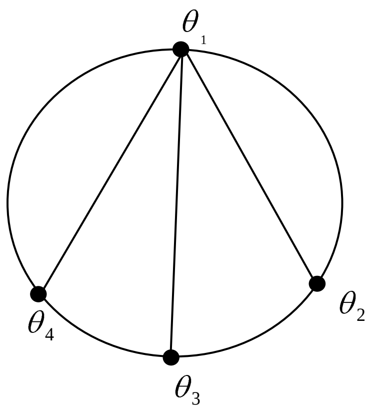


Figure 5.1: Illustration of chord lengths

Note that, $\max\{B_j\} \leq 2(n-1)$ for $j = 1,\ldots,n$. As θ_{ij} increases from 0 to π, $\sin\frac{\theta_{ij}}{2}$ increases from 0 to 1. When $\theta_{ij} = \pi$, $\sum_{i=1}^{n} \sin\frac{\theta_{ij}}{2} = n-1$, for $i = 1,\ldots,n$, while $\sum_{i=1}^{n} \sin\frac{\theta_{ij}}{2} = 0$ for $i = 1,\ldots,n$, when $\theta_{ij} = 0$. Therefore, the maximum value of B_j suggests that θ_j is a candidate of outlier. The proposed chord statistic is given by

$$Chord = \max_j \left\{ \frac{B_j}{2(n-1)} \right\}, \qquad j = 1,\ldots,n. \tag{5.4}$$

We will consider A and chord statistics for further discussion in the following sections. The discussion above can be extended to the case when $\theta_j \sim VM(\mu,\kappa)$, $j = 1,\ldots,n$. The next subsection describes and discusses the percentage points of the alternative statistics.

5.2.3 Percentage points of *A* and chord statistics

A simulation study is designed to find the percentage points of the null distribution of no outliers in circular data set for *A* and chord statistics. We consider sixteen values of concentration parameter in the range of 0.2 to 20 and different sample sizes ranging from 5 to 150. For each combination of sample size *n* and concentration parameter κ, we generate 5000 random samples of size *n* from a von Mises $VM(0,\kappa)$. The *A* and chord statistics in each generated random sample are calculated based on equations (5.2) and (5.4). We wish to estimate the percentage points of *A* and chord statistics at the 10, 5 and 1 percentages when no outlier is presented in the sample.

The simulation results are tabulated in Tables 5.1 and 5.2. For each sample size *n* and concentration parameter κ, 10, 5 and 1 percentages are given in the first, second and third rows, respectively. The following results are observed:

(i) In both statistics, it can be seen that the percentages have a peak value at around 1 for samples with small sizes ($n \leq 10$). For ($10 < n \leq 30$), the respective peak occurs when κ is around 2 , while for larger sample size ($n \geq 40$) the peak is at $\kappa = 3$. The pattern can be seen clearly through visual plot. For example, Figure 5.2 plots the 5 percentage points for *n* =20, where the peak is clearly at κ =2.

(ii) For small values of concentration parameter κ, the percentage points are a decreasing function of the sample size *n*, while for larger κ, the percentage point are increasing as illustrated in Figures 5.3.

(iii) In general, the percentage points of *A* statistic are smaller than the percentage points of chord statistic.

Table 5.1: The 10, 5 and 1 percentage points of the null distribution of A statistic

n	Perc.	κ															
		0.2	**0.5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**17**	**20**
	10%	0.86	0.86	0.87	0.79	0.58	0.45	0.34	0.29	0.25	0.22	0.20	0.18	0.15	0.12	0.10	0.08
5	5%	0.90	0.90	0.90	0.86	0.68	0.53	0.42	0.34	0.30	0.27	0.24	0.22	0.18	0.15	0.13	0.11
	1%	0.95	0.95	0.95	0.94	0.91	0.75	0.59	0.47	0.45	0.37	0.34	0.30	0.25	0.20	0.18	0.17
	10%	0.83	0.84	0.84	0.80	0.59	0.44	0.36	0.31	0.25	0.22	0.20	0.19	0.15	0.12	0.11	0.07
6	5%	0.86	0.87	0.88	0.86	0.69	0.54	0.43	0.36	0.30	0.28	0.25	0.23	0.18	0.15	0.13	0.10
	1%	0.92	0.93	0.94	0.94	0.89	0.72	0.59	0.50	0.44	0.36	0.33	0.31	0.23	0.20	0.18	0.16
	10%	0.82	0.83	0.84	0.80	0.60	0.45	0.36	0.31	0.26	0.23	0.21	0.18	0.16	0.12	0.11	0.07
7	5%	0.85	0.86	0.88	0.87	0.71	0.55	0.42	0.37	0.31	0.27	0.25	0.22	0.20	0.14	0.13	0.10
	1%	0.90	0.92	0.94	0.93	0.87	0.72	0.61	0.50	0.41	0.39	0.34	0.29	0.27	0.19	0.19	0.17
	10%	0.79	0.82	0.83	0.81	0.64	0.48	0.37	0.32	0.28	0.24	0.20	0.20	0.16	0.13	0.11	0.08
8	5%	0.82	0.85	0.87	0.87	0.75	0.58	0.45	0.37	0.33	0.28	0.25	0.23	0.20	0.15	0.13	0.10
	1%	0.88	0.90	0.93	0.94	0.89	0.79	0.61	0.53	0.43	0.38	0.32	0.32	0.27	0.20	0.17	0.15
	10%	0.79	0.80	0.83	0.82	0.64	0.49	0.39	0.33	0.27	0.24	0.21	0.20	0.16	0.13	0.12	0.10
9	5%	0.82	0.83	0.87	0.87	0.76	0.57	0.46	0.39	0.33	0.29	0.25	0.24	0.20	0.16	0.13	0.11
	1%	0.87	0.89	0.92	0.92	0.90	0.77	0.57	0.52	0.43	0.39	0.34	0.32	0.27	0.21	0.19	0.16
	10%	0.77	0.79	0.83	0.82	0.66	0.50	0.39	0.33	0.28	0.24	0.22	0.20	0.16	0.13	0.12	0.10
10	5%	0.80	0.82	0.86	0.87	0.76	0.59	0.46	0.38	0.33	0.29	0.26	0.24	0.19	0.16	0.14	0.12
	1%	0.85	0.88	0.92	0.92	0.91	0.79	0.61	0.51	0.42	0.41	0.36	0.32	0.26	0.22	0.19	0.16
	10%	0.75	0.78	0.82	0.83	0.67	0.50	0.39	0.35	0.30	0.26	0.23	0.20	0.17	0.14	0.12	0.10
12	5%	0.78	0.81	0.85	0.87	0.77	0.58	0.46	0.41	0.34	0.30	0.26	0.24	0.20	0.16	0.13	0.11
	1%	0.82	0.86	0.89	0.92	0.90	0.76	0.61	0.52	0.44	0.40	0.38	0.32	0.28	0.21	0.18	0.15
	10%	0.74	0.77	0.81	0.84	0.71	0.53	0.42	0.35	0.30	0.26	0.24	0.21	0.17	0.14	0.13	0.11
14	5%	0.76	0.79	0.84	0.87	0.79	0.62	0.48	0.41	0.34	0.29	0.28	0.24	0.20	0.16	0.15	0.12
	1%	0.81	0.84	0.89	0.91	0.90	0.80	0.66	0.53	0.46	0.39	0.36	0.32	0.27	0.22	0.20	0.16
	10%	0.73	0.76	0.81	0.83	0.73	0.53	0.42	0.36	0.31	0.27	0.24	0.21	0.18	0.14	0.13	0.11
16	5%	0.75	0.78	0.84	0.87	0.81	0.61	0.49	0.41	0.36	0.31	0.28	0.25	0.21	0.17	0.15	0.12
	1%	0.80	0.82	0.88	0.91	0.90	0.77	0.65	0.55	0.47	0.43	0.38	0.33	0.28	0.21	0.21	0.17
	10%	0.71	0.75	0.81	0.84	0.73	0.55	0.44	0.36	0.32	0.27	0.25	0.21	0.19	0.14	0.13	0.11
18	5%	0.73	0.78	0.83	0.87	0.83	0.64	0.51	0.41	0.35	0.31	0.28	0.25	0.21	0.17	0.15	0.12
	1%	0.77	0.81	0.87	0.91	0.91	0.81	0.63	0.54	0.46	0.42	0.35	0.31	0.28	0.23	0.22	0.18
	10%	0.70	0.74	0.80	0.84	0.75	0.56	0.45	0.38	0.32	0.28	0.25	0.22	0.19	0.15	0.13	0.11
20	5%	0.73	0.76	0.82	0.87	0.82	0.68	0.51	0.44	0.37	0.31	0.28	0.25	0.22	0.17	0.15	0.12
	1%	0.77	0.81	0.86	0.90	0.92	0.86	0.73	0.53	0.48	0.41	0.38	0.32	0.30	0.20	0.20	0.17
	10%	0.69	0.73	0.80	0.85	0.75	0.58	0.45	0.37	0.33	0.29	0.25	0.23	0.19	0.16	0.14	0.12
25	5%	0.71	0.76	0.82	0.88	0.83	0.69	0.54	0.44	0.37	0.33	0.29	0.27	0.22	0.17	0.16	0.13
	1%	0.75	0.79	0.85	0.90	0.92	0.85	0.66	0.54	0.53	0.43	0.35	0.32	0.28	0.24	0.20	0.17
	10%	0.67	0.72	0.80	0.86	0.77	0.60	0.49	0.38	0.34	0.29	0.26	0.23	0.20	0.16	0.14	0.12
30	5%	0.69	0.74	0.82	0.88	0.85	0.68	0.56	0.43	0.39	0.34	0.30	0.26	0.23	0.18	0.16	0.13
	1%	0.72	0.78	0.84	0.91	0.91	0.88	0.69	0.55	0.49	0.42	0.39	0.33	0.32	0.24	0.20	0.17

Table 5.1, continued.

		κ															
n	Perc.	**0.2**	**0.5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**17**	**20**
	10%	0.65	0.71	0.79	0.86	0.84	0.64	0.51	0.43	0.36	0.31	0.28	0.25	0.21	0.17	0.15	0.12
40	5%	0.67	0.73	0.80	0.87	0.89	0.72	0.58	0.48	0.42	0.36	0.32	0.28	0.24	0.19	0.16	0.13
	1%	0.70	0.76	0.83	0.90	0.92	0.91	0.77	0.59	0.55	0.46	0.39	0.35	0.30	0.24	0.21	0.17
	10%	0.64	0.70	0.78	0.86	0.86	0.65	0.51	0.43	0.37	0.32	0.29	0.25	0.22	0.18	0.16	0.13
50	5%	0.65	0.71	0.80	0.87	0.89	0.76	0.59	0.49	0.43	0.35	0.32	0.28	0.25	0.20	0.18	0.14
	1%	0.69	0.74	0.82	0.89	0.92	0.91	0.82	0.61	0.53	0.42	0.41	0.36	0.30	0.24	0.22	0.18
	10%	0.63	0.69	0.78	0.86	0.86	0.69	0.54	0.45	0.40	0.34	0.30	0.27	0.22	0.18	0.16	0.13
60	5%	0.64	0.70	0.79	0.87	0.89	0.77	0.61	0.50	0.44	0.38	0.34	0.30	0.24	0.21	0.17	0.14
	1%	0.67	0.72	0.81	0.89	0.92	0.88	0.78	0.62	0.55	0.46	0.41	0.38	0.30	0.26	0.22	0.18
	10%	0.62	0.68	0.77	0.86	0.86	0.69	0.56	0.46	0.39	0.34	0.30	0.27	0.23	0.19	0.17	0.14
70	5%	0.64	0.70	0.79	0.87	0.89	0.76	0.64	0.51	0.44	0.40	0.34	0.31	0.26	0.20	0.18	0.15
	1%	0.66	0.71	0.81	0.89	0.92	0.91	0.79	0.66	0.54	0.50	0.44	0.37	0.32	0.25	0.23	0.18
	10%	0.62	0.68	0.78	0.86	0.88	0.73	0.56	0.47	0.40	0.34	0.32	0.28	0.23	0.19	0.17	0.14
80	5%	0.63	0.70	0.79	0.88	0.90	0.80	0.63	0.52	0.44	0.38	0.35	0.31	0.26	0.21	0.19	0.16
	1%	0.66	0.72	0.80	0.89	0.92	0.92	0.78	0.67	0.57	0.49	0.43	0.39	0.32	0.25	0.23	0.18
	10%	0.61	0.68	0.77	0.86	0.88	0.75	0.58	0.47	0.41	0.36	0.32	0.29	0.24	0.19	0.17	0.14
90	5%	0.63	0.69	0.78	0.87	0.90	0.82	0.65	0.53	0.46	0.41	0.35	0.31	0.27	0.22	0.20	0.17
	1%	0.65	0.72	0.80	0.88	0.92	0.92	0.84	0.69	0.62	0.48	0.43	0.38	0.35	0.28	0.25	0.19
	10%	0.61	0.67	0.76	0.86	0.89	0.74	0.57	0.49	0.41	0.37	0.32	0.30	0.25	0.20	0.17	0.14
100	5%	0.62	0.69	0.77	0.87	0.90	0.82	0.64	0.55	0.47	0.40	0.35	0.32	0.27	0.22	0.20	0.17
	1%	0.64	0.70	0.79	0.88	0.92	0.92	0.85	0.66	0.56	0.49	0.46	0.38	0.32	0.26	0.24	0.18
	10%	0.60	0.67	0.77	0.86	0.89	0.78	0.59	0.49	0.42	0.37	0.33	0.30	0.24	0.20	0.17	0.14
110	5%	0.61	0.68	0.77	0.87	0.90	0.85	0.66	0.55	0.47	0.40	0.37	0.34	0.27	0.22	0.19	0.16
	1%	0.64	0.70	0.79	0.88	0.92	0.93	0.78	0.69	0.61	0.47	0.43	0.42	0.33	0.29	0.23	0.18
	10%	0.60	0.67	0.76	0.86	0.88	0.78	0.61	0.51	0.43	0.37	0.33	0.31	0.25	0.20	0.18	0.15
120	5%	0.61	0.68	0.78	0.87	0.90	0.84	0.68	0.57	0.48	0.41	0.38	0.34	0.28	0.22	0.20	0.17
	1%	0.63	0.70	0.79	0.88	0.92	0.93	0.85	0.73	0.59	0.51	0.47	0.43	0.34	0.26	0.26	0.19
	10%	0.60	0.67	0.76	0.86	0.89	0.78	0.61	0.51	0.42	0.38	0.34	0.30	0.25	0.20	0.18	0.15
130	5%	0.61	0.68	0.77	0.87	0.90	0.86	0.68	0.57	0.47	0.42	0.37	0.34	0.28	0.22	0.20	0.17
	1%	0.62	0.70	0.78	0.88	0.92	0.93	0.87	0.70	0.58	0.52	0.46	0.41	0.33	0.26	0.24	0.18
	10%	0.59	0.66	0.76	0.86	0.90	0.81	0.61	0.51	0.43	0.39	0.34	0.31	0.26	0.21	0.19	0.16
140	5%	0.61	0.67	0.77	0.87	0.91	0.87	0.69	0.58	0.47	0.42	0.37	0.34	0.28	0.22	0.20	0.17
	1%	0.63	0.69	0.78	0.88	0.92	0.93	0.82	0.74	0.55	0.48	0.48	0.42	0.33	0.27	0.26	0.19
	10%	0.59	0.66	0.76	0.86	0.89	0.77	0.63	0.52	0.44	0.39	0.34	0.32	0.26	0.20	0.19	0.16
150	5%	0.60	0.67	0.77	0.87	0.90	0.83	0.73	0.59	0.49	0.44	0.38	0.35	0.28	0.22	0.20	0.17
	1%	0.62	0.69	0.78	0.88	0.92	0.93	0.91	0.75	0.63	0.52	0.45	0.44	0.35	0.27	0.26	0.19

Table 5.2: The 10, 5 and 1 percentage points of the null distribution of chord statistic

n	Perc.	κ															
		0.2	**0.5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**17**	**20**
	10%	0.91	0.91	0.91	0.86	0.74	0.63	0.55	0.51	0.48	0.44	0.41	0.40	0.36	0.31	0.30	0.28
5	5%	0.94	0.94	0.94	0.91	0.81	0.69	0.61	0.56	0.53	0.50	0.46	0.45	0.40	0.36	0.33	0.31
	1%	0.96	0.96	0.96	0.96	0.94	0.84	0.75	0.69	0.64	0.60	0.56	0.54	0.49	0.44	0.40	0.38
	10%	0.90	0.90	0.90	0.88	0.73	0.64	0.58	0.52	0.48	0.44	0.42	0.40	0.36	0.32	0.30	0.28
6	5%	0.92	0.92	0.92	0.91	0.82	0.71	0.62	0.58	0.52	0.49	0.46	0.43	0.40	0.36	0.34	0.31
	1%	0.95	0.96	0.96	0.96	0.92	0.83	0.74	0.68	0.62	0.59	0.55	0.53	0.48	0.43	0.41	0.38
	10%	0.88	0.89	0.90	0.88	0.76	0.64	0.57	0.53	0.48	0.46	0.43	0.41	0.37	0.33	0.30	0.28
7	5%	0.90	0.91	0.92	0.91	0.83	0.71	0.63	0.58	0.53	0.50	0.47	0.44	0.41	0.37	0.34	0.32
	1%	0.93	0.95	0.96	0.96	0.92	0.85	0.75	0.68	0.68	0.61	0.56	0.54	0.49	0.43	0.40	0.37
	10%	0.87	0.88	0.89	0.88	0.77	0.66	0.57	0.53	0.49	0.45	0.42	0.40	0.37	0.33	0.31	0.29
8	5%	0.89	0.89	0.91	0.91	0.85	0.72	0.63	0.58	0.54	0.49	0.47	0.45	0.41	0.37	0.34	0.31
	1%	0.93	0.94	0.95	0.95	0.94	0.85	0.74	0.70	0.64	0.61	0.55	0.53	0.49	0.42	0.40	0.38
	10%	0.86	0.87	0.89	0.89	0.78	0.66	0.59	0.53	0.50	0.46	0.44	0.41	0.38	0.34	0.32	0.29
9	5%	0.88	0.89	0.91	0.92	0.84	0.73	0.64	0.59	0.54	0.51	0.47	0.45	0.42	0.37	0.34	0.32
	1%	0.92	0.92	0.95	0.96	0.95	0.88	0.74	0.69	0.63	0.60	0.57	0.53	0.47	0.44	0.41	0.38
	10%	0.85	0.86	0.89	0.89	0.80	0.69	0.60	0.54	0.51	0.48	0.44	0.42	0.39	0.34	0.33	0.30
10	5%	0.87	0.88	0.91	0.92	0.86	0.74	0.65	0.60	0.56	0.52	0.49	0.47	0.42	0.38	0.36	0.32
	1%	0.90	0.92	0.94	0.95	0.94	0.88	0.77	0.69	0.64	0.60	0.58	0.56	0.50	0.44	0.42	0.38
	10%	0.84	0.85	0.88	0.89	0.80	0.69	0.60	0.56	0.50	0.47	0.45	0.43	0.38	0.34	0.32	0.29
12	5%	0.85	0.87	0.91	0.92	0.86	0.75	0.66	0.60	0.55	0.52	0.48	0.46	0.43	0.37	0.35	0.32
	1%	0.88	0.91	0.93	0.95	0.94	0.86	0.79	0.69	0.63	0.61	0.58	0.53	0.49	0.44	0.40	0.37
	10%	0.82	0.84	0.88	0.89	0.81	0.69	0.61	0.56	0.52	0.48	0.45	0.43	0.39	0.34	0.32	0.30
14	5%	0.82	0.86	0.89	0.93	0.87	0.75	0.67	0.60	0.56	0.53	0.48	0.47	0.43	0.38	0.36	0.32
	1%	0.87	0.89	0.93	0.95	0.94	0.88	0.78	0.70	0.65	0.61	0.58	0.55	0.51	0.44	0.42	0.38
	10%	0.81	0.84	0.87	0.90	0.82	0.69	0.62	0.57	0.52	0.48	0.46	0.44	0.39	0.35	0.33	0.31
16	5%	0.83	0.85	0.90	0.92	0.89	0.76	0.67	0.61	0.57	0.52	0.49	0.47	0.43	0.38	0.36	0.33
	1%	0.86	0.89	0.92	0.94	0.94	0.89	0.80	0.71	0.65	0.61	0.58	0.54	0.49	0.44	0.42	0.39
	10%	0.80	0.84	0.88	0.90	0.83	0.70	0.64	0.57	0.53	0.49	0.47	0.43	0.40	0.36	0.34	0.31
18	5%	0.83	0.85	0.89	0.92	0.89	0.76	0.69	0.69	0.57	0.53	0.51	0.48	0.43	0.38	0.36	0.33
	1%	0.86	0.88	0.92	0.94	0.94	0.90	0.80	0.72	0.68	0.60	0.58	0.55	0.51	0.44	0.42	0.38
	10%	0.80	0.83	0.87	0.91	0.83	0.72	0.63	0.57	0.53	0.49	0.46	0.44	0.41	0.36	0.34	0.32
20	5%	0.82	0.84	0.88	0.92	0.89	0.78	0.68	0.62	0.58	0.54	0.51	0.47	0.44	0.39	0.37	0.34
	1%	0.85	0.87	0.91	0.94	0.94	0.89	0.80	0.72	0.66	0.62	0.60	0.55	0.51	0.45	0.43	0.39
	10%	0.78	0.82	0.86	0.91	0.86	0.74	0.65	0.58	0.54	0.51	0.48	0.45	0.41	0.36	0.34	0.31
25	5%	0.80	0.83	0.89	0.92	0.91	0.79	0.70	0.64	0.58	0.54	0.51	0.49	0.45	0.40	0.38	0.34
	1%	0.83	0.86	0.91	0.94	0.95	0.92	0.79	0.73	0.68	0.64	0.59	0.55	0.51	0.46	0.43	0.39
	10%	0.78	0.82	0.86	0.91	0.87	0.74	0.66	0.60	0.56	0.52	0.49	0.46	0.42	0.37	0.35	0.32
30	5%	0.79	0.83	0.88	0.92	0.91	0.80	0.71	0.64	0.60	0.56	0.52	0.50	0.46	0.40	0.37	0.35
	1%	0.82	0.85	0.90	0.94	0.94	0.92	0.83	0.73	0.67	0.63	0.60	0.57	0.53	0.46	0.43	0.40

Table 5.2, continued.

		κ															
n	Perc.	**0.2**	**0.5**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**12**	**15**	**17**	**20**
	10%	0.76	0.80	0.86	0.91	0.89	0.77	0.68	0.62	0.56	0.53	0.49	0.47	0.43	0.38	0.36	0.33
40	5%	0.77	0.81	0.87	0.92	0.92	0.82	0.73	0.67	0.62	0.56	0.53	0.51	0.46	0.41	0.38	0.36
	1%	0.80	0.84	0.89	0.93	0.94	0.93	0.83	0.76	0.70	0.64	0.61	0.57	0.53	0.46	0.44	0.40
	10%	0.74	0.80	0.86	0.91	0.91	0.79	0.69	0.62	0.58	0.54	0.51	0.48	0.44	0.39	0.37	0.34
50	5%	0.76	0.81	0.87	0.92	0.93	0.86	0.73	0.66	0.62	0.58	0.54	0.51	0.47	0.42	0.40	0.36
	1%	0.79	0.83	0.88	0.93	0.95	0.94	0.87	0.76	0.69	0.64	0.62	0.58	0.53	0.47	0.44	0.41
	10%	0.74	0.78	0.85	0.92	0.92	0.80	0.71	0.64	0.59	0.55	0.52	0.49	0.45	0.40	0.39	0.35
60	5%	0.75	0.80	0.86	0.92	0.93	0.86	0.76	0.68	0.63	0.59	0.55	0.52	0.48	0.43	0.40	0.37
	1%	0.77	0.82	0.87	0.93	0.95	0.95	0.86	0.77	0.71	0.67	0.62	0.59	0.54	0.49	0.46	0.41
	10%	0.74	0.79	0.86	0.92	0.92	0.83	0.73	0.66	0.61	0.58	0.54	0.50	0.47	0.41	0.39	0.36
70	5%	0.75	0.80	0.87	0.92	0.94	0.88	0.77	0.70	0.65	0.61	0.57	0.55	0.50	0.45	0.42	0.38
	1%	0.77	0.81	0.88	0.94	0.95	0.95	0.88	0.80	0.73	0.67	0.64	0.61	0.57	0.50	0.47	0.42
	10%	0.74	0.79	0.85	0.92	0.93	0.85	0.74	0.67	0.61	0.58	0.54	0.51	0.48	0.42	0.38	0.36
80	5%	0.75	0.80	0.86	0.93	0.94	0.90	0.80	0.72	0.66	0.62	0.57	0.55	0.50	0.45	0.42	0.38
	1%	0.77	0.81	0.87	0.94	0.95	0.96	0.88	0.79	0.72	0.69	0.65	0.62	0.55	0.50	0.47	0.43
	10%	0.73	0.79	0.86	0.92	0.93	0.85	0.75	0.68	0.62	0.58	0.55	0.52	0.48	0.42	0.39	0.37
90	5%	0.75	0.80	0.86	0.93	0.94	0.91	0.79	0.72	0.66	0.61	0.58	0.55	0.52	0.44	0.41	0.39
	1%	0.77	0.82	0.88	0.93	0.96	0.96	0.89	0.82	0.73	0.67	0.65	0.60	0.56	0.50	0.44	0.43
	10%	0.73	0.78	0.85	0.92	0.94	0.86	0.74	0.68	0.62	0.59	0.56	0.53	0.48	0.42	0.40	0.38
100	5%	0.74	0.79	0.86	0.93	0.95	0.91	0.78	0.73	0.67	0.62	0.60	0.56	0.52	0.44	0.42	0.40
	1%	0.76	0.81	0.87	0.93	0.96	0.96	0.87	0.81	0.73	0.69	0.68	0.60	0.59	0.49	0.48	0.44
	10%	0.73	0.78	0.85	0.92	0.94	0.87	0.76	0.69	0.62	0.60	0.56	0.52	0.47	0.42	0.40	0.39
110	5%	0.74	0.79	0.86	0.93	0.95	0.93	0.82	0.73	0.66	0.63	0.60	0.56	0.50	0.45	0.42	0.41
	1%	0.76	0.81	0.87	0.94	0.96	0.96	0.93	0.83	0.73	0.69	0.67	0.61	0.58	0.50	0.45	0.45
	10%	0.73	0.78	0.85	0.92	0.94	0.86	0.77	0.69	0.65	0.59	0.56	0.53	0.47	0.43	0.41	0.40
120	5%	0.74	0.79	0.86	0.93	0.95	0.92	0.82	0.73	0.69	0.63	0.59	0.57	0.50	0.45	0.44	0.42
	1%	0.75	0.80	0.87	0.93	0.96	0.96	0.93	0.83	0.76	0.70	0.65	0.65	0.58	0.51	0.47	0.46
	10%	0.73	0.78	0.85	0.92	0.94	0.87	0.76	0.69	0.63	0.59	0.56	0.53	0.48	0.43	0.41	0.41
130	5%	0.73	0.79	0.86	0.93	0.95	0.92	0.80	0.74	0.68	0.63	0.60	0.56	0.51	0.46	0.44	0.43
	1%	0.75	0.80	0.87	0.94	0.96	0.96	0.91	0.82	0.75	0.71	0.66	0.67	0.57	0.50	0.48	0.47
	10%	0.72	0.78	0.85	0.92	0.95	0.88	0.79	0.71	0.64	0.60	0.57	0.54	0.49	0.44	0.42	0.42
140	5%	0.73	0.79	0.86	0.93	0.95	0.93	0.83	0.75	0.67	0.64	0.61	0.57	0.52	0.47	0.45	0.44
	1%	0.75	0.80	0.87	0.94	0.96	0.96	0.92	0.82	0.76	0.72	0.66	0.67	0.58	0.51	0.49	0.48
	10%	0.71	0.78	0.85	0.93	0.95	0.88	0.79	0.70	0.65	0.61	0.57	0.54	0.49	0.44	0.41	0.43
150	5%	0.72	0.79	0.86	0.93	0.95	0.93	0.84	0.76	0.68	0.65	0.60	0.57	0.52	0.47	0.43	0.45
	1%	0.74	0.80	0.87	0.93	0.96	0.97	0.95	0.83	0.74	0.70	0.65	0.63	0.58	0.52	0.49	0.49

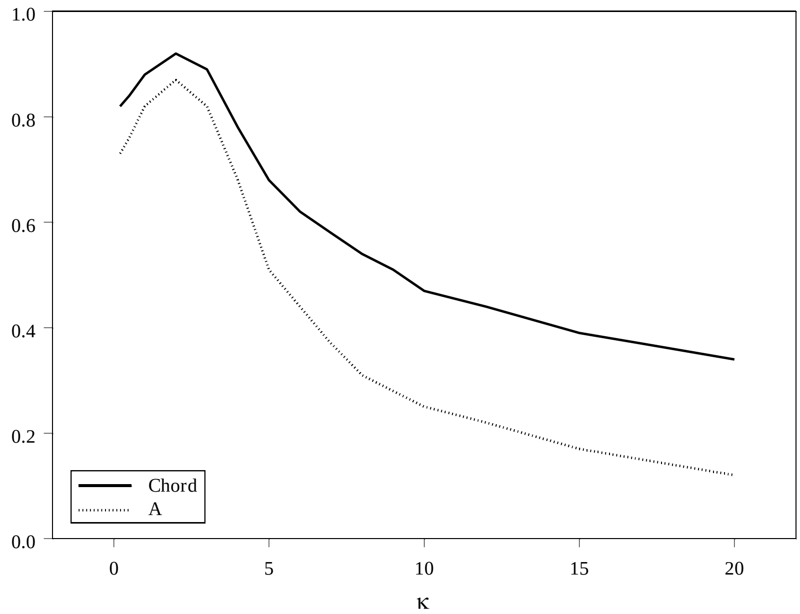


Figure 5.2: The percentage points of *A* and chord statistics for *n* = 20, $\alpha = 0.05$

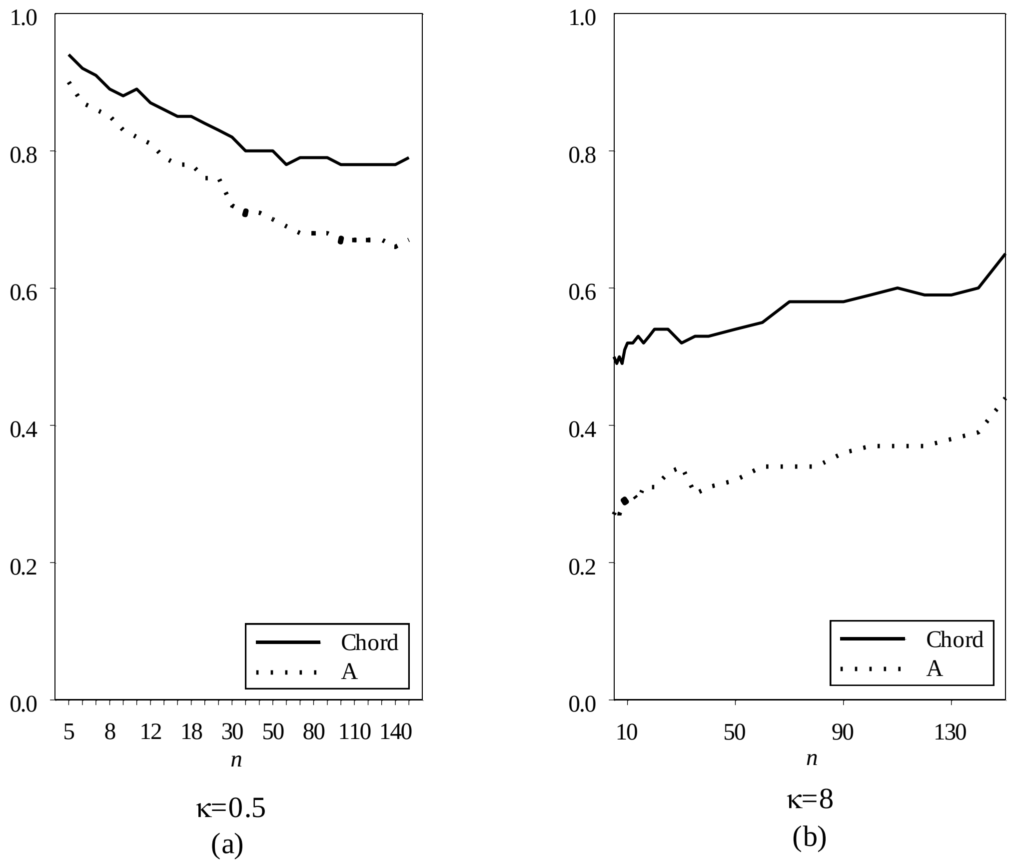


Figure 5.3: The percentage points for *A* and chord statistics at κ = 0.5 and 8, $\alpha = 0.05$

5.2.4 The performance of *A* and chord statistics compared to other tests

(i) Definitions and notations

Collett (1980) applied selected measures to test the performances of several statistics to detect an outlier in circular sample. In this subsection, we use similar measures to compare the performance of the *A* and chord statistics with *C*, *D* and *M* statistics which are reviewed in Section 2.3.

David (1970, p.185) and Barnett & Lewis (1984, p.132-135), state that a good test should have (i) a high power function, (ii) a high probability of identifying a contaminating value as an outlier when it is in fact an extreme value, and (iii) a low probability of wrongly identifying a good observation as discordant. In circular statistics context the extreme value is defined as a point with the maximum circular deviation.

Let P1=(1-β) be the power function where β is the type-II error, P3 the probability that the contaminant point is an extreme point and is identified as discordant, and P5 the probability that the contaminant point is identified as a discordant given that it is an extreme point. A good test is expected to have (i) high P1, (ii) high P5 and (iii) low P1-P3.

(ii) Description of simulation algorithm

To study the performance of the *A* and chord statistics, we use 2000 samples based on different sizes *n* = 5, 10, 20 and 50, and concentration parameter κ = 2, 5 and 7. The samples are generated in such a way that ($n-1$) of the observations come from $VM(\alpha,\kappa)$ and the remaining one observation comes from $VM(\alpha+\lambda\pi,\kappa)$, where λ is the degree of contamination and $0\le\lambda\le1$. When $n=5$, the contaminated point is placed at the third ordered position in the sample, whereas for the others, the contaminated point is set at the seventh ordered position in the sample. The *C*, *D*, *M*, *A* and chord statistics in each random sample are then calculated based on corresponding equations in Sections 2.3 and 5.2.

(iii) Discussion

Figure 5.4 displays the performance measure P3 against the degree of contamination λ using the 5 percentage points for the *A* and chord statistics. It is obvious that both statistics have similar performance for all cases. Figure 5.4(a) shows that, for $n = 20$, the performance of both statistics are better when higher value of concentration parameter is used. On the other hand, Figure 5.4(b) illustrates that, for $\kappa = 7$, the performance is lower when larger sample sizes are used.

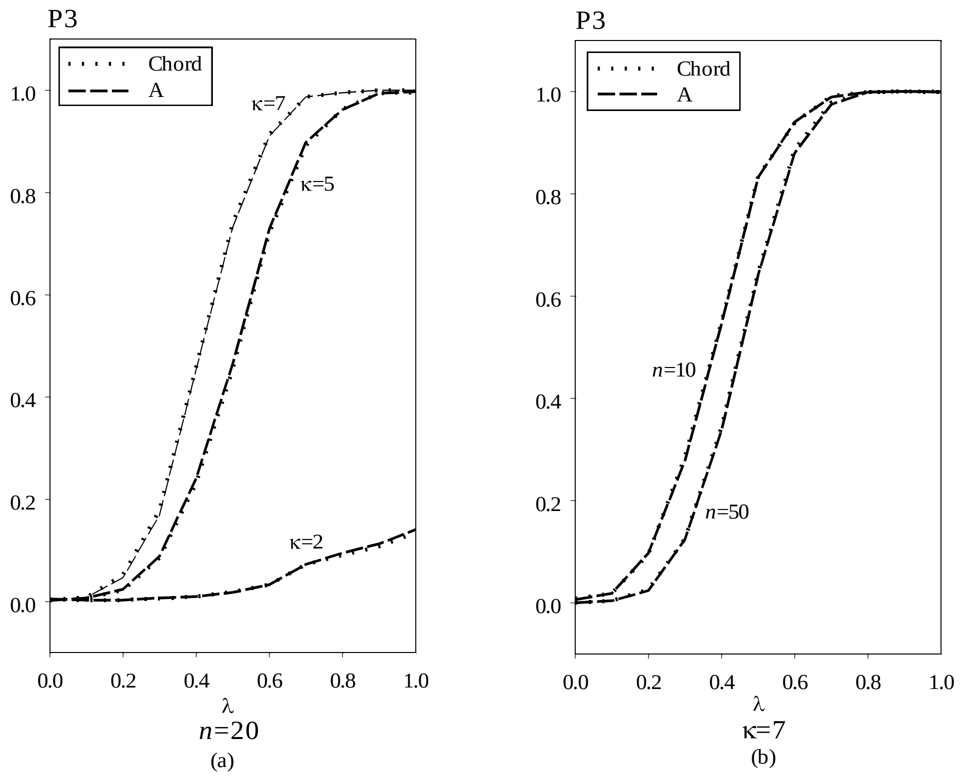


Figure 5.4: Power of performance for *A* and chord statistics

Figure 5.5 gives three selected graphs of the performance measures P1 and P3 against λ for *C, D, M, A* and Chord statistics. The following results are observed:

(i) In case of small sample size ($n = 5$) and small concentration parameter ($\kappa = 2$), the values of P1 are better for *M* statistic compared to others for all contamination λ levels, as illustrated in Figure 5.5(a). However, as *n* gets larger, *A* statistic performs better followed by chord statistic as shown in Figure 5.5(b). Similar trend is observed for P3 and P5.

(ii) For larger sample size *n* = 10, 15, 20 and 50 and larger concentration parameter, ($\kappa = 5$ and 7), *A* and chord statistics perform almost similar and slightly better in terms of P1, P3 and P5 compared to *C* and *D* statistics but they are much better than *M* statistic for $\lambda \geq 0.4$, as partially shown in Figure 5.5(c). For $\lambda < 0.4$ the performance for all statistics are similar.

(iii) The differences between P1 and P3 generally are very close to 0 for all cases.

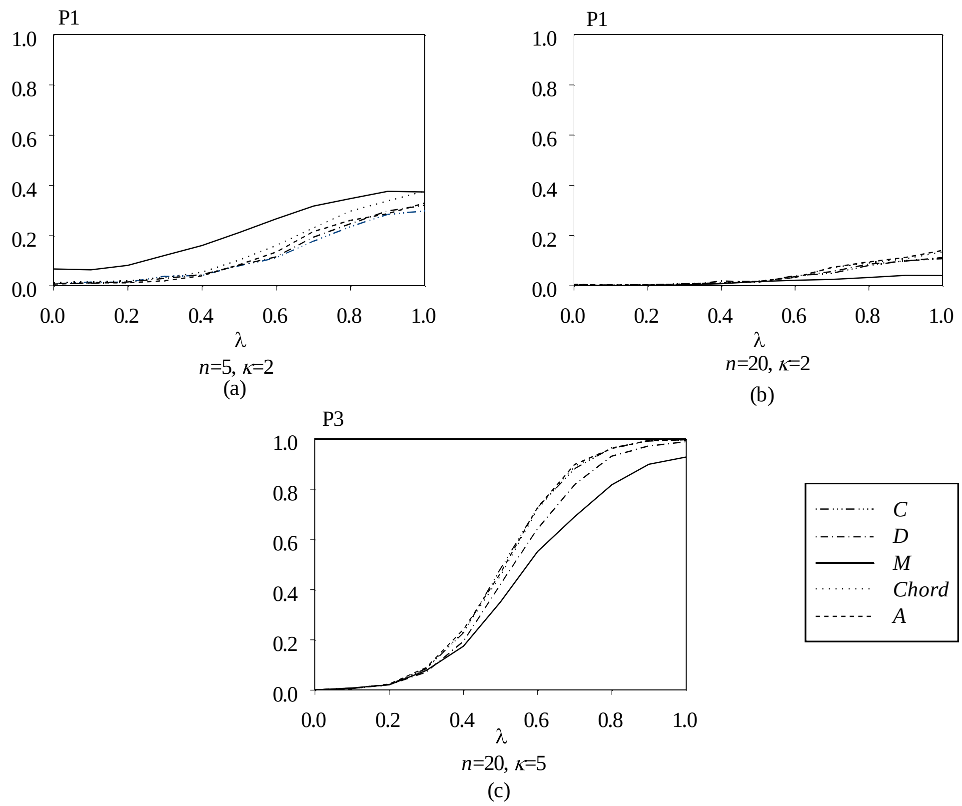


Figure 5.5: Relative performance of discordancy tests

Based on the results obtained from this simulation study, we conclude that, in general, the *A* and chord statistics have similar performance but perform better than the other tests of discordancy in a circular data set. Similar pattern is observed for the other cases and the complete results are available in Appendix (A.2).

5.3 On the approximate distribution of the alternative tests of discordancy

M statistic is the only known test of discordancy in circular samples that has an asymptotic distribution, under the assumption that the sample comes from von Mises distribution with a large concentration parameter κ. This section discusses on the approximate distribution of the alternative tests of discordancy which have been proposed in Section 5.2. Further, we propose a procedure to identify outliers in circular samples based on the approximate distribution of the circular distance in (5.1).

Result 5.1.

Let $\theta_1,...,\theta_n$ be (i.i.d) sample from von Mises distribution with mean direction μ and large concentration parameter κ. Then, for any θ_i and θ_j $i, j = 1,\dots,n, i \neq j$

$$\kappa[1-\cos(\theta_i-\theta_j)]\xrightarrow{d}\chi_1^2.$$

Proof:

Suppose $\Theta = \{\theta_1,...,\theta_n\}$ is a random circular sample from von Mises distribution with mean μ and large concentration parameter κ, $\Theta \sim VM(\mu,\kappa)$. For any observation $\theta_i \in \Theta$ and large κ, it has been shown in (4.1) that

$$\sqrt{\kappa}(\theta_i-\mu) \sim N(0,1).$$

Since θ_i and θ_j are independent observations then,

$$\sqrt{\frac{\kappa}{2}}(\theta_i - \theta_j) \sim N(0,1).$$

From the properties of standard normal distribution, we get

$$\frac{\kappa}{2}(\theta_i - \theta_j)^2 \sim \chi_1^2. \tag{5.5}$$

For large κ, the distribution is more concentrated. Thus the circular distance between any two points is relatively small. From the second order of Taylor series expansion, $\cos\phi \approx 1 - \frac{\phi^2}{2}$ or $\frac{\phi^2}{2} \approx 1 - \cos\phi$. Thus, by letting $\phi = (\theta_i - \theta_j)$, we have

$$\frac{1}{2}(\theta_i - \theta_j)^2 = [1 - \cos(\theta_i - \theta_j)].$$

Multiplying both sides by κ, then we get

$$\frac{\kappa}{2}(\theta_i - \theta_j)^2 = \kappa[1 - \cos(\theta_i - \theta_j)].$$

From (5.5), for $d_{ij} = 1 - \cos(\theta_i - \theta_j)$, we have

$$\kappa d_{ij} = \kappa[1 - \cos(\theta_i - \theta_j)] \sim \chi_1^2.$$

□

Result 5.2.

Let $\theta_1,...,\theta_n$ be (i.i.d) sample from von Mises distribution with mean direction μ and large concentration parameter κ. Then,

$$\boldsymbol{B_{ij}} = 2\kappa \sin^2\left(\frac{\theta_{ij}}{2}\right) \sim \chi_1^2, \ \text{for } i, j = 1,\dots,n \text{ and } i \neq j,$$

where θ_{ij} is the circular distance between θ_i and θ_j.

Proof:

From the definition of circular distance in Subsection 5.2.2 and the *cosine* function properties, we have

$$\cos(\theta_{ij}) = \cos(\pi - |\pi - |\theta_i - \theta_j||),$$

$$= -\cos(|\pi - |\theta_i - \theta_j||),$$

$$= \cos(\theta_i - \theta_j). \tag{5.6}$$

Consider the trigonometric identities $\cos\phi = 1 - 2\sin^2\left(\frac{\phi}{2}\right)$. Thus, for circular distance θ_{ij}, we have

$$1 - \cos(\theta_{ij}) = 2\sin^2\left(\frac{\theta_{ij}}{2}\right).$$

By multiplying both sides by κ, we get

$$\kappa[1 - \cos(\theta_{ij})] = 2\kappa\sin^2\left(\frac{\theta_{ij}}{2}\right).$$

From (5.6) and Result 5.1, we have

$$\boldsymbol{B_{ij}} = 2\kappa\, sin^2\left(\frac{\theta_{ij}}{2}\right) \sim \chi_1^2.$$

□

Since κd_{ij} and B_{ij} have the same approximate distribution, we can make use the approximate distribution of κd_{ij} given by Result 5.1 to identify outliers in circular samples.

Unfortunately, the statistic $\sum_{i=1}^{n} \kappa d_{ij}$ does not follow Chi-squares distribution with (n-1) degree of freedom due to the absence of independency. Alternatively, in order to optimize the usefulness of Result 5.1 towards the identification of outliers in circular samples, the count of κd_{ij} which exceed the critical values $\chi^2_{\alpha,1}$ at α level of significant for each $j = 1,\ldots,n$ is considered as an indicator of outlier existence. In other words, with respect to the *j*th observation, let P_j be the percentage of κd_{ij} that exceeds the critical value $\chi^2_{\alpha,1}$ for each observation θ_i, $i, j = 1,\ldots,n$. Then value of P_j close to 100% indicates that the jth observation is a candidate to be an outlier. It is obvious that for sample of size *n*, the maximum count of P_j could not exceed (*n*-1).

Simulation study is carried out to determine the size of $\max_j \{P_j\}$. Two main factors are considered: concentration parameter and sample size. A total of seven sample sizes *n* =10, 30, 50, 70, 100, 150 and 200, together with nine values of concentration parameter κ = 2, 5, 10, 30, 50, 70, 100 and 200 are considered. For each combination of sample size *n* and concentration parameter κ, we generated 2000 samples from von Mises distribution $VM(0, \kappa)$. The percentage of $\max_j \{P_j\}$ for each generated sample is specified and then ranked in an ascending order to obtain the percentiles.

For small concentration parameter κ, all κd_{ij} values will be less than $\chi^2_{\alpha,1}$, since $1 - \cos(\theta_i - \theta_j) \in [0,2]$. This suggests that κ should be at least $\frac{1}{2}\chi^2_{\alpha,1}$ where α is the level of significance for the observation to be identified as an outlier. Simulation results are given in Appendix (A.3) and show that the percentage (cut-off) points are consistent for $\kappa \geq 2$.

Table 5.3 presents the arithmetic mean and the standard deviations as given in parenthesis for the percentages for $\kappa \geq 2$. Results show that the cut-off points highly dependent on the sample size *n* and approach 100% for very large sample size *n*.

Further, the standard deviations is less than 6%, which suggest the consistent behaviour of the percentage points for $\kappa \geq 2$.

Table 5.3: The cut-off points for the percentage of the P_j for $\kappa \geq 2$

n	**10**	**30**	**50**	**70**	**100**	**150**	**200**
90%	40.00	54.67	62.60	66.57	69.70	74.27	76.65
	(0.00)	(1.72)	(2.67)	(2.45)	(4.00)	(4.79)	(5.43)
95%	41.00	63.00	70.80	73.71	76.10	80.07	81.95
	(3.16)	(3.31)	(3.91)	(3.03)	(4.15)	(4.84)	(5.89)
99%	60.00	79.00	83.40	85.29	87.20	88.80	90.20
	(0.00)	(5.22)	(3.53)	(3.23)	(3.82)	(4.58)	(4.36)

5.4 Summary

This chapter has proposed three alternative numerical tests of discordancy. The first is based on the circular distance while the second is based on the chord lengths. The other is an approach to detect outliers in circular samples based on the approximate distribution of circular distance. The proposed statistics are simple and easy to interpret by practitioners. It is found that the first two statistics have equal performance, but perform slightly better than the other known tests of discordancy.

CHAPTER SIX

LABELING OUTLIERS VIA CIRCULAR BOXPLOT

6.1 Introduction

Visual display is an easy and informative technique to describe any given data set, for example, histogram, pie chart, Q-Q plot and boxplot. Boxplot is a simple and flexible graphical tool in exploratory data analysis. It was developed by Tukey (1977) and it consists of five-number summaries which are the smallest observation, lower quartile Q_1, median, upper quartile Q_3 and largest observation. One of its main applications is to identify extreme values and outliers in univariate data sets.

Extensive research has been conducted on the labelling of outliers by using boxplot. To identify outliers in real line data sets, most studies use 1.5 as the value for resistant constant, ν, in the boxplot criterion, $\nu \times IQR$, where IQR is the interquartile range. In other words, any observation with value smaller than $(Q_1 - 1.5 \times IQR)$ or greater than $(Q_3 + 1.5 \times IQR)$ are labelled as "outlier". Hoaglin *et al.* (1986) investigated the performance of boxplot for outlier labelling by considering different values of ν. They concluded that $\nu = 1.5$ is the best choice in avoiding masking problems while choosing $\nu = 3$ is considered to be extremely conservative. On the other hand, Ingelfinger *et al.* (1983) suggested the use of $\nu = 2$ while Sim *et al.* (2005) demonstrated that the choice of resistant constant $\nu = 1.5$ or $\nu = 3$ is in general inappropriate for normal sample and is completely inappropriate for skewed distributions. This signifies the importance of choosing the best value of ν for different data set with different underlying distributions.

Fisher (1993) reviewed circular plots which went back to 1858 when Florence Nightingale drew a circular plot of the causes of mortality in the British Army during the Crimean War. This plot is also known as rose diagram or wind rose diagram. Graedel (1977) used boxplot to describe the wind speed in different sectors of the wind rose diagram. However, in general, boxplot is not suitable to be used directly on a circular data set. Meanwhile, Anderson (1993) described briefly a version of circular boxplot using five summary statistics as found in linear boxplot. However, the whiskers are fixed to map out the central 90% of the data for all cases. No attempt was made to determine the appropriate values of the resistant constants and other properties of the circular boxplot. In our works, we develop a more comprehensive theory of the circular boxplot with the main purpose of labelling outliers in the circular data.

Figure 6.1 displays the boxplot of the data given in (2.4). There is an isolated observation on the right side of the boxplot. However, the value of this point is actually consistent with the other values if it is treated as circular observation. Thus, the construction of a new boxplot for circular variables is really indispensable, which must be able to taken into account the periodicity of circular variables.

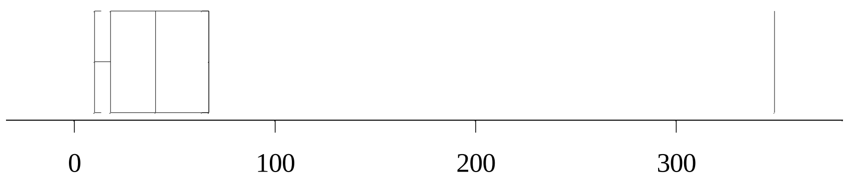


Figure 6.1: Boxplot of data in (2.4)

The objectives of this chapter are as follows:

(i) to propose a special boxplot version for circular data sets called circular boxplot,

(ii) to label possible outliers with the circular boxplot, and

(iii) to develop a subroutine in S-Plus environment to display the circular boxplot.

This chapter is organized as follows. The next section discusses the proposed construction of circular boxplot. Simulation and numerical studies are carried out in Sections 6.3 and 6.4 to estimate the appropriate values of resistant constant ν and to investigate the power of performance of the circular boxplot respectively. Practical example is given in Section 6.5.

6.2 Summary statistics for circular boxplot

Due to the unusual characteristics of circular variables, many relevant descriptive measures and display plots were developed, for example, mean direction, variance, circular histogram and stem-and-leaf diagrams. However, there is no known design of boxplot for circular variable.

The encountered difficulty in constructing the boxplot for circular variables arises from the complexity of determining the median. This is due to the bounded range of circular variables and the problem of overlapping which is highly expected to occur when the concentration parameter κ of circular sample is small.

The following subsections discuss the number summaries which are required to construct the circular boxplot.

6.2.1 Median direction and quartiles of circular variable

The definition of the median direction is given in Section 3.2. In the case of prior knowledge of the circular distribution, Mardia (1972) defined the median direction ϕ as the solution of

$$\int_{\phi}^{\phi+\pi} f(\theta)\, d\theta = \int_{\phi+\pi}^{\phi+2\pi} f(\theta)\, d\theta = 0.5\,,$$

where $f(\theta)$ is the probability density function of θ. The first and third quartile directions Q_1 and Q_3 is any solution of

$$\int_{\phi-Q_1}^{\phi} f(\theta) d\theta = 0.25 \quad \text{and} \quad \int_{\phi}^{\phi+Q_3} f(\theta) d\theta = 0.25$$

respectively. In most cases, the circular distribution is unknown. To date, no published literature is found on a nonparametric estimator of Q_1 and Q_3 for circular variables. However, it seems sensible to estimate Q_1 and Q_3 by classifying the sample observations into two groups based on their locations with respect to the sample median direction. Subsequently, Q_1 can be considered as the median of the first group and Q_3 as the median of the second.

If the value of Q_1 is larger than the value of Q_3 we simply interchange their labels. For simplicity and to avoid the confusion caused by the localization of Q_1 and Q_3, rotatable property of circular data by subtracting the estimated mean direction of the circular sample from each sample observation is used to make sure that the mean is in the zero direction. This rotation might be helpful to identify Q_1 and Q_3 in a more consistent way. That is, we can assume $(Q_1 - \bar{\theta}) \in [0, \pi]$ and $(Q_3 - \bar{\theta}) \in [\pi, 2\pi]$. The robustness of mean direction (see Wehrly and Shine (1981)) is a useful property which gives a fair assurance that the existence of any possible outlier will not have much effect on the estimated mean direction. Figure 6.2.(a) shows the quartiles for simulated circular data from von Mises distribution with mean direction $\pi/4$ and concentration parameter $\kappa = 4$. The first quartile $Q_1 = 33^\circ$, the median direction $\phi = 50^\circ$ and the third quartile $Q_3 = 69^\circ$.

6.2.2 Circular interquartiles range *CIQR* and fences

Analogues to the linear case, circular interquartiles range $CIQR$ is required to construct the circular boxplot. After the rotation of sample observation, $CIQR$ can be obtained by the following formula:

$$CIQR = 2\pi - Q_3 + Q_1.$$

For highly concentrated data it is possible to have quartiles and mean directions at the same point. Thus, the $CIQR = 0$. The upper and lower fences can be identified such as, lower fence $L_F = Q_1 + \nu \times CIQR$ and upper fence $U_F = Q_3 - \nu \times CIQR$, where ν is the resistant constant. Figure 6.2(b) illustrates a particular example of the proposed circular boxplot for symmetric simulated circular data.

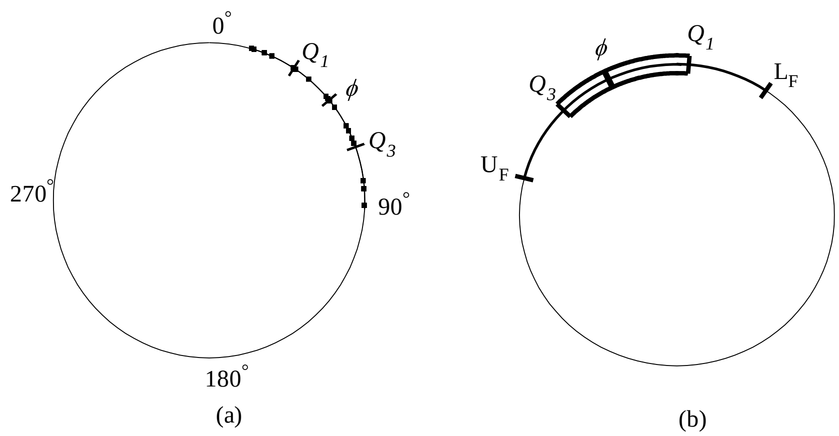


Figure 6.2: Proposed structure of circular boxplot

In the following section, numerical and simulation studies will be carried out in order to specify appropriate values of resistant constant ν.

6.3 Estimation of the resistant constant ν

In the real line case $1.5 \times IQR$ criterion is a popular choice for boxplot to identify outliers. The interest of investigating the appropriate values of the resistant constant ν was developed since the first construction of boxplot. Hoaglin *et al.* (1986), Ingelfinger *et al.* (1994) and Sim *et al.* (2005) discussed the appropriate values of resistant constant ν, which can be used to identify outliers in linear samples. It is not sensible to utilize similar resistant constant ν of linear boxplot in the case of circular data due to the bounded range of the circle. This is because the high possibility of overlapping problem between lower and upper fences for large resistant constant ν when the concentration parameter κ is small.

Hoaglin *et al.* (1986) used three different measures to investigate the behaviour of boxplot. In this section we extend their methodology to estimate the appropriate values of the resistant constant ν. The following notations are used. Let $A(\nu, n)$ denote the outside rate per observation in von Mises samples size *n*. Further, $B(\nu, n)$ denotes the probability of a sample of size *n* that contains no observation outside the interval (L_F, U_F), and $B3(\nu, n)$ is the probability that the von Mises sample of size *n* contains more than two observations outside the interval (L_F, U_F).

6.3.1 Simulation and numerical studies

In order to investigate the behaviour of circular variables with respect to five different summaries which are the median, Q_1, Q_3, L_F and U_F, series of simulation studies are carried out. Samples are generated from von Mises distribution $VM(\mu, \kappa)$, with different sizes n between 5 and 200. Different values of concentration parameter are considered, κ = 0.5 ,1 (1) 10. Further, various values of the resistant constant ν = 1(0.2)3 and 3.5 are utilized in order to obtain L_F and U_F .

A total of 3000 samples with sample size n and concentration parameter κ are generated from von Mises distribution $VM(\mu,\kappa)$. By using different values of the resistant constant ν, the following statistics which are *CIQR*, mean, median, Q_1, Q_3, L_F, U_F, $A(\nu,n)$, $B(\nu,n)$ and $B3(\nu,n)$ are obtained. There are huge amount of results and information obtained from this simulation studies. In the following subsection we will look into the properties of *CIQR*, circular distance between mean and median direction, overlapping problem and further descriptions of the three measures $A(\nu,n)$, $B(\nu,n)$ and $B3(\nu,n)$. A part of simulations results are given in Appendix (A.4).

6.3.2 Results and discussion

(i) Circular interquartiles range *CIQR*

CIQR can be estimated from the cumulative distribution function of any statistical distribution and it is defined as $CIQR = x_{75} - x_{25}$, where x_{25} and x_{75} are the solutions of $\int_0^{x_{25}} f(x)\,dx = 0.25$ and $\int_0^{x_{75}} f(x)\,dx = 0.75$, respectively.

By comparing the obtained *CIQR* from simulations with the *CIQR* based on the c.d.f. of the von Mises distribution table Batschelet (1981, p.322-331), close values are obtained especially for large sample size $n \geq 20$. For instance:

(1) For, $\kappa = 2$, $x_{25} = 149^\circ$ and $x_{75} = 211^\circ$, then $CIQR = 211^\circ - 149^\circ = 62^\circ$.

(2) For, $\kappa = 10$, $x_{25} = 167^\circ$ and $x_{75} = 193^\circ$, then $CIQR = 193^\circ - 167^\circ = 26^\circ$.

These values are close to the results found by simulation, which are 61.3° for $\kappa = 2$ and *n*=60 and 25° for $\kappa = 10$, *n* =40, (see Appendix (A.4)). Thus, the c.d.f. of any known distributions can be used to construct the boxplot, while for unknown distribution or at a stage of exploring the data, nonparametric methods are used to define the median and *CIQR*.

Fisher (1993, p.54) stated that " there is no circular distribution available with an associated measure of spread, which can rescale to have unit spread". This lack of "standardized" circular distribution especially with von Mises distribution causes difficulties as there is no standard von Mises distribution as analogues to the standard normal distribution. Consequently, it is rather difficult to find functional relationship for *CIQR*. An attempt has been made to find it via a simulation study.

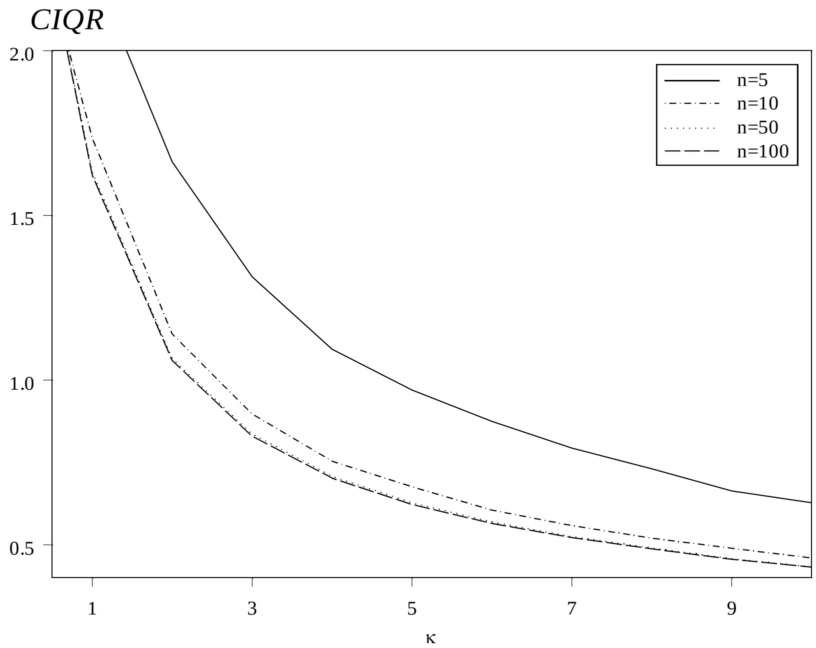


Figure 6.3: *CIQR* for different sample size *n* at various levels of κ

When the probability density function of von Mises distribution is used, Figure 6.3 shows that the *CIQR* is a decreasing function of concentration parameter κ and the plots of *CIQR* are very close to each other for large sample sizes ($n \geq 10$). Furthermore, we found that, for large samples ($n \geq 10$) of von Mises distribution with concentration parameter ($\kappa > 3$), there is a functional relationship between *CIQR* (in radian) and concentration parameter κ. It can be expressed in the following form:

$$CIQR \approx (\ln \kappa)^{-1}.$$

As shown in Figure 6.4, the *CIQR* values lie very close to the curve of $(\ln \kappa)^{-1}$.

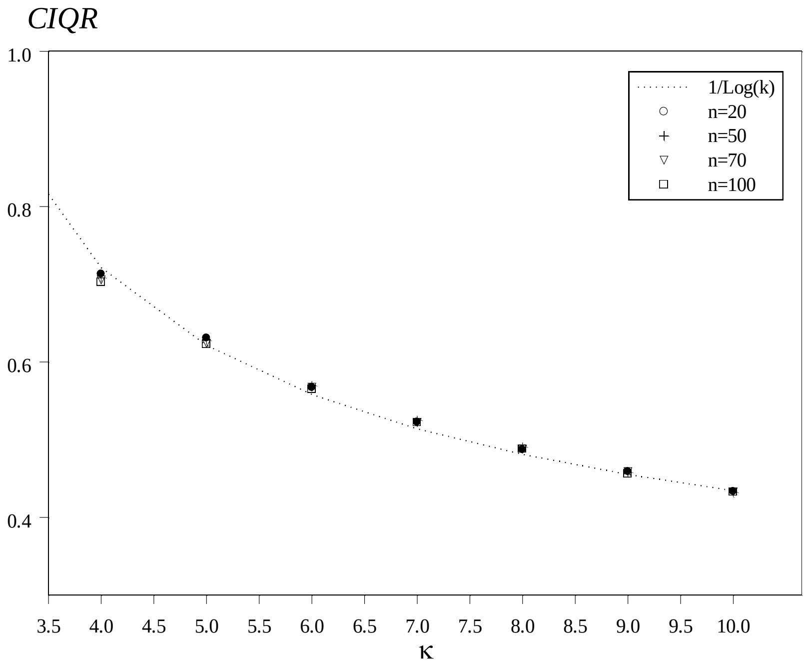


Figure 6.4: Functional relation between *CIQR* and concentration parameter κ

(ii) Circular distance between mean and median directions

The circular distance is defined as the smallest difference angle between any two angles. Figure 6.5 illustrates the behaviour of circular distance between mean and median direction for different values of concentration parameter κ and different sample size *n*. It shows that the circular distance is a decreasing function of concentration parameter κ and sample size *n*. For $n = 50$ the difference when $\kappa = 0.5$ is around 8.3°, while it is around 2.5° for $\kappa > 3$. These results confirm Wehrly and Shine (1981) conclusion about the robustness of the mean direction.

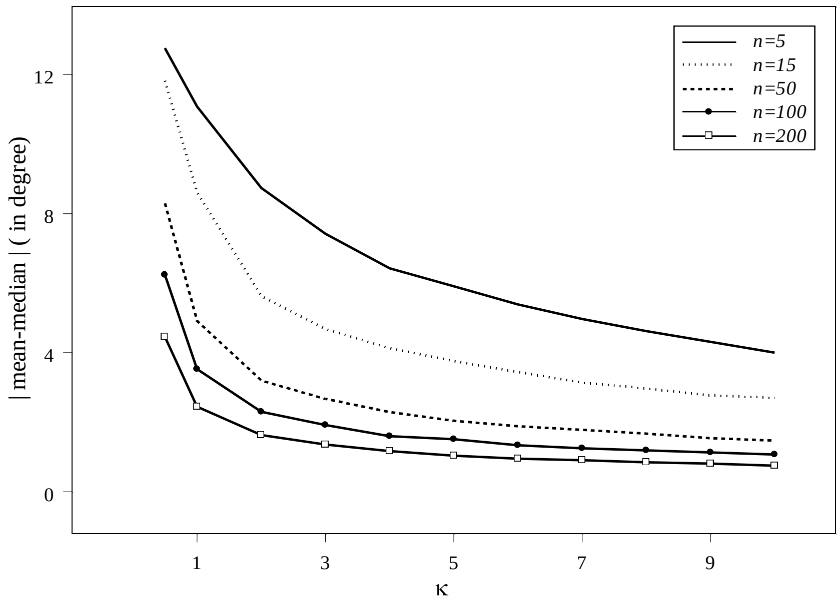


Figure 6.5: Circular distance between mean and median direction versus κ

(iii) Overlapping problem of the fences

Overlapping problem between the upper and lower fences is expected to occur at some values of resistant constant ν because of the bounded range of circular variables. Such problem has caused a messy boxplot structure and could lead to miss-identification of outliers.

Result 6.1

For large concentration parameter κ and sample size with $n \geq 10$, the upper and lower fences of circular boxplot are subject for overlapping if

$$\nu > \pi \ln(\kappa) - 0.5$$

where ν is the resistant constant.

Proof:

Overlapping problem occurs if $Q_1^* + \nu \times CIQR > \pi$. For symmetric samples $Q_1^* = \frac{1}{2} CIQR$, then overlapping occurs when

$$(0.5 + \nu) CIQR > \pi,$$

for large concentration parameter κ and large sample size *n*. However, $CIQR \approx (\ln \kappa)^{-1}$. Thus,

$$(0.5+\nu)\frac{1}{\ln(\kappa)} > \pi .$$

Hence, overlapping problem occurs if $\nu > \pi \ln(\kappa) - 0.5$.

□

As an example, for $n \geq 10$ with $\kappa = 4$, if ν is larger than 3.8 then the overlapping problem is expected to occur. The case will be more complicated for smaller concentration parameter $(\kappa < 3)$, where the overlapping will occur even for small values of resistant constant ν.

(iv) Description of B(ν,*n*) measure

$B(\nu,n)$ denotes the probability of no observation outside the interval (L_F, U_F) for von Mises sample of size *n* and resistant constant ν. Overlapping problem affects the behaviour of $B(\nu,n)$ measure. Figure 6.6 shows the behaviour of $B(\nu,50)$. For small concentration parameter ($\kappa < 2$). It is observed that $B(\nu,50)$ is non-monotone function of resistant constant ν, while it is an increasing function of ν for large concentration parameter ($\kappa \geq 2$), where it is not affected much by the increase of κ.

Simulation results of $B(\nu,n)$ are used to specify values of resistant constant ν which can be used to construct circular boxplot. It is more informative to interpret the results of simulation studies according to the mode of sample size *n* with respect to 4. Thus, the sample size can be clustered into one of 4 groups according to whether *n* has the form $4j, 4j+1, 4j+2$ or $4j+3$, where $j \in N$.

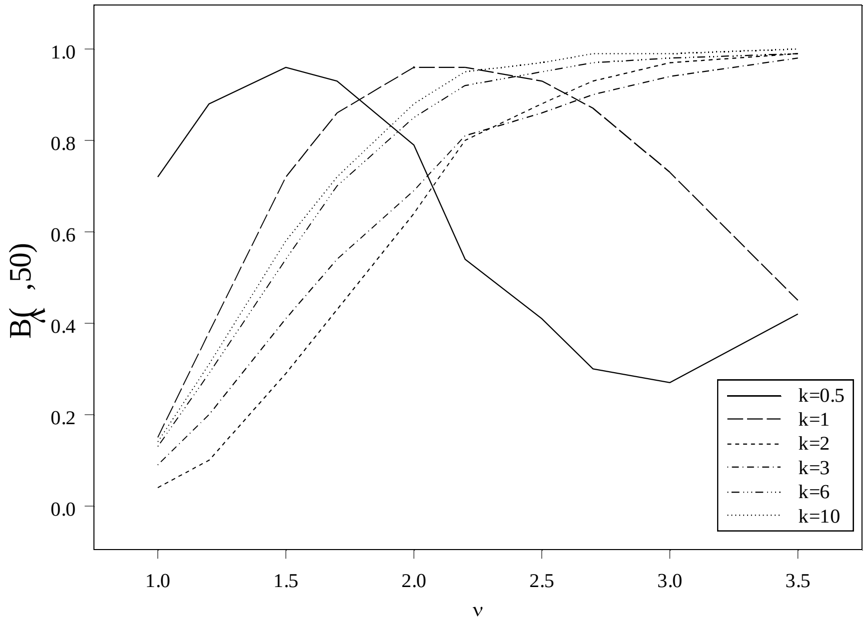


Figure 6.6: Behaviour of $B(\nu, n)$ measure for simulated data

Figure 6.7 shows the values of resistant constant ν for sample size $25 < n \leq 200$ and small concentration parameter $\kappa = 0.5$. At 0.05 significant level, the values of resistant constant ν vary between 1 and 1.7. Generally, at 0.1, 0.05 and 0.01 significant level all values of ν vary between 1 and 2. Thus, we recommend the use of $(1 \leq \nu \leq 2)$ for circular boxplot criterion when the concentration parameter is small.

Figures 6.8 and 6.9 show the percentile values of $B(\nu, n)$ measure for large concentration parameter κ =7 and κ =10, respectively. In both figures at 0.05 significant level, the values of resistant constant ν seem to be stationary for $5 \leq n \leq 55$ with respect to the remainder number after dividing sample size *n* by 4. Figures 6.8 and 6.9 suggest that it is appropriate to take into consideration the values of resistant constant to be between 2 and 2.7. Similar behaviour can be observed for $\alpha = 0.1$, but with the values of resistant constant ν vary within 1.5 and 2.2. The situation is different for $\alpha = 0.01$, where the cut points are fixed at $\nu = 3.5$ for $(5 < n < 25)$ and decreases to $\nu = 3$ or less for $(n \geq 25)$.

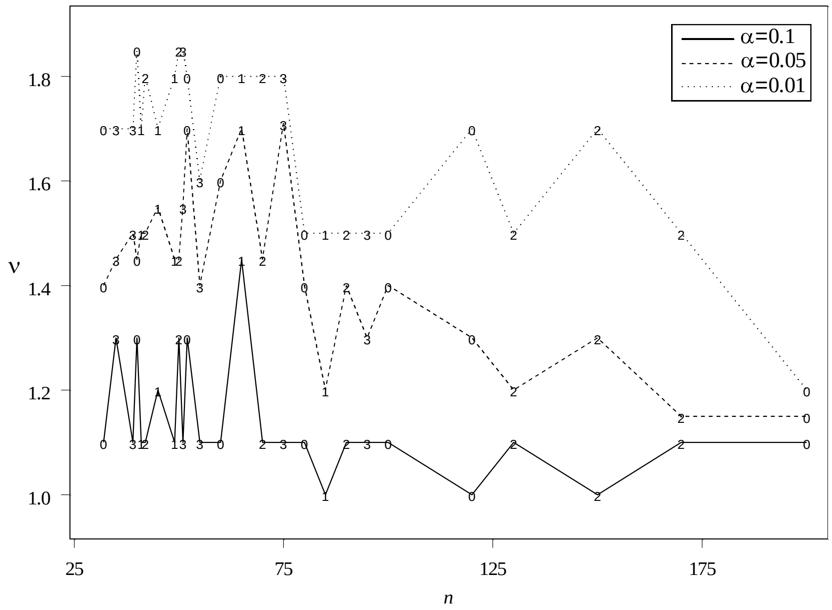


Figure 6.7: Percentiles of the resistant constant ν for different sample size, at $\kappa = 0.5$

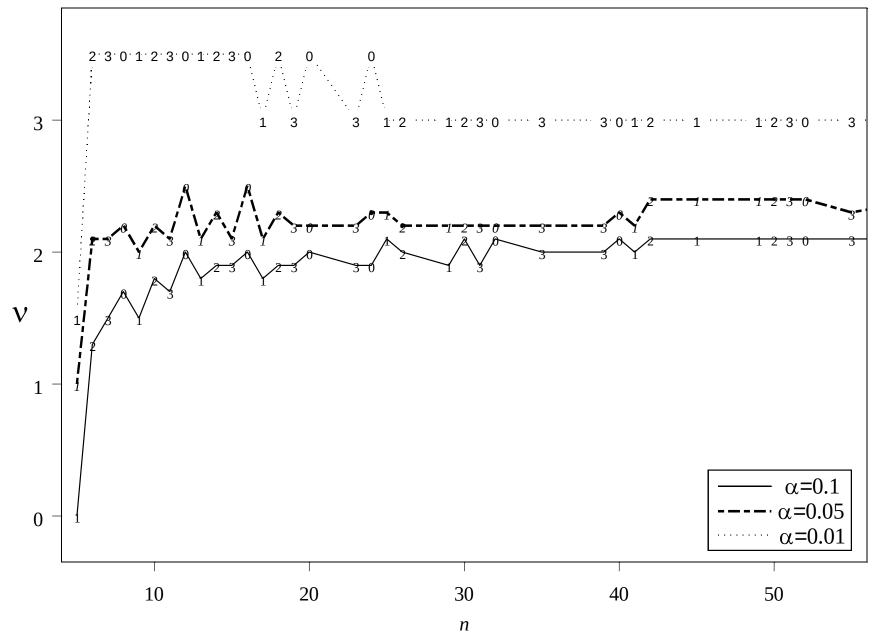


Figure 6.8: Percentile points for different sample size *n*, at κ =7

For further description of $B(\nu, n)$ measure, Figure 6.10 shows the scatter plot of $B(1.5, n)$ versus the sample size *n*. There is a decreasing linear relationship between the sample size *n* and $B(1.5, n)$ for $5 \le n \le 55$ with respect to each cluster of reminder after dividing sample size n by 4. Note that, the behaviour of $B(1.5, n)$ measure in Figure 6.10 is agreed with the discussion on the linear boxplot (see Hoaglin *et al.* (1986)).

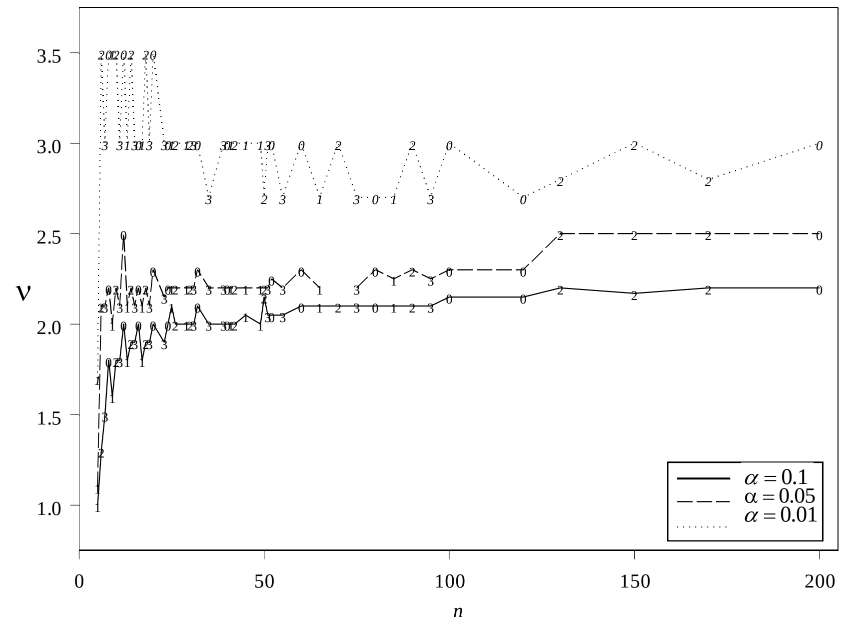


Figure 6.9: Percentile points for different sample size n, at κ =10

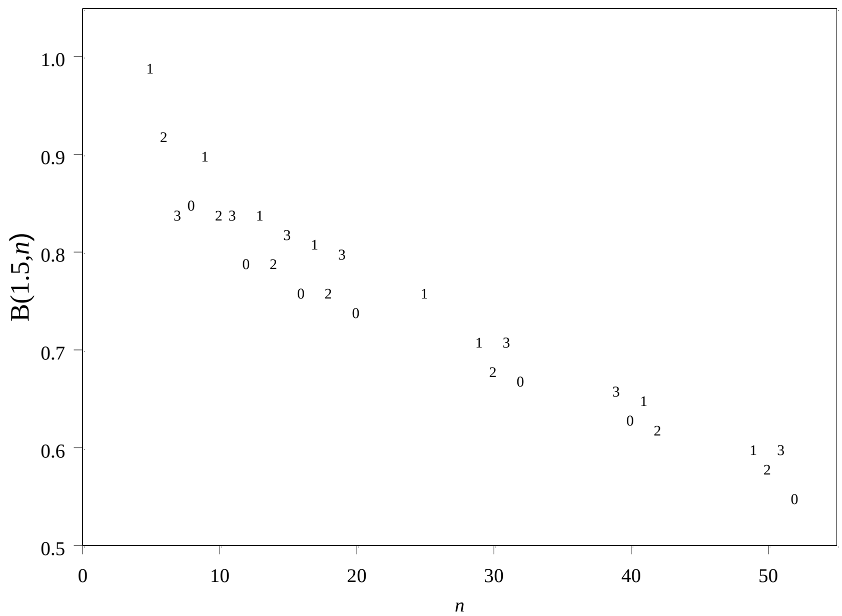


Figure 6.10: Simulation estimate of $B(1.5, n)$

(v) Description of B3(*v*,*n*) measure

$B3(\nu,n)$ is the probability that the von Mises sample of size n contains more than two observations outside the interval (L_F, U_F). Table 6.1 tabulate the probability of $B3(\nu,n)$ for different sample size n and concentration parameter κ =7. The probability of $B3(\nu,n)$ is displayed in Figure 6.11. It is obvious that $B3(\nu,n)$ is a decreasing function of the resistant constant ν. For small sample size ($n < 10$) the probability of $B3(\nu,n)$ is almost 0.

The results in Table 6.1 agree with the previous conclusion on the values of resistant constant ν for large concentration parameter and in order to identify more than two outliers, smaller values of resistant constant are recommended. Further, it is observed that $B3(1,n)$ approaches to asymptotic value 1 for a very large sample size *n*.

Table 6.1: $B3(\nu,n)$ measure for different sample size *n* at $\kappa = 7$

Resistant constant *v*	**Sample size *n***				
	7	**25**	**50**	**100**	**200**
1.0	0.0	0.15	0.46	0.81	0.98
1.2	0.0	0.07	0.22	0.49	0.83
1.5	0.0	0.02	0.06	0.14	0.37
1.7	0.0	0.01	0.02	0.05	0.14
2.0	0.0	0.0	0.01	0.01	0.02
2.2	0.0	0.0	0.0	0.0	0.0
2.5	0.0	0.0	0.0	0.0	0.0
2.7	0.0	0.0	0.0	0.0	0.0
3.0	0.0	0.0	0.0	0.0	0.0
3.5	0.0	0.0	0.0	0.0	0.0

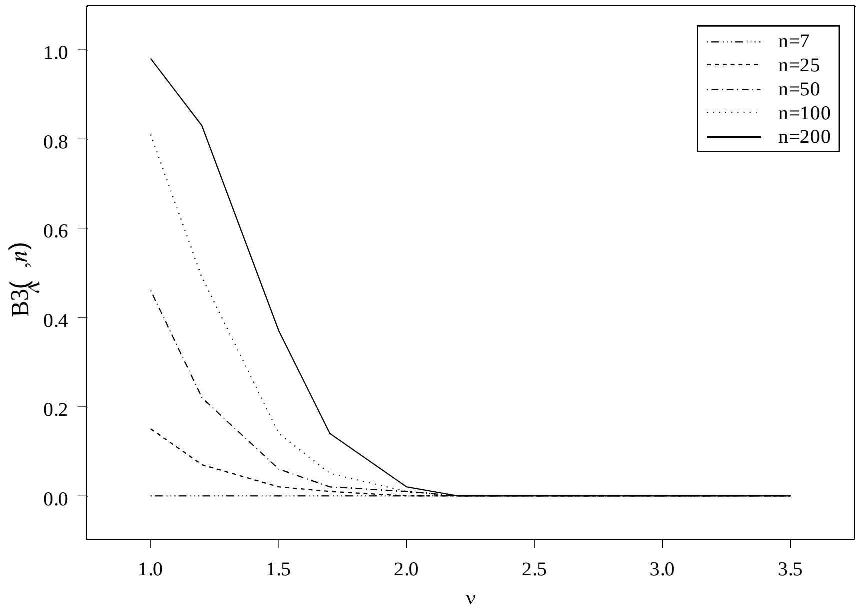


Figure 6.11: $B3(\nu, n)$ for different sample size *n* at concentration parameter, $\kappa = 7$

(vi) Description of A(*ν*,*n*) measure

$A(\nu, n)$ is the outside rate per observation in von Mises samples of size *n*. The simulation results (see Appendix A.4) show that $A(\nu, n)$ measure is a decreasing function of ν for large concentration parameter $\kappa > 3$. Due to overlapping problem $A(\nu, n)$ behaves similar to $1 - B(\nu, n)$ with respect to the values of resistant constant ν.

It is of interest to investigate the relationship between the function $A(\nu, n)$ and the sample size *n*. Figure 6.12 shows a nonlinear relationship between $A(1.5, n)$ and $(5 \le n \le 200)$. The function approaches to asymptotic value 0 for very large sample size. Figure 6.13 shows there is a linear relationship between simulation estimate of $A(1.5, -n^{-1})$ measure and $-n^{-1}$ for each reminder cluster.

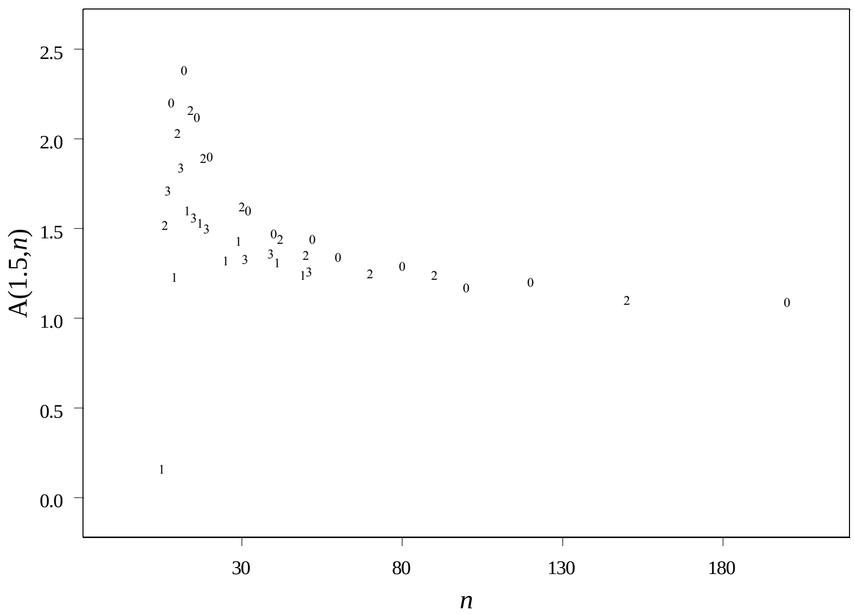


Figure 6.12: Simulation estimate of $A(1.5, n)$ measure

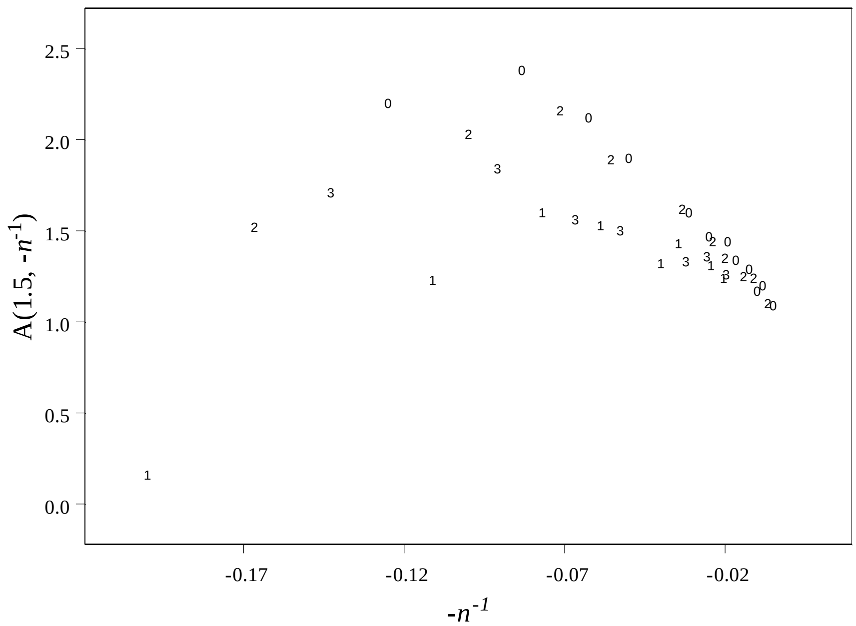


Figure 6.13: Simulation estimate of $A(1.5, -n^{-1})$ measure

6.4 Power of performance

The performance of any discordancy test can be examined by using the same approach described in Section 4.2. In order to study the performance of circular boxplot, 3000 samples based on different sample size n =5(5)20, 60 and 100, with concentration parameters κ =1, 5, 7 and 10 are considered. Samples are generated in such a way that $(n-1)$ observations come from $VM(\alpha,\kappa)$ and the other one observation generated from $VM(\alpha+\lambda\pi,\kappa)$, where λ is the degree of contamination and $0 \le \lambda \le 1$.

Based on simulation studies in Section 6.3, for small concentration parameter (i.e. κ =1), small values of ν are examined ($1 \le \nu \le 2$), while larger values of resistant value ν, ($2 \le \nu \le 2.7$) are used when κ is large, (i.e. κ =5, 7 and 10).

Figure 6.14(a) illustrates the plot of P1 for $n = 60$ and $\nu = 2$ for different concentration levels. It is shown that the power function P1 is an increasing function of concentration parameter κ. For small concentration parameter, the power of circular boxplot is weaker compared to the higher concentration parameter κ. This is due to the bounded range of the circle, and also, when the concentration parameter is small the observations tend to distribute uniformly. Hence, it becomes difficult to divide the circular sample into quartiles without covering all the circumferences of the circle which result in no outlier being detected. Figure 6.14(b) shows the plot of P1 for $\kappa = 10$ and $\nu = 2$ for different sample sizes, where the power of performance is also an increasing function of the sample size n. For small sample size $n = 5$, the power of circular boxplot does not overtake 50% at any level of concentration or any contamination level λ. The power of circular boxplot enhances gradually as the sample size n increases. Further, Figure 6.14 shows that the power of performance highly depends on the level of contamination λ. The complete results of the power of performance for circular boxplot are given in Appendix (A.5).

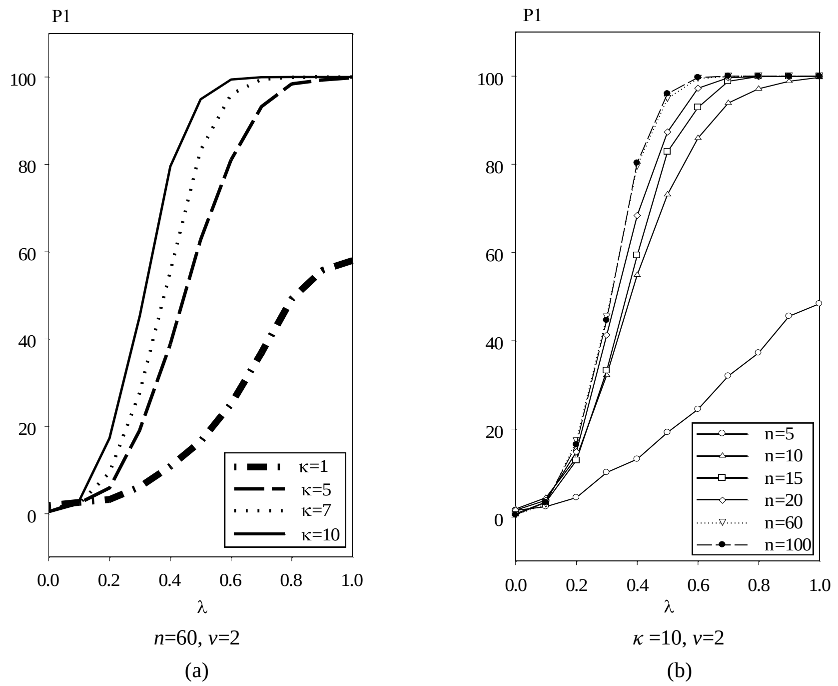


Figure 6.14: Power of performance of circular boxplot

6.5 Practical Example

In this section we consider frog direction data which was described in Section 3.3. We recall the summary statistics for frog data, The mean direction is $\bar{\theta} = 146.104^{\circ}$, the resultant length is $R = 10.1527$, the mean resultant length is $\bar{R} = 0.725$ and the maximum likelihood estimate of concentration parameter $\hat{\kappa} = 2.18$. Four tests which were discussed in previous sections will be applied to frog data to identify any possible outliers.

6.5.1 Identifying outliers by using *A* and Chord statistics

Table 6.2 summarizes the results of applying the *C, D* and *M* statistics which are the same as that found in Collett (1980). It can be seen that only *D* and *M* statistics have identified the observation with value 316° as an outlier. For *C* statistic, it is developed by considering the effect of outlier on the resultant length. Note that 316° is an extreme point since it has the largest circular deviation of magnitude 160.61°. The deletion of point 316° from original data changes the resultant length from $R = 10.15$ to $R_{(-14)} = 11.14$. Whereas, the deletion of any other single observation decreases the mean resultant length to around 9.2. The observed changes on the resultant length is however not large enough to be detected by the *C* statistic.

In the case of *A* statistic, the summation of all circular distances is $\sum_{j=1}^{14} D_j = 92.92$. By omitting point 316° from the data, the summation of all circular distances reduces to 68.92 while omitting any other single point reduces the summation of all circular distances to around 89. The greatest change is when 316° is omitted. Using equation (5.2), statistic $A = 9.23$. Comparing this value with the critical value as given in Table 5.1 (approximately 0.87), we reject the null hypothesis. It means that *A* statistic identifies observation 316° as an outlier. On the other hand, for chord statistic, the summation of all chord lengths is $\sum_{j=1}^{14} B_j = 149.18$. By omitting point 316° from the data, the summation of all chord lengths reduces to 124.23 while omitting any other single point reduces the summation of all chords lengths to around 140. The change is the greatest when 316° is omitted. Using equation (5.4) the chord statistic equals $\text{Chord} = 0.96$. By comparing the value with the critical value as given in Table 5.2

(approximately equals 0.93), thus, the null hypothesis is rejected and observation 316° is identified as an outlier.

Table 6.2: Results of applying different discordancy tests on the frog data

Statistic	Statistics' value	Observation	Critical value,95%	Conclusion
C	0.182	316°	0.2	Not outlier
D	0.78	316°	0.74	Outlier
M	0.52	316°	0.50	Outlier
Chord	0.96	316°	0.93	Outlier
A	0.92	316°	0.83	Outlier

6.5.2 Identifying outliers by using modified chord statistic

Table 6.3 gives the values of P_j for $j = 1,\ldots,14$. Observation 316° with corresponding number 14 has the maximum value of $N(\kappa d_j)$ with a percentage of 64.28% from its sample size, which can be considered as an outlier.

Table 6.3: The values of P_j for frog data set

Observation	1	2	3	4	5	6	7	8	9	10	11	12	13	14
P_j	7	7	7	7	7	7	7	7	7	0	0	0	0	64

6.5.3 Identifying outliers by using circular boxplot

The circular boxplot is used to detect possible outliers in the frog data. The estimated concentration parameter is $\hat{\kappa} = 2.18$, the estimated median direction is $\phi = 145^\circ$, first quartile is $Q_1 = 121^\circ$ the third quartile is $Q_3 = 192^\circ$, $CIQR = 71^\circ$ and the circular distance between the mean and median direction is 0.97° which confirms the robustness of the mean direction. Since $\hat{\kappa} = 2.18$, small values of resistant constant ν are used. Table 6.4 shows the observations detected as outliers in frog direction data by using different values of ν. Observation 316° is identified as an outlier when $\nu = 1$, 1.2, 1.5 and 1.7, while for larger values of ν none of the observations was identified as an outlier. Figure 6.15 shows the boxplot of frog direction data for $\nu = 1.5$, where the plot is obtained by using special subroutine developed in S-Plus environment.

Table 6.4: Summary of the outliers detected using several values of ν for frog data

ν	L_F	U_F	**Number of outliers**	**Outliers**
1.0	50.0°	263.0°	1	316°
1.2	35.8°	277.2°	1	316°
1.5	14.5°	289.5°	1	316°
1.7	0.3°	312.7°	1	316°
2.0	334.0°	339.0°	0	-
2.2	355.3°	317.7°	0	-
2.5	303.5°	9.5°	0	-
2.7	289.3°	23.7°	0	-
3.0	45.0°	268.0°	0	-
3.5	80.5°	232.5°	0	-

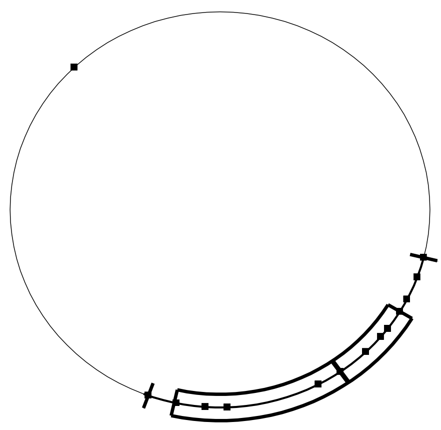


Figure 6.15: Circular boxplot of frogs direction, for ν =1.5

6.5.4 Discussion

A, chord, modified chord statistics and circular boxplot are able to identify observation value 316° as an outlier. These results are in agreement with the findings in

Collett (1980), where the D and M statistics succeeded in identifying observation 316° as an outlier, while C statistic failed to do so. Thus, we can conclude that the new proposed statistics provide alternative tests to identify outliers in circular sample.

6.6 Summary

The boxplot is a popular tool for explanatory data analysis. It was developed gradually over the past 40 years. There is no known structure of boxplot for circular variables. However, specifying the median direction, first and third quartiles solve part of the problem of constructing circular boxplot, while the determination of the upper and lower fences are more challenging because of the bounded range of the circle.

It is shown that the level of concentration parameter strongly affects the structure of circular boxplot. There are some interesting points being highlighted based on the simulation studies, such as the functional relationship between the *CIQR* and the large concentration parameter κ, which may be given by $CIQR \approx (\ln \kappa)^{-1}$ and the overlapping problem which may occur for $\nu > \pi \ln(\kappa) - 0.5$.

It is recommended to use different values of ν to identify possible outlier in circular variables. For samples with large concentration parameter $\kappa > 3$, it is appropriate to use different values of ν, which are $2 < \nu < 2.7$, while for samples with small concentration parameter $\kappa \leq 3$, the values of resistant rules can be chosen between 1 and 2. These values are comparable to the linear case in which ν equals to 1.5.

CHAPTER SEVEN

SIMPLE CIRCULAR REGRESSION MODEL AND ITS DIAGNOSTIC CHECKING

7.1 Introduction

In some cases the relationship between two circular variables can be fitted by a straight line. This chapter discusses the development of simple circular regression model, differences with other circular models, parameters estimates, its asymptotic properties and its applications. For diagnostic checking, we propose and examine a new practical definition of circular residuals based on circular distance. We apply different numerical tests and graphical tools to identify outliers in circular regression based on this circular residual.

7.2 Simple circular regression model

Gould (1969) proposed a regression model to predict the mean direction of circular response variable Θ from a vector of linear covariates $\boldsymbol{X} = (\boldsymbol{x_1}, \ldots, \boldsymbol{x_k})$, where Θ follows von Mises distribution with mean direction μ and concentration parameter κ. The model is given by

$$\mu = \mu_0 + \sum_{j=1}^{k} \beta_j x_j , \qquad j = 1, \ldots, k , \tag{7.1}$$

where μ_0 and β's are unknown parameters and $\boldsymbol{x}_j$ is a linear covariate. Mardia (1972) extended Gould's model (1969) by assuming θ_i, $i = 1, \ldots, n$, to be independently distributed from von Mises distribution with mean direction μ_i and unknown concentration parameter κ. The model is given by

$$\mu_i = \mu_0 + \beta t_i, \qquad i = 1,\ldots,n, \qquad (7.2)$$

where t_i are known numbers, while μ_0, β and κ are unknown parameters. It is important to mention that the explanatory variables in model (7.2) are linear variables.

Hussin *et al.* (2004) extended model (7.2) to the case when both response and explanatory variables are circular. For any circular observations $(x_1, y_1),\ldots,(x_n, y_n)$ of circular variables $\boldsymbol{X}$ and $\boldsymbol{Y}$ with a linear relationship between them, the proposed model is given by

$$y = \alpha + \beta x + \varepsilon \ (\bmod 2\pi), \qquad (7.3)$$

where ε is circular random error having a von Mises distribution with circular mean 0 and concentration parameter κ.

A restriction is imposed on the values of the parameter to ensure the practicality of the model. Consider the following four points (in radian): (0.10,0.90), (2.00,1.99), (4.30,4.63) and (6.25,6.24). The points are fairly close to the straight lines $y = x$ and $y = 4x \ (\bmod 2\pi)$. By maximizing the log likelihood function of model (7.3), there is a clear maximum at $\beta = 1$ in the interval $0.5 < \beta < 1.5$. Other local maximums are observed at $\beta = 4$ and $\beta = 129.7$. However, there is no practical interpretation for the last two values. Thus, the value close to one would be a logical and simpler choice. This model is considered in this study due to its simpler form compared to the other circular models.

7.2.1 Differences of other circular models and practical applications

Simple circular regression model is one of the circular regression models that consider the relationship between two circular variables. This model is specified for the case when the relationship between two circular variables is linear, with a set of assumptions about the circular error which are analogue to those in the linear regression.

Hussin *et al.* (2004) applied their model on the wind and wave direction data recorded by two different techniques, which are the HF radar system and the anchored wave buoy. The model can be used to fit the relationship between any two circular variables with a true linear relationship between them. Practically, it can be involved in the modelling of the relationship between any two instruments used to measure circular phenomena, or circular variables in different location or time. In such cases, whenever the comparison between two circular variables is required, diagnostic checking of the model and the detection of any possible outliers are necessary and important. The next subsection discusses the maximum likelihood estimates of the model parameters.

7.2.2 Maximum likelihood estimates and asymptotic properties of model parameters

Let $(x_1, y_1),..., (x_n, y_n)$ be pairs of circular observations, where $0 \le x_i, y_i < 2\pi$. Suppose the data is fitted using the simple circular regression model given by (7.3) and the resulting circular residuals follow von Mises distribution. The log likelihood function is given by

$$\log L(\alpha, \beta, \kappa; x_1,..., x_n, y_1,..., y_n) = -n\log(2\pi) - n\log I_0(\kappa) + \kappa \sum \cos(y_i - \alpha - \beta x_i).$$

The maximum likelihood estimate of parameter α is given by

$$\hat{\alpha} = \begin{cases} \tan^{-1}(S/C), & \text{if} \quad S \geq 0, C > 0, \\ \tan^{-1}(S/C) + \pi, & \text{if} \quad C < 0, \\ \tan^{-1}(S/C) + 2\pi, & \text{if} \quad S < 0, C \geq 0, \end{cases}$$

where $S = \sum \sin(y_i - \hat{\beta}x_i)$ and $C = \sum \cos(y_i - \hat{\beta}x_i)$, $\hat{\beta}$ is the MLE of β. Due to the nonlinear nature of the first partial derivative of the log likelihood function with respect to β, then it can be estimated iteratively according to the formula $\hat{\beta}_1 \approx \hat{\beta}_0 + \dfrac{\sum x_i \sin(y_i - \hat{\alpha} - \hat{\beta}_0 x_i)}{\sum x_i^2 \cos(y_i - \hat{\alpha} - \hat{\beta}_0 x_i)}$ by choosing a suitable initial value β_0. The estimation of concentration parameter is obtained by $\hat{\kappa} = A^{-1}\left(\dfrac{1}{n}\sum \cos(y_i - \hat{\alpha} - \hat{\beta}x_i)\right)$, where the function $A(.)$ is the ratio of the modified Bessel function for the first kind of order one, and first kind of order zero. One of the inverses of function $A(\omega)$ is approximated by Dobson (1978) and is given by $A^{-1}(\omega) \approx \dfrac{9 - 8\omega + 3\omega^2}{8(1-\omega)}$.

In order to assess the accuracy of the maximum likelihood estimators of model parameters, the asymptotic variances can be obtained by inverting the 3×3 Fisher information matrix and are given by

$$Var(\hat{\alpha}) = \frac{\sum x_i^2}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}},$$

$$Var(\hat{\beta}) = \frac{n}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}},$$

and
$$Var(\hat{\kappa}) = \frac{\hat{\kappa}}{n\{\hat{\kappa} - \hat{\kappa}A^2(\hat{\kappa}) - A(\hat{\kappa})\}}.$$

Further, the covariance of parameters are given by

$$\text{cov}(\hat{\alpha}, \hat{\beta}) = \frac{-\sum x_i}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}},$$

$$\text{cov}(\hat{\alpha}, \hat{\kappa}) = \text{cov}(\hat{\beta}, \hat{\kappa}) = 0 ,$$

where $A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\cos(y_i - \hat{\alpha} - \hat{\beta}x_i)$.

7.3 Circular error

The analysis of error term is considered to be as old as the subject of statistical modelling. Error or random disturbance is an essential term for any regression model. In the case of linear regression, errors are assumed to be random, independent, identically and normally distributed with mean zero and constant variance. Most of circular regression models do not give enough attention to the analysis of error term.

Hussin *et al.* (2004), assumed that the error is uncorrelated and has von Mises distribution with circular mean 0° and concentration parameter κ. This section proposes a new definition of circular residuals based on the circular distance. Numerical and simulation studies are carried out to investigate the properties of the proposed circular residuals.

The standard definition of residuals is given by $e_i = y_i - \hat{y}_i$, where y_i and $\hat{y}_i$ are the observed and predicted values respectively, which is no longer valid here because the variables are circular. For instance, let $y_i = 350^\circ$ and $\hat{y}_i = 5^\circ$, then $e_i = 350^\circ - 5^\circ = 345^\circ$, whereas the actual circular distance as defined in (2.9) is 15°.

Mardia (1972, p.128) defined the angular deviations of the observations from their fitted values by using the following statistic

$$1-\frac{1}{n}\sum_{i=1}^{n}\cos(y_i-\hat{y}_i)\,,$$

which correspond to the estimated residuals. The acceptance of this definition means that the estimation of circular residuals for the ith observation can be obtained by $e_i^*=1-\cos(y_i-\hat{y}_i)$. Note that e_i^* is linear and is bounded within the interval $[0,2]$. Hence, it is important to define new residuals of circular type so that the assumptions of error, such as whether they come from von Mises distribution, can be investigated.

7.3.1 New circular residuals

Consider the circular distance given by equation (2.9)

$$d_\circ(\phi,\theta)=\min(\phi-\theta,2\pi-(\phi-\theta))=\pi-\big|\pi-|\phi-\theta|\big|,$$

where $d_\circ(\theta,\hat{\theta})\in[0,\pi]$. The direct use of circular distance to obtain the circular residuals is not possible to satisfy every assumption of circular errors. For example, it is impossible to show the circular mean of such residuals to be zero. Moreover, the estimated concentration parameter also tends to increase as the residuals are distributed in the interval $[0,\pi]$ instead of the entire circumference.

In order to make the values of the residuals distributed around zero, we utilize the definition of absolute value

$$|x-z|=\begin{cases} x-z, & \text{if} \quad x\geq z,\\ -(x-z), & \text{if} \quad x<z.\end{cases}$$

Thus, we propose new circular residual based on circular distance as follows

$$r_{A_i}=\begin{cases} (\pi-|\,\pi-|\,y_i-\hat{y}_i\,||), & \text{if} \quad \hat{y}_i\leq y_i, y_i-\hat{y}_i\leq\pi \quad \text{or} \quad \hat{y}_i>y_i,\hat{y}_i-y_i>\pi,\\ -(\pi-|\,\pi-|\,y_i-\hat{y}_i\,||), & \text{if} \quad \hat{y}_i\leq y_i, y_i-\hat{y}_i>\pi \quad \text{or} \quad \hat{y}_i>y_i,\hat{y}_i-y_i\leq\pi.\end{cases} \tag{7.4}$$

Equation (7.4) can be written in a simpler form as follows

$$r_{A_i} = y_i - \hat{y}_i \,(\mathrm{mod}\, 2\pi). \tag{7.5}$$

Note that when $r_{A_i} > \pi \Rightarrow r_{A_i} = r_{A_i} - 2\pi$. It is obvious that from definition (7.4) and (7.5), the new residuals r_{A_i} are in the range $[-\pi, \pi]$.

Numerical and simulation studies are carried out in the following subsection in order to show that the circular distance residuals $\boldsymbol{r}_A$ are uncorrelated and follow von Mises distribution with circular mean 0 and concentration parameter κ.

7.3.3 Simulation study to investigate the circular error assumptions

Simulation studies are carried out to study the properties of the proposed circular distance residuals $\boldsymbol{r}_A$. Five different sample sizes are used: n =30, 50, 70,100 and 150 together with six different values of concentration parameter κ= 2, 5, 10, 30 and 50. For each sample size and concentration parameter, 2000 samples of circular errors ε are generated from von Mises distribution with circular mean 0° and concentration parameter κ. Another 2000 samples for $\boldsymbol{X}$ variable with similar sample size are generated from von Mises distribution with circular mean $\bar{\theta} = \pi/4$ and concentration parameter $\kappa = 1.5$. Without loos of generality, the parameters of model (7.3) are fixed at $\alpha = 0$ and $\beta = 1$. The observed values $\mathbf{Y}_s$ for each generated set $\mathbf{X}_s$ and ε_j are obtained based on model (7.3). The fitted values $\hat{\mathbf{Y}}_s$ and then the circular residuals, r_{A_s}, for $s = 1, \ldots, 2000$ are obtained based on (7.4). Then we estimate the mean direction $\bar{\theta}_s$ and the concentration parameter $\hat{\kappa}_s$ for all s.

Table 7.1 summarizes the results of simulation studies for all combinations of sample size n and concentration parameter κ. In each cell, the mean of circular error and concentration parameters are denoted by $\bar{\hat{\theta}}$ and $\bar{\hat{\kappa}}$, respectively together with their biases. The third row in each cell gives the proportion of residuals r_{A_s} that follow $VM(0,\kappa)$ at 0.05 level of significance based on U_n^2 statistic (see Section 3.3).

Further, we apply Durbin-Watson statistic (D.W.) to measure the autocorrelation between residuals for each r_{A_s}. The fourth row of each cell gives the proportion of r_{A_s} which is considered to be uncorrelated at 0.05 significant level.

Table 7.1: The results of simulations processes for circular residual $\boldsymbol{r}_A$.

n	κ	**2**	**5**	**10**	**30**	**50**
30	$\bar{\hat{\theta}}$ (bias)	6.282 (-0.000)	6.282 (-0.001)	6.283 (-0.001)	6.283 (0.000)	6.283 (0.000)
	$\bar{\hat{\kappa}}$ (bias)	2.112 (0.112)	5.152 (0.152)	11.259 (1.259)	31.976 (1.976)	51.875 (1.875)
	U_n^2	0.997	0.994	1.00	0.995	1.00
	D.W.	0.972	0.990	0.985	0.987	0.985
50	$\bar{\hat{\theta}}$ (bias)	6.283 (-0.000)	6.282 (-0.001)	6.283 (-0.001)	6.283 (0.000)	6.283 (0.000)
	$\bar{\hat{\kappa}}$ (bias)	1.939 (-0.061)	5.106 (0.106)	10.928 (0.928)	31.319 (1.319)	51.498 (1.498)
	U_n^2	0.992	0.995	0.991	1.00	0.993
	D.W.	0.959	0.993	0.979	0.986	0.987
70	$\bar{\hat{\theta}}$ (bias)	6.282 (-0.000)	6.282 (-0.002)	6.282 (-0.001)	6.284 (0.000)	6.283 (0.000)
	$\bar{\hat{\kappa}}$ (bias)	1.953 (-0.047)	4.918 (-0.082)	10.633 (0.633)	29.920 (-0.080)	50.881 (0.881)
	U_n^2	0.992	0.991	0.994	1.00	0.994
	D.W.	0.951	0.990	0.991	0.988	0.974
100	$\bar{\hat{\theta}}$ (bias)	6.282 (-0.000)	6.281 (-0.002)	6.282 (-0.001)	6.282 (-0.001)	6.283 (-0.001)
	$\bar{\hat{\kappa}}$ (bias)	1.99 (-0.01)	5.178 (0.178)	10.428 (0.428)	30.331 (0.331)	49.863 (-0.137)
	U_n^2	1.00	0.999	0.995	0.991	1.00
	D.W.	0.984	0.980	0.990	0.976	0.978
150	$\bar{\hat{\theta}}$ (bias)	0.001 (0.001)	6.284 (0.001)	6.281 (-0.002)	6.282 (-0.001)	6.285 (0.002)
	$\bar{\hat{\kappa}}$ (bias)	2.063 (0.063)	5.130 (0.130)	10.260 (0.260)	30.716 (0.716)	50.137 (0.137)
	U_n^2	1.00	0.995	1.00	0.994	1.00
	D.W.	0.974	0.996	0.991	0.988	0.984

Results in Table 7.1 show that all the estimated mean directions $\bar{\hat{\theta}}$ are very close to 0 (or equivalently 2π) and the biases vary between -0.002 to 0.002 radians (or -1.115° to 1.115°). The biases of the estimated mean of concentration parameter also vary between -0.136 to 1.875, where the large biases correspond to large values of the concentration parameter. These suggest that the estimated mean directions and concentration parameters do not differ much from the original values. On the other hand, the proportions of $\boldsymbol{r}_A$ that follows $VM(0,\kappa)$ based on U_n^2 statistic are close or equal to 1, while the proportions of $\boldsymbol{r}_A$ that have insignificant Durbin-Watson statistic are always greater than 0.95. These results suggest that the proposed circular residuals are uncorrelated and follow $VM(0,\kappa)$. In other words, the new circular residuals, $\boldsymbol{r}_A$, has sufficient properties to be used for simple circular regression model.

7.4 Goodness-of-fit for simple circular regression model

Lund (1999) assessed the goodness-of-fit of the least circular distance regression model described in Subsection (2.2.2) by using the function $A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\cos[y_i - \mu(\phi_i, \mathbf{X}_i, \hat{\beta}_1, \hat{\beta}_2)]$ as an analogue of residuals sums of squares in linear regression model. Thus, the $A(\hat{\kappa})$ for the simple circular regression model is given by

$$A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\cos(y_i - \hat{y}_i),$$

$$= \frac{1}{n}\sum_{i=1}^{n}\cos(y_i - \hat{\alpha} - \hat{\beta}x_i).$$

where $A(\hat{\kappa}) \in [-1,1]$. It is known that $\cos(y_i - \hat{y}_i)$ is a non-monotone function, where

$$\cos(y_i - \hat{y}_i) = \begin{cases} \geq 0, & \text{if} \quad |y_i - \hat{y}_i| \leq \frac{\pi}{2}, \\ < 0, & \text{if} \quad |y_i - \hat{y}_i| > \frac{\pi}{2}. \end{cases}$$

Due to the differences of signs, some of the terms may vanish which lead to losing some important information. Thus, we suggest to improve the goodness-of-fit test by squaring the $\cos(y_i - \hat{y}_i)$ function to be more consistent with residuals sums of squares. The modified statistic becomes

$$A^*(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}[\cos^2(y_i - \hat{\alpha} - \hat{\beta}x_i)]$$

$$= \frac{1}{n}\sum_{i=1}^{n}[\cos^2(y_i - \hat{y}_i)],$$

where $A^*(\hat{\kappa}) \in [0,1]$. Therefore, the closer $A^*(\hat{\kappa})$ to 1 indicates a better goodness-of-fit of the model. In addition, Lund (1999) also mentioned various circular correlation measures which could be applied to the observed and fitted values. For a random sample $(x_1, y_1), \ldots, (x_n, y_n)$, the simplest measure is proposed by Jammalamadaka & Sarma (1988) and given by

$$r_c = \frac{\sum_{i=1}^{n}\sin(x_i - \bar{x})\sin(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}\sin^2(x_i - \bar{x})\sin^2(y_i - \bar{y})}},$$

where $\bar{x}$ and $\bar{y}$ are the sample mean directions. Consequently, squaring r_c gives an analogy to the coefficient of determination, R^2, for linear regression.

7.5 Diagnostic checking of simple circular regression model

Residuals analysis has been widely used in investigating the adequacy of fitted model. In this section we extend it to the circular regression case by using different graphical and numerical methods.

7.5.1 Graphical tools

As shown in the literature, there are few techniques available for diagnostic checking in circular regression. Jammalamadaka & SenGupta (2001) suggested several graphical techniques for diagnostic checking in circular samples, such as, circular distance plot between the observations of circular sample, P-P plot and Q-Q plot. Furthermore, circular boxplot is proposed in Chapter 6 can play an important role to detect possible outliers in the circular residuals. Some of the techniques used in the linear regression can be extended to the circular case.

(i) Circular residuals on circumferences of circle

It could be easier to identify visually any points that are relatively far from the rest of other points by plotting the points on the circumferences of a circle.

(ii) Circular residuals versus the observation index

It is a linear plot for the circular residuals versus the observation index, which enable us to investigate the randomness property and to detect any possible outliers. Those points which are inconsistent with the other points are candidates to be considered as outliers.

(iii) Spoke plot of the fitted and observed values

The long line between the observed observation y_i and its fitted value $\hat{y}_i$ indicates a possibility of the *i*th observation being an outlier.

7.5.2 Numerical tests

Applying more than one test to detect outliers is recommended. There is no known published works on the detection of outliers in circular regression. Beside the graphical techniques described in the previous subsection, there are three numerical tests that can be used to identify outliers in univariate circular sample; *C*, *D* and *M* statistics. We can also use the proposed numerical tests presented in Chapter 5. These numerical tests may be extended to detect outliers in circular regression based on the circular residuals.

7.6 Numerical example (wind direction data)

The scatter plot of wind direction data in Figure 3.6 shows a linear relationship between the measurements of HF radar and anchored buoy with slope close to 1 and negligible intersect. Since both measurements of variables are circular, we fit the wind direction data using model (7.3).

7.6.1 Estimation of the model parameters and calculation of circular residuals

The maximum likelihood estimates of the parameters are obtained and given by $\hat{\alpha} = 0.164, \hat{\beta} = 0.973$ and $\hat{\kappa} = 7.338$. Hence, the fitted model is given by

$$\hat{y}_i = 0.164 + 0.973 x_i \quad (\text{mod } 2\pi).$$

The circular distance residuals r_{A_i} are obtained using (7.4). The estimated mean direction and concentration parameter of the residuals are $\hat{\mu} = 0.017$ and $\hat{\kappa} = 7.338$ respectively. The measures of the goodness-of-fit are $A(\hat{\kappa}) = 0.929$ and $A^*(\hat{\kappa}) = 0.908$, and the square of the circular correlation coefficient between the response and predicted values is $r_c^2 = 0.908$. These suggest that the model fits the data well.

7.6.2 Graphical tools

Figures 7.1 to 7.4 give the plots for diagnostic purposes. When the residuals are plotted on a circumference of a circle and index plot as shown in Figures 7.1 and 7.2 respectively, we can observed that two residuals points corresponding to observation 38 and 111 are inconsistent with the rest of the residuals. Plot of circular distance in Figure 7.3 also shows that the circular distance between observations 38 as well as 111 and their neighbours at both sides are longer than circular distance between any other observations and their neighbours. Similarly, all the quantile points in Figure 7.4 are close to the straight line, except for these two observations. While in Figure 7.5, there are two obvious long lines also correspond to the same observation crossing the inner ring of the spoke plot. Generally, the graphical tools suggest that observations numbered 38 and 111 as possible outliers.

In order to obtain the circular boxplot for the circular residuals, we calculate its summary statistics. The mean direction of circular residuals is $\hat{\mu} = 0.017$ and the estimated concentration parameter is $\hat{\kappa} = 7.338$. The estimated median direction is $\phi =$ 0.0072, the first quartile is $Q_1 = 0.202$, the third quartile is $Q_3 = 6.125$, giving the *CIQR* = 0.360. Since $\hat{\kappa} = 7.338$ is considered to be large enough, larger values of the resistant constant ν are recommended.

Table 7.2 shows the observations detected as outliers in wind direction data set by using different values of resistant constant ν. Observations numbered 38 and 111 are identified as outliers for all values of resistant constant ν considered. For smaller values of ν, other observations are also identified as outliers. Figure 7.6 shows the boxplot of circular residuals of wind direction data for $\nu = 2.5$. It is obvious that the two points 38 and 111 are identified as outliers.

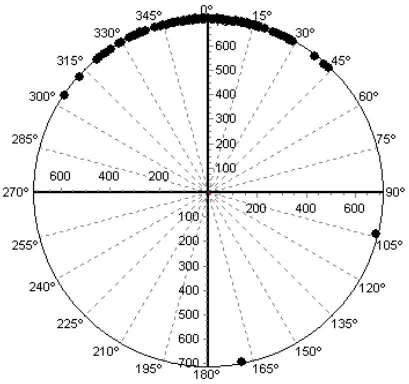


Figure 7.1: Circular residuals on circle circumferences

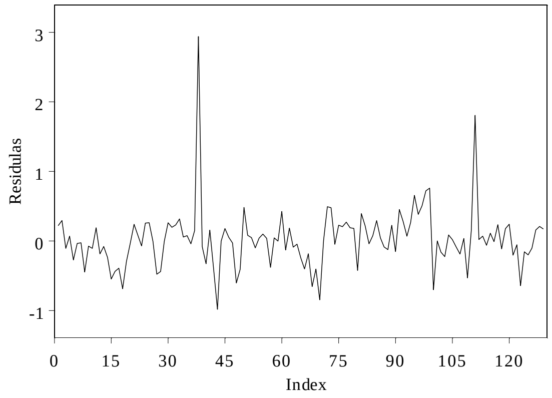


Figure 7.2: Circular residuals versus observations index

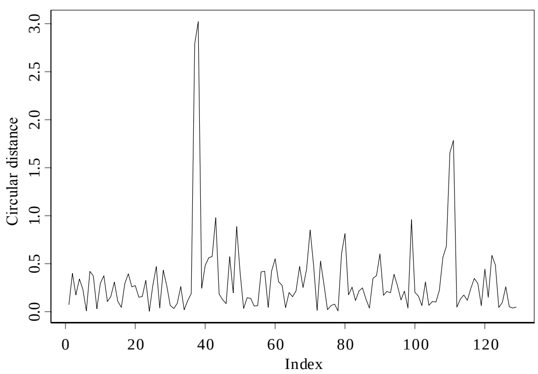


Figure 7.3: Circular distance between circular residuals

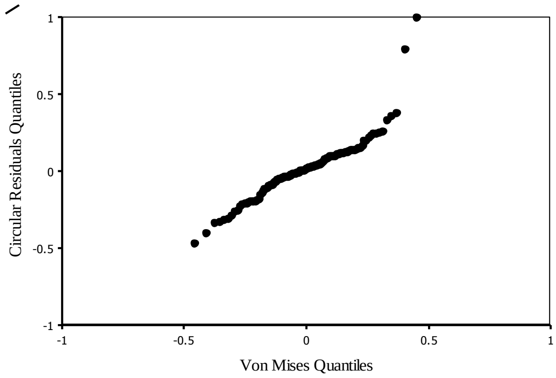


Figure 7.4: Q-Q plot for circular residuals

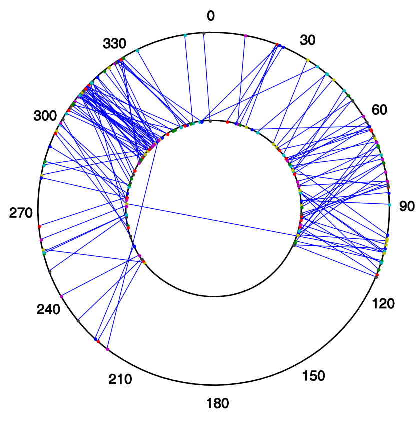


Figure 7.5: Spoke plot for the fitted and observed values

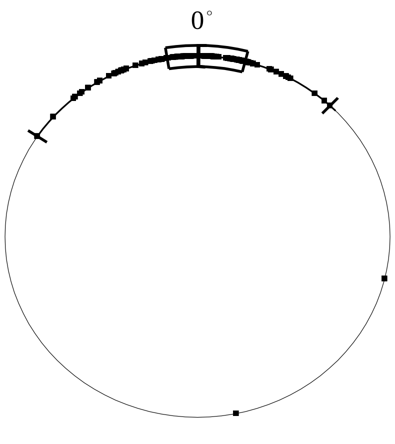


Figure 7.6: Circular boxplot of wind residuals for $\nu = 2.5$

Table 7.2: Summary of the outliers detecting by several values of ν for circular residuals of wind data

ν	L_F	U_F	**Number of outliers**	**Outliers**
1.0	0.561	5.765	14	15,18,38,43,48,68,70,95,98,99,100,109,111,123.
1.2	0.633	5.693	12	18,38,43,48,68,70,95,98,99,100,111,123.
1.5	0.741	5.586	6	38,43,70,99,100,111.
1.7	0.813	5.514	4	38,43,70,111.
2.0	0.921	5.406	3	38,43,111.
2.2	1.029	5.298	2	38,111.
2.5	1.101	5.226	2	38,111.
2.7	1.173	5.154	2	38,111.
3.0	1.281	5.046	2	38,111.
3.5	1.461	4.866	2	38,111.

7.6.3 Numerical methods

Table 7.3 summarizes the results by applying different discordancy tests on the circular distance residuals r_{A_i}. All of the six tests have successfully identified observation number 38 as the only outlier. After removing observation 38 from the data, only the C statistic does not detect observation 111 as an outlier. Figure 7.7 displays the values of P_j for the residuals. There are two points with the highest values of P_j corresponding to observations number 38 and 111.

Table 7.3: Results by applying different discordancy tests on wind direction data

Statistic	**Statistics' value**	**Observation**	**Critical value,95%**	**Conclusion**
C	0.0160	38	< 0.016	Outlier
	0.01	111	< 0.016	Not an outlier
D	0.480	38	0.13	Outlier
	0.299	111	0.13	Outlier
M	0.217	38	< 0.01	Outlier
	0.174	111	< 0.01	Outlier
A	0.963	38	0.47	Outlier
	0.616	111	0.37	Outlier
Chord	0.972	38	0.68	Outlier
	0.774	111	0.60	Outlier

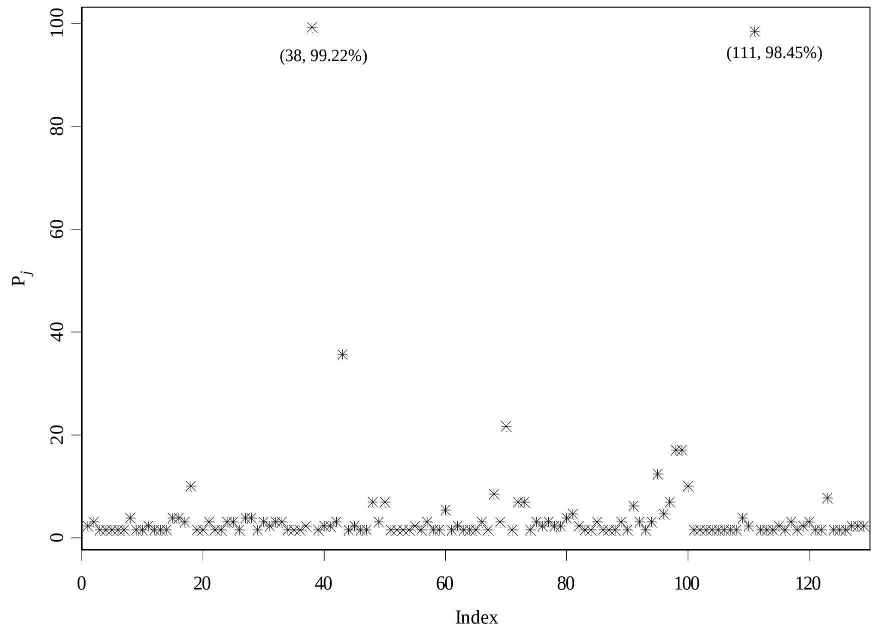


Figure 7.7: The values of P_j for the residuals of wind direction data

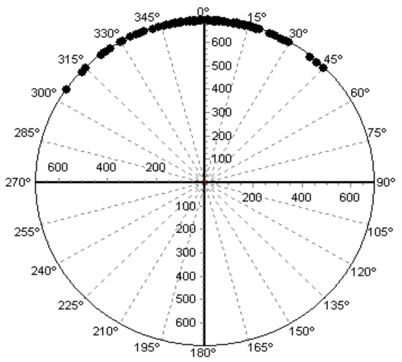
7.6.4 The effect of outliers on the estimation

Table 7.4 summarizes the effect of excluding the outliers on the parameter estimates. The removal of observation numbers 38 and 111 does not significantly change the value of $\hat{\alpha}$ and $\hat{\beta}$. However, both values are getting closer to 0 and 1, respectively. Furthermore, without the outlier, the estimated concentration parameter has increased from 7.338 to 11.010 and $A(\hat{\kappa})$ is increased from 0.929 to 0.953, as well as $A^*(\hat{\kappa})$ increased from 0.908 to 0.915. The r_c^2 increased from 0.908 to 0.955. Thus, the estimation is more accurate and we may have better model fitting for the data when observation 38 and 111 are excluded from the data set.

Figure 7.8 shows the diagnostic checking plots of circular residuals after removing observations number 38 and 111 from the wind direction data. The four plots suggest that the residuals points are distributed within an acceptable range.

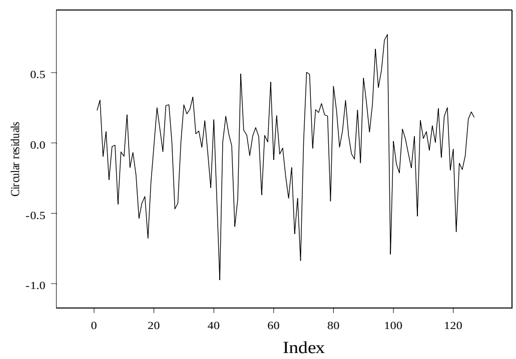
Table 7.4: Summary of the effect of outlier removal

	Full data	**Excluding observation number 38**	**Excluding observation numbers 38 and 111**
$\hat{\alpha}$	0.164	0.159	0.153
$\hat{\beta}$	0.973	0.974	0.974
$\hat{\kappa}$	7.338	9.229	11.01
r_c^2	0.908	0.954	0.955
$A(\hat{\kappa})$	0.929	0.944	0.953
$A^*(\hat{\kappa})$	0.908	0.909	0.915



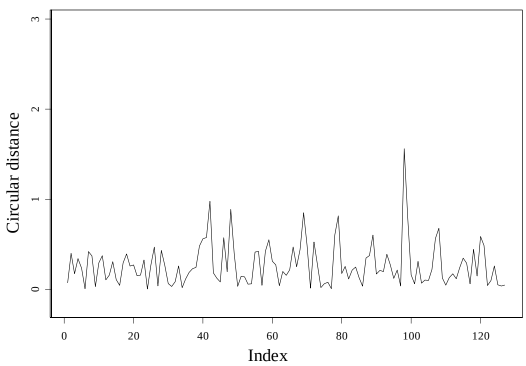
(a)

Circular residuals on circumferences



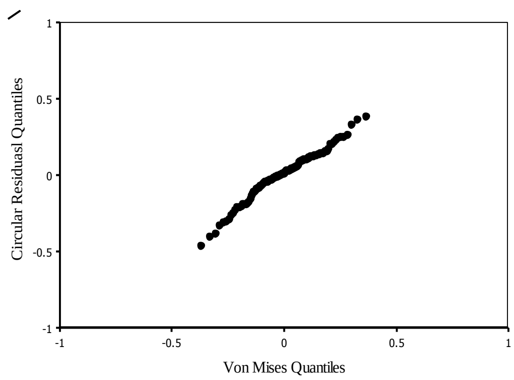
(b)

Circular residuals versus observations index



(c)

Circular distance



(d)

Q-Q plot

Figure 7.8: Diagnostic graphical tools for circular residuals without observations number 38 and 111

7.7 Summary

The simple circular regression model is considered in this chapter and new circular residuals are defined. The proposed residuals based on the circular distance can be used to check the adequacy of the fitted model by investigating the assumption made about the error. Several numerical tests and graphical techniques are utilized to identify outliers in circular regression based on the circular residuals. Observations number 38 and 111 have been identified as outliers when applied on wind direction data. The exclusion of these two observations from the original data set improves the goodness-of-fit for the model. In the following chapter, other statistics are proposed to identify outliers in simple circular regression using row deletion approach.

CHAPTER EIGHT

IDENTIFICATION OF INFLUENTIAL OBSERVATIONS IN SIMPLE CIRCULAR REGRESSION MODEL

8.1 Introduction

The existence of outliers in statistical data may indicate failure of the model or point to an unanticipated phenomenon. Hoaglin *et al.* (1986, p.991) stated that “It is still informative, however, and may be important, to examine samples and residuals for presence of outliers or exotic values”. In regression, interest focus on the outlier which is influential. There are extensive literatures available on statistical tests to identify influential observations in linear regression. On the other hand, the absence of such tests in circular case motivates us to develop numerical tests to identify influential observations in circular regression models. One of the possible ways is the row deletion approach, which was first developed by Belsley *et al.* (1980) for linear regression models. It investigates the impact of deleting one row at a time from both $\boldsymbol{X}$ matrix and $\boldsymbol{Y}$ vector on the estimated parameters, fitted values and residuals. Some of these statistics have been reviewed in Section 2.4.

This chapter presents five new numerical tests to detect the influential observations in simple circular regression model. Two of them are based on the difference between observed and fitted values. Another two are developed based on the approximate distribution for the mentioned statistics. The fifth numerical test is based on the *COVRATIO* statistic which is an analogy to the linear case.

8.2 Mean circular error statistics

8.2.1 Development of mean circular error statistics

Rao (1969) defined the circular distance between two circular observations θ_i and θ_j as $d_{ij} = 1 - \cos(\theta_i - \theta_j)$, where d_{ij} is a monotone increasing function of $(\theta_i - \theta_j)$ and $d_{ij} \in [0,2]$. Mardia (1972, p.128) defined the angular deviation of observations from their fitted values for circular regression model. In this section we use this statistic for the detection of influential observations in the simple circular regression model (7.3) by using row deletion approach. Let the statistic be known as mean circular error $MCEc$ and given by

$$MCEc = 1 - \frac{1}{n}\sum_{i=1}^{n}\cos(y_i - \hat{y}_i), \tag{8.1}$$

where n is the sample size and $MCEc \in [0,2]$.

If an observation y_i is an outlier then the circular distance between y_i and its associated fitted value $\hat{y}_i$ is expected to be relatively large. Thus, the existence of such observation in a data set will increase the summation of all circular distances as well as the value of $MCEc$ statistic. Consequently, the removal of the *i*th observation denoted by $MCEc_{(-i)}$ from the data set will decrease the value of the statistic. Let the maximum absolute difference between the value of the statistics for full and reduced data sets be

$$DMCEc = \max_i \{| MCEc - MCEc_{(-i)} |\}.$$

The *i*th observation is identified as an influential observation if $DMCEc$ exceeds a pre-specified cut-off points.

Using circular distance in (2.9) allows us to use *sine* function as an alternative measure of mean circular error, where *sine* is an increasing function on the interval $[0, \pi/2]$. Therefore, an alternative statistic is

$$MCEs = \frac{1}{n}\sum_{i=1}^{n}\sin\left(\frac{d_i}{2}\right), \tag{8.2}$$

where $d_i = \pi - \left|\pi - |y_i - \hat{y}_i|\right|$ is the circular distance between y_i and $\hat{y}_i$, with sample size *n* and $MCEs \in [0,1]$.

Analogy to the $MCEc$ statistic, if an observation y_i is an outlier, then the half of the circular distance $d_i/2$ is expected to be relatively large compared to other $d_i/2$'s. Thus, the existence of such observation increases the value of $MCEs$. Consequently, the removal of y_i decreases the value of $MCEs$ and denoted by $MCEs_{(-i)}$. Let the maximum absolute difference between the value of the statistics for full and reduced data sets be

$$DMCEs = \max_i \{| MCEs - MCEs_{(-i)} |\}.$$

The *i*th observation is identified as an influential observation if $DMCEs$ exceeds a pre-specified cut-off points.

$MCEc$ and $MCEs$ statistics are considered as a sort of arithmetic means which is not resistant to the existence of outliers. Thus, it can be used to detect possible influential observation in circular regression. It is expected that both statistics are more powerful for small sample size n, because the estimated mean of smaller samples is more sensitive to the existence of outlier rather than larger samples.

8.2.2 Percentage points of mean circular error statistics

(i) Description of simulation process

A series of simulation studies is carried out to find the percentage (cut-off) point of $DMCEc$ and $DMCEs$ statistics by using Monte Carlo methods. Fifteen different sample sizes are used which are n=10(10)150. For each sample size n, a set of circular random error from von Mises distribution with mean direction 0 and various values of concentration parameter κ are generated, where $\kappa =$ 5, 10, 30 and 50. Samples of von Mises distribution $VM(\pi/4, 1.5)$ with corresponding size n are generated to represent the values of $\boldsymbol{X}$ variable. Without loss of generality, the parameters of model (7.3) are fixed at $\alpha = 0$ and $\beta = 1$. Observed values of response variable $\boldsymbol{Y}$ are calculated based on model (7.3) and consequently the fitted values $\hat{\boldsymbol{Y}}$ are obtained.

We then compute the value of $MCEc$ and $MCEs$ statistics for full data set. Sequentially, we exclude the ith observation from the generated sample, for $i = 1, \ldots, n$. We refit the reduced data using model (7.3) and then calculate the values of $MCEc_{(-i)}$ and $MCEs_{(-i)}$ statistics as well as the values of $DMCEc$ and $DMCEs$ statistics. The process is repeated 2000 times for each combination of sample size n and concentration parameter κ.

The percentage points of $DMCEc$ and $DMCEs$ statistics for each sample size n and concentration parameter κ are tabulated in Tables 8.1 and 8.2, respectively. The 10, 5 and 1 percentage points are given in the first, second and third rows of each table respectively.

Table 8.1: The 10, 5 and 1 percentage points of the null distribution of $DMCEc$ statistic

n	Percentage	κ			
		5	**10**	**30**	**50**
10	10%	0.0666	0.0457	0.0215	0.0176
	5%	0.0997	0.0650	0.0259	0.0204
	1%	0.1279	0.0923	0.0318	0.0281
20	10%	0.0389	0.0181	0.0106	0.0091
	5%	0.0471	0.0217	0.0125	0.0098
	1%	0.0721	0.0302	0.0159	0.0126
30	10%	0.0302	0.0132	0.0078	0.0059
	5%	0.0358	0.0155	0.0087	0.0066
	1%	0.0515	0.0200	0.0098	0.0076
40	10%	0.0230	0.0112	0.0056	0.0045
	5%	0.0272	0.0126	0.0062	0.0049
	1%	0.0354	0.0166	0.0073	0.0056
50	10%	0.0186	0.0087	0.0046	0.0036
	5%	0.0212	0.0099	0.0051	0.0039
	1%	0.0269	0.0129	0.0062	0.0048
60	10%	0.0165	0.0076	0.0038	0.0031
	5%	0.0193	0.0086	0.0042	0.0034
	1%	0.0243	0.0109	0.0049	0.0039
70	10%	0.0149	0.0067	0.0034	0.0026
	5%	0.0178	0.0073	0.0038	0.0028
	1%	0.0219	0.0102	0.0048	0.0033
80	10%	0.0132	0.0060	0.0030	0.0023
	5%	0.0153	0.0068	0.0032	0.0025
	1%	0.0204	0.0087	0.0040	0.0028
90	10%	0.0111	0.0056	0.0026	0.0021
	5%	0.0128	0.0063	0.0028	0.0023
	1%	0.0158	0.0082	0.0034	0.0026
100	10%	0.0111	0.0052	0.0024	0.0019
	5%	0.0127	0.0062	0.0026	0.0020
	1%	0.0156	0.0071	0.0030	0.0023
110	10%	0.0095	0.0045	0.0022	0.0017
	5%	0.0111	0.0049	0.0024	0.0019
	1%	0.0137	0.0071	0.0027	0.0022
120	10%	0.0094	0.0044	0.0020	0.0016
	5%	0.0107	0.0049	0.0022	0.0017
	1%	0.0147	0.0058	0.0025	0.0020
130	10%	0.0086	0.0040	0.0019	0.0014
	5%	0.0100	0.0046	0.0020	0.0016
	1%	0.0130	0.0057	0.0024	0.0018

Table 8.1, continued.

n	Percentage	κ			
		5	**10**	**30**	**50**
	10%	0.0085	0.0037	0.0017	0.0014
140	5%	0.0099	0.0045	0.0018	0.0015
	1%	0.0132	0.0060	0.0022	0.0018
	10%	0.0075	0.0036	0.0016	0.0013
150	5%	0.0084	0.0040	0.0018	0.0014
	1%	0.0101	0.0048	0.0020	0.0016

(ii) Discussion

Tables 8.1 and 8.2 show the estimated percentage points for $DMCEc$ and $DMCEs$ statistics. The estimated percentage (cut-off) points for $DMCEc$ and $DMCEs$ statistics are decreasing functions of the concentration parameter κ. It is noticeable that the percentages are decreasing functions of the sample size n for any level of concentration parameter κ. Figure 8.1 shows the behaviour of $DMCEc$ and $DMCEs$ statistics for different sample size n when $\kappa = 10$ for $\alpha = 0.05$. The percentage points become almost constant for large sample size n, (i.e. $n > 100$).

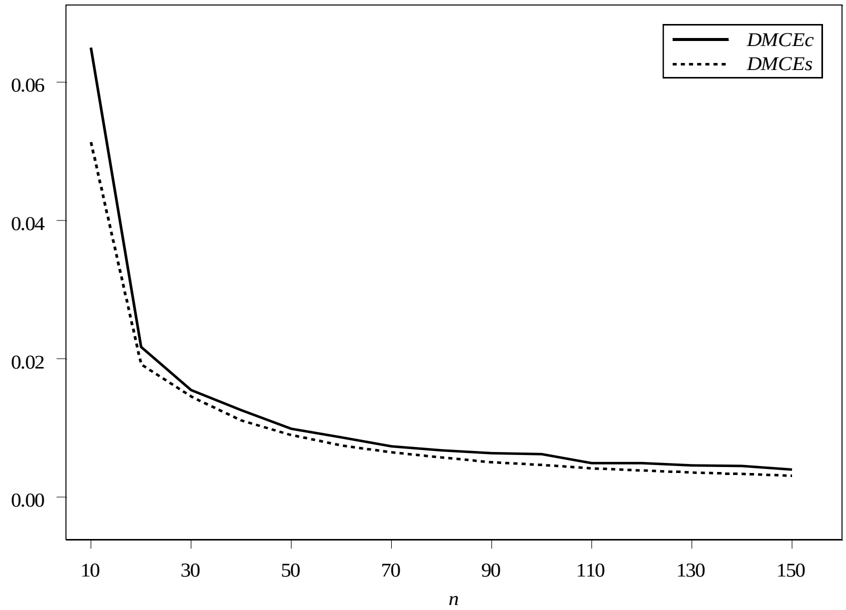


Figure 8.1: Percentage points of $DMCEc$ and $DMCEs$ statistics for $\kappa = 10$, $\alpha = 0.05$

Table 8.2: The 10, 5 and 1 percentage points of the null distribution of *DMCEs* statistic

		κ			
n	Percentage	**5**	**10**	**30**	**50**
	10%	0.0605	0.0365	0.0100	0.0060
10	5%	0.0766	0.0513	0.0125	0.0073
	1%	0.1068	0.0868	0.0483	0.0316
	10%	0.0286	0.0168	0.0057	0.0036
20	5%	0.0333	0.0192	0.0067	0.0044
	1%	0.0461	0.0277	0.0091	0.0063
	10%	0.0199	0.0128	0.0046	0.0026
30	5%	0.0219	0.0145	0.0055	0.0031
	1%	0.0266	0.0173	0.0064	0.0048
	10%	0.0144	0.0103	0.0033	0.0020
40	5%	0.0157	0.0111	0.0038	0.0022
	1%	0.0198	0.0136	0.0051	0.0032
	10%	0.0116	0.0078	0.0029	0.0017
50	5%	0.0129	0.0090	0.0033	0.0019
	1%	0.0150	0.0103	0.0041	0.0027
	10%	0.0099	0.0067	0.0025	0.0015
60	5%	0.0108	0.0075	0.0028	0.0017
	1%	0.0124	0.0085	0.0036	0.0022
	10%	0.0085	0.0058	0.0022	0.0013
70	5%	0.0095	0.0064	0.0026	0.0015
	1%	0.0105	0.0074	0.0034	0.0018
	10%	0.0077	0.0053	0.0020	0.0011
80	5%	0.0083	0.0057	0.0022	0.0013
	1%	0.0095	0.0064	0.0030	0.0016
	10%	0.0065	0.0047	0.0017	0.0011
90	5%	0.0071	0.0050	0.0020	0.0013
	1%	0.0080	0.0059	0.0025	0.0016
	10%	0.0062	0.0043	0.0016	0.0010
100	5%	0.0066	0.0046	0.0018	0.0011
	1%	0.0076	0.0054	0.0022	0.0013
	10%	0.0055	0.0038	0.0015	0.0009
110	5%	0.0060	0.0041	0.0017	0.0010
	1%	0.0068	0.0047	0.0021	0.0013
	10%	0.0053	0.0036	0.0014	0.0008
120	5%	0.0057	0.0038	0.0016	0.0009
	1%	0.0065	0.0045	0.0022	0.0012
	10%	0.0047	0.0033	0.0013	0.0008
130	5%	0.0050	0.0035	0.0015	0.0009
	1%	0.0058	0.0041	0.0019	0.0012

Table 8.2, continued.

n	Percentage	κ 5	10	30	50
140	10%	0.0045	0.0031	0.0012	0.0007
	5%	0.0051	0.0033	0.0014	0.0009
	1%	0.0058	0.0039	0.0016	0.0011
150	10%	0.0041	0.0029	0.0011	0.0007
	5%	0.0045	0.0031	0.0013	0.0008
	1%	0.0049	0.0036	0.0015	0.0010

For all studied cases, the percentage of $DMCEs$ are smaller than the percentiles of $DMCEc$ due to the range of $MCEs$ is smaller than the range of $MCEc$, as shown in Figure 8.1. However, the percentage points of $DMCEc$ and $DMCEs$ become closer for large concentration κ as well as for large sample size *n*.

8.2.3 The power of performance of mean circular error statistics

This subsection describes the numerical and simulation studies to investigate the power of performance of $DMCEc$ and $DMCEs$ statistics and subsequently discuss the obtained results.

(i) Description of simulation process

To investigate the power of performance of $DMCEc$ and $DMCEs$ statistics, four different sample sizes are considered, *n* = 30, 70, 100 and 150. We generate the data as described in Section 8.2.2. At position $[d]$ of the response variable $\boldsymbol{Y}$, the observation $y[d]$ is contaminated as follows

$$y^*[d] = y[d] + \lambda\pi \quad (\text{mod } 2\pi),$$

where $y^*[d]$ is the contaminated observation at position $[d]$ and λ is the degree of contamination in the range $0 \le \lambda \le 1$. When $\lambda = 0$, there is no contamination at position $[d]$, whereas when $\lambda = 1$, the observation $y^*[d]$ is located at the anti mode of its initial location. The generated data are fitted using model (7.3) and the values of $\hat{\boldsymbol{Y}}$ are obtained. Thus, values of $DMCEc$ and $DMCEs$ statistics are calculated for each generated data set.

The process is repeated for 2000 times. The power of performances of $DMCEc$ and $DMCEs$ statistics are investigated by computing the percentage of correctly detecting outlier as influential observation at position $[d]$.

(ii) Discussion

A part of the results are displayed in the following figures and the complete simulation results are given in Appendix (A.6). Figure 8.2 shows that the test based on $DMCEc$ statistic is superior compared to the $DMCEs$ statistic for all considered cases.

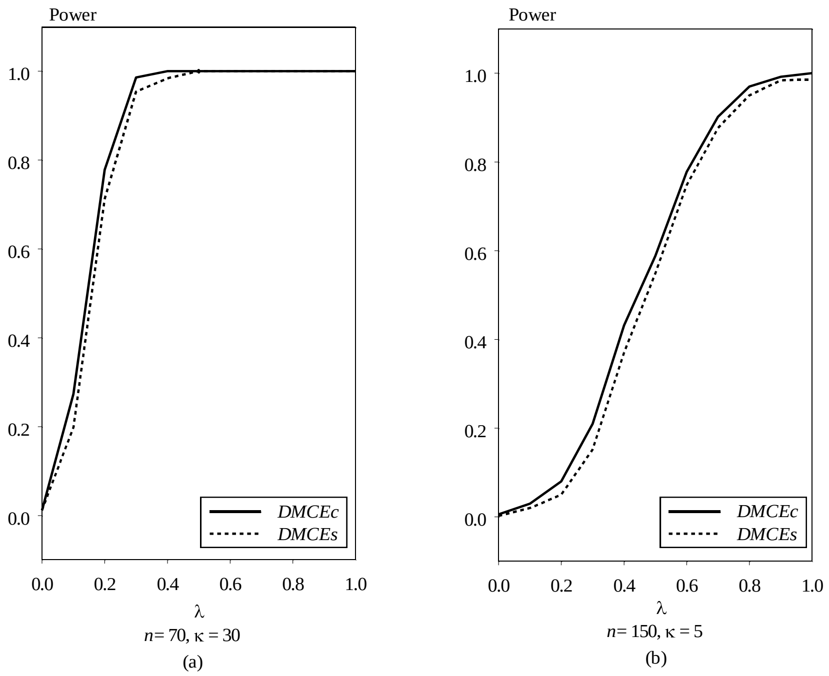


Figure 8.2: Power of performance of $DMCEc$ and $DMCEs$ statistics

Figure 8.3 shows the performance of $DMCEc$ and $DMCEs$ statistics for $n = 70$ for different values of concentration parameter κ. It is obvious that both statistics have similar behaviour. The performance of both statistics highly depend on the concentration parameter κ, where the power of performances are increasing functions of the concentration parameter κ.

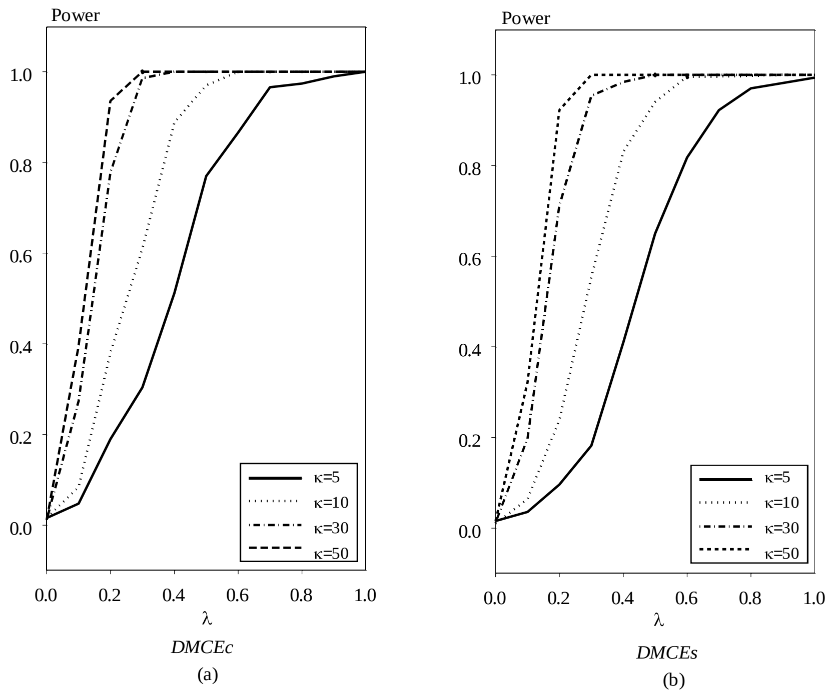


Figure 8.3: Power of performance of $DMCEc$ and $DMCEs$ statistics, for $n = 70$

Figure 8.4 shows the performance of $DMCEc$ and $DMCEs$ statistics at $\kappa = 10$ for different sample size *n*. For both statistics the power of performances are decreasing functions of sample size *n*. However, sample size has a slight effect on the performance of the $DMCEs$ statistic compare to the $DMCEc$ statistic.

In general, the power of performance is an increasing function of the contamination level λ, as shown in Figures 8.2, 8.3 and 8.4.

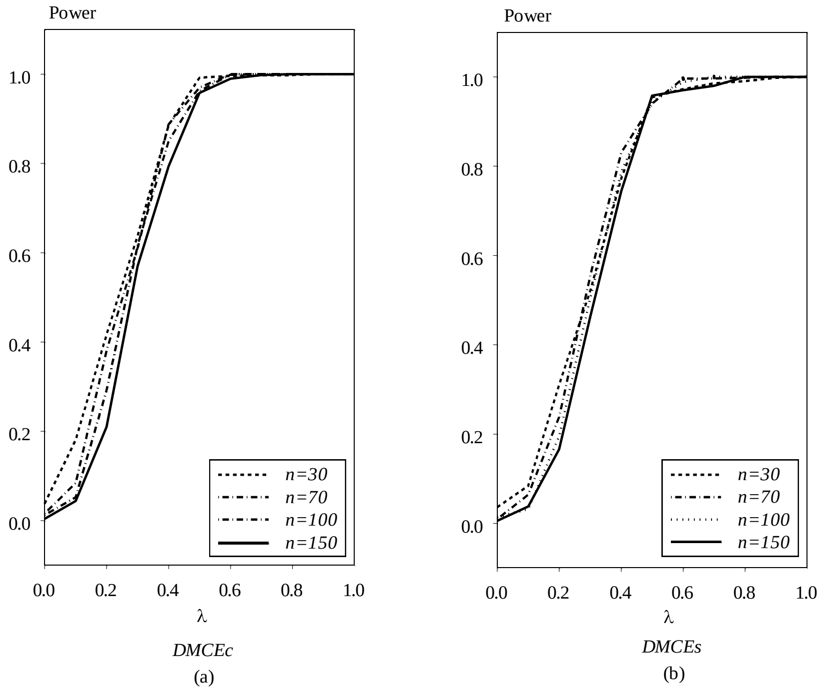


Figure 8.4: Power of performance of $DMCEc$ and $DMCEs$ statistics, for $\kappa = 10$

8.3 The approximate distribution of modified mean circular error statistics

In the previous section we discussed the mean circular error statistics. The cut-off points and the power of performance were obtained by using Monte Carlo simulations. In this section we discuss the approximate distribution of mean circular error statistics. The formula of the approximate distribution will slightly change from those proposed in Section 8.2. The interest here is to find the statistical distribution which enables us to use the available statistical distributions to identify possible influential observations in the circular regression models.

8.3.1 Approximate distribution of modified mean circular error $MDC_{(-j)}$

This subsection discusses the derivation of the approximate distribution of $DMCEc$ which will be denoted by $MDC_{(-j)}$, under the assumptions that $(y_i - \hat{y}_i)\ (\mathrm{mod}\ 2\pi)$ follow von Mises distribution with mean zero and large concentration parameter κ.

Result 8.1

For any circular regression model with $(y_i - \hat{y}_i)\ (\mathrm{mod}\ 2\pi)$ (i.i.d) follows the von Mises distribution with mean direction 0 and large concentration parameter κ, then the

$$MDC_{(-j)} = \left| MMCEc - MMCEc_{(-j)} \right| \sim \chi^2_{(1)}, \qquad j = 1,...,n$$

where y_i and $\hat{y}_i$ are the ith observed and expected values respectively, $MMCEc$ is the modified mean circular error and given by $MMCEc = 2\kappa \sum_{i=1}^{n}[1 - \cos(y_i - \hat{y}_i)]$ for full data set and $MMCEc_{(-j)} = 2\kappa_{(-j)} \sum_{i=1}^{n}[1 - \cos(y_i - \hat{y}_i)]$, $i \neq j$ is the modified mean circular error after removing the jth row.

Proof:

Suppose that $(y_i - \hat{y}_i)\ (\mathrm{mod}\ 2\pi) \sim VM(0,\kappa)$, with large κ. It has been shown in equation (4.1) that if $(y_i - \hat{y}_i) \sim VM(0,\kappa)$, then

$$\sqrt{\kappa}(y_i - \hat{y}_i) \sim N(0,1).$$

By using the standard normal distribution properties

$$\kappa(y_i - \hat{y}_i)^2 \sim \chi^2_{(1)} \tag{8.3}$$

For large concentration parameter κ, then small values of $(y_i - \hat{y}_i)$ is obtained. Thus, from the second order of Taylor series expansion we have

$$\cos(y_i - \hat{y}_i) \approx 1 - \frac{(y_i - \hat{y}_i)^2}{2}.$$

Thus,

$$(y_i - \hat{y}_i)^2 \approx 2[1 - \cos(y_i - \hat{y}_i)]. \qquad (8.4)$$

From equations (8.3) and (8.4) we get

$$2\kappa[1 - \cos(y_i - \hat{y}_i)] \sim \chi^2_{(1)}.$$

From the properties of Chi-square distribution, the modified mean circular error is given by

$$MMCEc = 2\kappa \sum_{i=1}^{n} [1 - \cos(y_i - \hat{y}_i)] \sim \chi^2_{(n)}.$$

By the removal of the *j*th row from full data set we have,

$$MMCEc_{(-j)} = 2\kappa_{(-j)} \sum_{i=1}^{n} [1 - \cos(y_i - \hat{y}_i)] \sim \chi^2_{(n-1)}, \text{ where } i \neq j$$

and $\kappa_{(-j)}$ is the concentration parameter of $(y_i - \hat{y}_i) \pmod{2\pi}$, after removing the *j*th row. The absolute difference between $MMCEc$ and $MMCEc_{(-j)}$ has the following distribution

$$MDC_{(-j)} = \left| MMCEc - MMCEc_{(-j)} \right| \sim \chi^2_{(1)}.$$

In other words, if the absolute difference between $MMCEc$ and $MMCEc_{(-j)}$ exceeds the tabulated value of Chi-square at degree of freedom 1 and desired level of significance, then observation at *j*th row is candidate to be considered as an influential observation.

8.3.2 Approximate distribution of modified mean circular error $MDS_{(-j)}$

Analogous to the approximate distribution of the mean circular error in terms of circular distance and similar to the Result 5.2 we may arrive at the following result.

Result 8.2

For any circular regression model with $(y_i - \hat{y}_i)\ (\text{mod } 2\pi)$ *(i.i.d) follows the von Mises distribution with mean direction 0 and large concentration parameter* κ *, then the*

$$MDS_{(-j)} = \left| MMCEs - MMCEs_{(-j)} \right| \sim \chi^2_{(1)}, \qquad j = 1, \ldots, n.$$

where y_i *and* $\hat{y}_i$ *are the ith observed and expected values respectively,* $MMCEs$ *is the modified mean circular error and given by* $MMCEs = 4\kappa \sum_{i=1}^{n} \sin^2\left(\frac{d_i}{2}\right)$ *for full data set and* $MMCEs_{(-j)} = 4\kappa_{(-j)} \sum_{i=1}^{n} \sin^2\left(\frac{d_i}{2}\right), i \neq j$ *is the modified mean circular error after removing the jth row.*

Proof:

From equations (5.6) we have $\cos(d_i) = \cos(y_i - \hat{y}_i)$ and by the trigonometric identities, we have $\cos\phi = 1 - 2\sin^2\left(\frac{\phi}{2}\right)$. Thus,

$$1 - \cos(d_i) = 2\sin^2\left(\frac{d_i}{2}\right).$$

By multiplying both sides by 2κ,

$$2\kappa[1 - \cos(y_i - \hat{y}_i)] = 4\kappa \sin^2\left(\frac{d_i}{2}\right).$$

Therefore,

$$2\kappa\sum_{i=1}^{n}[1-\cos(y_i-\hat{y}_i)] = 4\kappa\sum_{i=1}^{n}\sin^2\left(\frac{d_i}{2}\right).$$

From Result 8.1 we get

$$MMCEs = 4\kappa\sum_{i=1}^{n}\sin^2\left(\frac{d_i}{2}\right) \sim \chi^2_{(n)}.$$

Removing the *j*th row from full data set then,

$$MMCEs_{(-j)} = 4\kappa_{(-j)}\sum_{i=1}^{n}\sin^2\left(\frac{d_i}{2}\right) \sim \chi^2_{(n-1)}, \text{ where } i \neq j.$$

The absolute difference between $MMCEs$ and $MMCEs_{(-j)}$ has the following distribution

$$MDS_{(-j)} = |MMCEs - MMCEs_{(-j)}| \sim \chi^2_{(1)}.$$

In other words, if the absolute difference between $MMCEs$ and $MMCEs_{(-j)}$ exceeds the tabulated value of Chi-squares at degree of freedom 1, then observation at *j*th row is candidate to be considered as an influential observation.

8.4 COVRATIO statistic for simple circular regression model

In Section 2.4 we reviewed some of the available tests to identify influential observations in linear regression. One of them is *COVRATIO* statistic which is the ratio of the estimated covariance matrix of the estimated coefficients using all available data with estimated covariance matrix that results when *i*th observation is deleted.

Belsley *et al.* (1980) suggested a comparison based on the determinantal ratio which is given by

$$COVRATIO_{(-i)} = \frac{|COV_{(-i)}|}{|COV|}.$$

where $|COV|$ is the determinant covariance matrix of coefficients for full data set and $|COV_{(-i)}|$ is for the reduced data set by excluding the *i*th row. If the ratio is close to the unity, then there is no significant difference between the covariance matrices. In other words, the *i*th observation is consistent with the other observations. Alternatively, if the value of $|COVRATIO_{(-i)} - 1|$ is close or larger than $(3p/n)$ then it indicates that the *i*th observation is a candidate to be an influential observation, where p is the number of estimated coefficients and n is the sample size.

This section discusses the extension of *COVRATIO* statistic to the circular case. The covariance matrix, cut-off points and the power of performance are discussed subsequently.

8.4.1 Covariance matrix of simple circular regression model

Subsection 7.2.2 has discussed the asymptotic variance and covariance for the parameters of simple circular regression model. The covariance matrix is given by

$$\begin{bmatrix} \frac{\sum x_i^2}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}} & \frac{-\sum x_i}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}} & 0 \\ \frac{-\sum x_i}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}} & \frac{n}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}} & 0 \\ 0 & 0 & \frac{\hat{\kappa}}{n\{\hat{\kappa} - \hat{\kappa}A^2(\hat{\kappa}) - A(\hat{\kappa})\}} \end{bmatrix},$$

where $A(\hat{\kappa}) = \frac{1}{n}\sum_{i=1}^{n}\cos(y_i - \hat{\alpha} - \hat{\beta}x_i)$ and $\hat{\kappa}$ is the estimated concentration parameter of circular random error. To apply the *COVRATIO* statistic, in analogy to the linear case we will consider the coefficients covariance matrix, which is given by

$$COV = \frac{1}{\hat{\kappa}A(\hat{\kappa})\{n\sum x_i^2 - (\sum x_i)^2\}} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}.$$

It can be shown that the determinant of covariance matrix for model (7.3) is given by

$$|COV| = \frac{1}{\hat{\kappa}A(\hat{\kappa})}. \tag{8.5}$$

Thus, the *COVRATIO* statistic for the *i*th row is given by

$$COVRATIO_{(-i)} = \frac{|COV_{(-i)}|}{|COV|} = \frac{\hat{\kappa}_{(-i)}A(\hat{\kappa}_{(-i)})}{\hat{\kappa}A(\hat{\kappa})}. \tag{8.6}$$

The above statistic is simple and easy to be obtained. Any observation with $|COVRATIO_{(-i)} - 1|$ exceeding the cut-off point which will be calculated in the following section will be identified as an influential observation.

8.4.2 Percentage points of COVRATIO statistic

(i) Description of simulation process

The percentage points are obtained by using Monte Carlo simulation method. Fifteen different sample sizes of n =10(10)150 are used. For each sample size n, a set of circular random error from von Mises distribution with mean direction 0 and various values of concentration parameter κ are generated, where κ = 5, 10, 30 and 50 as follows:

Step 1. Generate $\boldsymbol{X}$ variable of size n from $VM(\pi/4, 1.5)$. The parameters of simple circular regression model (7.3) are fixed at $\alpha = 0$ and $\beta = 1$.

Step 2. Calculate the observed values of the response variable $\boldsymbol{Y}$ based on model (7.3).

Step 3. Fit the generated circular data by using model (7.3).

Step 4. Calculate $|COV|$ by using equation (8.5)

Step 5. Exclude the *i*th row from the generated sample, where i=1,...,n. Repeat Steps 3 to 5 to obtain $|COV_{(-i)}|$ for all i.

Step 6. Compute $COVRATIO_{(-i)}$ by using equation (8.6) and then obtain the values of

$|COVRATIO_{(-i)} - 1|$ for all *i*.

Step 7. Specify the maximum value of $|COVRATIO_{(-i)} - 1|$.

The process is repeated 2000 times for each combination of sample size *n* and concentration parameter κ. Then the 10th, 5th and 1st upper percentiles of the maximum values of $|COVRATIO_{(-i)} - 1|$ are calculated. The percentiles are tabulated in Table 8.3. For each sample size *n* and concentration parameter κ, where 10, 5 and 1 percentages are given in the first , second and third rows, respectively.

(i) Discussion

Results in Table 8.3 show that the cut points of $|COVRATIO_{(-i)} - 1|$ statistic are independent of the concentration parameter κ. Figure 8.5 illustrates the values of cut-off points versus the concentration parameter κ for *n* =50 where similar results are obtained for other sample size *n*.

In order to estimate the cut-off points for each sample size at different percentiles we suggest calculating the arithmetic mean of the simulated cut-off points for each sample size *n*. Table 8.4 gives the cut-off points and the corresponding standard deviations as given in parenthesis for various sample size *n*. The results show that the cut-off points is a decreasing functions of sample size *n*. The values of the standard deviations are also very small indicating the independency of percentiles values on the concentration parameter.

Table 8.3: The percentage points of the null distribution of $|COVRATIO_{(-i)} - 1|$

		κ			
n	Percentage	**5**	**10**	**30**	**50**
	10%	1.035	1.432	1.035	1.143
10	5%	1.353	1.501	1.383	1.446
	1%	1.611	1.747	1.889	1.481
	10%	0.569	0.547	0.504	0.536
20	5%	0.769	0.696	0.606	0.671
	1%	0.808	0.800	0.847	0.848
	10%	0.364	0.338	0.375	0.344
30	5%	0.459	0.444	0.450	0.452
	1%	0.727	0.552	0.559	0.586
	10%	0.276	0.263	0.258	0.245
40	5%	0.337	0.346	0.340	0.348
	1%	0.463	0.435	0.410	0.423
	10%	0.225	0.200	0.199	0.193
50	5%	0.262	0.263	0.265	0.268
	1%	0.335	0.312	0.297	0.335
	10%	0.187	0.182	0.170	0.173
60	5%	0.230	0.232	0.236	0.239
	1%	0.292	0.261	0.254	0.254
	10%	0.167	0.150	0.151	0.146
70	5%	0.196	0.181	0.199	0.180
	1%	0.270	0.223	0.235	0.203
	10%	0.146	0.130	0.128	0.127
80	5%	0.173	0.167	0.171	0.174
	1%	0.222	0.201	0.198	0.172
	10%	0.120	0.128	0.118	0.122
90	5%	0.141	0.146	0.136	0.136
	1%	0.166	0.184	0.179	0.184
	10%	0.120	0.113	0.108	0.106
100	5%	0.139	0.132	0.122	0.119
	1%	0.186	0.163	0.152	0.149
	10%	0.104	0.098	0.099	0.096
110	5%	0.122	0.126	0.121	0.119
	1%	0.142	0.138	0.134	0.144
	10%	0.104	0.096	0.089	0.089
120	5%	0.118	0.106	0.105	0.115
	1%	0.151	0.132	0.134	0.123
	10%	0.094	0.088	0.086	0.087
130	5%	0.105	0.102	0.110	0.105
	1%	0.126	0.127	0.138	0.128

Table 8.3, continued.

n	Percentage	κ 5	10	30	50
140	10%	0.091	0.081	0.081	0.077
	5%	0.105	0.095	0.110	0.091
	1%	0.129	0.132	0.108	0.130
150	10%	0.079	0.076	0.072	0.075
	5%	0.090	0.086	0.089	0.095
	1%	0.109	0.113	0.101	0.101

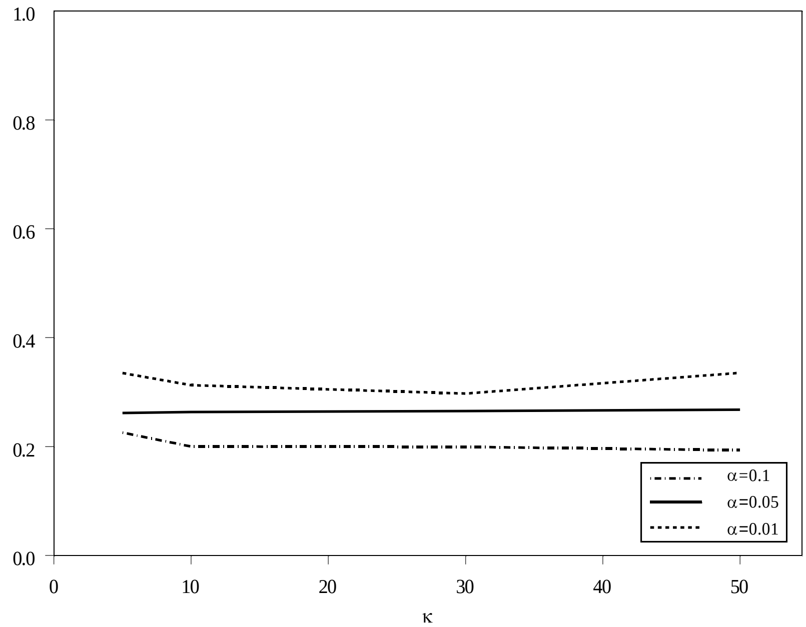


Figure 8.5: The percentage points of $|COVRATIO_{(-i)} - 1|$ statistic, for $n = 50$

For the linear case Belsley *et al.* (1980) stated that $(3p/n)$ is an appropriate cut-off points of $|COVRATIO_{(-i)} - 1|$ statistic at 0.05 significant level. It is of interest to find such formula for circular case. We found that the cut-off points for any sample size are very close to the values $(7p/n)$ where $p = 2$ and n is the sample size. The exact values of $(7p/n)$ are given in the last row of Table 8.4, followed by the bias of the cut-off points at 0.05 significant level in parenthesis. The biases in all cases are less than 0.021. Thus, the approximated value $(7p/n)$ can be used as the cut-off points of $|COVRATIO_{(-i)} - 1|$ statistic at 0.05 level of significant for any sample of size n and $p = 2$.

Table 8.4: Means of the percentage points for the null distribution of $|COVRATIO_{(-i)} - 1|$

n	10		20		30	
90%	1.161	0.188	0.539	0.027	0.355	0.017
95%	1.421	0.066	0.685	0.067	0.451	0.006
99%	1.682	0.175	0.826	0.026	0.606	0.082
(7*p*/*n*)	1.4	(0.021)*	0.7	(-0.015)*	0.467	(-0.015)*
n	40		50		60	
90%	0.261	0.013	0.204	0.014	0.178	0.008
95%	0.343	0.005	0.264	0.003	0.234	0.004
99%	0.433	0.023	0.320	0.018	0.265	0.018
(7*p*/*n*)	0.35	(-0.007)*	0.28	(-0.016)*	0.233	(0.001)*
n	70		80		90	
90%	0.154	0.009	0.133	0.009	0.122	0.004
95%	0.189	0.010	0.171	0.003	0.139	0.005
99%	0.233	0.028	0.198	0.020	0.178	0.009
(7*p*/*n*)	0.2	(-0.011)*	0.175	(-0.004)*	0.156	(-0.016)*
n	100		110		120	
90%	0.112	0.006	0.099	0.003	0.095	0.007
95%	0.128	0.009	0.122	0.003	0.111	0.006
99%	0.162	0.017	0.140	0.004	0.135	0.012
(7*p*/*n*)	0.14	(-0.012)*	0.127	(-0.005)*	0.117	(-0.006)*
n	130		140		150	
90%	0.089	0.003	0.083	0.006	0.076	0.003
95%	0.105	0.003	0.100	0.009	0.090	0.004
99%	0.130	0.006	0.125	0.011	0.106	0.006
(7*p*/*n*)	0.108	(-0.002)*	0.1	(0.000)	0.093	(-0.003)*

* The bias of 95% cut-off-points from the corresponding value of (7*p*/*n*).

8.4.3 Power of performance of COVRATIO statistic

This subsection examines the performance of $|COVRATIO_{(-i)} - 1|$ statistic through numerical and simulation studies. Part (i) describes the used algorithm and part (ii) discusses the obtained results.

(i) Description of simulation process

Monte Carlo simulation method is used to examine the performance of $|COVRATIO_{(-i)} - 1|$ statistic for detecting influential observations in the simple circular regression model. Samples of five different sizes n = 30, 70, 100 and 150 are used. We follow similar procedures described in Subsection 8.4.2(a) to generate the data. In addition, we let the observation at position $[d]$, say $y[d]$, be contaminated such that

$$y^*[d] = y[d] + \lambda\pi, \quad (\bmod 2\pi),$$

where $y^*[d]$ is the value of $y[d]$ after contamination, λ is the degree of contamination in the range $0 \le \lambda \le 1$. The generated data of $\boldsymbol{X}$ and $\boldsymbol{Y}$ are then fitted by using model (7.3) and $|COV|$ is calculated using equation (8.5). Consequently, by excluding the *i*th row from sample, for $i = 1,\ldots,n$ and refitting the reduced data we calculate $COVRATIO_{(-i)}$ by using equation (8.6). Finally, we specify the maximum value of $|COVRATIO_{(-i)} - 1|$ statistic.

The process is repeated for 2000 times. The power of performance of $|COVRATIO_{(-i)} - 1|$ statistic is examined by computing the percentage of correct detection of the contaminated observation at position $[d]$.

(ii) Discussion

Three main factors are considered in the discussion on the power of performance of $|COVRATIO_{(-i)} - 1|$ statistic, namely, the level of contamination λ, concentration parameter κ and the sample size *n*. The complete results of the power of performance are given in Appendix (A.7).

Figure 8.6 illustrates the power of performance of $|COVRATIO_{(-i)} - 1|$ statistic for $n = 70$ and four values of the concentration parameter $\kappa = 5$, 10, 30 and 50. It is shown that the power of performance is an increasing function of concentration parameter κ. As the concentration parameter increases, the power of performance also increases. The power of performance highly depends on the level of contamination λ. On the other hand, the power of performance is a decreasing function of sample size n as shown in Figure 8.7.

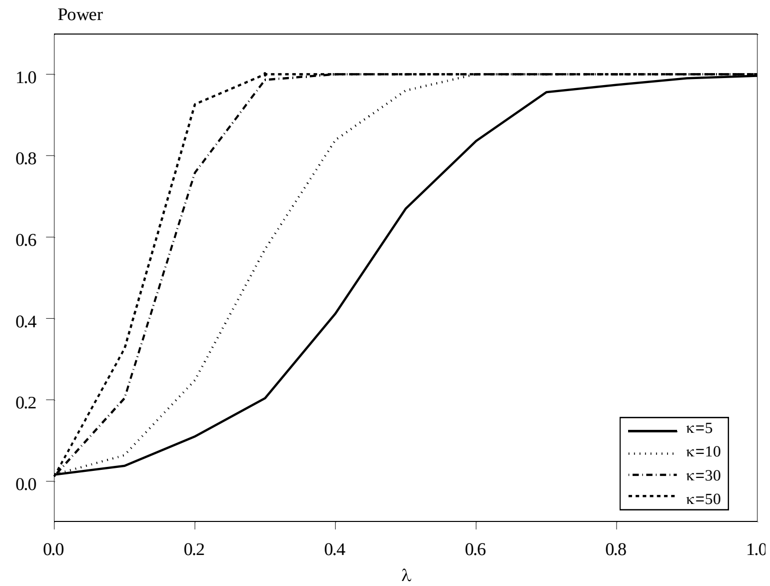


Figure 8.6: Power of performance for $|COVRATIO_{(-i)} - 1|$ statistic, for $n = 70$

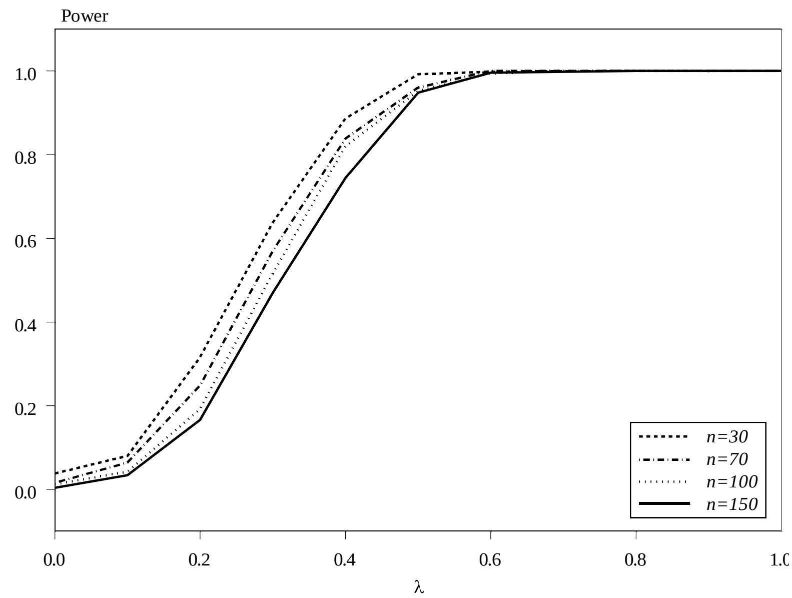


Figure 8.7: Power of performance for $|COVRATIO_{(-i)} - 1|$ statistic, for $\kappa = 10$

8.5 Practical example

This section considers the wind direction data, which has been described in Chapter 3 and fitted in Chapter 7. Several numerical and graphical methods have identified observations numbered 38 and 111 as outliers. The numerical methods which are discussed in this chapter will be applied to the data in order to identify possible influential observations.

8.5.1 Mean circular error

The estimated concentration parameter is $\hat{\kappa} = 7.338$ and the sample size $n = 129$. The mean circular error for full data set is $MCEc$ = 0.071 and $MCEs$ = 0.127. From Tables 8.1 and 8.2, the corresponding cut-off points at 0.05 significant level for $DMCEc$ and $DMCEs$ statistics are 0.005 and 0.004, respectively. The $|MCEc - MCEc_{(-i)}|$ and $|MCEs - MCEs_{(-i)}|$ statistics are calculated and the results are plotted in Figures 8.8 and 8.9 respectively. It is obvious that observations number 38 and 111 exceed the cut-off points as shown by dash line in Figures 8.8 and 8.9.

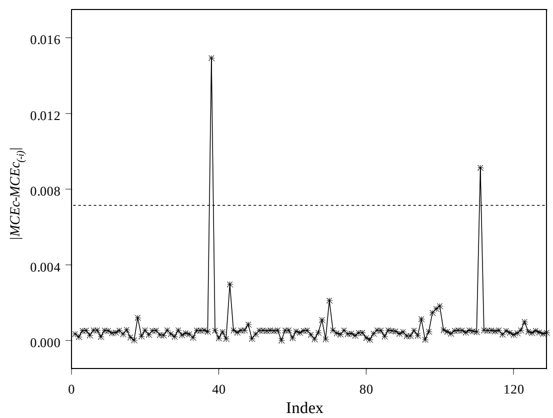


Figure 8.8: $DMCEc$ statistic for wind data

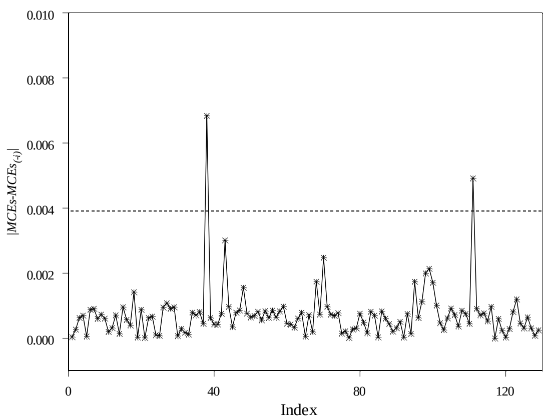


Figure 8.9: $DMCEs$ statistic for wind data

8.5.2 Approximate distribution of mean circular error statistics

The $MDC_{(-j)}$ statistic is calculated for wind direction data and the values of statistic are plotted in Figures 8.10. The cut off point is $\chi^2_{1,0.05} = 3.841$ as shown by the dash line. Hence, we conclude that the $MDC_{(-j)}$ statistic has successfully identified observations numbered 38 and 111 as influential observations. The results of $MDS_{(-j)}$ is similar to the $MDC_{(-j)}$.

8.5.3 COVRATIO statistic

The determinant of coefficients covariance matrix for full data set $|COV|$ is 2.89×10^{-7} and the corresponding cut-off point is 0.108. The $|COVRATIO_{(-i)}-1|$ statistic values for wind direction data are plotted in Figure 8.11. It can be seen that the $|COVRATIO_{(-i)}-1|$ statistic values for observations numbered 38 and 111 exceed the cut-off point.

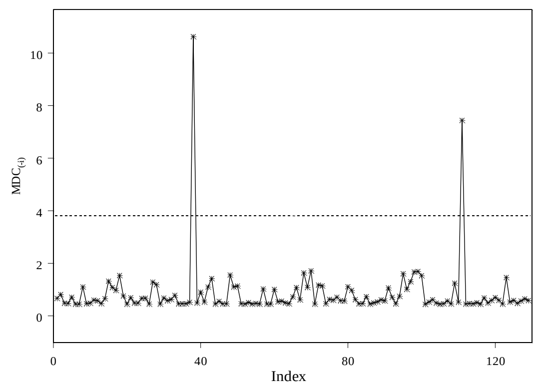


Figure 8.10: $MDC_{(-i)}$ statistic for wind data

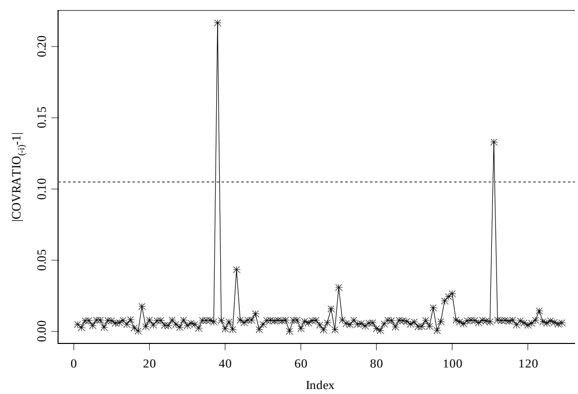


Figure 8.11: $|COVRATIO_{(-i)}-1|$ statistic for wind data

8.6 Summary

Five statistics to identify influential observations in circular regression are proposed using row deletion approach. These statistics can be extended to the case of multiple influential observations by obtaining the appropriate cut-off points or by considering the difference by the approximate distribution of the full and reduced data follow Chi-squares with the reduction as degree of freedom.

Generally, the proposed tests perform competitively well and they are able to identify the same points identified in Chapter 7 as outliers.

CHAPTER NINE

CONCLUSIONS

9.1 Summary

This study aims to shed the light on some problems of outliers in circular data. The lack of published work in this area motivates the researcher to propose new techniques for detecting outliers. We present important and significant works on the detection of outliers in circular samples and simple circular regression via several numerical and graphical techniques.

9.2 Significance of the study

This study has involved new methodological developments in three main aspects:

- Application of the approximation technique of circular variable into linear variables.
- Development of alternative procedures to detect outliers in univariate circular samples.
- Proposal on new techniques to identify possible outliers in circular regression.

Firstly, through simulation studies, it was found that the approximation of samples from von Mises distribution by normal distribution depends on the sample size n and concentration parameter κ. For small sample size ($n < 20$), samples are approximated by normal distribution if the concentration parameter $\kappa > 2$. For larger samples, the concentration parameter κ should be larger than 4.

Secondly, two new tests of discordancy based on circular distance and chord lengths were proposed. The cut-off points and the power of performance were obtained. The simulation results showed that the new tests performed better than other known available tests. Moreover, they are easier to be used and interpreted by the practitioners. Discussion on the approximate distribution of the tests for samples from von Mises distribution with a large concentration parameter was presented.

Thirdly, circular boxplot was developed to identify possible outliers in circular samples. Five circular summary statistics are used. The circular median is obtained using the definition given by Fisher (1993) and subsequently was extended to obtain the values of the first and third quartiles. Extensive simulation work was used to find suitable circular boxplot criterion $\nu \times CIQR$ where ν is the resistant constant. Several interesting results were observed:

(i) There is a functional relationship between $CIQR$ (in radians) and concentration parameter κ such that $CIQR \approx (\ln \kappa)^{-1}$ for large κ.

(ii) The whiskers of the circular boxplot overlap if $\nu > \pi \ln(\kappa) - 0.5$.

(iii) The circular boxplot criterion depends on the concentration parameter. For large concentration $\kappa > 3$ the appropriate resistant constant is between 2 and 2.7, while for small concentration $\kappa \leq 3$ the values of resistant constant can be $1 < \nu < 2$.

Furthermore, we investigated the power of performances of the proposed circular boxplot. The results suggested that the proposed procedure was more effective for large concentration parameter since the observations for small κ tend to be distributed uniformly. The power of the circular boxplot also increases gradually as the sample size *n* increases. We then developed the diagrammatical representation of the circular boxplot in S-Plus environment and applied on two real circular data sets.

Fourthly, the problem of identifying outliers in simple regression model for circular variable was considered. A new definition of circular residuals $\boldsymbol{r}_A$ was proposed based on the circular distance. Simulations studies showed that $\boldsymbol{r}_A$ satisfied the model assumptions. The circular residuals $\boldsymbol{r}_A$ successfully identified possible outliers in the model by applying several numerical tests and graphical techniques described in this study.

Fifthly, two alternative statistics based on the circular distance between the observed and fitted values were proposed. Row deletion approach was used to investigate the effect of excluding one observation at a time. The cut-off points for both statistics are obtained and discussed, and the power of performance were investigated through simulation. It was found that the power of performance is a decreasing function of sample size n but an increasing function of the concentration parameter. The approximated distributions for both statistics are shown to follow Chi-square distribution with one degree of freedom.

Sixthly, the *COVRATIO* statistic was extended to the circular case. The percentiles were obtained through simulations. It was found that $(7p/n)$ is an appropriate estimated cut-off points of $\left|COVRATIO_{(-i)} - 1\right|$ statistic at 0.05 level of significance, where p is the number of terms and n is the sample size. The power of performance is an increasing function of concentration parameter κ, while it is a decreasing function of the sample size n.

Finally, throughout this study, we used two real data sets for illustration. The proposed detection methods were able to detect outliers in the circular data.

9.3 Further research

There are various possibilities for further research in this area. Some suggestions are given as follows:

(i) To extend the procedures of the detection of outliers to other circular regression models.

(ii) To extend the procedures of the detection of outliers to the circular functional relationship model.

(iii) To carry out more studies on the diagnostics checking on the circular regression models.

(iv) To develop some effective procedures to detect multiple outliers as in circular regression models.

(v) To develop comprehensive and easy-to-use software for circular data analysis.

We recognize that there are still many problems ready to be explored in circular statistics, and it is fascinating for statisticians to work on them.

REFERENCES

Anderson, C. M. (1993). Graphical methods for circular and cylindrical data. *IIQR Research Report*, RR-93-02.

Anscombe, F. J. (1960). Rejection of outliers. *Technometrics*, **2** (2), 123-147.

Bagchi, P. and Guttman, I. (1990). Spuriosity and outliers in directional data. *Journal of Applied Statistics*, **17**, 341-350.

Barnett, V. (1978). The study of outliers: Purpose and model. *Applied Statistics*, **27** (3), 242-250.

Barnett, V. and Lewis, T. (1984). *Outliers in Statistical Data*. 2nd ed., John Wiley & Sons, Chichesters.

Batschelet, E. (1981). *Circular Statistics in Biology*. Academic Press, London.

Beckman, R. J. and Cook, R. D. (1983). Outlier……….s. *Technometrics*, **25** (2), 119-149.

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

Bernoulli, D. (1777). The most probable choice between several discrepant observations and the formation therefrom of the most likely induction, in Allen, C. G. (1961). *Biometika*, **48** (1), 3-18.

Best, D. J. and Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. *Communication in Statistics – Simulations and Computations*. **B10** (5), 394-502.

Bowtell, P. and Patefield, M. (1999). Circular functional relationships. *Communication in Statistics- Theory and Methods*, **28** (9), 2103-2118.

Caires, S. and Wyatt, L. R. (2003). A linear functional relationship model for circular data with an application to the assessment of ocean wave measurements. *Journal of Agricultural, Biological and Environmental Statistics*, **8** (2), 153-169.

Chatterjee, S., Hadi, A. S. and Price, B. (2000). *Regression Analysis by Example*. 3rd ed., John Wiley & Sons, New York.

Collett, D. (1980). Outliers in circular data. *Applied Statistics*, **29** (1), 50-57.

Cook, R. D. (1977). Detection of influential observations in linear regression. *Technometrics*, **19** (1), 15-19.

Cook, R. D. (1979). Influential observations in linear regression. *Journal of the American Statistical Association*, **74** (365), 169-174.

David, H. A. (1970). *Order Statistics*. Wiley, New York and London.

Dean R. B. and Dixon, W. J. (1951). Simplified statistics for small numbers of observations. *Anal. Chem*., **23**, 636-638.

Dixon, W. J. (1950). Analysis of extreme values. *The Annals of Mathematical Statistics,* **21** (4), 488-506.

Dixon, W. J. (1951). Ratios involving extreme values. *The Annals of Mathematical Statistics,* **22** (1), 68-78.

Dobson, A. J. (1978). Simple approximations for the von Mises concentration statistic. *Applied Statistics*, **27** (3), 345-347.

Downs, T. D. and Mardia, K. V. (2002). Circular regression. *Biometrika*, **89** (3), 683-697.

Ferguson, D. E., Landreth, H. F. and Mckeown, J. P. (1967). Sun compass orientation of the northern cricket frog, Acris crepitans. *Animal Behaviour,* **15** (1), 45-53.

Fisher, N. I., Lewis, T. and Embleton, B. J. J. (1987). *Statistical Analysis of Spherical Data.* Cambridge University Press, Cambridge.

Fisher, N. I. and Lee, A. J. (1992). Regression models for an angular response. *Biometrics,* **48**, 665-677.

Fisher, N. I. (1993). *Statistical Analysis of Circular Data*. Cambridge University Press, London.

Gatto, R. and Jammalamadaka, S. R. (2007). The generalized von Mises distribution. *Statistical Methodology*, **4**, 341-353.

Gill, J. and Hangartner, D. (2009). Circular data in political science and how to handle it. Paper presented at the annual meeting of the MPSA Annual National Conference, Palmer House Hotel, Hilton, Chicago. 2009-05-23, from http://www.allacademic.com/meta/p265709_index.html.

Goldsmith, P. L. and Boddy, R. (1973). Critical analysis of factorial experiments and orthogonal fractions. *Applied Statistics*, **22** (2), 141-160.

Gordon, A. D., Jupp, P. E. and Byrne, R. W. (1989). Construction and assessment of mental maps. *British Journal of Mathematical and Statistical Psychology*, **42**,169-182.

Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, **25**, 683-700.

Graedel, T. E. (1977). The wind boxplot: an improved wind rose. *Journal of Applied Meteorology*, **16** (4), 448-450.

Grubbs, F. E. (1969). Procedures for detecting outlying observations in samples. *Technometrics,* **11** (1), 1-21.

Hoaglin, D. C., Iglewicz, B. and Tukey, J. W. (1986). Performance of some resistant rules for outlier labeling. *Journal of the American Statistical Association*, **81** (396), 991-999.

Huber, P. J. (1981). *Robust Statistics*. John Wiley & Sons, New York.

Hussin, A. G., Fieller, N. R. J. and Stillman, E. C. (2004). Linear regression for circular variables with application to directional data. *Journal of Applied Science & Technology,* **9,** (1 & 2), 1-6.

Hussin, A. G. (1997). *Pesudo-replication in Functional Relationships with Environmental Applications*. Ph.D. thesis, University of Sheffield.

Hussain F, Zubairi, Y. Z. and Hussin, A. G. (2007). An alternative diagrammatical representation of wind data. *Proceedings of the WSEAS Conference on Applied Mathematics*. December, Egypt.

Ingelfinger, J. A., Mosteller, F., Thibodeau, L. A., and Ware, J. H. (1983). *Biostatistics in Clinical Medicine*. Macmillan, New York.

Jammalamadaka, S. R., Bhadra, N., Chaturvedi, D., Kutty, T. K., Majumdar, P. P. and Poduval, G. (1986). Functional assessment of knee and ankle during level walking. In Matusita, K., editor. *Data Analysis in Life Science,* 21-54. Indian Statistical Institute, Calcutta, India.

Jammalamadaka, S. R. and Sarma, Y. R. (1988). A correlation coefficient for angular variables. *Statistical Theory and Data Analysis II*, 349-364, North Holand, Amesterdam.

Jammalamadaka, S. R. and Sarma, Y. R. (1993). Circular regression. In Matusita, K., editor, *Statistical Theory and Data Analysis,* 109-128. VSP, Utrecht.

Jammalamadaka, S. R. and SenGupta, A. (2001). *Topics in Circular Statistics*. World Scientific Press, Singapore.

Johnson, R. A. and Wehrly, T. E. (1977). Measures and models for angular correlation and angular- linear correlation. *Journal of the Royal Statistical Society, Series B*, **39**, 222-229.

Johnson, R. A. and Wehrly, T. E. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association,* **73** (363), 602-606.

Jones, T. A. and James, W. R. (1969). Analysis of bimodal orientation data. *Mathematical Geology*, **1** (2), 129-135.

Jupp, P. E. and Mardia, K. V. (1989). A unified view of the theory of directional statistics, 1975-1988. *Internat. Statist. Rev.*, **57**, 261-294.

Kato, S., Shimizu, K. and Shieh, G. S. (2008). A circular-circular regression model. *Statistica Sinica*, **18** (2), 633-643.

Kopf, E. W. (1916), Florence Nightingale as statistician. *Journal of the American Statistical Association*, **15**, 388 - 404.

Laycock, P. J. (1975). Optimal design: Regression model for directions. *Biometrika*, **62** (2), 305-311.

Lenth, R. V. (1981). On finding the source of a signal. *Technometrics*, **23** (2), 149-154.

Lund, U. (1999). Least circular distance regression for directional data. *Journal of Applied Statistics*, **26** (6), 723-733.

Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, London.

Mardia, K. V. (1975). Statistics of directional data. *Journal of the Royal Statistical Society, Series* B, **37**, 349-393.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*. John Wiley & Sons, London.

Mickey, M. R. (1974). Detecting outliers with stepwise regression. *Communications-UCLA Health Science Facility*, **1**, 1-9.

Montgomery, D. C. and Peck, E. A. (1992). *Introduction to Linear Regression Analysis,* 2nd ed., John Wiley & Sons, New York.

Murdoch, J. and Barnes, J. A. (1998). *Statistical Tables for Students of Science, Psychology, Engineering, Business, Management and Finance*. 4th ed., Palgrave Macmillan.

Pearson, E. S. and Hartley, H. O. (1966). *Biometrika Tables for Statisticians*.Vol.1, 3rd ed., Cambridge University Press, London.

Peirce, B. (1852). Criterion for the rejection of doubtful observations. *Astronomical Journal*, **2**, 161-163.

Rao, J. S. (1969). *Some Contributions to the Analysis of Circular Data.* Ph.D. thesis, Indian Statistical Institute, Calcutta, India.

Rivest, L.-P. (1997). A decentred predictor for circular –circular regression. *Biometrika,* **84** (3), 717-726.

Siew, H. Y., Kato, S. and Shimizu, K. (2008). The generalized t-distribution on the circle. *Japanese Journal of Applied Statistics*, **37** (1), 1-16.

Sim, C. H., Gan, F. F. and Chang, T. C. (2005). Outlier labeling with boxplot procedures. *Journal of the American Statistical Association*, **100** (470), 642-652.

Stephens, M. A. (1964). The distribution of the goodness-of-fit statistic U_N^2. II. *Biometrika*, **51** (3 & 4), 393-397.

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

Upton, G. (1993). Testing for outliers in circular data. *Journal of Applied Statistics*, **20** (2), 229-235.

Von Mises, R. (1918). Uber die "Ganzzahligkeit" der atomgewichte und vewandte fragen. *Physikal. Z.* , **19**,490-500.

Watson, G. S. and Williams, E. J. (1956). On the construction of significance tests on the circle and the sphere. *Biometrika*, **43**, 344-352.

Watson, G. S. (1961). Goodness-of-fit tests on a circle. *Biometrika*, **48** (1 & 2), 109-114.

Wehrly, T. E. and Shine, E. P. (1981). Influence curves of estimators for directional data. *Biometrika*, **68** (1), 334-335.

Wright, T. W. (1884). *A Treatise on the Adjustment of Observations with Applications to geodetic work and other measures of precision*. Van Nostrand, New York.

Wu, J.-W. and Lee, W. (2006). Computational algorithm of least absolute deviation method for determining number of outliers under normality. *Applied Mathematics and Computation*, **175**, 609-617.

Appendix 1
Wind Direction Data

Obs. No.	Radar		Anchored Buoy		Obs. No.	Radar		Anchored Buoy	
	Time	Obs.	Time	Obs.		Time	Obs.	Time	Obs.
1	1.615	0.79	1.618	1.154	33	3.823	5.406	3.826	5.744
2	1.656	0.715	1.66	1.154	34	3.865	5.472	3.868	5.547
3	1.698	0.975	1.701	1.007	35	3.906	5.401	3.91	5.498
4	1.74	0.97	1.743	1.178	36	3.948	5.42	3.951	5.4
5	1.781	0.993	1.785	0.859	37	3.99	5.276	3.993	5.449
6	1.823	0.902	1.826	1.007	38	4.031	1.728	4.035	4.786
7	1.837	0.943	1.847	1.056	39	4.406	5.512	4.41	5.449
8	2.406	1.728	2.41	1.4	40	4.448	5.486	4.451	5.178
9	2.448	1.445	2.451	1.497	41	4.49	5.444	4.493	5.62
10	2.49	1.679	2.493	1.693	42	4.531	5.518	4.535	5.13
11	2.531	1.703	2.535	2.012	43	4.559	5.505	4.576	4.541
12	2.573	1.862	2.576	1.792	44	9.573	5.558	9.576	5.571
13	2.615	1.726	2.618	1.766	45	9.615	5.42	9.618	5.62
14	2.656	1.79	2.66	1.669	46	9.656	5.398	9.66	5.473
15	2.698	1.831	2.701	1.4	47	9.698	5.334	9.701	5.327
16	2.726	1.719	2.743	1.4	48	9.781	5.418	9.785	4.835
17	2.781	1.646	2.785	1.375	49	9.823	5.418	9.826	5.032
18	2.823	1.622	2.826	1.056	50	9.892	5.338	9.91	5.842
19	2.865	1.342	2.868	1.178	51	9.948	5.47	9.951	5.571
20	2.906	1.176	2.91	1.276	52	9.99	5.455	9.993	5.522
21	2.948	1.325	2.951	1.693	53	10.073	5.555	10.076	5.473
22	2.99	1.103	2.993	1.325	54	10.115	5.462	10.118	5.522
23	3.406	6.131	3.41	6.062	55	10.156	5.401	10.16	5.522
24	3.448	5.719	3.451	5.988	56	10.198	5.316	10.201	5.376
25	3.49	5.713	3.493	5.988	57	10.24	5.439	10.243	5.081
26	3.531	5.487	3.535	5.498	58	10.406	5.408	10.41	5.473
27	3.573	5.742	3.576	5.276	59	10.448	5.431	10.451	5.449
28	3.615	5.728	3.618	5.302	60	10.49	5.473	10.493	5.915
29	3.656	5.61	3.66	5.62	61	10.531	5.46	10.535	5.351
30	3.698	5.463	3.701	5.744	62	10.573	5.364	10.576	5.571
31	3.74	5.427	3.743	5.644	63	10.615	5.444	10.618	5.376
32	3.781	5.418	3.785	5.669	64	10.656	5.35	10.66	5.327
65	10.698	5.202	10.701	4.983	70	10.906	5.238	10.91	4.417
66	10.74	5.161	10.743	4.786	71	10.948	4.97	10.951	5.007
67	10.781	5.062	10.785	4.908	72	10.99	4.947	10.993	5.473
68	10.823	5.145	10.826	4.517	73	11.031	4.887	11.035	5.4
69	10.865	5.212	10.868	4.835	74	11.073	4.872	11.076	4.859

Obs. No.	Radar		Anchored Buoy		Obs. No.	Radar		Anchored Buoy	
	Time	Obs.	Time	Obs.		Time	Obs.	Time	Obs.
75	11.115	4.589	11.118	4.859	103	20.906	0.237	20.91	0.171
76	11.156	4.51	11.16	4.761	104	20.948	0.045	20.951	0.295
77	11.281	4.319	11.285	4.639	105	20.99	6.241	20.993	6.259
78	11.323	4.427	11.326	4.664	106	21.031	0.248	21.035	0.319
79	11.337	4.436	11.347	4.664	107	21.073	0.578	21.076	0.539
80	11.406	4.451	11.41	4.074	108	21.087	0.627	21.097	0.81
81	12.198	3.84	12.201	4.295	109	21.406	0.251	21.41	6.161
82	12.24	3.819	12.243	4.098	110	21.448	5.299	21.451	5.473
83	12.281	4.159	12.285	4.173	111	21.49	3.749	21.493	5.62
84	12.323	3.987	12.326	4.122	112	21.531	1.876	21.535	2.012
85	19.823	5.506	19.826	5.817	113	21.573	1.776	21.576	1.963
86	19.865	5.509	19.868	5.571	114	21.615	1.786	21.618	1.841
87	19.906	5.643	19.91	5.571	115	21.656	1.658	21.66	1.89
88	19.948	5.707	19.951	5.596	116	21.684	1.377	21.701	1.497
89	19.99	5.727	19.993	5.964	117	21.74	1.305	21.743	1.669
90	20.031	5.685	20.035	5.547	118	21.781	1.309	21.785	1.325
91	20.073	5.696	20.076	6.161	119	21.823	1.337	21.826	1.644
92	20.115	5.745	20.118	6.037	120	21.865	1.198	21.868	1.571
93	20.142	5.837	20.16	5.915	121	21.906	1.15	21.91	1.08
94	20.531	1.146	20.535	1.546	122	21.948	1.047	21.951	1.129
95	20.573	1.074	20.576	1.866	123	21.99	0.97	21.993	0.466
96	20.615	1.201	20.618	1.717	124	22.031	0.998	22.035	0.981
97	20.656	1.253	20.66	1.89	125	22.073	1.071	22.076	1.007
98	20.698	1.032	20.701	1.89	126	22.531	0.793	22.535	0.834
99	20.74	1.093	20.743	1.988	127	22.573	0.753	22.576	1.056
100	20.781	0.505	20.785	6.137	128	22.615	0.573	22.618	0.932
101	20.823	0.234	20.826	0.393	129	22.656	0.437	22.66	0.761
102	20.865	0.275	20.868	0.271					

Appendix 2
Power of Performance of Discordancy Statistics

n=5		*P1*					*P3*					*P5*					*P1-P3*				
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.01	0.01	0.07	0.01	0.01	0.01	0.01	0.06	0.01	0.01	0.04	0.04	0.30	0.06	0.04	0.00	0.00	0.00	0.00	0.00
	0.1	0.01	0.01	0.06	0.01	0.01	0.01	0.01	0.07	0.01	0.01	0.05	0.05	0.31	0.06	0.05	0.00	0.00	0.00	0.00	0.00
	0.2	0.02	0.02	0.08	0.01	0.01	0.01	0.02	0.08	0.01	0.01	0.07	0.06	0.32	0.07	0.05	0.00	0.00	0.00	0.00	0.00
	0.3	0.04	0.03	0.12	0.03	0.02	0.04	0.03	0.12	0.03	0.02	0.10	0.08	0.33	0.09	0.06	0.00	0.00	0.00	0.00	0.00
	0.4	0.04	0.05	0.16	0.05	0.04	0.04	0.04	0.16	0.05	0.04	0.10	0.10	0.37	0.11	0.08	0.00	0.00	0.00	0.00	0.00
κ=2	0.5	0.08	0.08	0.21	0.08	0.08	0.08	0.08	0.21	0.08	0.08	0.15	0.15	0.39	0.14	0.15	0.00	0.00	0.00	0.00	0.00
	0.6	0.11	0.12	0.27	0.12	0.13	0.11	0.11	0.27	0.12	0.13	0.17	0.18	0.41	0.19	0.21	0.00	0.00	0.00	0.00	0.00
	0.7	0.18	0.19	0.32	0.19	0.22	0.18	0.19	0.32	0.19	0.22	0.25	0.27	0.45	0.27	0.30	0.00	0.00	0.00	0.00	0.00
	0.8	0.23	0.25	0.35	0.25	0.26	0.23	0.24	0.35	0.25	0.26	0.31	0.32	0.46	0.33	0.34	0.00	0.00	0.00	0.00	0.00
	0.9	0.28	0.30	0.38	0.29	0.29	0.28	0.29	0.37	0.29	0.29	0.35	0.37	0.47	0.36	0.35	0.00	0.00	0.00	0.00	0.00
	1	0.30	0.32	0.37	0.31	0.33	0.30	0.32	0.38	0.31	0.33	0.36	0.39	0.48	0.38	0.42	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.01	0.01	0.07	0.01	0.01	0.01	0.01	0.07	0.01	0.01	0.04	0.04	0.31	0.04	0.03	0.00	0.00	0.00	0.00	0.00
	0.1	0.02	0.02	0.09	0.02	0.02	0.02	0.02	0.09	0.02	0.02	0.07	0.09	0.35	0.08	0.09	0.00	0.00	0.00	0.00	0.00
	0.2	0.06	0.07	0.16	0.07	0.07	0.06	0.07	0.16	0.07	0.07	0.16	0.17	0.41	0.17	0.16	0.00	0.00	0.00	0.00	0.00
	0.3	0.18	0.19	0.29	0.18	0.21	0.18	0.19	0.29	0.18	0.21	0.30	0.31	0.48	0.31	0.35	0.00	0.00	0.00	0.00	0.00
	0.4	0.37	0.36	0.42	0.38	0.36	0.37	0.35	0.42	0.38	0.36	0.50	0.47	0.57	0.50	0.48	0.00	0.00	0.00	0.00	0.00
κ=5	0.5	0.62	0.59	0.57	0.62	0.61	0.62	0.59	0.57	0.62	0.61	0.70	0.66	0.64	0.70	0.70	0.00	0.00	0.00	0.00	0.00
	0.6	0.80	0.77	0.68	0.80	0.77	0.80	0.77	0.68	0.80	0.77	0.85	0.81	0.72	0.85	0.83	0.00	0.00	0.00	0.00	0.00
	0.7	0.92	0.90	0.77	0.92	0.92	0.92	0.90	0.77	0.92	0.92	0.94	0.92	0.78	0.94	0.95	0.00	0.00	0.00	0.00	0.00
	0.8	0.97	0.96	0.83	0.97	0.98	0.97	0.96	0.83	0.97	0.98	0.98	0.97	0.84	0.98	0.99	0.00	0.00	0.00	0.00	0.00
	0.9	0.99	0.99	0.88	0.99	0.99	0.99	0.99	0.88	0.99	0.99	1.00	0.99	0.88	1.00	0.99	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.99	0.88	1.00	1.00	1.00	0.99	0.88	1.00	1.00	1.00	0.99	0.88	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.01	0.01	0.06	0.01	0.02	0.01	0.01	0.06	0.01	0.02	0.04	0.03	0.31	0.03	0.09	0.00	0.00	0.00	0.00	0.00
	0.1	0.03	0.02	0.10	0.03	0.04	0.03	0.02	0.10	0.03	0.04	0.10	0.09	0.35	0.10	0.13	0.00	0.00	0.00	0.00	0.00
	0.2	0.13	0.12	0.21	0.13	0.11	0.13	0.12	0.21	0.13	0.11	0.26	0.24	0.44	0.26	0.23	0.00	0.00	0.00	0.00	0.00
	0.3	0.32	0.28	0.38	0.32	0.32	0.32	0.28	0.38	0.32	0.32	0.47	0.40	0.56	0.46	0.44	0.00	0.00	0.00	0.00	0.00
	0.4	0.60	0.55	0.55	0.60	0.57	0.60	0.54	0.55	0.60	0.57	0.69	0.62	0.63	0.69	0.67	0.00	0.00	0.00	0.00	0.00
κ=7	0.5	0.85	0.81	0.75	0.84	0.84	0.85	0.81	0.75	0.84	0.84	0.88	0.84	0.79	0.88	0.88	0.00	0.00	0.00	0.00	0.00
	0.6	0.95	0.93	0.83	0.95	0.94	0.95	0.93	0.83	0.95	0.94	0.97	0.94	0.84	0.97	0.97	0.00	0.00	0.00	0.00	0.00
	0.7	0.99	0.98	0.89	0.98	0.99	0.99	0.98	0.89	0.98	0.99	0.99	0.98	0.90	0.99	0.99	0.00	0.00	0.00	0.00	0.00
	0.8	1.00	0.99	0.95	1.00	1.00	1.00	0.99	0.95	1.00	1.00	1.00	0.99	0.96	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	0.9	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.00	0.00	0.00	0.00	0.00

Appendix 2, continued.

n=10		*P1*					*P3*					*P5*					*P1-P3*				
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*Chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.04	0.03	0.09	0.05	0.03	0.00	0.00	0.00	0.00	0.00
	0.1	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.01	0.01	0.01	0.04	0.04	0.10	0.06	0.04	0.00	0.00	0.00	0.00	0.00
	0.2	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.05	0.05	0.10	0.06	0.05	0.00	0.00	0.00	0.00	0.00
	0.3	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.07	0.05	0.13	0.07	0.06	0.00	0.00	0.00	0.00	0.00
	0.4	0.03	0.02	0.04	0.02	0.02	0.03	0.02	0.04	0.02	0.02	0.08	0.06	0.12	0.07	0.07	0.00	0.00	0.00	0.00	0.00
κ=2	0.5	0.04	0.03	0.05	0.03	0.03	0.04	0.03	0.05	0.03	0.03	0.09	0.07	0.12	0.08	0.08	0.00	0.00	0.00	0.00	0.00
	0.6	0.07	0.06	0.07	0.06	0.07	0.07	0.06	0.07	0.07	0.07	0.13	0.11	0.13	0.12	0.13	0.00	0.00	0.00	0.00	0.00
	0.7	0.12	0.10	0.10	0.10	0.10	0.12	0.09	0.10	0.10	0.10	0.18	0.14	0.15	0.16	0.15	0.00	0.00	0.00	0.00	0.00
	0.8	0.14	0.13	0.11	0.15	0.14	0.14	0.12	0.11	0.15	0.14	0.19	0.17	0.14	0.20	0.19	0.00	0.00	0.00	0.00	0.00
	0.9	0.21	0.17	0.13	0.19	0.18	0.21	0.16	0.13	0.19	0.18	0.26	0.20	0.16	0.23	0.24	0.00	0.00	0.00	0.00	0.00
	1	0.21	0.18	0.13	0.20	0.21	0.21	0.18	0.13	0.20	0.21	0.26	0.22	0.17	0.25	0.26	0.00	0.01	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.07	0.05	0.12	0.06	0.05	0.00	0.00	0.00	0.00	0.00
	0.1	0.01	0.01	0.03	0.01	0.01	0.01	0.01	0.03	0.01	0.01	0.08	0.08	0.17	0.08	0.08	0.00	0.00	0.00	0.00	0.00
	0.2	0.04	0.04	0.06	0.04	0.04	0.04	0.04	0.06	0.04	0.04	0.15	0.14	0.19	0.14	0.13	0.00	0.00	0.00	0.00	0.00
	0.3	0.15	0.13	0.14	0.14	0.11	0.15	0.13	0.14	0.14	0.11	0.28	0.25	0.27	0.28	0.22	0.00	0.00	0.00	0.00	0.00
	0.4	0.31	0.29	0.23	0.30	0.31	0.31	0.28	0.23	0.30	0.31	0.44	0.41	0.33	0.43	0.44	0.00	0.00	0.00	0.00	0.00
κ=5	0.5	0.57	0.51	0.40	0.57	0.55	0.57	0.51	0.40	0.57	0.56	0.66	0.58	0.46	0.65	0.65	0.00	0.00	0.00	0.00	0.00
	0.6	0.80	0.74	0.56	0.79	0.78	0.80	0.74	0.56	0.79	0.78	0.85	0.79	0.59	0.84	0.83	0.00	0.00	0.00	0.00	0.00
	0.7	0.92	0.87	0.68	0.92	0.92	0.92	0.87	0.68	0.92	0.92	0.94	0.88	0.69	0.93	0.93	0.00	0.00	0.00	0.00	0.00
	0.8	0.97	0.95	0.79	0.97	0.98	0.97	0.95	0.79	0.97	0.98	0.98	0.96	0.80	0.98	0.99	0.00	0.00	0.00	0.00	0.00
	0.9	0.99	0.98	0.84	0.99	0.99	0.99	0.98	0.84	0.99	0.99	0.99	0.99	0.84	0.99	1.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.99	0.85	1.00	1.00	1.00	0.99	0.85	1.00	1.00	1.00	0.99	0.85	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.03	0.03	0.12	0.04	0.07	0.00	0.00	0.00	0.00	0.00
	0.1	0.02	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.11	0.08	0.14	0.12	0.09	0.00	0.00	0.00	0.00	0.00
	0.2	0.08	0.06	0.09	0.09	0.10	0.08	0.06	0.09	0.09	0.10	0.21	0.16	0.23	0.22	0.25	0.00	0.00	0.00	0.00	0.00
	0.3	0.25	0.21	0.21	0.27	0.28	0.25	0.21	0.21	0.27	0.28	0.40	0.32	0.32	0.42	0.43	0.00	0.00	0.00	0.00	0.00
	0.4	0.55	0.46	0.39	0.56	0.55	0.55	0.46	0.39	0.56	0.55	0.65	0.55	0.47	0.67	0.65	0.00	0.00	0.00	0.00	0.00
κ=7	0.5	0.80	0.73	0.60	0.81	0.83	0.80	0.72	0.60	0.81	0.83	0.86	0.78	0.65	0.87	0.88	0.00	0.00	0.00	0.00	0.00
	0.6	0.95	0.92	0.78	0.95	0.94	0.95	0.92	0.78	0.95	0.94	0.96	0.93	0.79	0.97	0.96	0.00	0.00	0.00	0.00	0.00
	0.7	0.99	0.97	0.89	0.99	0.99	0.99	0.97	0.89	0.99	0.99	0.99	0.98	0.89	0.99	0.99	0.00	0.00	0.00	0.00	0.00
	0.8	1.00	0.99	0.95	1.00	1.00	1.00	0.99	0.95	1.00	1.00	1.00	0.99	0.95	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	0.9	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	1.00	1.00	0.97	1.00	1.00	0.00	0.00	0.00	0.00	0.00

Appendix 2, continued.

n=20		*P1*					*P3*					*P5*					*P1-P3*				
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.05	0.02	0.02	0.02	0.03	0.00	0.00	0.00	0.00	0.00
	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.02	0.02	0.04	0.05	0.00	0.00	0.00	0.00	0.00
	0.2	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.05	0.04	0.02	0.05	0.05	0.00	0.00	0.00	0.00	0.00
	0.3	0.01	0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.05	0.05	0.04	0.07	0.06	0.00	0.00	0.00	0.00	0.00
	0.4	0.01	0.02	0.01	0.02	0.01	0.01	0.01	0.01	0.02	0.01	0.06	0.05	0.04	0.08	0.06	0.00	0.00	0.00	0.00	0.00
κ=2	0.5	0.01	0.02	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.02	0.05	0.08	0.05	0.08	0.08	0.00	0.00	0.00	0.00	0.00
	0.6	0.04	0.04	0.02	0.03	0.03	0.04	0.03	0.02	0.03	0.03	0.09	0.09	0.05	0.08	0.10	0.00	0.00	0.00	0.00	0.00
	0.7	0.05	0.06	0.03	0.06	0.07	0.05	0.05	0.03	0.06	0.07	0.10	0.10	0.05	0.11	0.15	0.00	0.00	0.00	0.00	0.00
	0.8	0.08	0.09	0.03	0.09	0.09	0.08	0.08	0.03	0.09	0.10	0.13	0.13	0.05	0.14	0.16	0.00	0.00	0.00	0.00	0.00
	0.9	0.10	0.10	0.04	0.11	0.11	0.10	0.10	0.04	0.11	0.11	0.15	0.14	0.06	0.16	0.16	0.00	0.00	0.00	0.00	0.00
	1	0.11	0.11	0.04	0.12	0.14	0.11	0.11	0.04	0.12	0.14	0.15	0.15	0.06	0.17	0.20	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.01	0.02	0.02	0.05	0.00	0.00	0.00	0.00	0.00
	0.1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.09	0.07	0.10	0.09	0.08	0.00	0.00	0.00	0.00	0.00
	0.2	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.11	0.11	0.11	0.11	0.11	0.00	0.00	0.00	0.00	0.00
	0.3	0.08	0.07	0.08	0.08	0.09	0.08	0.07	0.08	0.08	0.09	0.21	0.18	0.20	0.19	0.21	0.00	0.00	0.00	0.00	0.00
	0.4	0.23	0.20	0.17	0.21	0.24	0.23	0.19	0.17	0.21	0.24	0.37	0.31	0.28	0.34	0.37	0.00	0.00	0.00	0.00	0.00
κ=5	0.5	0.48	0.43	0.35	0.47	0.47	0.48	0.42	0.35	0.47	0.47	0.59	0.52	0.43	0.57	0.57	0.00	0.00	0.00	0.00	0.00
	0.6	0.73	0.64	0.55	0.72	0.73	0.73	0.64	0.55	0.72	0.73	0.79	0.70	0.60	0.78	0.78	0.00	0.00	0.00	0.00	0.00
	0.7	0.88	0.82	0.69	0.87	0.90	0.88	0.82	0.69	0.87	0.90	0.91	0.84	0.71	0.90	0.92	0.00	0.00	0.00	0.00	0.00
	0.8	0.96	0.93	0.82	0.96	0.96	0.96	0.93	0.82	0.96	0.96	0.98	0.94	0.83	0.97	0.97	0.00	0.00	0.00	0.00	0.00
	0.9	0.99	0.97	0.90	0.99	0.99	0.99	0.97	0.90	0.99	0.99	0.99	0.97	0.90	0.99	0.99	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.99	0.93	1.00	1.00	1.00	0.99	0.93	1.00	1.00	1.00	0.99	0.93	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.07	0.04	0.06	0.00	0.00	0.00	0.00	0.00
	0.1	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.08	0.08	0.13	0.09	0.06	0.00	0.00	0.00	0.00	0.00
	0.2	0.06	0.05	0.06	0.07	0.05	0.06	0.05	0.06	0.07	0.05	0.20	0.15	0.19	0.21	0.17	0.00	0.00	0.00	0.00	0.00
	0.3	0.20	0.14	0.16	0.21	0.17	0.20	0.14	0.16	0.21	0.17	0.35	0.25	0.28	0.36	0.31	0.00	0.00	0.00	0.00	0.00
	0.4	0.46	0.37	0.36	0.47	0.45	0.46	0.37	0.36	0.47	0.45	0.58	0.47	0.46	0.60	0.58	0.00	0.00	0.00	0.00	0.00
κ=7	0.5	0.77	0.67	0.62	0.78	0.73	0.77	0.67	0.62	0.78	0.73	0.82	0.71	0.66	0.83	0.78	0.00	0.00	0.00	0.00	0.00
	0.6	0.92	0.86	0.82	0.92	0.91	0.92	0.86	0.82	0.92	0.91	0.94	0.87	0.83	0.94	0.93	0.00	0.00	0.00	0.00	0.00
	0.7	0.98	0.97	0.94	0.99	0.99	0.98	0.96	0.94	0.98	0.99	0.99	0.97	0.94	0.99	0.99	0.00	0.00	0.00	0.00	0.00
	0.8	1.00	0.99	0.97	1.00	1.00	1.00	0.99	0.97	1.00	1.00	1.00	0.99	0.97	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	0.9	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00	1.00	1.00	0.99	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.00	0.00	0.00	0.00	0.00

Appendix 2, continued.

n=50		*P1*					*P3*					*P5*					*P1-P3*				
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.04	0.00	0.00	0.00	0.00	0.00
	0.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.05	0.00	0.00	0.00	0.00	0.00
	0.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.00
	0.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.05	0.00	0.00	0.00	0.00	0.00
	0.4	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.01	0.00	0.07	0.07	0.00	0.00	0.00	0.00	0.00
κ=2	0.5	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.03	0.01	0.00	0.07	0.07	0.00	0.00	0.00	0.00	0.00
	0.6	0.00	0.00	0.00	0.02	0.02	0.00	0.00	0.00	0.01	0.02	0.05	0.02	0.01	0.08	0.08	0.00	0.00	0.00	0.00	0.00
	0.7	0.00	0.00	0.00	0.03	0.03	0.00	0.00	0.00	0.03	0.03	0.08	0.05	0.01	0.08	0.09	0.00	0.00	0.00	0.00	0.00
	0.8	0.01	0.00	0.00	0.04	0.04	0.01	0.00	0.00	0.03	0.04	0.09	0.07	0.02	0.09	0.09	0.00	0.00	0.00	0.00	0.00
	0.9	0.01	0.01	0.00	0.04	0.05	0.01	0.01	0.00	0.04	0.05	0.09	0.08	0.03	0.10	0.09	0.00	0.00	0.00	0.00	0.00
	1	0.02	0.01	0.01	0.05	0.05	0.02	0.01	0.01	0.05	0.04	0.10	0.09	0.05	0.10	0.10	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.00	0.00	0.10	0.10	0.00	0.01	0.00	0.00	0.00
	0.1	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.03	0.00	0.01	0.00	0.00	0.00
	0.2	0.01	0.01	0.00	0.01	0.01	0.01	0.00	0.00	0.01	0.01	0.05	0.00	0.00	0.08	0.08	0.00	0.01	0.00	0.00	0.00
	0.3	0.03	0.01	0.00	0.05	0.05	0.03	0.01	0.00	0.05	0.05	0.10	0.02	0.00	0.17	0.16	0.00	0.01	0.00	0.00	0.00
	0.4	0.09	0.05	0.01	0.14	0.14	0.09	0.02	0.00	0.14	0.14	0.17	0.05	0.00	0.26	0.26	0.00	0.01	0.00	0.00	0.00
κ=5	0.5	0.23	0.19	0.13	0.34	0.34	0.23	0.19	0.13	0.34	0.34	0.31	0.17	0.10	0.46	0.46	0.00	0.01	0.00	0.00	0.00
	0.6	0.45	0.37	0.23	0.57	0.57	0.45	0.37	0.23	0.57	0.57	0.51	0.35	0.25	0.65	0.64	0.00	0.00	0.00	0.00	0.00
	0.7	0.72	0.68	0.41	0.81	0.81	0.72	0.68	0.41	0.81	0.81	0.76	0.70	0.45	0.85	0.85	0.00	0.00	0.00	0.00	0.00
	0.8	0.87	0.72	0.49	0.93	0.93	0.87	0.72	0.49	0.93	0.93	0.89	0.76	0.50	0.95	0.95	0.00	0.00	0.00	0.00	0.00
	0.9	0.96	0.86	0.56	0.97	0.97	0.96	0.86	0.56	0.97	0.97	0.98	0.88	0.55	0.99	0.99	0.00	0.00	0.00	0.00	0.00
	1	0.99	0.95	0.78	0.99	0.99	0.99	0.95	0.78	0.99	0.99	0.99	0.96	0.79	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	λ	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*	*C*	*D*	*M*	*chord*	*A*
	0	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
	0.1	0.01	0.01	0.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00	0.10	0.01	0.00	0.07	0.07	0.00	0.00	0.00	0.00	0.00
	0.2	0.04	0.02	0.01	0.03	0.02	0.04	0.01	0.01	0.03	0.02	0.20	0.02	0.01	0.13	0.12	0.00	0.01	0.00	0.00	0.00
	0.3	0.17	0.10	0.05	0.13	0.12	0.17	0.10	0.05	0.13	0.12	0.37	0.10	0.04	0.27	0.27	0.00	0.01	0.00	0.00	0.00
	0.4	0.41	0.35	0.09	0.34	0.34	0.41	0.34	0.09	0.34	0.34	0.57	0.34	0.11	0.48	0.47	0.00	0.01	0.00	0.00	0.00
κ=7	0.5	0.70	0.59	0.20	0.65	0.64	0.70	0.59	0.20	0.65	0.64	0.79	0.58	0.22	0.73	0.72	0.00	0.00	0.00	0.00	0.00
	0.6	0.92	0.80	0.32	0.89	0.88	0.92	0.80	0.32	0.89	0.88	0.94	0.81	0.30	0.91	0.90	0.00	0.00	0.00	0.00	0.00
	0.7	0.98	0.89	0.36	0.98	0.97	0.98	0.89	0.36	0.98	0.97	0.99	0.88	0.37	0.99	0.98	0.00	0.00	0.00	0.00	0.00
	0.8	1.00	0.95	0.53	1.00	1.00	1.00	0.95	0.53	1.00	1.00	1.00	0.96	0.50	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	0.9	1.00	0.99	0.76	1.00	1.00	1.00	0.99	0.76	1.00	1.00	1.00	0.99	0.73	1.00	1.00	0.00	0.00	0.00	0.00	0.00
	1	1.00	0.99	0.89	1.00	1.00	1.00	0.99	0.89	1.00	1.00	1.00	1.00	0.85	0.00	1.00	0.00	0.00	0.00	0.00	0.00

Appendix 3
Percentage points for $N(\kappa d_j)$

		n						
κ	Percentages	**10**	**30**	**50**	**70**	**100**	**150**	**200**
1	90%	10	0.0	0	0.0	0	0.0	0.0
	95%	20	0.0	0	0.0	0	0.0	0.0
	99%	50	0.0	0	0.0	0	0.0	0.0
2	90%	40	56.7	66	65.7	64	66.0	65.5
	95%	40	70.0	76	74.3	71	72.0	70.5
	99%	60	90.0	86	85.7	84	80.7	80.5
5	90%	40	56.7	68	72.9	79	85.3	88.0
	95%	40	66.7	78	81.4	86	91.3	94.5
	99%	60	86.7	92	92.9	97	98.7	98.0
10	90%	40	56.7	64	67.1	73	76.7	79.5
	95%	40	63.3	74	74.3	79	82.7	84.5
	99%	60	76.7	84	87.1	88	92.7	93.5
30	90%	40	53.3	60	65.7	70	75.3	75.5
	95%	40	63.3	68	74.3	77	81.3	80.0
	99%	60	76.7	80	87.1	89	90.0	90.5
50	90%	40	53.3	62	65.7	70	73.3	75.5
	95%	40	60.0	70	71.4	76	78.7	81.5
	99%	60	80.0	82	84.3	87	87.3	88.5
70	90%	40	53.3	62	67.1	69	72.0	76.5
	95%	40	60.0	68	74.3	75	78.0	81.5
	99%	60	76.7	82	81.4	84	88.7	90.0
100	90%	40	56.7	62	64.3	68	73.3	76.0
	95%	40	63.3	70	71.4	73	79.3	80.0
	99%	60	76.7	82	84.3	86	88.0	89.5
200	90%	40	53.3	60	64.3	69	74.0	77.0
	95%	40	60.0	68	71.4	76	80.0	83.5
	99%	60	73.3	82	82.9	85	87.3	91.0
500	90%	40	53.3	62	65.7	68	73.3	76.0
	95%	40	63.3	70	72.9	75	78.7	80.0
	99%	60	76.7	84	84.3	87	87.3	89.5
1000	90%	40	53.3	60	67.1	67	73.3	77.0
	95%	50	60.0	66	71.4	73	78.7	83.5
	99%	60	76.7	80	82.9	85	87.3	91.0

Appendix 4
Measures of Circular Boxplot

***n*=5**

κ	*CIQR*	*d* \ υ	1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
			B(υ,*n*)									
0.5	144.1	12.8	0.8	0.8	0.8	0.8	0.7	0.7	0.7	0.7	0.7	0.7
1	129.2	11.1	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8	0.8
2	95.3	8.7	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.8	0.9
3	75.3	7.4	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9	0.9	0.9
4	62.7	6.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9	0.9
5	55.6	5.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.9
6	50.1	5.4	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
7	45.5	5.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
8	41.9	4.6	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
9	38.0	4.3	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
10	36.0	4.0	0.9	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

κ	B3(υ,*n*)									
0.5	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
1	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1	0.1	0.1
2	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.1	0.1
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*)									
0.5	6.4	7.0	8.3	10.8	13.4	12.6	13.1	11.6	11.1	11.3
1	4.7	5.0	6.6	9.4	10.8	10.6	10.8	11.1	10.2	9.9
2	1.6	1.6	2.2	2.7	4.3	5.9	6.3	6.8	7.5	7.2
3	1.1	0.7	0.4	0.7	1.7	2.2	3.4	4.0	5.1	5.8
4	0.9	0.4	0.3	0.3	0.6	0.7	1.3	2.0	2.9	3.5
5	0.8	0.3	0.2	0.1	0.3	0.3	0.5	0.7	1.5	2.8
6	1.0	0.5	0.3	0.1	0.3	0.1	0.2	0.5	0.9	1.7
7	1.0	0.4	0.2	0.1	0.2	0.0	0.1	0.1	0.4	1.1
8	1.2	0.6	0.3	0.2	0.2	0.0	0.1	0.1	0.4	0.7
9	1.5	0.6	0.3	0.1	0.3	0.0	0.0	0.0	0.2	0.4
10	1.5	0.6	0.3	0.2	0.4	0.0	0.0	0.0	0.2	0.2

***n*=6**

κ	*CIQR*	*d* \ υ	1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
			B(υ,*n*)									
0.5	125.4	13.2	0.7	0.8	0.8	0.8	0.7	0.6	0.6	0.6	0.6	0.6
1	99.4	10.1	0.6	0.7	0.8	0.8	0.8	0.8	0.8	0.7	0.7	0.7
2	65.3	7.0	0.6	0.7	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9
3	51.4	5.8	0.6	0.7	0.8	0.8	0.9	0.9	0.9	0.9	0.9	0.9
4	43.2	5.1	0.6	0.7	0.8	0.9	0.9	0.9	0.9	1.0	1.0	1.0
5	38.7	4.7	0.6	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.0
6	34.7	4.2	0.6	0.8	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
7	32.0	3.8	0.7	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
8	29.8	3.6	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
9	28.1	3.5	0.6	0.8	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
10	26.4	3.3	0.7	0.8	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*)									
0.5	0.0	0.0	0.1	0.1	0.2	0.2	0.2	0.2	0.2	0.2
1	0.0	0.0	0.0	0.0	0.1	0.1	0.1	0.2	0.2	0.2
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*)									
0.5	4.5	3.7	5.4	9.1	13.8	16.1	16.5	15.5	14.5	14.1
1	5.8	4.2	3.5	4.5	6.8	8.2	9.9	11.5	13.2	13.2
2	6.4	4.6	3.0	2.3	1.8	1.8	2.0	2.7	3.6	6.1
3	6.0	4.3	2.6	1.9	1.4	1.0	0.8	0.9	1.0	1.7
4	5.5	3.9	2.3	1.8	1.2	0.7	0.6	0.5	0.4	0.5
5	5.3	3.5	2.0	1.5	1.0	0.6	0.5	0.4	0.4	0.2
6	5.1	3.5	2.1	1.5	0.9	0.6	0.4	0.3	0.3	0.1
7	5.1	3.5	2.0	1.5	1.0	0.6	0.5	0.4	0.4	0.1
8	5.2	3.5	2.2	1.5	1.0	0.5	0.4	0.3	0.2	0.1
9	5.0	3.3	2.0	1.5	1.0	0.6	0.4	0.3	0.3	0.1
10	5.2	3.4	2.1	1.5	1.1	0.6	0.5	0.4	0.3	0.1

d is the circular distance between the mean and median in degrees

Appendix 4, continued.

***n*=15**		**υ**	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
κ	***CIQR***	***d***					B(υ,*n*)					
0.5	130.0	11.8	0.8	0.8	0.9	0.8	0.7	0.6	0.6	0.5	0.6	0.6
1	99.9	8.6	0.6	0.7	0.8	0.9	0.9	0.8	0.8	0.7	0.7	0.6
2	65.1	5.6	0.4	0.6	0.7	0.8	0.9	0.9	0.9	0.9	0.9	0.9
3	50.4	4.7	0.5	0.6	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0
4	43.0	4.1	0.5	0.6	0.8	0.8	0.9	0.9	1.0	1.0	1.0	1.0
5	38.0	3.8	0.5	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.0
6	34.4	3.4	0.5	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0	1.0
7	31.9	3.1	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
8	29.6	3.0	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
9	27.8	2.8	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
10	26.3	2.7	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0	1.0
κ							B3(υ,*n*)					
0.5			0.1	0.0	0.1	0.1	0.3	0.3	0.4	0.4	0.3	0.3
1			0.1	0.1	0.0	0.0	0.1	0.1	0.2	0.2	0.3	0.3
2			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
3			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10			0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
κ							A(υ,*n*)					
0.5			2.9	2.2	3.9	9.1	17.1	19.9	19.9	19.6	17.4	15.3
1			5.5	3.6	2.2	2.6	4.8	8.4	10.2	13.6	15.6	17.0
2			6.2	4.4	2.6	1.9	1.3	0.8	0.9	1.2	2.0	4.8
3			5.5	3.9	2.3	1.6	0.9	0.6	0.4	0.3	0.3	0.6
4			5.0	3.4	1.9	1.3	0.8	0.5	0.3	0.2	0.2	0.1
5			4.9	3.3	1.8	1.2	0.7	0.4	0.3	0.2	0.1	0.0
6			4.9	3.1	1.7	1.2	0.7	0.4	0.3	0.2	0.1	0.1
7			4.8	3.1	1.6	1.1	0.6	0.3	0.2	0.2	0.1	0.0
8			4.7	3.1	1.6	1.1	0.6	0.3	0.3	0.2	0.1	0.0
9			4.8	3.1	1.7	1.1	0.6	0.4	0.3	0.2	0.1	0.0
10			4.6	2.9	1.6	1.1	0.6	0.3	0.2	0.1	0.1	0.0

***n*=20**		**υ**	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
κ	***CIQR***	***d***					B(υ,*n*)					
0.5	126.7	11.0	0.7	0.8	0.9	0.8	0.7	0.6	0.5	0.5	0.5	0.5
1	95.5	7.6	0.4	0.6	0.8	0.8	0.9	0.9	0.8	0.8	0.7	0.5
2	62.4	4.9	0.3	0.4	0.6	0.7	0.8	0.9	0.9	0.9	0.9	0.9
3	48.9	4.0	0.3	0.5	0.7	0.7	0.8	0.9	0.9	0.9	1.0	1.0
4	40.9	3.5	0.4	0.5	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0
5	36.1	3.3	0.4	0.5	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0
6	32.5	3.0	0.4	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
7	29.9	2.7	0.4	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
8	27.9	2.5	0.4	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
9	26.3	2.5	0.4	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
10	24.8	2.3	0.4	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
κ							B3(υ,*n*)					
0.5			0.1	0.1	0.1	0.1	0.2	0.4	0.4	0.4	0.4	0.4
1			0.2	0.1	0.0	0.0	0.0	0.1	0.1	0.2	0.3	0.4
2			0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1
3			0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4			0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5			0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6			0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7			0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8			0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9			0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10			0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
κ							A(υ,*n*)					
0.5			3.1	2.0	2.2	6.4	15.2	21.2	23.2	23.1	20.3	15.6
1			6.1	3.8	2.0	1.6	2.6	5.7	8.6	11.7	14.7	19.3
2			7.2	5.1	3.0	2.2	1.3	0.8	0.6	0.7	1.0	3.5
3			6.2	4.4	2.4	1.8	1.1	0.7	0.5	0.3	0.3	0.2
4			6.0	4.1	2.2	1.5	0.9	0.5	0.4	0.2	0.2	0.1
5			5.9	3.9	2.0	1.3	0.8	0.4	0.3	0.2	0.1	0.1
6			5.9	3.8	2.0	1.3	0.7	0.4	0.3	0.2	0.1	0.0
7			5.8	3.7	1.9	1.2	0.7	0.4	0.2	0.2	0.1	0.0
8			5.6	3.6	1.9	1.2	0.6	0.3	0.2	0.1	0.1	0.0
9			5.7	3.7	1.9	1.2	0.6	0.3	0.2	0.2	0.1	0.0
10			5.7	3.7	1.9	1.2	0.7	0.4	0.3	0.2	0.1	0.0

Appendix 4, continued.

n=25		υ	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
κ	***CIQR***	***d***	B(υ,*n*)									
0.5	127.3	7.7	0.7	0.9	1.0	1.0	0.8	0.6	0.4	0.3	0.2	0.4
1	93.3	4.5	0.1	0.3	0.7	0.9	1.0	1.0	1.0	0.9	0.8	0.4
2	61.3	2.9	0.0	0.1	0.2	0.4	0.6	0.8	0.9	0.9	1.0	1.0
3	47.5	2.5	0.0	0.1	0.3	0.5	0.6	0.8	0.8	0.9	0.9	1.0
4	40.6	2.1	0.1	0.2	0.4	0.6	0.8	0.9	0.9	0.9	1.0	1.0
5	36.0	1.9	0.1	0.2	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0
6	32.5	1.7	0.1	0.2	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0
7	29.9	1.6	0.1	0.2	0.5	0.7	0.8	0.9	1.0	1.0	1.0	1.0
8	27.8	1.5	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0
9	26.2	1.4	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0
10	24.7	1.4	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0

n=30	υ	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
CIQR	***d***	B(υ,*n*)									
127.3	10.1	0.7	0.8	0.9	0.9	0.7	0.6	0.5	0.4	0.4	0.5
94.2	6.4	0.3	0.5	0.7	0.8	0.9	0.9	0.9	0.8	0.7	0.5
61.5	4.1	0.1	0.3	0.5	0.6	0.7	0.8	0.9	0.9	1.0	1.0
48.2	3.3	0.2	0.3	0.5	0.7	0.8	0.9	0.9	0.9	1.0	1.0
40.9	2.9	0.3	0.4	0.6	0.7	0.9	0.9	0.9	1.0	1.0	1.0
36.1	2.6	0.3	0.4	0.7	0.8	0.9	0.9	0.9	1.0	1.0	1.0
32.6	2.4	0.3	0.4	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
30.2	2.3	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
28.0	2.1	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
26.5	2.0	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0
25.0	1.9	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*) (*n*=25)										B3(υ,*n*) (*n*=30)									
0.5	0.2	0.0	0.0	0.0	0.2	0.4	0.6	0.7	0.8	0.6	0.1	0.1	0.0	0.1	0.2	0.4	0.5	0.5	0.5	0.4
1	0.7	0.4	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.5	0.3	0.2	0.1	0.0	0.0	0.1	0.1	0.2	0.3	0.5
2	0.8	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.7	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*) (*n*=25)										A(υ,*n*) (*n*=30)									
0.5	1.6	0.5	0.3	1.6	11.5	26.8	34.7	36.3	31.2	16.1	2.6	1.3	1.4	4.4	14.9	24.0	26.1	26.2	23.6	16.1
1	6.8	3.8	1.1	0.4	0.2	1.1	2.8	6.2	14.6	28.8	6.5	3.8	1.7	1.0	1.2	3.6	6.2	9.8	16.3	22.0
2	7.3	5.1	2.9	2.0	1.0	0.5	0.3	0.1	0.1	0.1	7.3	5.1	3.0	2.1	1.2	0.7	0.4	0.3	0.3	1.6
3	6.3	4.2	2.2	1.5	0.8	0.5	0.3	0.2	0.1	0.0	6.4	4.2	2.3	1.6	0.9	0.5	0.3	0.2	0.1	0.1
4	5.7	3.5	1.7	1.1	0.5	0.3	0.2	0.1	0.1	0.0	5.8	3.7	1.9	1.2	0.6	0.4	0.3	0.2	0.1	0.0
5	5.4	3.3	1.5	0.8	0.4	0.2	0.1	0.1	0.0	0.0	5.6	3.6	1.8	1.2	0.6	0.3	0.2	0.1	0.1	0.0
6	5.4	3.2	1.4	0.8	0.4	0.2	0.1	0.1	0.0	0.0	5.7	3.5	1.7	1.0	0.5	0.2	0.2	0.1	0.1	0.0
7	5.3	3.1	1.4	0.8	0.3	0.1	0.1	0.0	0.0	0.0	5.4	3.3	1.6	1.0	0.5	0.2	0.1	0.1	0.0	0.0
8	5.1	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0	5.5	3.3	1.6	1.0	0.5	0.2	0.1	0.1	0.0	0.0
9	5.2	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0	5.1	3.1	1.5	0.9	0.4	0.2	0.1	0.1	0.0	0.0
10	5.1	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0	5.4	3.3	1.6	1.0	0.5	0.2	0.1	0.1	0.0	0.0

Appendix 4, continued.

n=40

κ	*CIQR*	*d*	υ = 1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
			B(υ,*n*)									
0.5	126.7	8.9	0.7	0.9	0.9	0.9	0.8	0.6	0.4	0.4	0.3	0.4
1	93.7	5.5	0.2	0.4	0.7	0.9	0.9	0.9	0.9	0.9	0.7	0.5
2	61.4	3.5	0.1	0.2	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0
3	47.6	2.9	0.1	0.2	0.4	0.6	0.7	0.8	0.9	0.9	0.9	1.0
4	40.5	2.6	0.2	0.3	0.5	0.7	0.8	0.9	0.9	0.9	1.0	1.0
5	35.8	2.3	0.2	0.3	0.6	0.7	0.8	0.9	0.9	1.0	1.0	1.0
6	32.3	2.1	0.2	0.3	0.6	0.7	0.9	0.9	1.0	1.0	1.0	1.0
7	30.0	2.0	0.2	0.4	0.6	0.7	0.9	0.9	1.0	1.0	1.0	1.0
8	27.7	1.9	0.2	0.4	0.6	0.8	0.9	0.9	1.0	1.0	1.0	1.0
9	26.0	1.7	0.2	0.4	0.6	0.8	0.9	0.9	1.0	1.0	1.0	1.0
10	25.0	1.6	0.2	0.4	0.7	0.8	0.9	1.0	1.0	1.0	1.0	1.0

κ	B3(υ,*n*) 1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
0.5	0.1	0.1	0.0	0.1	0.2	0.4	0.5	0.6	0.6	0.5
1	0.5	0.3	0.1	0.0	0.0	0.0	0.1	0.1	0.3	0.5
2	0.5	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
3	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*) 1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
0.5	2.2	0.9	0.7	2.8	13.1	24.8	29.5	29.9	27.3	16.9
1	6.8	3.8	1.4	0.7	0.7	2.4	4.3	8.0	15.4	24.5
2	7.5	5.2	3.0	2.1	1.2	0.6	0.4	0.3	0.2	0.7
3	6.4	4.2	2.3	1.6	0.9	0.5	0.4	0.3	0.1	0.1
4	6.0	3.7	1.9	1.2	0.6	0.4	0.2	0.2	0.1	0.0
5	5.6	3.5	1.7	1.0	0.5	0.3	0.2	0.1	0.1	0.0
6	5.5	3.3	1.6	0.9	0.4	0.2	0.1	0.1	0.0	0.0
7	5.3	3.2	1.5	0.9	0.4	0.2	0.1	0.1	0.0	0.0
8	5.5	3.3	1.5	0.9	0.4	0.2	0.1	0.1	0.0	0.0
9	5.6	3.3	1.5	0.9	0.4	0.2	0.1	0.1	0.0	0.0
10	5.2	3.1	1.4	0.8	0.3	0.1	0.1	0.1	0.0	0.0

n=50

κ	*CIQR*	*d*	υ = 1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
			B(υ,*n*)									
0.5	127.6	8.3	0.7	0.9	1.0	0.9	0.8	0.5	0.4	0.3	0.3	0.4
1	93.5	4.9	0.1	0.4	0.7	0.9	1.0	1.0	0.9	0.9	0.7	0.4
2	61.1	3.2	0.0	0.1	0.3	0.4	0.6	0.8	0.9	0.9	1.0	1.0
3	47.9	2.7	0.1	0.2	0.4	0.5	0.7	0.8	0.9	0.9	0.9	1.0
4	40.6	2.3	0.1	0.2	0.5	0.6	0.8	0.9	0.9	0.9	1.0	1.0
5	36.0	2.0	0.1	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0
6	32.6	1.9	0.1	0.3	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0
7	30.1	1.8	0.1	0.3	0.6	0.7	0.9	0.9	1.0	1.0	1.0	1.0
8	28.1	1.7	0.2	0.3	0.6	0.7	0.9	0.9	1.0	1.0	1.0	1.0
9	26.2	1.5	0.1	0.3	0.6	0.7	0.9	0.9	1.0	1.0	1.0	1.0
10	24.8	1.5	0.1	0.3	0.6	0.7	0.9	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*) 1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
0.5	0.2	0.1	0.0	0.1	0.2	0.4	0.6	0.7	0.7	0.5
1	0.6	0.3	0.1	0.0	0.0	0.0	0.1	0.1	0.3	0.5
2	0.7	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
3	0.6	0.3	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
4	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*) 1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
0.5	1.9	0.7	0.5	2.5	12.8	27.0	31.5	33.3	28.0	16.6
1	6.8	3.8	1.3	0.6	0.4	1.4	3.6	7.8	15.8	27.3
2	7.4	5.1	2.9	2.0	1.1	0.5	0.3	0.2	0.1	0.4
3	6.2	4.0	2.1	1.4	0.8	0.5	0.3	0.2	0.1	0.0
4	5.7	3.6	1.8	1.1	0.6	0.3	0.2	0.1	0.1	0.0
5	5.3	3.2	1.5	0.9	0.4	0.2	0.1	0.1	0.0	0.0
6	5.3	3.2	1.5	0.8	0.4	0.2	0.1	0.1	0.0	0.0
7	5.3	3.1	1.4	0.8	0.4	0.1	0.1	0.0	0.0	0.0
8	5.0	2.9	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
9	5.2	3.1	1.4	0.8	0.3	0.2	0.1	0.0	0.0	0.0
10	5.2	3.1	1.4	0.8	0.3	0.1	0.1	0.0	0.0	0.0

Appendix 4, continued.

***n*=60** κ	***CIQR***	υ / ***d***	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
			B(υ,*n*)									
0.5	127.3	7.7	0.7	0.9	1.0	1.0	0.8	0.6	0.4	0.3	0.2	0.4
1	93.3	4.5	0.1	0.3	0.7	0.9	1.0	1.0	1.0	0.9	0.8	0.4
2	61.3	2.9	0.0	0.1	0.2	0.4	0.6	0.8	0.9	0.9	1.0	1.0
3	47.5	2.5	0.0	0.1	0.3	0.5	0.6	0.8	0.8	0.9	0.9	1.0
4	40.6	2.1	0.1	0.2	0.4	0.6	0.8	0.9	0.9	0.9	1.0	1.0
5	36.0	1.9	0.1	0.2	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0
6	32.5	1.7	0.1	0.2	0.5	0.7	0.8	0.9	0.9	1.0	1.0	1.0
7	29.9	1.6	0.1	0.2	0.5	0.7	0.8	0.9	1.0	1.0	1.0	1.0
8	27.8	1.5	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0
9	26.2	1.4	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0
10	24.7	1.4	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0

***n*=70** κ	***CIQR***	υ / ***d***	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
			B(υ,*n*)									
0.5	127.2	7.3	0.7	0.9	1.0	1.0	0.8	0.6	0.4	0.2	0.2	0.4
1	93.4	4.2	0.1	0.3	0.7	0.9	1.0	1.0	1.0	0.9	0.8	0.4
2	60.8	2.7	0.0	0.0	0.2	0.3	0.5	0.8	0.9	0.9	1.0	1.0
3	47.3	2.2	0.0	0.1	0.3	0.4	0.6	0.7	0.8	0.9	0.9	1.0
4	40.4	2.0	0.1	0.1	0.4	0.5	0.7	0.8	0.9	0.9	1.0	1.0
5	35.7	1.8	0.1	0.2	0.4	0.6	0.8	0.9	0.9	1.0	1.0	1.0
6	32.5	1.6	0.1	0.2	0.5	0.6	0.8	0.9	0.9	1.0	1.0	1.0
7	29.9	1.5	0.1	0.2	0.5	0.7	0.8	0.9	1.0	1.0	1.0	1.0
8	27.9	1.4	0.1	0.2	0.5	0.7	0.8	0.9	1.0	1.0	1.0	1.0
9	26.3	1.3	0.1	0.2	0.5	0.7	0.9	0.9	1.0	1.0	1.0	1.0
10	24.8	1.2	0.1	0.2	0.5	0.7	0.8	0.9	1.0	1.0	1.0	1.0

***n*=60** κ	B3(υ,*n*)									
0.5	0.2	0.0	0.0	0.0	0.2	0.4	0.6	0.7	0.8	0.6
1	0.7	0.4	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.5
2	0.8	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
3	0.7	0.4	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
4	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

***n*=70** κ	B3(υ,*n*)									
0.5	0.2	0.0	0.0	0.0	0.2	0.4	0.6	0.7	0.8	0.6
1	0.7	0.4	0.1	0.0	0.0	0.0	0.0	0.1	0.2	0.6
2	0.9	0.6	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0
3	0.8	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
4	0.7	0.4	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
5	0.6	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

***n*=60** κ	A(υ,*n*)									
0.5	1.6	0.5	0.3	1.6	11.5	26.8	34.7	36.3	31.2	16.1
1	6.8	3.8	1.1	0.4	0.2	1.1	2.8	6.2	14.6	28.8
2	7.3	5.1	2.9	2.0	1.0	0.5	0.3	0.1	0.1	0.1
3	6.3	4.2	2.2	1.5	0.8	0.5	0.3	0.2	0.1	0.0
4	5.7	3.5	1.7	1.1	0.5	0.3	0.2	0.1	0.1	0.0
5	5.4	3.3	1.5	0.8	0.4	0.2	0.1	0.1	0.0	0.0
6	5.4	3.2	1.4	0.8	0.4	0.2	0.1	0.1	0.0	0.0
7	5.3	3.1	1.4	0.8	0.3	0.1	0.1	0.0	0.0	0.0
8	5.1	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
9	5.2	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
10	5.1	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0

***n*=70** κ	A(υ,*n*)									
0.5	1.5	0.4	0.2	1.3	10.2	27.1	36.7	38.7	32.1	16.5
1	6.7	3.6	1.1	0.4	0.2	0.5	2.1	6.3	15.3	30.6
2	7.4	5.1	2.9	2.0	1.1	0.5	0.3	0.1	0.1	0.1
3	6.4	4.1	2.2	1.4	0.8	0.5	0.3	0.2	0.1	0.0
4	5.6	3.5	1.7	1.0	0.5	0.3	0.2	0.1	0.1	0.0
5	5.4	3.2	1.4	0.8	0.4	0.2	0.1	0.1	0.0	0.0
6	5.3	3.1	1.4	0.8	0.3	0.1	0.1	0.1	0.0	0.0
7	5.1	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
8	5.1	2.9	1.2	0.7	0.3	0.1	0.1	0.0	0.0	0.0
9	5.1	2.9	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
10	5.0	2.9	1.2	0.7	0.3	0.1	0.1	0.0	0.0	0.0

Appendix 4, continued.

***n*=100**		υ	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
κ	***CIQR***	***d***	B(υ,*n*)									
0.5	127.6	6.2	0.7	0.9	1.0	1.0	0.9	0.6	0.3	0.2	0.1	0.3
1	92.9	3.5	0.0	0.2	0.7	0.9	1.0	1.0	1.0	1.0	0.8	0.4
2	60.7	2.3	0.0	0.0	0.1	0.2	0.4	0.7	0.8	0.9	1.0	1.0
3	47.6	1.9	0.0	0.0	0.2	0.3	0.5	0.7	0.8	0.8	0.9	1.0
4	40.3	1.6	0.0	0.1	0.3	0.4	0.7	0.8	0.9	0.9	1.0	1.0
5	35.7	1.5	0.0	0.1	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0
6	32.4	1.3	0.0	0.1	0.3	0.5	0.8	0.9	0.9	1.0	1.0	1.0
7	29.9	1.3	0.0	0.1	0.4	0.6	0.8	0.9	1.0	1.0	1.0	1.0
8	27.9	1.2	0.0	0.1	0.4	0.6	0.8	0.9	1.0	1.0	1.0	1.0
9	26.1	1.1	0.0	0.1	0.4	0.6	0.8	0.9	1.0	1.0	1.0	1.0
10	24.8	1.1	0.0	0.1	0.4	0.6	0.8	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*)									
0.5	0.2	0.0	0.0	0.0	0.1	0.4	0.7	0.8	0.9	0.6
1	0.9	0.6	0.1	0.0	0.0	0.0	0.0	0.0	0.2	0.6
2	1.0	0.8	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0
3	0.9	0.7	0.3	0.2	0.1	0.0	0.0	0.0	0.0	0.0
4	0.9	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
5	0.8	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
6	0.8	0.5	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
7	0.8	0.5	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
8	0.8	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.8	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.8	0.5	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*)									
0.5	1.2	0.2	0.0	0.9	7.9	30.1	41.3	44.1	35.0	14.5
1	6.9	3.6	0.9	0.3	0.0	0.0	0.5	3.2	12.9	33.3
2	7.4	5.1	2.9	2.0	1.0	0.5	0.2	0.1	0.0	0.0
3	6.2	3.9	2.1	1.4	0.7	0.4	0.3	0.2	0.1	0.0
4	5.6	3.4	1.6	0.9	0.5	0.2	0.1	0.1	0.0	0.0
5	5.4	3.2	1.4	0.8	0.3	0.2	0.1	0.1	0.0	0.0
6	5.3	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
7	5.1	2.9	1.2	0.6	0.2	0.1	0.0	0.0	0.0	0.0
8	5.0	2.8	1.2	0.6	0.2	0.1	0.0	0.0	0.0	0.0
9	5.1	2.9	1.2	0.6	0.2	0.1	0.0	0.0	0.0	0.0
10	5.0	2.8	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0

***n*=130**		υ	**1.00**	**1.20**	**1.50**	**1.70**	**2.00**	**2.20**	**2.50**	**2.70**	**3.00**	**3.50**
κ	***CIQR***	***d***	B(υ,*n*)									
0.5	127.4	5.5	0.7	1.0	1.0	1.0	0.9	0.6	0.3	0.1	0.1	0.3
1	92.8	3.1	0.0	0.2	0.7	0.9	1.0	1.0	1.0	1.0	0.8	0.4
2	60.6	2.0	0.0	0.0	0.0	0.1	0.3	0.7	0.8	0.9	1.0	1.0
3	47.5	1.7	0.0	0.0	0.1	0.2	0.4	0.6	0.7	0.8	0.9	1.0
4	40.3	1.4	0.0	0.0	0.2	0.3	0.6	0.7	0.8	0.9	0.9	1.0
5	35.8	1.3	0.0	0.0	0.2	0.4	0.7	0.8	0.9	0.9	1.0	1.0
6	32.2	1.2	0.0	0.0	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0
7	29.9	1.1	0.0	0.1	0.3	0.5	0.8	0.9	0.9	1.0	1.0	1.0
8	27.9	1.0	0.0	0.1	0.3	0.5	0.8	0.9	0.9	1.0	1.0	1.0
9	26.3	1.0	0.0	0.1	0.3	0.6	0.8	0.9	1.0	1.0	1.0	1.0
10	24.9	0.9	0.0	0.1	0.3	0.6	0.8	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*)									
0.5	0.2	0.0	0.0	0.0	0.1	0.4	0.7	0.8	0.9	0.7
1	1.0	0.7	0.2	0.0	0.0	0.0	0.0	0.0	0.2	0.6
2	1.0	0.9	0.7	0.4	0.2	0.0	0.0	0.0	0.0	0.0
3	1.0	0.8	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0
4	0.9	0.8	0.3	0.2	0.0	0.0	0.0	0.0	0.0	0.0
5	0.9	0.7	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
6	0.9	0.7	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
7	0.9	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
8	0.9	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
9	0.9	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
10	0.9	0.6	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*)									
0.5	0.9	0.1	0.0	0.2	7.1	29.7	44.0	47.2	36.7	13.8
1	6.9	3.6	0.8	0.2	0.0	0.0	0.4	1.9	11.2	35.6
2	7.3	5.1	2.9	2.0	1.0	0.4	0.2	0.1	0.0	0.0
3	6.1	3.9	2.0	1.3	0.7	0.4	0.3	0.2	0.1	0.0
4	5.7	3.4	1.6	1.0	0.5	0.2	0.2	0.1	0.1	0.0
5	5.2	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
6	5.2	3.0	1.2	0.7	0.3	0.1	0.1	0.0	0.0	0.0
7	5.1	2.9	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0
8	4.9	2.8	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0
9	4.8	2.7	1.0	0.5	0.2	0.1	0.0	0.0	0.0	0.0
10	4.8	2.7	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0

Appendix 4, continued.

***n*=150**

κ	*CIQR*	*d* (υ →)	1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
			B(υ,*n*)									
0.5	127.5	5.1	0.8	1.0	1.0	1.0	0.9	0.6	0.3	0.1	0.0	0.3
1	92.9	2.9	0.0	0.1	0.7	0.9	1.0	1.0	1.0	1.0	0.9	0.4
2	60.6	1.9	0.0	0.0	0.0	0.1	0.3	0.6	0.8	0.9	1.0	1.0
3	47.5	1.6	0.0	0.0	0.1	0.2	0.4	0.6	0.7	0.8	0.9	1.0
4	40.4	1.4	0.0	0.0	0.2	0.3	0.5	0.7	0.8	0.9	0.9	1.0
5	35.7	1.2	0.0	0.0	0.2	0.4	0.6	0.8	0.9	0.9	1.0	1.0
6	32.4	1.1	0.0	0.0	0.2	0.4	0.7	0.9	0.9	1.0	1.0	1.0
7	30.0	1.0	0.0	0.0	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0
8	27.9	1.0	0.0	0.0	0.3	0.5	0.7	0.9	0.9	1.0	1.0	1.0
9	26.2	0.9	0.0	0.0	0.3	0.5	0.8	0.9	1.0	1.0	1.0	1.0
10	24.8	0.9	0.0	0.0	0.3	0.5	0.8	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*)									
0.5	0.2	0.0	0.0	0.0	0.1	0.4	0.7	0.9	0.9	0.7
1	1.0	0.7	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.6
2	1.0	1.0	0.8	0.5	0.2	0.1	0.0	0.0	0.0	0.0
3	1.0	0.9	0.5	0.3	0.1	0.0	0.0	0.0	0.0	0.0
4	1.0	0.8	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0
5	1.0	0.8	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
6	0.9	0.7	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
7	0.9	0.7	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
8	0.9	0.7	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
9	0.9	0.7	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0
10	0.9	0.7	0.2	0.1	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*)									
0.5	0.8	0.1	0.0	0.3	6.2	30.1	45.8	49.7	37.3	13.3
1	7.0	3.6	0.7	0.2	0.0	0.0	0.2	1.5	10.4	38.2
2	7.3	5.0	2.9	2.0	1.0	0.4	0.2	0.1	0.0	0.0
3	6.2	3.9	2.0	1.3	0.7	0.4	0.3	0.2	0.1	0.0
4	5.5	3.3	1.5	0.9	0.4	0.2	0.1	0.1	0.1	0.0
5	5.3	3.1	1.3	0.8	0.3	0.1	0.1	0.1	0.0	0.0
6	5.0	2.8	1.2	0.6	0.2	0.1	0.1	0.0	0.0	0.0
7	5.0	2.8	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0
8	5.0	2.8	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0
9	4.9	2.7	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0
10	4.8	2.6	1.0	0.5	0.2	0.1	0.0	0.0	0.0	0.0

***n*=200**

κ	*CIQR*	*d* (υ →)	1.00	1.20	1.50	1.70	2.00	2.20	2.50	2.70	3.00	3.50
			B(υ,*n*)									
0.5	127.6	4.5	0.8	1.0	1.0	1.0	0.9	0.6	0.3	0.1	0.0	0.3
1	92.8	2.4	0.0	0.1	0.7	0.9	1.0	1.0	1.0	1.0	0.9	0.4
2	60.7	1.6	0.0	0.0	0.0	0.0	0.2	0.6	0.8	0.9	1.0	1.0
3	47.5	1.4	0.0	0.0	0.0	0.1	0.3	0.5	0.6	0.7	0.8	1.0
4	40.3	1.2	0.0	0.0	0.1	0.2	0.5	0.7	0.8	0.8	0.9	1.0
5	35.8	1.0	0.0	0.0	0.1	0.3	0.6	0.8	0.9	0.9	1.0	1.0
6	32.4	1.0	0.0	0.0	0.1	0.3	0.7	0.8	0.9	1.0	1.0	1.0
7	29.9	0.9	0.0	0.0	0.2	0.4	0.7	0.9	0.9	1.0	1.0	1.0
8	27.9	0.9	0.0	0.0	0.2	0.4	0.7	0.9	0.9	1.0	1.0	1.0
9	26.1	0.8	0.0	0.0	0.2	0.4	0.7	0.9	0.9	1.0	1.0	1.0
10	24.9	0.7	0.0	0.0	0.2	0.4	0.7	0.9	1.0	1.0	1.0	1.0

κ	B3(υ,*n*)									
0.5	0.2	0.0	0.0	0.0	0.1	0.4	0.7	0.9	1.0	0.7
1	1.0	0.8	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.6
2	1.0	1.0	0.9	0.7	0.4	0.1	0.0	0.0	0.0	0.0
3	1.0	1.0	0.7	0.5	0.2	0.1	0.0	0.0	0.0	0.0
4	1.0	0.9	0.6	0.3	0.1	0.0	0.0	0.0	0.0	0.0
5	1.0	0.9	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0
6	1.0	0.9	0.4	0.2	0.0	0.0	0.0	0.0	0.0	0.0
7	1.0	0.8	0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0
8	1.0	0.8	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
9	1.0	0.8	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0
10	1.0	0.8	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0

κ	A(υ,*n*)									
0.5	0.6	0.0	0.0	0.0	4.1	32.3	51.0	53.4	37.8	12.0
1	7.1	3.6	0.6	0.1	0.0	0.0	0.1	0.7	6.8	42.1
2	7.3	5.0	2.9	2.0	1.0	0.4	0.2	0.0	0.0	0.0
3	6.1	3.8	1.9	1.3	0.7	0.4	0.3	0.2	0.1	0.0
4	5.5	3.3	1.5	0.9	0.4	0.2	0.1	0.1	0.1	0.0
5	5.2	3.0	1.3	0.7	0.3	0.1	0.1	0.0	0.0	0.0
6	5.0	2.9	1.2	0.6	0.2	0.1	0.0	0.0	0.0	0.0
7	4.9	2.8	1.1	0.6	0.2	0.1	0.0	0.0	0.0	0.0
8	4.8	2.7	1.0	0.5	0.2	0.1	0.0	0.0	0.0	0.0
9	4.9	2.7	1.0	0.5	0.2	0.1	0.0	0.0	0.0	0.0
10	4.8	2.6	1.0	0.5	0.2	0.1	0.0	0.0	0.0	0.0

Appendix 5
Power of Performance of Circular Boxplot

n=5, κ=2

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.4	0.0
0.1	0.1	0.1	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.4	0.0
0.2	0.1	0.1	0.2	0.0	0.1	0.1	0.3	0.0	0.1	0.0	0.3	0.0	0.1	0.0	0.3	0.0	0.1	0.0	0.4	0.0
0.3	0.1	0.1	0.2	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.1	0.0	0.4	0.0
0.4	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0
0.5	0.2	0.2	0.3	0.0	0.2	0.1	0.3	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0
0.6	0.2	0.2	0.3	0.0	0.2	0.2	0.3	0.0	0.2	0.1	0.4	0.0	0.2	0.1	0.4	0.0	0.2	0.1	0.5	0.0
0.7	0.3	0.2	0.3	0.0	0.2	0.2	0.3	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.5	0.0
0.8	0.3	0.3	0.3	0.0	0.3	0.2	0.3	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.5	0.0
0.9	0.3	0.3	0.3	0.0	0.3	0.3	0.4	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.4	0.0	0.3	0.2	0.5	0.0
1.0	0.3	0.3	0.3	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.4	0.0	0.3	0.2	0.5	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.4	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.5	0.0
0.1	0.1	0.0	0.4	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.6	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.6	0.0
0.2	0.1	0.0	0.5	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.6	0.0	0.1	0.0	0.6	0.0	0.1	0.0	0.6	0.0
0.3	0.1	0.1	0.5	0.0	0.1	0.1	0.5	0.0	0.1	0.1	0.6	0.0	0.1	0.0	0.6	0.0	0.1	0.1	0.6	0.0
0.4	0.1	0.1	0.5	0.0	0.1	0.1	0.6	0.0	0.1	0.1	0.6	0.0	0.1	0.1	0.6	0.1	0.1	0.1	0.6	0.1
0.5	0.2	0.1	0.5	0.0	0.2	0.1	0.6	0.0	0.1	0.1	0.6	0.1	0.1	0.1	0.6	0.1	0.1	0.1	0.6	0.1
0.6	0.2	0.1	0.5	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0
0.7	0.2	0.2	0.5	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0
0.8	0.3	0.2	0.5	0.0	0.2	0.2	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.2	0.6	0.0
0.9	0.3	0.2	0.5	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.6	0.0	0.2	0.1	0.6	0.0	0.2	0.2	0.6	0.0
1	0.3	0.3	0.5	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.7	0.0	0.2	0.2	0.6	0.0

n=5, κ=5

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.2	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.2	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0
0.3	0.2	0.2	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0
0.4	0.3	0.3	0.2	0.0	0.2	0.2	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0
0.5	0.3	0.3	0.2	0.0	0.2	0.2	0.2	0.0	0.2	0.2	0.2	0.0	0.2	0.2	0.2	0.0	0.2	0.2	0.2	0.0
0.6	0.4	0.4	0.2	0.0	0.3	0.3	0.2	0.0	0.2	0.2	0.2	0.0	0.2	0.2	0.2	0.0	0.2	0.2	0.3	0.0
0.7	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.3	0.3	0.2	0.0	0.3	0.3	0.2	0.0	0.3	0.3	0.3	0.0
0.8	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.3	0.0
0.9	0.5	0.5	0.3	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.3	0.0
1	0.5	0.5	0.3	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.5	0.5	0.3	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0
0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0
0.2	0.1	0.1	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.4	0.0	0.1	0.0	0.4	0.0
0.3	0.1	0.1	0.2	0.0	0.1	0.0	0.3	0.0	0.1	0.0	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0
0.4	0.2	0.2	0.3	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0
0.5	0.2	0.2	0.3	0.0	0.2	0.1	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.5	0.0	0.2	0.2	0.5	0.0
0.6	0.3	0.3	0.3	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.5	0.0	0.2	0.2	0.5	0.0	0.2	0.2	0.5	0.0
0.7	0.4	0.4	0.3	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.5	0.0	0.2	0.2	0.5	0.0	0.2	0.2	0.5	0.0
0.8	0.4	0.4	0.3	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.5	0.0	0.2	0.2	0.5	0.0	0.3	0.3	0.5	0.0
0.9	0.5	0.5	0.4	0.0	0.4	0.4	0.5	0.0	0.3	0.3	0.5	0.0	0.3	0.3	0.6	0.0	0.3	0.3	0.5	0.0
1	0.5	0.5	0.3	0.0	0.4	0.4	0.5	0.0	0.3	0.3	0.5	0.0	0.3	0.3	0.6	0.0	0.3	0.3	0.5	0.0

**dif* is the rounded P1-P3

Appendix 5, continued.

n=5, κ=7

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.2	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.0	0.1	0.0
0.3	0.2	0.2	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0
0.4	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.1	0.1	0.2	0.0
0.5	0.4	0.4	0.1	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.2	0.0
0.6	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.2	0.0
0.7	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.3	0.3	0.2	0.0	0.3	0.3	0.2	0.0	0.3	0.3	0.2	0.0
0.8	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0
0.9	0.5	0.5	0.3	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.5	0.5	0.2	0.0
1	0.5	0.5	0.3	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.5	0.5	0.2	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0
0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.2	0.0
0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.2	0.0
0.3	0.1	0.1	0.1	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.1	0.1	0.3	0.0
0.4	0.2	0.2	0.2	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.2	0.1	0.3	0.0
0.5	0.3	0.2	0.2	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.3	0.0	0.2	0.2	0.3	0.0
0.6	0.3	0.3	0.2	0.0	0.2	0.2	0.3	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.3	0.0
0.7	0.4	0.4	0.3	0.0	0.3	0.3	0.3	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.5	0.0	0.3	0.3	0.4	0.0
0.8	0.5	0.5	0.3	0.0	0.4	0.4	0.4	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.5	0.0	0.3	0.3	0.4	0.0
0.9	0.5	0.5	0.3	0.0	0.4	0.4	0.4	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.5	0.0	0.3	0.3	0.4	0.0
1	0.5	0.5	0.3	0.0	0.4	0.4	0.4	0.0	0.4	0.4	0.5	0.0	0.3	0.3	0.5	0.0	0.3	0.3	0.5	0.0

n=5, κ=10

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.1	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.2	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0
0.3	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0
0.4	0.4	0.4	0.2	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.1	0.1	0.1	0.0
0.5	0.4	0.4	0.2	0.0	0.3	0.3	0.2	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.2	0.0
0.6	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.2	0.0
0.7	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.2	0.0
0.8	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0
0.9	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.2	0.0	0.4	0.4	0.2	0.0	0.5	0.5	0.2	0.0
1	0.5	0.5	0.3	0.0	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.5	0.5	0.3	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0
0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.2	0.0
0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.2	0.0
0.3	0.2	0.2	0.1	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.0	0.3	0.0	0.1	0.1	0.2	0.0
0.4	0.2	0.2	0.1	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.2	0.2	0.2	0.0
0.5	0.3	0.3	0.2	0.0	0.2	0.2	0.3	0.0	0.2	0.2	0.3	0.0	0.2	0.2	0.4	0.0	0.3	0.3	0.2	0.0
0.6	0.4	0.4	0.2	0.0	0.2	0.2	0.3	0.0	0.2	0.2	0.4	0.0	0.2	0.2	0.4	0.0	0.3	0.3	0.3	0.0
0.7	0.5	0.5	0.3	0.0	0.3	0.3	0.3	0.0	0.3	0.3	0.4	0.0	0.3	0.3	0.4	0.0	0.4	0.4	0.3	0.0
0.8	0.5	0.5	0.3	0.0	0.4	0.4	0.3	0.0	0.4	0.4	0.4	0.0	0.3	0.3	0.5	0.0	0.4	0.4	0.4	0.0
0.9	0.6	0.6	0.3	0.0	0.5	0.5	0.4	0.0	0.4	0.4	0.4	0.0	0.4	0.4	0.5	0.0	0.4	0.4	0.4	0.0
1	0.6	0.6	0.3	0.0	0.5	0.5	0.4	0.0	0.4	0.4	0.4	0.0	0.4	0.4	0.5	0.0	0.4	0.4	0.4	0.0

Appendix 5, continued.

n=10, κ=2

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.3	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0
0.1	0.1	0.1	0.3	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0
0.2	0.1	0.1	0.3	0.0	0.1	0.1	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0
0.3	0.2	0.1	0.3	0.1	0.1	0.1	0.3	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0
0.4	0.2	0.2	0.3	0.1	0.2	0.2	0.3	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0
0.5	0.3	0.2	0.3	0.1	0.2	0.2	0.3	0.0	0.2	0.1	0.2	0.0	0.2	0.1	0.2	0.0	0.1	0.1	0.2	0.0
0.6	0.4	0.3	0.3	0.1	0.3	0.3	0.3	0.1	0.3	0.2	0.3	0.0	0.3	0.2	0.3	0.0	0.2	0.2	0.2	0.0
0.7	0.5	0.4	0.4	0.1	0.4	0.4	0.3	0.1	0.3	0.3	0.3	0.0	0.3	0.3	0.3	0.0	0.3	0.2	0.2	0.0
0.8	0.6	0.5	0.4	0.1	0.5	0.5	0.3	0.1	0.4	0.3	0.3	0.0	0.4	0.3	0.3	0.0	0.3	0.3	0.3	0.0
0.9	0.6	0.6	0.4	0.1	0.6	0.5	0.3	0.0	0.4	0.4	0.4	0.0	0.4	0.4	0.4	0.0	0.4	0.3	0.3	0.0
1	0.7	0.6	0.4	0.1	0.6	0.5	0.4	0.0	0.4	0.4	0.4	0.0	0.4	0.4	0.4	0.0	0.4	0.3	0.3	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.7	0.0
0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.7	0.0
0.2	0.0	0.0	0.3	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.6	0.0	0.1	0.0	0.7	0.0
0.3	0.1	0.0	0.3	0.0	0.1	0.0	0.4	0.0	0.0	0.0	0.5	0.0	0.1	0.0	0.6	0.0	0.1	0.0	0.7	0.0
0.4	0.1	0.1	0.3	0.0	0.1	0.0	0.4	0.0	0.1	0.0	0.5	0.0	0.1	0.0	0.6	0.0	0.1	0.0	0.7	0.1
0.5	0.1	0.1	0.3	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.5	0.0	0.1	0.1	0.6	0.0	0.1	0.1	0.7	0.1
0.6	0.2	0.1	0.3	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.5	0.0	0.1	0.1	0.6	0.0	0.1	0.1	0.7	0.1
0.7	0.2	0.2	0.3	0.0	0.2	0.1	0.4	0.0	0.2	0.1	0.5	0.0	0.2	0.1	0.6	0.0	0.2	0.1	0.7	0.0
0.8	0.2	0.2	0.3	0.0	0.2	0.2	0.5	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.8	0.0
0.9	0.3	0.3	0.4	0.0	0.2	0.2	0.5	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.7	0.0	0.2	0.2	0.8	0.0
1	0.3	0.3	0.4	0.0	0.2	0.2	0.6	0.0	0.2	0.2	0.7	0.0	0.2	0.2	0.7	0.0	0.2	0.2	0.8	0.0

n=10, κ=5

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.1	0.1	0.1	0.3	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.2	0.2	0.1	0.3	0.0	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0
0.3	0.3	0.3	0.3	0.0	0.3	0.2	0.2	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.1	0.1	0.0
0.4	0.5	0.5	0.3	0.0	0.4	0.4	0.2	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0
0.5	0.7	0.7	0.3	0.0	0.6	0.6	0.2	0.0	0.5	0.5	0.2	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.1	0.0
0.6	0.9	0.8	0.3	0.0	0.8	0.8	0.2	0.0	0.7	0.7	0.2	0.0	0.7	0.7	0.2	0.0	0.6	0.6	0.2	0.0
0.7	0.9	0.9	0.3	0.0	0.9	0.9	0.2	0.0	0.8	0.8	0.2	0.0	0.8	0.8	0.2	0.0	0.7	0.7	0.2	0.0
0.8	1.0	1.0	0.3	0.0	0.9	0.9	0.3	0.0	0.9	0.9	0.2	0.0	0.9	0.9	0.2	0.0	0.8	0.8	0.2	0.0
0.9	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	0.9	0.9	0.2	0.0	0.9	0.9	0.2	0.0	0.9	0.9	0.2	0.0
1	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	0.9	0.9	0.4	0.0	0.9	0.9	0.4	0.0	0.9	0.9	0.3	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0
0.4	0.2	0.2	0.1	0.0	0.2	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.1	0.0
0.5	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0
0.6	0.5	0.5	0.2	0.0	0.4	0.4	0.1	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0
0.7	0.6	0.6	0.2	0.0	0.5	0.5	0.1	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.1	0.0	0.3	0.3	0.1	0.0
0.8	0.7	0.7	0.2	0.0	0.6	0.6	0.1	0.0	0.6	0.6	0.1	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.1	0.0
0.9	0.8	0.8	0.2	0.0	0.7	0.7	0.1	0.0	0.7	0.7	0.2	0.0	0.6	0.6	0.2	0.0	0.5	0.5	0.2	0.0
1	0.9	0.9	0.3	0.0	0.8	0.8	0.3	0.0	0.7	0.7	0.3	0.0	0.6	0.6	0.4	0.0	0.6	0.6	0.4	0.0

Appendix 5, continued.

n=10, κ=7

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.1	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.2	0.2	0.2	0.2	0.0	0.2	0.2	0.2	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0
0.3	0.4	0.4	0.2	0.0	0.4	0.3	0.2	0.0	0.3	0.3	0.1	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0
0.4	0.7	0.7	0.3	0.0	0.6	0.6	0.2	0.0	0.5	0.5	0.1	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.1	0.0
0.5	0.8	0.8	0.3	0.0	0.8	0.8	0.2	0.0	0.7	0.6	0.1	0.0	0.7	0.6	0.1	0.0	0.6	0.6	0.1	0.0
0.6	0.9	0.9	0.3	0.0	0.9	0.9	0.2	0.0	0.8	0.8	0.2	0.0	0.8	0.8	0.2	0.0	0.7	0.7	0.1	0.0
0.7	1.0	1.0	0.3	0.0	1.0	0.9	0.2	0.0	0.9	0.9	0.2	0.0	0.9	0.9	0.2	0.0	0.9	0.9	0.1	0.0
0.8	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	0.9	0.9	0.2	0.0
0.9	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0
1	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.2	0.2	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0
0.4	0.3	0.3	0.1	0.0	0.2	0.2	0.0	0.0	0.2	0.2	0.0	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0
0.5	0.5	0.5	0.1	0.0	0.4	0.4	0.1	0.0	0.3	0.3	0.0	0.0	0.3	0.3	0.0	0.0	0.2	0.2	0.0	0.0
0.6	0.6	0.6	0.1	0.0	0.5	0.5	0.1	0.0	0.5	0.5	0.0	0.0	0.4	0.4	0.0	0.0	0.4	0.4	0.1	0.0
0.7	0.8	0.8	0.1	0.0	0.7	0.7	0.1	0.0	0.6	0.6	0.1	0.0	0.6	0.6	0.1	0.0	0.5	0.5	0.1	0.0
0.8	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.1	0.0	0.6	0.6	0.1	0.0
0.9	0.9	0.9	0.2	0.0	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.1	0.0
1	1.0	1.0	0.3	0.0	0.9	0.9	0.3	0.0	0.9	0.9	0.3	0.0	0.8	0.8	0.2	0.0	0.8	0.8	0.3	0.0

n=10, κ=10

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.1	0.1	0.1	0.2	0.0	0.1	0.1	0.2	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0
0.2	0.3	0.3	0.2	0.0	0.3	0.2	0.2	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.1	0.0	0.1	0.1	0.1	0.0
0.3	0.6	0.6	0.2	0.0	0.5	0.5	0.2	0.0	0.4	0.4	0.1	0.0	0.4	0.4	0.1	0.0	0.3	0.3	0.1	0.0
0.4	0.8	0.8	0.2	0.0	0.7	0.7	0.2	0.0	0.6	0.6	0.1	0.0	0.6	0.6	0.1	0.0	0.5	0.5	0.1	0.0
0.5	0.9	0.9	0.2	0.0	0.9	0.9	0.2	0.0	0.8	0.8	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.1	0.0
0.6	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0
0.7	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	0.9	0.9	0.1	0.0
0.8	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0
0.9	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.2	0.0
1	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.2	0.2	0.1	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0
0.4	0.5	0.5	0.1	0.0	0.4	0.3	0.0	0.0	0.3	0.3	0.0	0.0	0.3	0.3	0.0	0.0	0.2	0.2	0.0	0.0
0.5	0.6	0.6	0.1	0.0	0.5	0.5	0.0	0.0	0.5	0.5	0.0	0.0	0.4	0.4	0.0	0.0	0.3	0.3	0.0	0.0
0.6	0.8	0.8	0.1	0.0	0.7	0.7	0.1	0.0	0.6	0.6	0.0	0.0	0.6	0.6	0.0	0.0	0.5	0.5	0.1	0.0
0.7	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.0	0.0	0.7	0.7	0.1	0.0
0.8	1.0	1.0	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0
0.9	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0
1	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	1.0	1.0	0.3	0.0	0.9	0.9	0.4	0.0	0.9	0.9	0.4	0.0

Appendix 5, continued.

n=60, κ=7

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.7	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.2	0.0
0.1	0.1	0.0	0.7	0.1	0.1	0.0	0.6	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0
0.2	0.3	0.2	0.7	0.2	0.2	0.1	0.6	0.1	0.1	0.1	0.4	0.0	0.1	0.1	0.4	0.0	0.1	0.1	0.3	0.0
0.3	0.6	0.4	0.7	0.2	0.5	0.4	0.6	0.2	0.4	0.3	0.4	0.1	0.4	0.3	0.4	0.1	0.3	0.2	0.3	0.0
0.4	0.9	0.7	0.7	0.2	0.8	0.6	0.6	0.2	0.7	0.6	0.4	0.1	0.7	0.6	0.4	0.1	0.6	0.5	0.3	0.1
0.5	1.0	0.9	0.7	0.1	0.9	0.9	0.6	0.1	0.9	0.8	0.4	0.0	0.9	0.8	0.4	0.0	0.8	0.8	0.3	0.0
0.6	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	0.9	0.3	0.0
0.7	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
0.8	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
0.9	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
1	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.3	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.2	0.2	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.4	0.4	0.4	0.1	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0
0.5	0.7	0.7	0.1	0.0	0.6	0.6	0.1	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.0	0.0	0.3	0.3	0.0	0.0
0.6	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0	0.8	0.8	0.0	0.0	0.7	0.7	0.0	0.0	0.5	0.5	0.0	0.0
0.7	1.0	1.0	0.2	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.0	0.0	0.9	0.9	0.0	0.0	0.8	0.8	0.0	0.0
0.8	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.9	0.9	0.0	0.0
0.9	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
1	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.2	0.0

n=60, κ=10

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.7	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.2	0.0
0.1	0.2	0.1	0.7	0.1	0.1	0.1	0.6	0.1	0.1	0.0	0.4	0.0	0.1	0.0	0.4	0.0	0.0	0.0	0.2	0.0
0.2	0.5	0.3	0.7	0.2	0.4	0.2	0.6	0.1	0.2	0.2	0.4	0.1	0.2	0.2	0.4	0.1	0.2	0.2	0.3	0.0
0.3	0.8	0.6	0.7	0.2	0.7	0.6	0.6	0.1	0.6	0.5	0.4	0.1	0.6	0.5	0.4	0.1	0.5	0.4	0.3	0.0
0.4	1.0	0.9	0.7	0.1	0.9	0.9	0.6	0.1	0.9	0.8	0.4	0.0	0.9	0.8	0.4	0.0	0.8	0.8	0.3	0.0
0.5	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	0.9	0.3	0.0
0.6	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
0.7	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
0.8	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
0.9	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0
1	1.0	1.0	0.7	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.6	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0
0.4	0.7	0.7	0.1	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.0	0.0	0.3	0.3	0.0	0.0	0.2	0.2	0.0	0.0
0.5	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.0	0.0	0.7	0.7	0.0	0.0	0.5	0.5	0.0	0.0
0.6	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	0.9	0.9	0.0	0.0	0.9	0.9	0.0	0.0	0.8	0.8	0.0	0.0
0.7	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	0.9	0.9	0.0	0.0
0.8	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
0.9	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
1	1.0	1.0	0.4	0.0	1.0	1.0	0.3	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0

Appendix 5, continued.

n=100, κ=2

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.9	0.1	0.1	0.0	0.8	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.5	0.0
0.1	0.1	0.0	0.9	0.1	0.1	0.0	0.8	0.1	0.0	0.0	0.6	0.0	0.0	0.0	0.6	0.0	0.0	0.0	0.6	0.0
0.2	0.2	0.0	0.9	0.1	0.1	0.0	0.8	0.1	0.1	0.0	0.7	0.0	0.1	0.0	0.7	0.0	0.0	0.0	0.6	0.0
0.3	0.2	0.0	0.9	0.2	0.2	0.0	0.8	0.1	0.1	0.0	0.7	0.1	0.1	0.0	0.7	0.1	0.1	0.0	0.6	0.0
0.4	0.3	0.1	0.9	0.3	0.2	0.1	0.8	0.2	0.1	0.0	0.7	0.1	0.1	0.0	0.7	0.1	0.1	0.0	0.6	0.1
0.5	0.5	0.1	0.9	0.4	0.4	0.1	0.8	0.3	0.2	0.1	0.7	0.2	0.2	0.1	0.7	0.2	0.2	0.1	0.6	0.1
0.6	0.6	0.1	0.9	0.5	0.5	0.1	0.8	0.4	0.4	0.1	0.7	0.2	0.4	0.1	0.7	0.2	0.3	0.1	0.6	0.2
0.7	0.8	0.2	0.9	0.5	0.7	0.2	0.8	0.4	0.5	0.2	0.7	0.3	0.5	0.2	0.7	0.3	0.4	0.2	0.6	0.2
0.8	0.8	0.3	0.9	0.6	0.8	0.3	0.8	0.5	0.6	0.3	0.7	0.4	0.6	0.3	0.7	0.4	0.5	0.2	0.6	0.2
0.9	0.9	0.4	0.9	0.5	0.8	0.4	0.8	0.5	0.7	0.3	0.7	0.4	0.7	0.3	0.7	0.4	0.6	0.3	0.6	0.3
1	0.9	0.4	0.9	0.5	0.9	0.4	0.8	0.5	0.8	0.4	0.7	0.4	0.8	0.4	0.7	0.4	0.6	0.3	0.7	0.3

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.4	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.4	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.5	0.1	0.0	0.4	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0
0.6	0.1	0.1	0.4	0.1	0.1	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.1	0.0
0.7	0.2	0.1	0.5	0.1	0.1	0.1	0.3	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0
0.8	0.3	0.2	0.5	0.1	0.1	0.1	0.4	0.0	0.1	0.1	0.3	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.2	0.0
0.9	0.4	0.2	0.5	0.1	0.2	0.1	0.4	0.1	0.1	0.1	0.3	0.0	0.1	0.0	0.3	0.0	0.0	0.0	0.4	0.0
1	0.4	0.3	0.6	0.2	0.2	0.1	0.5	0.1	0.1	0.1	0.4	0.0	0.1	0.0	0.2	0.0	0.0	0.0	0.5	0.0

n=100, κ=5

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.8	0.1	0.0	0.0	0.7	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.3	0.0
0.1	0.1	0.0	0.8	0.1	0.1	0.0	0.7	0.1	0.0	0.0	0.5	0.0	0.0	0.0	0.5	0.0	0.0	0.0	0.3	0.0
0.2	0.3	0.1	0.8	0.2	0.2	0.1	0.7	0.1	0.1	0.1	0.5	0.0	0.1	0.1	0.5	0.0	0.1	0.1	0.4	0.0
0.3	0.5	0.2	0.8	0.3	0.4	0.2	0.7	0.2	0.3	0.2	0.5	0.1	0.3	0.2	0.5	0.1	0.2	0.1	0.4	0.1
0.4	0.7	0.4	0.8	0.3	0.6	0.4	0.7	0.2	0.5	0.4	0.6	0.1	0.5	0.4	0.6	0.1	0.4	0.3	0.4	0.1
0.5	0.9	0.6	0.8	0.3	0.8	0.6	0.7	0.2	0.7	0.6	0.6	0.1	0.7	0.6	0.6	0.1	0.6	0.6	0.4	0.1
0.6	1.0	0.8	0.8	0.2	1.0	0.8	0.7	0.1	0.9	0.8	0.6	0.1	0.9	0.8	0.6	0.1	0.9	0.8	0.4	0.1
0.7	1.0	0.9	0.8	0.1	1.0	0.9	0.8	0.1	1.0	0.9	0.6	0.1	1.0	0.9	0.6	0.1	1.0	0.9	0.4	0.1
0.8	1.0	1.0	0.8	0.0	1.0	1.0	0.8	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0
0.9	1.0	1.0	0.8	0.0	1.0	1.0	0.8	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0
1	1.0	1.0	0.9	0.0	1.0	1.0	0.8	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.6	0.0	1.0	1.0	0.4	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.1	0.1	0.2	0.0	0.1	0.1	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0
0.4	0.2	0.2	0.2	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.1	0.0
0.5	0.5	0.4	0.2	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.1	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0
0.6	0.7	0.7	0.2	0.1	0.6	0.6	0.1	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.1	0.0	0.2	0.2	0.0	0.0
0.7	0.9	0.9	0.2	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.1	0.0	0.6	0.6	0.1	0.0	0.5	0.5	0.0	0.0
0.8	1.0	0.9	0.2	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.0	0.0
0.9	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.0	0.0
1	1.0	1.0	0.3	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	0.9	0.9	0.0	0.0

Appendix 5, continued.

n=100, κ=7

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.8	0.1	0.0	0.0	0.6	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.2	0.0
0.1	0.1	0.0	0.8	0.1	0.1	0.0	0.7	0.1	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0
0.2	0.3	0.1	0.8	0.2	0.2	0.1	0.7	0.1	0.1	0.1	0.5	0.1	0.1	0.1	0.5	0.1	0.1	0.1	0.3	0.0
0.3	0.6	0.3	0.8	0.3	0.5	0.3	0.7	0.2	0.4	0.3	0.5	0.1	0.4	0.3	0.5	0.1	0.3	0.2	0.4	0.1
0.4	0.9	0.6	0.8	0.3	0.8	0.6	0.7	0.2	0.7	0.6	0.5	0.1	0.7	0.6	0.5	0.1	0.6	0.5	0.4	0.1
0.5	1.0	0.8	0.8	0.1	0.9	0.8	0.7	0.1	0.9	0.8	0.5	0.1	0.9	0.8	0.5	0.1	0.8	0.8	0.4	0.0
0.6	1.0	1.0	0.8	0.1	1.0	1.0	0.7	0.0	1.0	0.9	0.5	0.0	1.0	0.9	0.5	0.0	1.0	0.9	0.4	0.0
0.7	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
0.8	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
0.9	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
1	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.2	0.2	0.2	0.0	0.1	0.1	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.0
0.4	0.4	0.4	0.2	0.0	0.3	0.3	0.1	0.0	0.2	0.2	0.0	0.0	0.2	0.2	0.0	0.0	0.1	0.1	0.0	0.0
0.5	0.7	0.7	0.2	0.0	0.6	0.6	0.1	0.0	0.5	0.5	0.0	0.0	0.4	0.4	0.0	0.0	0.3	0.3	0.0	0.0
0.6	0.9	0.9	0.2	0.0	0.8	0.8	0.1	0.0	0.8	0.8	0.1	0.0	0.7	0.7	0.0	0.0	0.5	0.5	0.0	0.0
0.7	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	0.9	0.9	0.1	0.0	0.9	0.9	0.0	0.0	0.8	0.8	0.0	0.0
0.8	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	0.9	0.9	0.0	0.0
0.9	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
1	1.0	1.0	0.4	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0

n=100, κ=10

ν	1				1.2				1.5				1.7				2			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.1	0.0	0.8	0.0	0.0	0.0	0.7	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.3	0.0
0.1	0.2	0.1	0.8	0.1	0.1	0.1	0.7	0.1	0.1	0.0	0.4	0.0	0.1	0.0	0.4	0.0	0.0	0.0	0.3	0.0
0.2	0.5	0.2	0.8	0.3	0.4	0.2	0.7	0.2	0.2	0.2	0.5	0.1	0.2	0.2	0.5	0.1	0.2	0.1	0.3	0.0
0.3	0.8	0.5	0.8	0.3	0.7	0.5	0.7	0.2	0.5	0.4	0.5	0.1	0.5	0.4	0.5	0.1	0.5	0.4	0.3	0.1
0.4	1.0	0.8	0.8	0.1	0.9	0.8	0.7	0.1	0.9	0.8	0.5	0.1	0.9	0.8	0.5	0.1	0.8	0.8	0.3	0.0
0.5	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	0.9	0.3	0.0
0.6	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
0.7	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
0.8	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
0.9	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0
1	1.0	1.0	0.8	0.0	1.0	1.0	0.7	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.5	0.0	1.0	1.0	0.4	0.0

ν	2.3				2.5				2.7				3				3.5			
λ	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*	P1	P3	P5	*dif*
0	0.0	0.0	0.1	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.2	0.1	0.1	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
0.3	0.3	0.3	0.2	0.0	0.2	0.2	0.1	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.1	0.0	0.0
0.4	0.7	0.7	0.2	0.0	0.5	0.5	0.1	0.0	0.4	0.4	0.0	0.0	0.4	0.3	0.0	0.0	0.2	0.2	0.0	0.0
0.5	0.9	0.9	0.2	0.0	0.8	0.8	0.1	0.0	0.8	0.7	0.0	0.0	0.7	0.7	0.0	0.0	0.5	0.5	0.0	0.0
0.6	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	0.9	0.9	0.0	0.0	0.9	0.9	0.0	0.0	0.8	0.8	0.0	0.0
0.7	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
0.8	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
0.9	1.0	1.0	0.2	0.0	1.0	1.0	0.1	0.0	1.0	1.0	0.0	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0
1.0	1.0	1.0	0.3	0.0	1.0	1.0	0.2	0.0	1.0	1.0	0.1	1.0	1.0	1.0	0.0	0.0	1.0	1.0	0.0	0.0

Appendix 6
Power of Performance for Row Deletion Statistics

		n=30			*n*=70			*n*=100			*n*=150		
κ	λ	*DM CEc*	*DM CEs*	*COV*	*DM CEc*	*DM CEs*	*COV*	*DM CEc*	*DM CEs*	*COV*	*DM CEc*	*DM CEs*	*COV*
	0.0	0.03	0.03	0.03	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.00	0.01
	0.1	0.08	0.05	0.07	0.05	0.04	0.04	0.04	0.02	0.03	0.03	0.02	0.02
	0.2	0.24	0.14	0.14	0.19	0.10	0.11	0.10	0.08	0.09	0.08	0.05	0.05
	0.3	0.42	0.29	0.32	0.30	0.18	0.20	0.28	0.18	0.18	0.21	0.15	0.15
	0.4	0.69	0.56	0.59	0.51	0.41	0.41	0.50	0.39	0.40	0.43	0.37	0.38
5	0.5	0.76	0.73	0.74	0.77	0.65	0.67	0.69	0.60	0.61	0.59	0.55	0.55
	0.6	0.91	0.88	0.91	0.87	0.82	0.84	0.87	0.78	0.79	0.78	0.75	0.75
	0.7	0.95	0.94	0.95	0.97	0.92	0.96	0.95	0.91	0.92	0.90	0.88	0.90
	0.8	0.98	0.97	0.98	0.97	0.97	0.97	0.97	0.96	0.97	0.97	0.95	0.96
	0.9	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99	0.98	0.98
	1.0	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	1.00	1.00	0.99	0.99
	0.0	0.04	0.04	0.04	0.02	0.01	0.02	0.01	0.01	0.01	0.00	0.01	0.00
	0.1	0.18	0.08	0.08	0.08	0.06	0.06	0.05	0.03	0.04	0.04	0.04	0.03
	0.2	0.42	0.31	0.32	0.38	0.24	0.25	0.29	0.19	0.19	0.21	0.17	0.17
	0.3	0.64	0.52	0.64	0.61	0.55	0.57	0.62	0.50	0.52	0.57	0.46	0.47
	0.4	0.89	0.77	0.89	0.89	0.83	0.84	0.85	0.79	0.82	0.79	0.74	0.74
10	0.5	0.99	0.95	0.99	0.97	0.94	0.96	0.96	0.94	0.95	0.96	0.96	0.95
	0.6	1.00	0.97	1.00	1.00	1.00	1.00	1.00	0.99	0.99	0.99	0.97	1.00
	0.7	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.98	1.00
	0.8	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0	0.03	0.02	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	0.1	0.49	0.28	0.29	0.27	0.20	0.20	0.21	0.16	0.15	0.17	0.09	0.11
	0.2	0.88	0.76	0.78	0.78	0.71	0.76	0.75	0.70	0.71	0.72	0.67	0.69
	0.3	0.98	0.97	0.98	0.99	0.95	0.99	0.98	0.97	0.98	0.98	0.95	0.98
	0.4	1.00	1.00	1.00	1.00	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
30	0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.0	0.04	0.04	0.04	0.01	0.02	0.01	0.01	0.01	0.01	0.00	0.00	0.00
	0.1	0.50	0.36	0.40	0.40	0.32	0.33	0.34	0.27	0.30	0.30	0.21	0.21
	0.2	0.97	0.94	0.95	0.94	0.92	0.93	0.92	0.91	0.92	0.92	0.91	0.92
	0.3	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.4	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
50	0.5	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.6	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.8	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	1.0	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Appendix 7
S-Plus Subroutine for Tests of Discordancy in Circular Data

```
Discordancy<-function(x,type){

      x<-na.exclude(x)
      n<-length(x)
      if (type==1){x<-x*pi/180}               # change from degree to radian

      C<-sum(cos(x))
      S<-sum(sin(x))
      R<-(C^2+S^2)^(0.5)                          # Resultant length
      Rbar<-R/n                                   # Sample mean Resultant
   length

      Ci<-matrix(0,nrow=n)
      Si<-matrix(0,nrow=n)
      Ri<-matrix(0,nrow=n)

      Rbari<-matrix(0,nrow=n)
       Cst<-matrix(0,nrow=n)
      ki<-matrix(0,nrow=n)

       #------- Estimate the mean direction--------------------#

             if(S>0 && C>0){Mu<-atan(S/C)}
                else if(C<0){Mu<-atan(S/C)+ pi}
                   else if(S<0 && C>0){Mu<-atan(S/C)+2*pi}

       #------ Estime the concentration parameter -------------#

      if (Rbar<0.53){k<-2*(Rbar)+(Rbar)^3+(5/6)*(Rbar)^5}
         else if (Rbar>=0.53 &&Rbar<0.85){k<--0.4+1.39*(Rbar)+(0.43/(1-
   Rbar))}
                   else if (Rbar>=0.85){k<-1/((Rbar)^3-4*(Rbar)^2+3*(Rbar))}

      #---------------------------------------------#
      #               C statistic                   #
      #---------------------------------------------#
         for (i in 1:n){
                Ci[i]<-sum(cos(x))-cos(x[i])
                Si[i]<-sum(sin(x))-sin(x[i])
                Ri[i]<-(Ci[i]^2+Si[i]^2)^(0.5)
                Rbari[i]<- Ri[i]/(n-1)
                Cst[i]<-(Rbari[i]-Rbar)/(Rbar)
                       }
                 Cmax<-max(Cst)

                for (i in 1:n){if(Cst[i]==Cmax){obC<-i}}
               Csta<-cbind(obC,Cmax)

      #---------------------------------------------#
      #               D statistic                   #
      #---------------------------------------------#
       theta1<-matrix(0,nrow=n-1,ncol=n)

       t1<-matrix(0,nrow=n-1,ncol=n)
       D1<-matrix(0,nrow=n-1,ncol=n)

      maxt1<-matrix(0,nrow=n)
       t<-matrix(0, nrow=n)
      D<-matrix(0, nrow=n)

    theta<-sort(x)
         # remoteness of each observation

       for ( i in 1:n){theta1[,i]<-remove.row(theta,i,1)}
```

```
# find arc length after the remoteness of each observation one by one

    for (j in 1:n){
        for (i in 1:n-2){
                    t1[i,j]<-theta1[i+1,j]-theta1[i,j]
                     D1[i,j]<-(t1[i,j]/t1[i-1,j])
                        }

            t1[n-1,j]<-2*pi-theta1[n-1,j]+theta1[1,j]
            D1[n-1,j]<-(t1[n-1,j]/t1[n-2,j])
            D1[1,j]<-(t1[1,j]/t1[n-1,j])
                  }


for( i in 1:n){maxt1[i]<-max(t1[,i])}  #find the greatest arc within columns

     MaxT1<-max(maxt1)                  # find the maximum among all
 columns

      for (i in 1:n){if(maxt1[i]==MaxT1){obd<-i}}

  # Apply the test on the whole data set

            for (i in 1:n-1){
                    t[i]<-theta[i+1]-theta[i]
                     D[i]<-(t[i]/t[i-1])
                         }

            t[n]<-2*pi-theta[n]+theta[1]
            D[n]<-(t[n]/t[n-1])
            D[1]<-(t[1]/t[n])

        for (i in 1:n){if(x[i]==theta[obd]){obD<-i}}

        D1<-(1/D)
       Df<-min(D[obd],D1[obd])

       Dsta<-cbind(obD,Df)

    #----------------------------------------------#
     #               M statistic                   #
     #----------------------------------------------#

         M<-matrix(0,nrow=n)

       for(i in 1:n){M[i]<-((Ri[i]-R+1)/(n-R))}
       Mmax<-max(M)

        for (i in 1:n){if(M[i]==Mmax){obM<-i}}

        Msta<-cbind(obM,Mmax)
        m<-cbind(M)

    #----------------------------------------------#
    #               L statistic                    #
    #----------------------------------------------#

        # Kuppa estimation with deleting one observation

    for(i in 1:n){
       if (Rbari[i]<0.53){ki[i]<-
 2*(Rbari[i])+(Rbari[i])^3+(5/6)*(Rbari[i])^5}
 else if (Rbari[i]>=0.53 &&Rbari[i]<0.85){ki[i]<-
 0.4+1.39*(Rbari[i])+(0.43/(1- Rbari[i]))}
 else if (Rbari[i]>=0.85){ki[i]<-1/((Rbari[i])^3-
 4*(Rbari[i])^2+3*(Rbari[i]))}
                        }

L<-matrix(0, nrow=n)

for(i in 1:n)
```

```
{L[i]<-(Ri[i]+1)*ki[i]-k*R
n*log((exp(ki[i])/(2*pi*ki[i])))/(exp(k)/(2*pi*k))}

     Lmax<-max(L)

     for (i in 1:n){if(L[i]==Lmax){obL<-i}}
     Lsta<-cbind(obL,Lmax)

    Summary<-cbind(n,C,S,R,Rbar,Mu,k)

    #---------------------------------------------#
    #              A and Chord statistic          #
    #---------------------------------------------#
   thetasif<-matrix(0,n,n)

    A<-matrix(0,n,n)
    Chord<-matrix(0,n,n)

    sumA<-matrix(0, nrow=n)
    sumChord<-matrix(0, nrow=n)

    for(i in 1:n){
        for(j in 1:n){
          thetasif[i,j]<-pi-abs(pi-abs(x[i]-x[j]))
                A[i,j]<-(1-cos(x[i]-x[j]))
               Chord[i,j]<-2*sin(thetasif[i,j]/2)
                 }}
   for(i in 1:n){
         sumA[i]<-(sum(A[,i]))/(2*(n-1))
         sumChord[i]<-(sum(Chord[,i]))/(2*(n-1))}

   Amax<-max(sumA)
   Chordmax<-max(sumChord)


    for (i in 1:n){if(sumA[i]==Amax){obA<-i}}
    for (i in 1:n){if(sumChord[i]==Chordmax){obChord<-i}}

    Asta<-cbind(obA,Amax)
    Chordsta<-cbind(obChord,Chordmax)

    #---------------------------------------------#
    #            Asymptotic distribution          #
    #---------------------------------------------#
    Distance<-matrix(0, nrow=n, ncol=n)
    ScoreChi<-matrix(0, nrow=n, ncol=n)

      for(i in 1:n){
         for(j in 1:n){
            Distance[i,j]<-k*(1-cos(x[i]-x[j]))
            if(Distance[i,j]>qchisq(0.95,1)){ScoreChi[i,j]<-1}
               }}
      Points<-apply(ScoreChi,1,sum)
      plot(Points)

list(Summary=Summary,Csta=Csta,Dsta=Dsta,Msta=Msta,Lsta=Lsta,Chordsta=Chordsta
  ,Asta=Asta,Points=Points)

#Discordancy(x,1)   if type isdegree=1, radian = other wise
}
```

Appendix 8
S-Plus subroutine for Circular Boxplot

```
CircularBoxplot<-function(x,v,r,type){

   if(type==1){x<-x*pi/180}
x1<-as.matrix(x)
x<-sort(x1)
n<-length(x)

                  #Mean and Concentration
                  #---------------------

               CircMeanCon<-function(x){
                  C<-sum (cos(x))
                   S<-sum(sin(x))
                  n<-length(x)
                  R<-(C^2+S^2)^(0.5)           # Resultant length
                  Rbar<-R/n                    # Sample mean Resultant
   length

                   # Mu Ml estimation             (Fisher pp.31 (2.9))

                     if(S>0 && C>0){Mu<-atan(S/C)}
                       else if(C<0){Mu<-atan(S/C)+ pi}
                        else if(S<0 && C>0){Mu<-atan(S/C)+2*pi}

                  # Kappa estimation          (Fisher pp.88 (4.40))

      if (Rbar<0.53){k<-2*(Rbar)+(Rbar)^3+(5/6)*(Rbar)^5}
         else if (Rbar>=0.53 &&Rbar<0.85){k<--0.4+1.39*(Rbar)+(0.43/(1-
   Rbar))}
                     else if (Rbar>=0.85){k<-1/((Rbar)^3-
   4*(Rbar)^2+3*(Rbar))}

                  list(Mu=Mu,k=k)}
      #------------------------------------------------------------#

                    #Mod Programme (Radian)
                    #---------------------

                   RadMod<-function(x){
                   for(i in 1:length(x)){
                     m<-as.integer(x[i]/(2*pi))
                       if (x[i]>=2*pi){x[i]<-(x[i]-(m*2*pi))}
                          else if (x[i]<0 &&x[i]>(-2*pi)){x[i]<-x[i]+(2*pi)}
                          else if (x[i]<=(-2*pi)){x[i]<-
   x[i]+((abs(m)+1)*2*pi)}
                                 else(x[i]<-x[i])
                                          }
                     list(x=x)}
      #------------------------------------------------------#

               # Programme To Estimate the median
               #--------------------------------
                   Cirmedian<-function(x){
                      n<-length(x)
                     Obs<-matrix(0, nrow=n)

                        #Circular distance
                        #----------------
                        CirDis<-function(a,x,n){
                        d<-matrix(0, nrow=n)
                           for (i in 1:n){d[i]<-abs(pi-abs(x[i]-a))}

                        sumd<-sum(d)
                        list(d=d,sumd=sumd)}

               #---------------------------------------------------#
```

```
                    #Estimation of Mean
                    #------------------

                    CircMean<-function(x){
                    C<-sum (cos(x))
                    S<-sum(sin(x))
                    # Mu Ml estimation             (Fisher pp.31 (2.9))
                       if(S>0 && C>0){Mu<-atan(S/C)}
                        else if(C<0){Mu<-atan(S/C)+ pi}
                        else if(S<0 && C>0){Mu<-atan(S/C)+2*pi}
                      list(Mu=Mu)
                             }
          #-----------------------------------------------------------#

           for (i in 1:n){
                 Obs[i]<-pi-(CirDis(x[i],x,n)$sumd)/n}

             CirMedSort<-sort(Obs)

          CirMedOdd<-CirMedSort[1]  #Odd case take the minimun
          CirMedEven<-CirMedSort[2] # Even case we look for less two
values


          # To specify the location of points

             for (i in 1:n){
                if(Obs[i]==CirMedOdd){PosOdd<-i}
                if(Obs[i]==CirMedEven){PosEven<-i}
                        }

          # To specify the value of meadian

           if((n%%2)==1){CirMed<-x[PosOdd]}
           if((n%%2)==0){TwoPoints<-c(x[PosOdd],x[PosEven])
             CirMed<-CircMean(TwoPoints)$Mu}

          list(CirMed=CirMed)

                       }
#------------------------------------------------------------------------#

 Mu<-CircMeanCon(x)$Mu   # To calculate the mean
kappa<-CircMeanCon(x)$k

# Rotation
#----------
xR<-RadMod(x-Mu)$x

FirstHalf<-matrix(NA,nrow=n)
secondHalf<-matrix(NA,nrow=n)

          CirMedian<-Cirmedian(xR)$CirMed

               for(i in 1:n){
             if(xR[i]>CirMedian && (xR[i]-CirMedian)<pi){secondHalf[i]<-xR[i]}
             if(xR[i]>CirMedian && (xR[i]-CirMedian)>pi){FirstHalf[i]<-xR[i]}
             if(xR[i]<CirMedian && (CirMedian-xR[i])<pi){FirstHalf[i]<-xR[i]}
             if(xR[i]<CirMedian && (CirMedian-xR[i])>pi){secondHalf[i]<-xR[i]}
                               }

               FirstHalf<-na.exclude(FirstHalf)
               secondHalf<-na.exclude(secondHalf)

          Q3<-Cirmedian(FirstHalf)$CirMed
          Q1<-Cirmedian(secondHalf)$CirMed

          CIQR<-2*pi-Q3+Q1

       WhiskerUpv<-RadMod(Q3-v*(CIQR))$x           #mode
```

```
       WhiskerDownv<-RadMod(Q1+v*(CIQR))$x          #mode


# To fix the whisker on the last point
#-------------------------------------

ddd<-matrix(10,nrow=n, ncol=2)

       for(i in 1:n){
          if (xR[i]<=WhiskerDownv){ddd[i,1]<-pi-abs(pi-abs(xR[i]-
   WhiskerDownv))}
          if (xR[i]>=WhiskerUpv){ddd[i,2]<-pi-abs(pi-abs(xR[i]-WhiskerUpv))}
                              }

                    for(i in 1:n){
                          if(ddd[i,1]==min(ddd[,1])){WhiskerDownv<-xR[i]}
                          if(ddd[i,2]==min(ddd[,2])){WhiskerUpv<-xR[i]}}

# Number of outliers
#------------------

OutlierCount<-0
OutlierValues<-matrix(NA,nrow=n)

       for(i in 1:n){if(xR[i]>WhiskerDownv &&xR[i]<WhiskerUpv)
            {OutlierCount<OutlierCount+1
             OutlierValues[i]<-xR[i]}}


#Re-rotate the observation
#-------------------------

Q1<-RadMod(Q1+Mu)$x
Med<-RadMod(CirMedian+Mu)$x
Q3<-RadMod(Q3+Mu)$x
Ul<-RadMod(WhiskerDownv+Mu)$x
Uf<-RadMod(WhiskerUpv+Mu)$x
OutlierValues<-RadMod(na.exclude(OutlierValues)+Mu)$x

outliers<-(as.vector(na.exclude(OutlierValues)))
Results<-cbind(WhiskerDownv,WhiskerUpv,OutlierCount,outliers)


#------------------------------------
# Start Drawing the Circular box plot
#------------------------------------
d<-x
       # Construct the Outer box
          plot(c(-r,r), c(-r,r), xlab="", ylab="",xaxt="n",yaxt="n",type="n")
          title("Circular Boxplot of")

       # Construct the main circle (middle circle)
          xp <- NULL
          yp <- NULL

             for(i in 0:100) {
                a <- (2 * pi * i)/100
                x <- (0.85*r) * cos(a)
                y <- (0.85*r) * sin(a)
                xp <- c(xp, x )
                yp <- c(yp, y )
                                   }
   lines(yp, xp, type = "l",lwd=1)

   #Plot data set on the circle

          dx<-(0.85*r)*cos(d)
          dy<-(0.85*r)*sin(d)

       points(dy, dx, type = "p",col=2,pch=15)
```

```
# Plot the Q1,Q3,Ul,Uf and Median on the Circle

          m<-c(Q1,Q3,Ul,Uf,Med)

              # Construct the inner circle points
                    mx1<-((0.85*r)-0.1)*cos(m)
                    my1<-((0.85*r)-0.1)*sin(m)

          # Construct the outer circle points
                    mx2<-((0.85*r)+0.1)*cos(m)
                    my2<-((0.85*r)+0.1)*sin(m)

              segments(my1[3], mx1[3], my2[3], mx2[3],lwd=5)       # Draw  Ul
              segments(my1[4], mx1[4], my2[4], mx2[4],lwd=5)      # Draw  Uf
              segments(my1[5], mx1[5], my2[5], mx2[5],lwd=7)      # Draw
   Median


# To plot the inner and outer circles

Dif<-matrix(0,nrow=1000)
Dif[1]<-Q3

if(Q1>Q3){             #Case when the median far from the zero direction

              for(i in 2:1000) {
                 Dif[i]<-Dif[i-1]+0.05
                 if(Dif[i]>Q1|Dif[i]==Q1){final<-i
                                         break}}

               Dif<-Dif[1:final]
                               }


             if(Q1<Q3){     #Case when the median close from the zero direction
              for(i in 2:1000) {
                 Dif[i]<-Dif[i-1]+0.005
                        if(Dif[i]>2*pi|Dif[i]==2*pi){final1<-i
                                              break}}
              Dif[final1+1]<-0
              for(i in (final1+2):1000) {
                       Dif[i]<-Dif[i-1]+0.005
                        if(Dif[i]>Q1|Dif[i]==Q1){final<-i
                                              break}}

               Dif<-(Dif[1:final])
         }

###########################

    Difmid<-matrix(0,nrow=1000)
    Difmid[1]<-Uf

       if(Ul>Uf){              #Case when the median far from the zero
    direction

              for(i in 2:1000) {
                 Difmid[i]<-Difmid[i-1]+0.05
                 if(Difmid[i]>Ul|Difmid[i]==Ul){finalmid<-i
                                         break}}
                 Difmid<-Difmid[1:finalmid]
                               }



       if(Ul<Uf){          #Case when the median close from the zero direction

              for(i in 2:1000) {
                 Difmid[i]<-Difmid[i-1]+0.05
                        if(Difmid[i]>2*pi|Difmid[i]==2*pi){finalmid<-i
```

```
                                break}}

                      Difmid[finalmid+1]<-0

                      for(i in (finalmid+2):1000) {
                       Difmid[i]<-Difmid[i-1]+0.05
                        if(Difmid[i]>Ul|Difmid[i]==Ul){finalmid<-i
                                               break}}

                      Difmid<-(Difmid[1:finalmid])
          }

# Construct the inner circle points
                xin <- ((0.85*r)-0.1) * cos(Dif)
                yin <- ((0.85*r)-0.1)* sin(Dif)

# Construct the outer circle points
                xout <-((0.85*r)+0.1) * cos(Dif)
                yout <-((0.85*r)+0.1)* sin(Dif)

# Construct the midd circle points
                xmid <-((0.85*r)) * cos(Difmid)
                ymid <-((0.85*r))* sin(Difmid)

# Draw the inner and outer circle
       lines(yin, xin, type = "l",lwd=5)
       lines(yout, xout, type = "l",lwd=5)
       lines(ymid, xmid, type = "l",lwd=3)
       segments(yin[1], xin[1], yout[1], xout[1],lwd=5)                  # Draw
   Q3
       segments(yin[final], xin[final], yout[final], xout[final],lwd=5)  # Draw
   Q1

Mean<-Mu
Median<-RadMod(CirMedian+Mu)$x
Concentration<-kappa
OutlierCountD<-OutlierCount*pi/180

SSRadian<-cbind(Mean,Median,Concentration,
   Q1,Q3,CIQR,WhiskerDownv,WhiskerUpv,OutlierCount)
SSDegree<cbind(Mean,Median,Concentration=Concentration*pi/180,Q1,Q3,CIQR,Whisk
   erDownv,WhiskerUpv,OutlierCountD)*180/pi

         OutliersDegree<-outliers*180/pi

list("Summary Statistics Radian"=SSRadian,outliers=outliers, "Summary
   Statistics in Degree"=SSDegree,OutliersDegree=OutliersDegree)

}
#CirQun(x,v,r,type)    type=1 degree      else degree
```

Appendix 9
S-Plus Subroutine to Fit the Simple Circular Regression Model

```
SCRM<-function(x,y,iter){
        # To obtain the initial values
   #----------------------------

    n<-length(x)

     dat.lm<-lm(y~x)

     dat.lm<-summary(dat.lm)

     dat.lm<-as.vector(dat.lm)

     dat.lm<-dat.lm$coefficients

     dat.lm<-as.vector(dat.lm)

      iniA<-dat.lm[1]

      iniB<-dat.lm[2]

   # The begining of the iterative procedure
   #---------------------------------------

      bb<-matrix(0,nrow=iter)
      alpha<-matrix(0,nrow=iter)

      bb[1]<-iniB
      alpha[1]<-iniA

      for (i in 2:iter){

             S<-sum(sin(y-bb[i-1]*x))
             C<-sum(cos(y-bb[i-1]*x))

          if (S>0 && C>0){alpha[i]<-atan(S/C)}
          else if (C<0) {alpha[i]<-atan(S/C)+pi}
          else if (S<0 && C>0){alpha[i]<-atan(S/C)+2*pi}


          K1<-sum(x*sin(y-alpha[i]-bb[i-1]*x))
          K2<-sum((x^2)*cos(y-alpha[i]-bb[i-1]*x))
          bb[i]<-bb[i-1]+(K1/K2)
          final<-i
          if (abs(bb[i-1]-bb[i])<0.001) break
            }

          alphaEst<-alpha[final]
          betaEst<-bb[final]

          w<-(sum(cos(y-alpha[final]-betaEst*x)))/n
          KappaEst<-(9-8*w+3*(w^2))/(8*(1-w))


          A2w<-(sum(cos(y-alpha[final]-betaEst*x)^2))/n




   #Asymptotic properties
   #---------------------

   pp<-n*(sum(x^2))-(sum(x))^2

   VarAlphaHead<-(sum(x^2))/(KappaEst*w*pp)
```

```
VarBetaEst<-n/(KappaEst*w*pp)
VarKappaEst<-KappaEst/(n*(KappaEst-KappaEst*(w^2)-w))

 sealpha<-sqrt(VarAlphaHead)
 sebeta<-sqrt(VarBetaEst)
 sekappa<-sqrt(VarKappaEst)


# fitted values
#---------------

    yhat<-alpha[final]+betaEst*x

        for(i in 1:length(yhat)){
         m<-as.integer(yhat[i]/(2*pi))
         if (yhat[i]>=2*pi){yhat[i]<-(yhat[i]-(m*2*pi))}
              else if (yhat[i]<0 &&yhat[i]>(-2*pi)){yhat[i]<-yhat[i]+(2*pi)}
                  else if (yhat[i]<=(-2*pi)){yhat[i]<-
yhat[i]+((abs(m)+1)*2*pi)}
                              else(yhat[i]<-yhat[i])
                               }


# Calaculate the residuals
#-------------------------

   res<-matrix(0, nrow=n)

   for (i in 1:n){
if(((yhat[i])<y[i])&&(y[i]-yhat[i])<=pi){res[i]<-(pi-abs(pi-abs(y[i]-
yhat[i])))}
if(((yhat[i])<y[i])&&(y[i]-yhat[i])>pi){res[i]<--(pi-abs(pi-abs(y[i]-
yhat[i])))}
if(((yhat[i])>y[i])&&(yhat[i]-y[i])<=pi){res[i]<--(pi-abs(pi-abs(y[i]-
yhat[i])))}
 if (((yhat[i])>y[i])&&(yhat[i]-y[i])>pi){res[i]<-(pi-abs(pi-abs(y[i]-
yhat[i])))}
                              }
# Calculate the circular correlation
#-----------------------------------
       Corr<-function (x,y){

       MeanCon<-function(CC){
          n<-length(CC)
          C<-sum(cos(CC))
          S<-sum(sin(CC))
          R<-(C^2+S^2)^(0.5)                    # Resultant length
          Rbar<-R/n                             # Sample mean Resultant
length

          if(S>=0 && C>0){Mu<-atan(S/C)}
             else if(C<0){Mu<-atan(S/C)+ pi}
                else if(S<0 && C>=0){Mu<-atan(S/C)+2*pi}

                   list(Mu=Mu)}

                xbar<-MeanCon(x)$Mu
                ybar<-MeanCon(y)$Mu

                A<-sin(x-xbar)
                B<-sin(y-ybar)

             r<-sum(A*B)/(sum(A^2)*sum(B^2))^0.5
             r2<-r^2
                list(r=r,r2=r2)
                     }

# Circular Correlation between X and Y
CorrXY<-Corr(x,y)$r

#Circular correlation between y and yaht
```

```
CorrYhatY<-Corr(y,yhat)$r
R2<-(CorrYhatY)^2

output1<-cbind(alpha[final],betaEst,KappaEst)
output2<-cbind(sealpha,sebeta,sekappa)
output3<-cbind(CorrXY,w,A2w,CorrYhatY,R2)

par(mfrow=c(1,2))
plot(x,y)
plot(res)


list("Estimation of Model parameters"=output1,"Standard Error"=output2,
   "Goodness of fit"=output3, res=res)

   }
#SCRM(x,y,iter)
#SCRM(wind$R,wind$A,50)
```

Appendix 10
S-Plus Subroutine to Obtain Row Deletion Statistics for Simple Circular Regression Model

```
RowDeletionProcedures<-function(x,y,iter){

   #---------------- Calculation of the statistics ----------------------#


         SCRM<-function(x,y,iter){

             # To obtain the initial values
            #-----------------------------

                n<-length(x)

               dat.lm<-lm(y~x)

               dat.lm<-summary(dat.lm)

              dat.lm<-as.vector(dat.lm)

              dat.lm<-dat.lm$coefficients

              dat.lm<-as.vector(dat.lm)

              iniA<-dat.lm[1]

              iniB<-dat.lm[2]


            # The begining of the iterative procedure
            #----------------------------------------

            bb<-matrix(0,nrow=iter)
            alpha<-matrix(0,nrow=iter)

            bb[1]<-iniB
            alpha[1]<-iniA

               for (i in 2:iter){

                  S<-sum(sin(y-bb[i-1]*x))
                  C<-sum(cos(y-bb[i-1]*x))

               if (S>0 && C>0){alpha[i]<-atan(S/C)}
               else if (C<0) {alpha[i]<-atan(S/C)+pi}
               else if (S<0 && C>0){alpha[i]<-atan(S/C)+2*pi}


               K1<-sum(x*sin(y-alpha[i]-bb[i-1]*x))
               K2<-sum((x^2)*cos(y-alpha[i]-bb[i-1]*x))
               bb[i]<-bb[i-1]+(K1/K2)
               final<-i
               if (abs(bb[i-1]-bb[i])<0.001) break
                 }

               alphaEst<-alpha[final]
               betaEst<-bb[final]

               w<-(sum(cos(y-alpha[final]-betaEst*x)))/n
               KappaEst<-(9-8*w+3*(w^2))/(8*(1-w))


         # fitted values
   #---------------

      yhat<-alpha[final]+betaEst*x
```

```
        for(i in 1:length(yhat)){
         m<-as.integer(yhat[i]/(2*pi))
          if (yhat[i]>=2*pi){yhat[i]<-(yhat[i]-(m*2*pi))}
                else if (yhat[i]<0 &&yhat[i]>(-2*pi)){yhat[i]<-yhat[i]+(2*pi)}
                    else if (yhat[i]<=(-2*pi)){yhat[i]<-
yhat[i]+((abs(m)+1)*2*pi)}
                                else(yhat[i]<-yhat[i])
                                 }


         #circular distance
         #-----------------

        d<-pi-abs(pi-abs(y-yhat))


        # Mean circular error statistics
        #-------------------------------

        MCEc<-sum(1-cos(y-yhat))/n
        MCEs<-sum(sin(d/2))/n


        # Estimation of Concentration parmater
        #-------------------------------------
                Concentration<-function(CC){
            n<-length(CC)
            C<-sum(cos(CC))
            S<-sum(sin(CC))
            Rbar<-((C^2+S^2)^(0.5))/n

        if (Rbar<0.53){k<-2*(Rbar)+(Rbar)^3+(5/6)*(Rbar)^5}
        else if (Rbar>=0.53 &&Rbar<0.85){k<-(-0.4+1.39*(Rbar)+(0.43/(1-
Rbar)))}
            else if (Rbar>=0.85){k<-1/((Rbar)^3-4*(Rbar)^2+3*(Rbar))}
                     list(k=k)
                               }

    kappa<-Concentration(y-yhat)$k

        # The Modified mean circular error statistic
        #-------------------------------------------

        MDC<-2*sum(1-cos(sqrt(kappa)*(y-yhat)))


         # The determinant
         #----------------

        determenant<-(1/(KappaEst*w))



Result<-cbind(MCEc,MCEs,MDC,determenant)

 list(Result=Result)
           }



#------------------- Row deletion approach ------------#

 n<-length(x)
 ResultAll<-matrix(0,nrow=4)
 ResultAll<-SCRM(x,y,100)$Result
 DeletionRowResult<-matrix(0,nrow=n, ncol=4)
 Statistics<-matrix(0,nrow=n, ncol=4)
```

```
 Combinexy<-data.frame(x,y)

    for(i in 1:n){
      Newdata<-remove.row(Combinexy,i,1)
       DeletionRowResult[i,]<-SCRM(Newdata$x,Newdata$y,iter)$Result
                         }

DMCEc<-abs(DeletionRowResult[,1]-ResultAll[1])
DMCEs<-abs(DeletionRowResult[,2]-ResultAll[2])
MDCi<-abs(DeletionRowResult[,3]-ResultAll[3])
COVRATIO<-abs(1-(DeletionRowResult[,4]/ResultAll[4]))

Statistics<-cbind(DMCEc,DMCEs,MDCi,COVRATIO)


# Graphical representation of the results
#----------------------------------------

par(mfrow=c(2,2))
plot(DMCEc)
plot(DMCEs)
plot(MDCi)
plot(COVRATIO)

list(ResultAll=ResultAll,Statistics=Statistics)
                                  }

#RowDeletionProcedures(x,y,iter)
```